EXPLORING THE RELATIONSHIP BETWEEN GEOLOCATED SOCIAL NETWORK SERVICE TEXT AND CRIME

EON KIM

A PH.D. IN SECURITY AND CRIME SCIENCE
DEPARTMENT OF SECURITY AND CRIME SCIENCE
UNIVERSITY COLLEGE LONDON

DECLARATION

I, Eon Kim, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Eon Kim

ABSTRACT

The purpose of this thesis is to explore the potential utility of Geocoded Social Network Service datasets, especially geolocated Twitter, to help understand criminal activities in urban settings. In particular, the thesis is concerned with how these datasets can advance understanding of crime events in their spatiotemporal context. Such data has potential to reveal usually unobserved underlying conditions of crime events as a complementary source of knowledge for advancing the development of crime prevention strategies. In pursuit of these goals, the thesis comprises three case studies conducted using data for New York City in the United States. It examines: 1) Mobile dynamic population patterns 2) the emotion patterns of the mobile population, and 3) the attribute patterns of the mobile population (i.e. online-traits such as type and topic of discussion). With the combined results from the three case studies, this thesis explores how adding this new type of data improves existing knowledge of crime patterns. Results demonstrated that tracking spatial-temporal fluctuations of populations, along with their emotions and concerns has potential for explaining patterns across different crime types.

IMPACT STATEMENT

In this thesis, the potential utility of georeferenced posts collected from social media especially from Twitter data has been studied so as to offer a potential new method of crime prediction. The findings from the current study would be useful to a number of stakeholders especially to academic and political groups and police for setting future directions of crime research and building more effective and fair/less-biased policing strategies.

The contents of the thesis offer the potential for future conversations with stakeholders on a number of topics. The literature review of current research highlights that there is a lack of empirical studies that reflect a firm theoretical argument for using social media data to study crime. As emphasised in the thesis, most existing research has not fully contemplated the conceptual framework linking certain behaviours and criminal situations and has been limited in suggesting computational methods or presenting statistical evidence. In past analytic research in other fields, big data use has revealed some critical problems when analysed without careful consideration of how it represents social phenomena and theory-based hypothesis. Taking lessons from these past mistakes, the thesis suggests a criminological framework for how social media data can help understand and potentially predict crime. As the first study examines the utility of social media data for a series of distinct 'crime situations' by crime types and place contexts, future research could be developed based upon the findings of current research for more in-depth study for each crime type. For example, presence of others has a differential relationship with assault and burglary, which depends upon the specific place context. Unlike traditional analysis which often relies on the demographic and socioeconomic characteristics of population to predict crime risk, this study employs new data which is ageless, financialess, raceless, and genderless and represents thoughts and emotions of population. This thesis suggests that approaches to identifying and preventing crime strategies could be expanded and modified through the use of social media data. The second and third studies demonstrate that social media data can indicate the situational atmosphere of place-time locations in terms of emotions and points of view of present populations. These have varying, sometimes significant, relationships with crime and might lead to new types of mitigation and intervention. This thesis provides the first comprehensive study on the potential utility of social media data with the potential for multiple academic publications. The thesis identifies a variety of avenues for future crime prediction research to expand on these findings.

ACKNOWLEDGEMENTS

This thesis would not exist at all without Professors Kate J. Bowers and Shane D. Johnson. During my doctoral studies, they always listened to, guided and encouraged me with their profound knowledge of the subject.

Professor Bowers has become my motivation and inspiration since the first time we met. Her knowledge and attitude toward research and colleagues has showed me the kind of researcher and person I wish to be. I would like to thank Professor Johnson. At every important moment, he provided endless support, encouraging me and his guidance was deeply influential and inspiring to this work.

I would like to thank Professor Ronald V. Clarke who trusted me even when I doubted myself. Having him and his support on my side was an incredible honour. Professor Mangai Natarajan and Professor Jason Barr who must be acknowledged for believing in me and my research. I would particularly like to thank Professor Richard Wortley and Professor Gloria Laycock at the Security and Crime Science department who were provided me with generous and stimulating support in relation to my work.

Lastly, many thanks to Dr Jay Kwon for being such a good friend during this journey and Professor KyungHwan Kim, my dad, who taught me the joy of learning and sacrificed alot for me and family. I hope this little achievement makes him smile.

I have been able to complete my thesis with the genuine support and solicitude from the professors who lead me here. I owe a lot of things to them and I hope I can pass on the support that I have received from them to others in the next stage of my career as a dedicated researcher.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 THESIS SUMMARY

The purpose of this thesis is to explore the potential utility of Geocoded Social Network Service datasets, specifically geolocated Twitter, to help understand criminal activities in urban settings. In particular, the thesis is concerned with how these data sets can advance understanding of crime events in their spatiotemporal context and also reveal unobserved underlying conditions of the events as a complementary source of crime knowledge for advancing the development of crime prevention strategies.

The structure and content of the thesis are as follows. Chapter 1 provides a general overview of the aims of the thesis and the structure of empirical studies proposed. Chapter 2 provides theoretical foundation by reviewing the relevant place-based literature and Chapter 3 lays out the potential of using big data and the requisite data science approaches for criminological research. Chapter 4 outlines the methodology adopted for the thesis. Chapter 5 to 7 contains three empirical studies and Chapter 8 concludes with discussions which identify potential future work and recommendations building on findings from the preceding empirical chapters.

1.2 THESIS AIMS AND THEORETICAL FRAMEWORK

To be clear, as an exploratory study to gain better understanding of the new data and its utility, the geolocated contents of Twitter posts were explored within the perspective of environmental criminology, especially routine activity theory, which emphasises the role of situations in the creation of opportunities for a crime to occur (Cohen and Felson, 1979). As the theory conceptualised, criminal activities are seen here as the consequence of interactions between motivated offenders, suitable targets and capable guardians in space and time. Crucially, previous data used in empirical studies of the theory has been unable to fully incorporate the dynamic interactions of these central actors. Geolocated Twitter data, however, which was employed in this thesis, has the potential to provide a measure that captures the convergence of potential offenders, targets and third parties as guardians in space and time.

As a consequence of peoples' daily routine activities, active population patterns change dramatically in urban areas throughout the day. Population activity patterns, which include travel from home to work, or engagement in leisure activities, cannot be precisely measured by traditional population census data. The routine activity perspective, however, suggests that crime results from the convergence in space and time of the actors described above. Consequently, estimating how population movements vary over time and by location is crucial in crime analysis and testing theories of crime pattern formation. Furthermore, it is not just important where people are but what kinds of people are present. Therefore, an accurate estimation of the characteristics of crime-prone or crime-vulnerable groups in a given place is also important, since different types of crime will be associated with different groups of people due to inherent differences in underlying causal mechanisms of crimes.

Georeferenced data from social media contains a vast amount of spatiotemporal information which has not been exploited in previous crime risk measurement exercises. Although there were some efforts to understand crimes using the geolocated social media messages, they were limited to explore the relationship between the volume of crimes and mobile populations measured from the messages and the findings were not robust. (see Malleson and Andresen, 2015b and discussions in Chapter 3 for more detail). Therefore, the study seeks to provide the flexibility to future explore area - which currently is a relatively new and limited information available - to crime scientists. Through social media, individuals leave behind footprints of their movement fluctuations, but in addition traces of their thinking, personality and identity- they broadcast emotions, and activities using an informal language which is closer to natural speech than many formally written forms. By virtue of such characteristics of the data, as is evident from the review that follows, it has recently been utilised in various academic fields to mine people's self-expressed emotions, and to provide an alternative estimate of the ambient population to the census. In this vein, this thesis employed crowd-sourced data from geotagged social media posts in an attempt to generate a better understanding of crime problems.

One of the major benefits of geolocated messages from social media is that they enable us to develop a better understanding of the criminal atmosphere at given places. As noted in Elijah Anderson (1999)'s ethnographic works, all bars are not risky places even though bars are well known as a crime generator in many studies (Anderson, 1999). Bars tend to be located at places which share similar characteristics in terms of social economic status, and are typically surrounded by similar physical urban landscapes. However, whether a bar is a risky location or not depends on what types of groups occupy the place, how they become aware of them, and what interactions they have in them. Places have their own 'code'. The code is shared by the locals, the place occupiers, and sets a threshold of behaviours which are allowed in the place. Such codes, however, have not been adequately captured by quantitative measurement approaches applied in the field of crime science so far. This is a significant limitation of the research and may explain why there exist inconsistent findings across studies. Beyond conceiving of Twitter as a different form of online communication platform, it should also be considered as a new lens through which to increase understanding of dynamic social activities and responses to criminogenic settings by offenders and victims. Collectively, geolocated Twitter data has shown a lot of promise (Jansen et al., 2009; Lohr, 2012) as an alternative to traditional data sources and could be of use in understanding the conditions under which crime takes place.

1.3 STRUCTURE OF EMPIRICAL RESEARCH

The thesis focuses on how geolocated messages can be used to 1) model spatio-temporal patterns of motivated offenders, suitable targets and capable guardians as core elements for criminal opportunity identification; and, 2) examine how georeferenced social media data could improve effective crime preventive interventions by suggesting a new type of surveillance and prediction method to the police based on the attributes of social media messages.

In pursuit of these goals, the thesis comprises three case studies conducted using data for New York City in the United States. It examines: 1) Mobile population patterns, 2) Mobile population emotion patterns, and 3) mobile population attribute patterns of geolocated Twitter activity (i.e. online-traits especially shown interests and online-activities) (Figure 1).

Mobile ambient population patterns

The relationship between crime patterns and variation in the size of the mobile population at particular locations, estimated using georeferenced social media data, is the focus of the first study. It explores the possibility of using geolocated data to estimate mobile populations and to explore their association with the clustering of crime in space and time. To understand the relationship, the movement and fluctuation of estimated ambient population derived from social media data is analysed spatiotemporally. In other words, the relationship between the number of crime events and Tweets in each unit is examined to test the possibility of the data as an appropriate denominator for crime opportunity.

Mobile population emotion patterns

Taking this approach further, rather than looking at variation in the mere presence of people in an area, the second study essentially is concerned with variation in the 'emotions' of the mobile population across time and location. It seeks to determine whether criminal atmosphere could be captured by emotions of social media texts. To answer this, the emotional patterns associated with texts are compared with crime patterns in space and time.

Mobile population attribute patterns

The third case study explores the relationship between levels of social cohesion and crime based on activities in online environments. To explain, by analysing the texts of the microblog messages, types of topic words or online behaviours associated with crime behaviours and crime locations are examined. A tweet has a large spectrum of topics, including stories about daily life and social and political issues. So far, researchers have found that some people use Twitter for narcissistic purposes while others use it for showing their social and political interests (Bakliwal et al., 2013; Parmelee, 2013). The third empirical chapter takes a more in-depth look at the content of Tweets. The messages are analysed according to online-trait categories associated with personality and identity of the mobile population and compared with crime patterns.

With the combined results from the three case studies, the aim of the thesis is to explore how adding this new type of data improves existing knowledge of crime patterns. To date, where models have incorporated contextual data on neighbourhood characteristics, they have been simply based on the static residential populations and physical characteristics or the urban landscape. The aim of the thesis would be to use dynamic mobile population activities and the attributes of them in crime forecasting approaches.

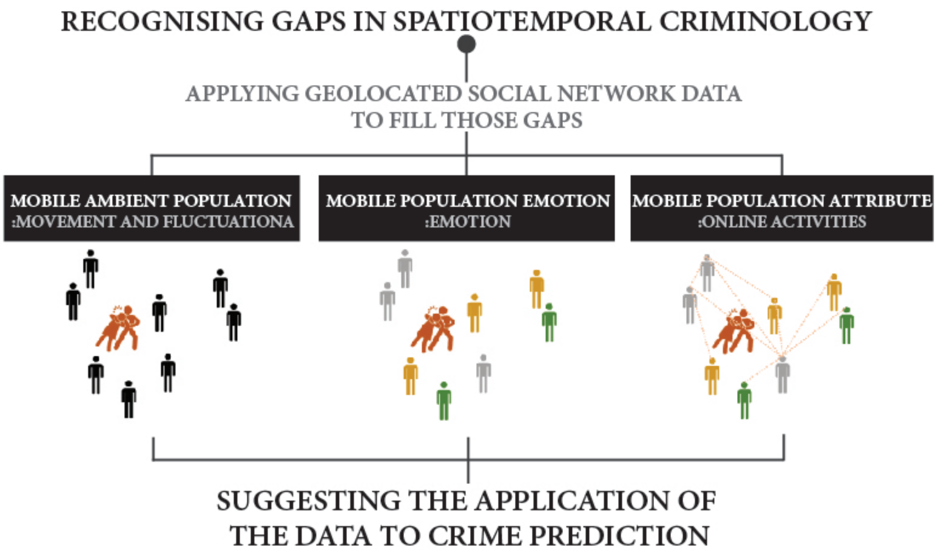


Figure 1: Thesis flowchart

2

THEORETICAL FRAMEWORK

The rationale of this thesis builds upon a particular theoretical perspective, the so-called environmental criminology. This chapter begins by detailing theories of neighbourhood effects and situational opportunities. It then goes on to discuss the challenge and promise of the integration of the two perspectives and finally proposes a theoretical model for the thesis. Note that this chapter provides a general framework for the empirical chapters and motivates the use of Twitter data in studying social issues. However, each of the three contributions also require a more specific review of the salient literature which, for convenience, are provided within the chapters themselves.

2.1 CRIME AND PLACENALITY

That crime is concentrated at very small places is now uncontested (see Johnson, 2010). Although the statement has been repeatedly supported in the past empirical research, traditional criminologists obstinately hold the idea that 'the density of offenders explains crime density' (Weisburd et al., 2016. p.42). While researchers interested in the characteristics of individuals study why some people commit crime and others do not, place-based theorists ask instead why some places and neighbourhoods have a high crime rate and others do not. In response to this question, routine activity theory provides a very straightforward and simple formulation, 'a different distribution of criminal opportunity makes a different crime rate' (Cohen and Felson, 1979), while social disorganisation theory proposes 'a difference of structural and cultural characteristics of places makes a different level of deviance' (Shaw and McKay, 1942). Although these two approaches propose different casual mechanisms and consider different theoretical constructs, both recognise that the characteristics of places matter and can that different compositions can make for radically different criminal atmospheres of places.

With respect to routine activity theory, Lawrence Cohen and Marcus Felson (1979) stated that "unlike many criminological inquires, we do not examine why individuals or groups are inclined criminally, but rather we take criminal inclination as given and examine the manner in which the spatio-temporal organisation of social activities helps people to translate their criminal inclinations into action"(Cohen and Felson, 1979, p.589). Similarly, Charis Kubrin and Ronald Weitzer (2003) note that "unlike theories centred on 'kind of people' explanations for crime, social disorganisation theory focuses on the effects of 'kind of places' – specifically, different types of neighbourhoods – in creating conditions of favourable or unfavourable to crime" (Kubrin and Weitzer, 2003, p.374).

That is, both perspectives explain the occurrence of crime events as the result of the interactions that occur as a function of human ecological processes. As both theories locate their focus in the characteristics of places, and nothing in either theory contradicts the mechanisms invoked by the other (Weisburd et al., 2016), they are compatible and, indeed, mutually supportive. It is perhaps surprising then, that research that seeks to integrate the two approaches is the minority. The following sections now review literature that seek to explain spatial patterns of crime.

2.2 NEIGHBOURHOOD INFLUENCES ON CRIME

2.2.1 *Juvenile delinquency and urban area*

The search for the root causes of crime in neighbourhood characteristics has enjoyed a long tradition within criminological and sociological research. Research concerned with social disorganisation focuses on "neighbourhood effects" and how they influence crime and human behaviour more generally (Markowitz et al., 2001; Browing et al., 2004; Harcourt and Ludwig, 2006). This perspective was initially discussed by Shaw and McKay (1942) in their book, Juvenile delinquency and urban area. Clifford Shaw and Henry McKay (1942) considered the production of delinquency and focused on the residential location of delinquents, rather than the location at which they offended. They argued that the structural characteristics of neighbourhoods, such as low social economic status, the mobility of residents, racial and ethnic heterogeneity, family disruption, and urbanisation, lead to social disorganisation at neighbourhood level that eventually leads to high crime rates (Shaw and McKay, 1942). They saw structural problems in a community as weakening the ability of residents to provide the informal control through which areas are regulated by those who live in them. Based on their theory, 'social disorganisation' occurs when a community does not share common values, and/or is unable to regulate order to attain those goals which are shared by the community (Bursik, 1988; Sampson, Raudenbush, and Earls, 1997).

According to Shaw and McKay (1942), the zone in transition is characterised by high rates of poverty, poor living conditions, and single parents and also has more crimes, issues of drug, alcohol, abortion, and prostitution (see Burgess, 1928). They said the zones are occupied with people who fail to adjust themselves and were also called "Ghetto" or "Slum" (Burgess, 1928; Shaw and McKay, 1942).

In *the truly disadvantage*, William Julius Wilson (1987) proposed that structurally disadvantaged areas create the truly disadvantaged underclass (Wilson, 1987). He said hot spots of crime – small areas with an unusually high crime rate - are characterised by the concentration of multiple layers of disadvantage. Between the late 1960's and the 1980's of Chicago, after the middle class moved out from city centres, the centres experienced impoverished conditions and decline (Wilson, 1987). According to Wilson (1987), the area was eventually populated by truly disadvantaged classes for whom unemployment, illegitimacy, single parenthood, and violence are typical characteristics.

Land et al. (1990) empirically supported Wilson's theory (1987), researching social isolation in areas of concentrated poverty, from 1970's to 1980's in large cities in America. They found that concentration effects of negative social indicators – per cent poverty, per cent black, and per cent of children under 18 not living with both parents - are noticeably associated with the economic deprivation indicators that are strongly correlated to homicide crimes (Land et al., 1990). They all accentuated that the individuals in isolated and disadvantaged areas are not abnormal biologically and mentally. Instead, the behaviours including criminal activities are the obviously anticipated results of the abnormal environment they have to cope with.

2.2.2 *Collective efficacy and social cohesion*

Beyond a traditional fixation on the influence of concentrated disadvantage, some researchers have sought to more explicitly articulate the mechanism through which social disorganisation might lead to crime (Bursik, 1988; Bursik and Grasmik, 1993; Sampson et al., 1999). They argued that a neighbourhood is not only a physical space where people live, but a living organism in which community members connect and interact. Robert Bursik and Harold Grasmick (1993) stressed that Shaw and McKay did not explicate the causal link between social disorganisation and crime. They noted that while social disorganisation might be associated with crime, the mechanism through which social disorganisation influences crime is informal social control. That is, concentrated disadvantage per se does not cause crime, but it may create conditions that impede the informal social control through which residents might ordinarily influence the behaviour of those in the neighbourhood.

Likewise, Sampson and Groves (1989) note that the indicators of residential mobility, heterogeneity, and poverty used in Shaw and McKay's work were used only to measure the level of disorganisation, and there was no explanation as to how these factors might influence crime rates. In their study, Sampson and Groves (1989) sought to articulate these mechanisms more explicitly and to test their association with crime empirically. Using data from the British Crime Survey for 238 electoral wards, they constructed measures of social control using data on local friendship networks, organisational participation, and the supervision of teenager peer groups. Using data on residential mobility, class and family disorganisation they derived measures of exogenous factors that might influence social disorganisation. They then correlated these variables with area-based crimes rates. They found a stronger relationship between crime rates and the measures of informal social control than the exogenous factors (Sampson and Groves, 1989).

While the early research focused on relatively static features of neighbourhood composition, more recent work has focused on the social processes through which neighbourhood structure might influence the formation of and sustainability of social ties. The collective efficacy framework of Sampson and his colleagues suggests that neighbourhood compositions influences the ability of residents of a neighbourhood to act collectively to intervene to deter crime. They elucidate collective efficacy as the reciprocal belief that residents can control the likelihood of undesirable behaviours effectively (Sampson, Morenoff, and Earls, 1999).

They also found collective efficacy can be gained through intergenerational networks, mutually transferral advices, information about child rearing, and, expectations for the joint informal control (Sampson, Morenoff, and Earls, 1999). According to this framework, crime occurs in a neighbourhood as a response to a lack of social control because criminals perceive the neighbourhood to be one in which crime can be committed with little risk (Sampson and Groves, 1989; Wilson and Kelling, 1982; Morenoff et al., 2001). A considerable body of empirical research provides support for these ideas, showing that community organisation (Sampson, 1988), residential interaction (Sampson, 1991), the size of friendship networks (Sampson and Groves, 1989), network stability (Bursik and Grasmick, 1993), and the extent of social ties (Warner and Roundtree, 1997) are negatively associated with neighbourhood crime rates.

2.2.3 *Neighbourhood cultural context*

However, these findings are not unequivocal. For example, studies have found that depending on the culture and context of a neighbourhood, social interactions and social ties may or may not lead to effective informal social control (Browning et al., 2004; St.Jean, 2007). Mary Pattillo-McCoy (1999) argues in her book, *Black Picket Fences* that social ties can impede efforts to construct social control in certain community contexts. According to Pattilo-McCoy (1999), even when a community has strong social cohesion, it could experience high crime rates because social ties play not only a surveillance role but also bring negative repercussion effects depending on neighbourhood cultural context (Pattillo-McCoy, 1999). In another study, Bellair (1997) argues that the connection between social ties and neighbourhood crime rates is far from clear since neighbourhoods can exhibit a high level of control ability without social ties (Bellair, 1997). He also points out that weak (rather than strong) social ties can reduce crime rates because they limit the interaction between community members and hence opportunities for confrontation (Bellair, 1997).

Browning et al. (2004) found high levels of community networking can reduce the regularity effect of collective efficacy in violent crime. In other words, neighbourhoods with strong social ties could have more crimes because strong social ties reduce their ability to regulate and control law-violating activities as they have a very tolerant view towards their neighbours (Browning et al. 2004). St.Jean (2007) notes that stability is not always good and mobility is not always bad (St. Jean, 2007). They both argue that neighbourhoods with a very low level of social ties could have a very low crime rate if residents trust each other and share similar expectations. In contrast, neighbourhoods with strong social ties could have very high rates of crime if residents do not share the same willingness to intervene for the common good because they do not trust one another (Warner and Rountree, 1997; Pattillo-McCoy, 1999; Browning et al., 2004; St.Jean, 2007).

In summary, the studies reviewed suggest that neighbourhood composition can influence the likelihood that residents will share common values, form social ties and act collectively to deter crime. However it appears that the neighbourhood composition including the social and physical structures of neighbourhoods do not solely determine the level of social control or resident's willingness to intervene, because the culture and context of places influence the formation and quality of ties that may form (Sampson et al. 1997, Kurbin and Weitzer, 2003; St.Jean, 2007). More importantly, the studies discussed suggest that these cultural contexts do not always match with prevalent measurements of social disorganisation such as demographic and socioeconomic status of neighbourhoods.

2.3 CRIMINOGENIC OPPORTUNITIES AND CRIME

Rather than explaining how criminality is shaped, opportunity theorists seek to answer long-standing questions such as; why does an individual choose to commit crime in a specific place and at specific time but not in other places and at other times? How might urban form shape crime patterns? To answer such questions, they take an ecological perspective which seeks to explain how the environmental backcloth (the urban fabric of towns and cities) and everyday activity might create opportunities for crime at particular locations at particular times (Brantingham and Brantingham, 1993). A number of related theories or perspectives have been formulated including routine activity theory (Cohen and Felson, 1979; Felson, 1987; Sherman et al., 1989), the rational choice perspective (Clarke and Cornish, 1985; Cornish and Clarke, 1986; Brantingham and Brantingham, 1978), crime pattern theory (Brantingham and Brantingham, 1991; 1993), lifestyle theory (Hindelang et al., 1978) and various crime prevention strategies and approaches (see Jeffery, 1971; Newman, 1972; Clarke, 1980; Willson and Kelling, 1982).

2.3.1 *Routine activity theory*

Cohen and Felson (1979) consider crime opportunities to emerge as a result of the patterns of activity that people engage in during everyday life. When people are at work their homes are left unattended, and hence vulnerable to burglary, while they travel from one place to the next they are exposed to the risk of crime against themselves (Cohen and Felson, 1979). As such, Cohen and Felson (1979) state "structural changes in routine activity patterns can influence crime rates by affecting the convergence in space and time of the three minimal elements of direct-contact predatory violation; motivated offenders, suitable targets, and the absence of capable guardians"(Cohen and Felson, 1979, p.589). This suggests that three key factors of crime events and three key concepts; offenders, targets, and guardians; and structural changes, space and time, and direct contact.

Who are offenders, targets, and guardians? How have they been conceptualised and operationalised in the previous works? How do routine activities vary at individual-level routine and aggregate-level routine? How do structural changes influence routine activities? In this section, the answers of these questions will be discussed.

In their original paper, Cohen and Felson (1979) emphasise the idea that 'an opportunity makes a crime' (Cohen and Felson, 1979). Crime does not happen when there is a willing offender alone. Crime can only happen when there is an opportunity- when a willing offender encounters a suitable target at a specific location in the absence of a capable guardian (Cohen and Felson, 1979). According to the theory, the emergence of opportunities varies by space and time, and hence crime opportunities are not ubiquitous and do not exist everywhere at all times. Rather, crime opportunities vary across urban settings over time. From this perspective, a key question concerns the way in which structural factors influence individuals' day-to-day routine activities, which in turn shape the likely convergence of people at particular places at particular times, which in turn influences the distribution of crime opportunity.

2.3.2 *Human ecology and routine activities*

Cohen and Felson's idea developed from studies of human ecology, and particularly the work of Hawley (1950). Hawley (1950) did not conceive of places as spatially bounded territories. Rather, he considered a place as 'an organisation of symbiotic and communalistic relationships' where there is mutual dependence among functionally dissimilar and similar organisms (Hawley, 1950). Cohen and Felson applied these concepts in the context of crime to explain crime events as being a natural by-product of the "mutual relationships between predator and prey affecting the sustenance of both" (Felson and Cohen, 1980). Putting it in a criminological context, they interpreted the ideas in the following way 'routine activities which feed upon other routine activities' or 'illegal activities must feed upon legal activities' (Madero-Hernandez and Fisher, 2012).

In his later work, Felson (2006) discussed and explained symbiosis as 'living together' and crime symbiosis as "providing illicit benefit to at least one of them" (Felson, 2006, p.163-164). Felson argues crime as a system cannot survive by itself; burglars share information on-their targets, a legal prostitute sells illegal drugs, and drug traders interplay within legal chains like bars, liquor stores, homeless centres, and convenient stores. Simply put, even though a criminal and their counterpart (the victim) may not have any direct connectivity, they could influence each other because crime is always built on complex chains of causation (Felson, 2006). In a crime mutualism perspective, Felson explains that mutualism refers to the 'net benefit' that results from interactions such as those that take place between a prostitute and a john, and, a drug seller and a drug dealer (Felson, 2006). Here, the criminals reciprocally exchange their resources against mutual enemies together and, spread and reproduce criminal activities as both parties benefit. Similarly, stolen goods should re-sale with cooperation with other criminals, offenders rely on each other to prevent themselves from being arrested, and prostitutes need a legitimate places to find customers (Felson, 2006). In all, criminals cannot be alone. They can only thrive with and exist through mutual assistance.

Another fundamental influence from Hawley's theoretical concept on routine activity theory concerns the structure of human routines over space and time. Hawley (1950) emphasised three temporal components - rhythm, tempo, and timing. Felson and Cohen (1980) employ these components to explain the temporal structure of routine activities. They argue that daily patterns of human activity are shaped in rhythm, tempo, and timing; "the regular periodicity with which events occur (rhythm), the number of events per unit of time (tempo), and the coordination among different activities which are more or less interdependent (timing)" (Felson and Cohen, 1980, p.391). By describing the activity of offenders, targets, and guardians these components contribute to our understanding of the frequency with which conditions converge to produce opportunities for crime (see Cohen and Felson, 1979, p.590).

2.3.3 *Regular routines in society structure*

As discussed above, according to Routine activity theory, crime patterns emerge as a by-product of regular routines (Clarke and Felson, 1993). Cohen and Felson (1979) define routine activities as "any recurrent and prevalent activities which provide for basic population and individual needs, whatever their biological or cultural origin" (Cohen and Felson, 1979, p.593). In criminology, routine activity has broadly been defined in terms of generalised social activities that occur at home, at work or in the pursuit of leisure activities away from home. A notable characteristic of individuals' activities concerns their spatiotemporal repetition. Activities are not random and have a clear cycle within dynamic spatiotemporal dimensions. Each individual engages in a layer of activity and the accumulated layers create concrete situational patterns of activity in time and space which can be thought of as 'routines of place' and 'routines of time'. Across hourly, daily, weekly, monthly, and yearly routines, people are constantly exposed to, interact with, and respond to the situations they are confronted with. If the situations they encounter are judged as safe to commit crime, motivated criminals may spontaneously offend. Therefore, if regularities are identified in individual routines, and accumulated routines are found to reveal underlying crime-prone situations, crime patterns may consequently be predictable. As Jeffrey Brantingham, an anthropologist, said, "The naysayers want you to believe that humans are too complex and too random—that this sort of math can't be done... but humans are not nearly as random as we think. In a sense, crime is just a physical process, and if you can explain how offenders move and how they mix with their victims, you can understand an incredible amount."(Perry et al., 2013). Put differently, understanding patternised routine and situations created by accumulated routines at place and time is key to explaining crime concentration (Weisburd et al., 2016).

As outlined above, understanding the factors that shape routine activities may provide valuable insight into crime pattern formation. Researchers thus began to pay particular attention to rethinking individuals' routine activities and the influential factors on it both conceptually and practically. Some of them focused on activity levels, at a micro-situational or a macro-context level (Felson, 2002), whereas others were interested in expected criminal routines such as where they live and where and when they commit crime (Brantingham and Brantingham, 1984). To sum up, routine activity theory provides an ecological framework for thinking about how the activities of motivated offenders, potential targets and guardians shape the opportunity for crime over space and time, how changes to peoples' routine activities might be influenced by changes of social conditions, and what variations in space and time could shape routine activities of offenders, victims, and guardians.

2.4 integration of theories

While traditional criminology has (at least historically) primarily dealt with criminal motivations, routine activity theory and social disorganisation theory seek to understand the root causes of crime in terms of places and how they differ either with respect to social structure, or the routine activity patterns associated with places. As noted above, while the theories discuss different mechanisms, they are not incompatible. In this section, their complementarity and compatibility will be discussed.

2.4.1 *Society changes and lifestyle routine changes*

Shaw and McKay did not posit a direct influence of social disorganisation on rates of delinquency (Bursik, 1988). Rather, neighbourhoods in transitional zones - generally characterised by poor economic status, high racial heterogeneity, and higher residential mobility – were assumed to produce higher levels of social disorganisation, in which the community is less able to assert control (Bursik, 1988). The motivation for the theory can be better understood through consideration of the time at which it was proposed. The theory was developed soon after World War I, which was a period of substantial change in the United States, including the development of cities such as Chicago. According to the theory, the rise in crime can be understood as an unexpected and unintended side effect of the economic developments that took place at the time. Amongst other things, these changes produced large-scale urban relocations of newcomers from suburban areas for a job which were accompanied by an excessive escalation in crime and disorder in the areas to which they migrated. Park, Burgess, and McKenzie (1984) sought to answer the question of how individuals were affected by the growth of the city and how the expansion of the city might lead to urban problems.

While social disorganisation theory looked for the answer of rising crime in terms of the process of urbanisation after the end of World War I, routine activity theory tried to explain increasing crime rates in terms of the social and economic changes that took place after World War II. The salient transformation/modification of people's daily lives after the wars inspired criminologists to consider consequences of the changes physically (i.e. residential settings and land use) and socially (i.e. population composition and employment status) on crime opportunities. During this period, Cohen and Felson aimed to understand steeply growing crime rates after the war, reconceptualising the root of crime. They applied economic and social factors on the crime rate which could not be fully explained by the U.S. Census Bureau data showing improved social economic conditions. They conducted an analysis on social changes in metropolitan areas such as the increasing married females in the workforce, unsupervised homes during days and social pattern changes by higher college entrance and found that these social changes are associated with crime rates rather than criminality (Felson and Cohen, 1980).

As discussed, a key concept of routine activity theory is that the structure of individuals' routines is influenced by societal norms. Spatial and temporal patterns of activity revolve around the home, employment, and recreation. Different places will represent different manifestations of populations undertaking different activities. An offender may, for example, encounter countless convergences that produce suitable crime opportunities if they engage in activities in socially disadvantaged areas. Although routines are a primary concept in explaining crime exposure situations of individuals (those routines which mean how, where, and when individuals move) are never explained without considering an individual's social position which naturally determines their routines.

Crime settings: who are the offenders?

In both theoretical perspectives described above, motivations to commit crime are ignored and appear to be in sharp contrast to other theories that focus on criminality. They both pay attention to how the environment and situation might shape the criminogenic character of an area, but ignore the reasons of criminal motivation. Routine activity theory substitutes the word motivation with inclination (Clarke and Felson, 1993) while social disorganisation theory uses the word 'dispositions' (Sutherland, Brunton-Smith, and Jackson, 2013). In common with routine activity theory, social disorganisation theory suggests that delinquency is more likely in areas where social control is weak and unlikely to be asserted. Cohen (1955) criticised social disorganisation theory in *Delinquency boys*, arguing that "It is wholly negative. It accounts for the presence of delinquency by the absence of effective constraints" (Cohen, 1955, p.33).

To summarise, both perspectives cautiously avoid explaining how crime motivation is created. They assume when a control level is low with either a low level of social and/or physical control, crime follows naturally. How the two theories differ is in terms of how they operationalise the concept of social control. In the case of routine activity theory, guardianship can be provided by anyone who (or anything that) is present and has the capacity to deter crime in some way. In the case of social disorganisation (see Sampson et al., 1997), the social fabric of the neighbourhood matters and influences the likelihood that residents will act collectively to deter crime.

Given the difficulties of measuring the presence of motivated offenders, most spatial studies have indirectly measured the likelihood of appearance of them. For example, indicators such as residing in crime hotspot areas, residing close to victims, neighbourhood social disorganised levels, and neighbourhood crime rates have been assumed as increasing the risk of potential offender emergence (Cohen and Cantor, 1981; Sampson and Wooldredge, 1987). Although most studies look for indicators of offenders in social structure contexts, some studies have studied offenders more directly within situations of crime. For example, Wortley (2001) proposes the role of immediate situational precipitators induced by the person-situation interaction such as crowding, being jostled and rude treatment (Wortley, 2001; 2013). He argues that those precipitators can provoke violent behaviours. Finkelhor and Asdigian (1996) also emphasise the importance of offender's emotional status. They found that the offender's anger or destructive impulses are aroused before the offenders decide to commit crime and this risk was significant in sexual assault crime among the young (Finkelhor and Asdigian, 1996). In *Warriors and Peacemakers: How Third Parties Shape Violence*, Cooney (1998) points out emotional status in criminogenic situations. Although Cooney (1998) mainly focuses on the role of a third party in violence crime, he strongly suggests violent crime situations are requisitely followed by emotional agitation such as swearing and cursing (Cooney, 1998). In short, despite the importance of directly measuring the presence of the offender in crime conditions, only a few studies include immediate situational indicators and others tend to use more general contextual indicators.

Crime settings: who are the victims?

One of the most widespread theoretical frameworks in victimisation research is routine activity theory (Hinderlang et al., 1978; McNeeley and Wilcox; 2015). The theory views the reason for victimisation as a function of differential routine behaviours of victim-prone individuals (Cohen and Felson, 1979). In the framework of the theory, operationalising the likelihood of victimisation can be divided into two concepts; an exposure level to crime-prone situations such as the accessibility and visibility of targets, and an attractiveness level in symbolic and economic values (Clarke and Felson, 1993, Madero-Hernandez and Fisher, 2012). In the case of social disorganisation, less attention is paid to the characteristics of victims. Instead the theory notes a potential victim as one who does not have enough supportive relationships with family and community members. In the theory, victims are described as individuals who reside in socially disorganised communities and lack financial and social networks (Sampson and Groves 1989; Bursik and Grasmik, 1993) and they eventually are exposed by existing in an uncooperative atmosphere (Tewksbury, Mustaine, and Covington, 2010).

A routine-based exposure level to criminogenic situations is measured by the likelihood of being victimised by lifestyles; especially consideration of where and how long individuals spend their times (Mustaine and Tewksbury, 1998). For example, the risk activity patterns, the number of nights spent outside, the number of time spent in public such as bars, parties and working have been tested to measure the level of exposure to crime situations (see Mieth et al., 1987; Kennedy and Forde, 1990; Miethe and Meier, 1990; Sampson and Lauritsen, 1990; Miethe and McDowall, 1993; Fisher et al., 1998; Mustaine and Tewksbury, 1998). With respect to target attractiveness, victimisation likelihood is estimated using households with a high income, jewellery worn in public, cash carried openly (Lynch, 1987; Miethe, Stafford and Long. 1987, Sampson and Wooldredge, 1987; Messner et al., 2007), housing value, socio economic status of households (Mieth and McDowall, 1993; Bowers, Johnson, and Pease, 2005; Zhang, Messner and Liu, 2007), and previous crime experience given that offenders have more information on prior targets (Wright and Decker, 1994; see Johnson and Bowers, 2004). Cohen's early work (see Cohen, 1981) with his colleagues used sociodemographic indicators to estimate population at risk of victimisation by levels of non-household activities such as marital status and occupation (Cohen and Cantor, 1981; Cohen, Kluegel and Land, 1981).

Similar to measurement of the concentration of offenders, the concentration of victims has mostly been measured indirectly in studies of social disorganisation. Most studies estimate the likelihood of being victimised by structural factors of neighbourhoods including per cent of black and average income (Fraser and Kilbride, 1980), perceived disorder (Melde et al, 2009), and collective efficacy (Fox, Lane, and Akers, 2010).

Crime settings: who are the guardians?

As mentioned above, control is a core concept in both theoretical perspectives, albeit it is conceptualised differently. The empirical research based on social disorganisation theory consistently indicates that a high level of informal social control is negatively associated with neighbourhood levels of crime rates through weakening of a level of informal surveillance, common governing rules of neighbourhoods, and direct interventions (see Greenberg et al., 1982; see also Bursik and Grasmick 1993; Bellair, 2000); Informal surveillance refers to casual but vigilant activity occurring at places (see Jacobs, 1961), common governing rules refers to widely accepted understanding among the members of neighbourhoods such as particularly unsafe areas in their neighbourhoods, and direct intervention denotes willingness to intervene in unusual activities, unacceptable behaviours by the juveniles and informing parents or guardians of misbehaviour (Bellair, 2000). Social disorganisation theory highlights an important role of social ties and social control in neighbourhood crime rates as a tool for crime prevention. Many studies that adopt this perspective, suggest that broken social control in neighbourhoods provides opportunities for offending. Bursik (1988) asserts that in neighbourhoods with strong commitment, crime is kept in check by the residents as a group standard, but where it is deficient, crime spirals out of control (Bursik, 1988). At an operationalisation level, social ties and control are largely measured by membership, influence, sharing of values, fulfilment of needs, shared emotional connections (Glynn, 1981; Chavis et al., 1986; Sampson and Groves, 1989; Perkins et al., 1990, Warner and Roundtree, 1977; Bellair, 2000; Markowitz et al., 2001; Hipp, 2007).

In routine activity theory, a guardian is anyone that has the capacity to control crime opportunity successfully and is anyone who can 'keep an eye on potential crime targets' discouraging crime (Reynald, 2009; 2010; 2011). Guardians exist in daily routine in the form of friends, relatives, neighbours, families, or the owner of the targeted property (Felson and Cohen, 1982; Clarke and Felson, 1993; Reynald, 2011; Eck, 2015). In previous research, guardianship has been categorised by its form such as social or physical, intensity, and levels considering both situational or contextual crime conditions. Some researchers examine social guardianship in terms of the number of adults living at home, networks among neighbours, participations in neighbourhood activities, or parental attachment (Mieth and McDowall, 1993; Tseloni et al. 2004; Schreck, Stweart, and Fisher, 2006; Zhang, Messner, and Liu, 2007). Others focus on physical guardianship with target hardening indicators including security alarm, door locks, carrying a weapon, owning a dog or undertaking self-protection training (Wilcox Roundtree, Land and Miethe., 1994; Fisher et al., 1998; Mustaine and Tewksbury, 1998; Outlaw, Ruback, and Britt, 2002). In conceptual explorations, Wilcox, Madensen, and Tillyer (2007) propose two different levels of guardianship; individual-level guardianship and aggregated-level guardianship (Wilcox, Madensen, and Tillyer, 2007) and Reynald (2009) argues guardianship is categorised in four groups by intensity– invisible, available, capable and intervening (Reynald, 2009). In the four stage model, available refers to where a guardian is present and visible, but may not be capable. Capable refers to where a guardian can be expected to fulfil a guardianshiop role in potential crime conditions but may not intervene in situations willingly. In the final stage, intervening is where a guardian is present and actively mediates in situations as a controller (Reynald, 2009).

In sum, social disorganisation theory suggests that guardianship is created through community structures, and argues that a community would regulate itself if it has collective action and a high level of civic engagement (Sampson and Groves, 1989; Carr, 2005) or intervention ability (Morenoff et al., 2001). Routine activity theory, on the other hand, focuses on either the immediate situation or accumulated situations underlining the appearance of capable guardians to deter crime. It views the absence of capable guardians as a crucial component of crime occurrence, and defines guardians as anything that can provide direct or indirect supervision in crime conditions. To summarise, although the two theories differ in terms of the processes and the conditions necessary for effective guardianship to occur, both see the protective action of residents as essential in understanding area-level crime rates.

2.4.2 *Limitations of the theories and big data*

Routine activity theory and social disorganisation theory have had a profound impact on the field of criminology and policy. Despite this, there remain unanswered questions, and critically, the data used to empirically test these theories is often imprecise and often represents only a proxy of the core constructs it seeks to represent. The criminal opportunity approach afforded by routine activity theory has been enthusiastically embraced by practitioners (for example in the form of situational crime prevention, see Felson and Clarke (1998) and below), but has also attracted criticism . In particular, limitations have been suggested with respect to theoretical indeterminacy and the use of broadly defined measures (Madero-Hernandez and Fisher, 2012). On the other hand, social disorganisation theory has been criticised regarding ambiguities and the complexities of the mechanisms proposed. Bursik (1988) notes that Shaw and McKay did not clarify the definition of social disorganisation, or the causal link between social disorganisation and neighbourhood crime rates, and eventually these limitations cause empirical deficiencies of the mechanism (Bursik, 1988). This thesis therefore focuses on four issues which include the conceptual clarity of the theories, and the measurement of key constructs – not properly measured hitherto - through the use of big data, such as geolocated Twitter messages.

First, neither theory considers the cultural contexts associated with places. Empirical studies have not sufficiently dealt with this issue as the characteristics of places associated with culture, context and code are not easily captured and computed. Consequently, the criticism on a lack of cultural context and its impact on crime rate have been raised with respect to both theories. Researchers have tried to measure a cultural code which could account for the paradox sometimes evident in social disorganisation; the neighbourhoods with a low crime rate but a high social disadvantage level. This cultural code would not be observable using demographic characteristics or social economic status. Considering routine activities, variables such as age, sex, race, income, and level of education are vital parameters that likely influence people's routine activities, and research has tended to rely on them to date. However, these are simple "proxy" variables that cannot provide accurate insight into the intricate and complex patterns of activity people engage in (Degarmo, 2011). Similarly, Garland (1999) argues that all situations are different since they are created in the cultural context of the place. As Anderson(1999) notes the place has its own culture and code, it cannot be easily measured in quantitative terms using traditional census data or simple interviews. All (non-cyber) crime is local to somewhere, and cultural variations will always exist in criminal situations (Felson, 1998; Garland, 1999). Thus, traditional forms of data are unlikely to capture the cultural dynamics of specific places at specific times. Big data might afford this opportunity.

Second, Cohen and Felson (1979) assume that criminals and non-criminals are "functionally dissimilar" but miss an explanation of the predispositions of the dissimilarity between them (Tillyer, 2010). Likewise, Shaw and McKay do not elucidate why only some people who live in the same or similarly structured neighbourhoods engage in crime. Routine activity theory neither clarifies the root to criminal motivation nor describes why criminal motivation varies in degree among individuals. This lack of conceptual clarity equally applies to the definition of victims. As explained by Clarke and Felson (1979), in many cases, victims are not clearly distinguished from criminals (Clarke and Felson, 1979). Scholars have not drawn clear boundaries between them and have not addressed the question 'are drug users or a drunken person a motivated offender (Sampson and Lauritsen, 1990) or an easy crime target (Spano and Nagy, 2005)?' The existing criminological literature suggests that victims and offenders share similar demographic characteristics, especially for violent crime (Gottfredson, 1981; Jensen and Brownfield, 1986; Daday et al., 2005; Martinez, Rosenfeld and Mares, 2008). As both offenders and victims reside in similar socially and physically structured places, it is perhaps not surprising that they have been found to have similar risky routine activities and situational characteristics (Sampson and Lauritsen, 1990; Singer, 1981. Daday et al., 2016). However, the differences between three crime elements, in particular offenders and victims, could be observed when they are analysed by big data presenting things that are usually unseen ostensibly such as emotions and personality.

Third, the theories also suffer from limitations associated with empirical measurement. In their original paper, Cohen and Felson (1979) put an emphasis on 'direct contact' in terms of the convergence of the three elements necessary for crime to occur. However, absent of direct measures of people's activities, routine activity patterns were estimated using indirect "proxy" measures. For instance, although most violent crimes are not instrumental and happen spontaneously (Mieth, Stafford, and Lond, 1897), in the previous measures this temporal crime atmosphere could not be captured the situation. The measurements are substituted by indirect ways which show higher likelihood of the commission of this type of crime (e.g. the presence of bars serving alcohol, see Graham and Homel (2008)). Additionally, a reliance on proxy and alternative measurements is problematic because the associated inaccuracy of the data weakens the extent to which the validity of the theories can be tested (Weisburd and Piquero, 2008) and crime specific dynamics can be captured (Madero-Hernandez and Fisher, 2012). To illustrate, in previous research concerned with routine activity theory, guardians and guardianship have not been measured directly. Although the formulation of the theory is noticeably simple, the research could rely on proxy and indirect measures of the variables, which represents an important methodological limitation of such research (Wooldredge, Cullen and Latessa, 1992). In the current study, using big data, guardianship is estimated more directly capturing the immediate moments of crime situations.

Following on from this, a fourth issue is the lack of temporal information concerning routine activities and the use of alternative and proxy measures which make difficult to develop the theories and to examine their predictive validity. Groff (2006) argues that although routine activity theory has a well-developed framework, its empirical validity is still not confirmed because of the difficulties associated with obtaining individual-level data. The analysis of Big Data offers considerable potential for the testing of criminological theories such as those discussed here. This is because it is often in a format which is both spatially and temporally specific. The data thus offer the possibility of measuring crime and its underlying situation in new ways and the discovery of activity patterns not previously possible.

These four gaps in the theoretical concept – 1) a lack of explanation of contextual variations in crime situation and 2) unclear clarification of characteristics of offenders and victims - and in empirical measurements – 3) missing temporal data in the previous empirical research and 4) indirect measurement of crime situation – can, it is argued here, be filled by the data which provides dynamic interactions of offenders, victims, and guardians and their patterns spatiotemporally. At the risk of repeating previous points, the big data to be used has countless layers of information including people's activities and will help to address questions more directly than has been possible in previous research.

2.5 THEORETICAL MODEL FOR THE THESIS

Although the conceptual framework for the thesis is built upon theoretical foundations of routine activity theory and social disorganisation theory, in the thesis, crime events will be mainly understood within an opportunity perspective, especially the routine activity framework. To be clear, the thesis principally views crime events as the result of the convergence of three crime elements' emergence, - routine activity theory - and the emergence patterns as the results of social and physical structure of place - social disorganisation theory (Figure 2).

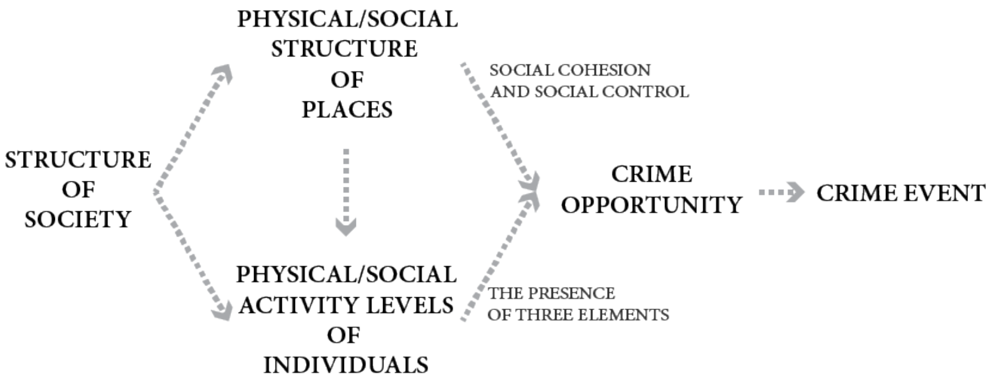


Figure 2: Theories synthesis model

More importantly, unlike previous research, the work reported in this thesis will pay particular attention to the immediate (and dynamic) conditions in which crimes occur. While research concerned with routine activities has mostly examined aggregate behaviours estimated using proxy variables, scholars have tried to employ the idea on an individual situation level. According to routine activity theory, a potential offender engages in criminal activity only when suitable targets come together with an absence of a guardian and, this turning moment is immediate. Therefore, the thesis concerns not only the characteristics of place with crime conducive conditions, but also the specific conditions of the turning moments and their patterns. While other theories target the desire to offend, routine activity theory addresses the possibility of offending taking place as a result of the dynamic interactions that take place between motivated offenders, suitable targets and capable guardians (Gottfredson, 1981). An offender alone is not an absolute condition for a crime event. Hence, the thesis will focus on criminals but pay the same attention to other players in the crime equation as they all create crime facilitative conditions together.

Undeniably, crime occurs at a specific moment that occurs during the course of day-to-day regular routine activities which themselves are framed by societal structure and norms, routine activities should thus be understood at both micro and macro levels. In other words, the kinds of situations an individual faces in their day-to-day routine activities are a barometer of the risk of crime to which individuals are exposed, and the kinds of situations that emerge in a neighbourhood are an indicator of the crime involvement rates of the neighbourhood (Wikström, 2009). The fusion of routine activity theory with social disorganisation theory provides a way to understand routine activities at an individual level and a group level.

3 BIG DATA AND CRIME

This chapter devotes considerable attention to introducing big data as an innovative source for crime modelling and prediction. In this chapter, the basics of big data, particularly that captured through social media platforms such as Twitter will be outlined. This will be followed by a discussion of the potential uses to which such data might be put in criminological enquiry, and a critical assessment of these forms of data.

3.1 WHAT IS BIG DATA

3.1.1 *Definition of big data*

In the documentary the human fact of big data, Michael Coren, a journalist, described the era of big data as "Every century, a new technology – steam power, electricity, atomic energy, or microprocessors – has swept away the old world with the vision for a new one. Today, we seem to be entering the era of Big Data". The term, big data, which delivers the most necessary information about itself by means of its name, has been a major key word in the technologies world in the most recent five years.

Since the term emerged in the 1980s, its definition has been modified and has evolved progressively with the exponential growth in interest in it (Gandomi and Haider, 2015). A dictionary definition of big data is "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges" (Oxford English Dictionary, 2016). In a more practical description, Gartner defined big data as "high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Gartner IT Glossary, 2016). Beyond the most authoritative definitions, in academia, big data is generally refers to massive amounts of structured and unstructured data that cannot be captured, processed, stored, searched, and analysed by traditional database tools (Kaisler, Armour, and Espinosa 2013; Sagirouglu and Sinanc, 2013).

Big data is growing everyday and exists everywhere in our daily lives. It comes in numerous forms such as social media posts, weather forecast information, mobile phone GPS signals, credit card or oyster card records and traces of peoples' activity on the internet. Because of its unique characteristics, scientists, but also practitioners, politicians and businesspersons, are facing challenges in finding a certain type of big data which could be applied on their own field for gaining better knowledge. While there are many types of big data available, social scientists are likely to learn most from the data coming with a rich set of human behaviour knowledge. Data with this quality also advances understanding of how people move and how they think because the knowledge offers qualitative insights into our daily lives.

3.1.2 *Characteristics of Big Data*

Because the volume, tempo and complexity of big data are massive, the patterns that exist within the data may not be identified easily. For this reason, the identification of potentially hidden patterns requires 'big data analytics' which also could be referred to as a revolutionary step forward from traditional data analysis (Sagirouglu and Sinanc, 2013). Regarding its size, speed and format, sometimes referred to as 3V, volume, velocity or variety (Doug Laney, 2011; Gerhardt, Griffin and Klemann, 2012; Schneider, 2012; Singh and Singla; 2015), the analysis of big data has required new scientific methods of capturing and analysing themes in the data.

Each characteristic of 3V describes one attribute of big data. First, volume refers to the amount of data. According to IBM, 2.5 exabytes, of data are created every day. An exabyte is equal to one billion gigabytes of data. Moreover, IBM estimate that 90 per cent of the data that exists in the world today was created in the last two-years. Velocity refers to the increasing speed at which data is created. Every single minute, 227,000 tweets are created, 216,000 Instagram posts are uploaded, 204,000,000 emails are sent, and 72 hours of video footage is uploaded to Youtube (V and C, 2015). Lastly, variety refers to the structure of the data. The data is categorised into three types of data; structured, semi-structured, and unstructured. Structured data include texts or numbers, while semi-structured data include text such as eXtensible Markup Language (XML), Application Programming Interface (API), Electronic Interchange Message (EDI) and so on. The majority of the data is generated in an unstructured format which includes documents, photos, videos and audio files or free texts such as microblogging messages which have most significant potential to be used in crime research as they are created with the natural speech comprised personality, identity, and emotion status of users.

3.1.3 *Big Data analytics*

Because big data does not fit into the traditional data paradigm, it is not analysed using traditional analytic methods. The core difference between big and traditional data (e.g. that collected from sample surveys), sometimes called small data, is that big data is captured, stored, and aggregated in different ways. These differences lead to a noticeable obstacle in the application of conventional methods of analysis for big data. In order to discover meaningful patterns in big data, previous methods of analysis cannot simply be applied to big data. For example, 90 per cent of big data - such as emails, pictures, videos, posts from social media - is unstructured. As such, it cannot be easily recorded in a spreadsheet, as would be the case for most traditional datasets. In addition, since the formats of data keep continuously changing and evolving on the top of unprecedented streaming speed and enormous volume, big data analytics have lagged behind, and still being developed to keep up the evolution until recently. At the beginning of the big data revolution, researchers were mainly interested in working out how to mine or manage the accumulated datasets, but they have quickly changed their concerns to how to advance data analysis tools and how to capture hidden patterns in big data efficiently. Currently, there are four common big data analytic techniques: text mining, opinion mining, social network analysis, and clustering analysis, and these analytics are mainly used to extract and discover concealed information in the data.

3.2 Big data hubris

Boasting it knows us better than we know ourselves, the related technologies are growing but, in reality, big data itself, and its analytics are not invincible. Isolating the signal from the noise in big data remains unsolved and, advances in analytic capability have not kept pace with the technologies of collecting and storing data. Sometimes big data overestimated results or suggested irrelevant results (i.e. cases from Google flu trend tracker, Facebook headlines or election polling of Eric Cantor, see Lazer et al., 2014). Simply put, the predictive capabilities of big data analytics are questionable and still in development in many fields. As the warning 'Big data hubris' says, the data should not be seen as a substitute for, but rather a supplement to, traditional data collection and analysis.

Because big data (like other forms of secondary data) were not initially designed for a specific research purpose, they ignore some foundational caveats of primary data collection methodology including measurement, validity and reliability, and dependencies (Lazer et al, 2014). In addition to issues associated with the data itself, in many cases, the data is simply analysed without a particular theoretical motivation. The data should be understood within the paradigm of social science and human behaviour-based approaches, as it is created by people who interact with given environments, and who are constantly influenced by social and cultural factors.

Taking a more empirical perspective, sampling issues in social media data, which is one of most prevalent forms of big data, should also be discussed. Nagler and Tucker (2015) discuss two sampling issues in social media big data; representativeness and generalisation (Nagler and Tucker, 2015). Representativeness issues arise when the sample from which big data are drawn does not match the population it is considered to represent. This occurs if the groups of people who use social media do not represent the whole population (Nagler and Tucker, 2015). Generalisation concerns the issue of whether the sample collected from different social media platform can be generalised to broader populations of interest. As the users of social media are concentrated in certain age groups or specific types of people who share the same interests, researchers should cautiously review characteristics of platforms which will be used to data collection considering research purposes.

Despite these noticeable caveats, the potential rewards and enormous scientific possibilities of using big data cannot be denied, since these datasets have the potential to capture interesting accounts of human behaviours that cannot easily be obtained in other ways. Analytic technologies are becoming stronger in various fields and the needs of big data are hugely increasing. While some areas of research are changing gradually in their capacity for big data analytics, others, such as medicine and economics, are radically developing and already have shown impressive results on how this influences changes in everyday life. For example, social media technologies has been used to monitor behavioural and psychological patterns and its influences on mental illness (Guntuku et al., 2017), HIV (Young, Rivers, and Lewis, 2014), or purchase intentions (Want, Yu and Wei, 2012). Big data cannot be merely defined as a large dataset. It contains qualitative, real-time, and spatial knowledge of human behaviours which could not easily be collected using more traditional methods. It gives us more sophisticated and dynamic information about people than demographic data alone, and provides more natural facts than survey data. For the reasons discussed above, while it is clear that big datasets cannot (presently at least) be used as a substitute for more traditional forms of primary data (collected for a particular purpose), there is little doubt that they can be used as a supplement to provide insights that would otherwise likely be impossible.

3.3 WHAT IS TWITTER?

3.3.1 *Facts about Twitter*

By human nature, people always want to be connected, network and communicate with other people in the society where they belong to. The numerous social networking platforms have been built upon this human nature and in 2018, around two billion users were on social networking sites (statistic.com, 2018). With traction among the users, the figure is continuously increasing and with high demand from people who want new and advanced ways of social networking, the usage of the platforms is becoming diverse. The scale and nature of social network interactions put the resulting data very much in the category of big data. According to Twitter statistics published in 2013, 18 per cent of US adults use the platform and 42 per cent of them are on it daily. This platform was mainly used by adults ages above 18, male, and urban residents (see Pew Research Center, 2013)

Each social networking platform specialises in different needs or is designed for different functions. Unlike other leading social networks platforms such as Facebook or Instgram, Twitter does not focus on interaction between friends and family using photos or videos. Twitter is all about rapid and instant communication with the world (statistic.com, 2018) and this platform highlights the function of broadcasting and conveying a message to others in a limited number of characters up to 140, called Tweet. This format makes the social networking simple and practical and this simplicity is the largest advantage of Twitter comparing with other platforms on the market (Raamkumar et al., 2018). It basically has a tweet and re-tweet structure that allows to people 'tweet' messages on their own bulletin board and share the thoughts with followers as well as with the followers of their followers. A broad spectrum of interests is broadcast in real time through the platform. Across all subjects and across all types of people, anyone or any sector can make a comment on political controversies, commercial items, their everyday life and so on.

3.3.2 *Benefits of Twitter*

Twitter might not be the most suitable social networking platform for collecting people's emotion or personality which is casually and naturally expressed in intimacy relationships. This is because the platform is not optimised for the purpose sharing daily activities and life with friends or family using photos or videos like Instagram. Since the platform was released in 2006, the users have kept developing its usage and defined how to use Twitter based on its characteristics. As a result, the platform ended up establishing; a platform for informative social networking and are been using for interacting and exchanging information with others. An analysis on Twitter traffic found that the messages such as text-based, list-based, how-to, breaking news, or social activities tends to get more retweets and the messages contains immediate and opinionative response to topics including news, politics, current events, and technology (Ahmed, Bath, and Demartini, 2017).

Although this social media could face the critics for not being used in intimate social networking such as Facebook or Instagram, the messages from it are still incredibly valuable as it elicits feedback with emotions on the original post. Twitter has become an important data source as it is freely and openly accessible via its Application Programming Interfaces (APIs), which consequently increases the value of the data as a research tool (Ahmed, 2015; Solymosi and Bowers, 2018). The data are very trendy, therefore, the messages contain the latest and more current subjects showing people's interests with an insight into their views, values, and thoughts on the topics.

Location tagging is the most valuable strength of Twitter to researchers and anyone else who use the data to understand the value of the data with geographic information. According to the Twitter policy 2013, users can choose to publish their locations with posts by selecting an option at the device or the website, and once the option is selected, real-time locations will be recorded/announced to every posts unless the users remove the options on their accounts [1]. Most importantly for crime research, Twitter has valuable metadata, locations from which Tweet was created (Ahmed, 2015). By achieving geospatial information, Twitter data can provide significant aids to crime patterns and prediction studies providing real-time and -place clues of crime situations (Solymosi and Bowers, 2018).

However, it is advisable that the likelihood of behavioural biases in the use of geotagging service with biases in the use of the platform, Twitter, as well as the use of online needs to be carefully considered in the context of analysis being performed. According to Sloan and

1 see the location information polity of Twitter: https:/twitter.com/en/privacy. With a user agreement, the post was tagged with a precise location (i.e.latitude and longitude).

Morgan (2015), the tendency to enable Twitter location services varies by demographic characteristics such as gender, age and languages. In particular, unlike differences in gender and age show significant but minor gaps between groups, the study found users who tweet in Russian less likely to use geotagging services than users who use Portuguese and Indonesian (Sloan and Morgan, 2015). As they pointed out, demographic differences may be small, it is important to acknowledge other factors such as socioeconomic characteristics of individuals would generate more significant behaviour differences in enabling geotagging services. To conclude, it should be clearly noted that there, may be tolerable, are limitations in terms of range of applicability.

3.4 BIG DATA IN CRIME SCIENCE

Although "Measurement is the basis of all science" (Ferraro and LaGrange 1987, p.70), criminologists face inherent difficulties in measuring the phenomenon which is crime (Osgood, McMorris, and Potenza, 2002; Sullivan and McGloin, 2014). The central theories in criminology such as deterrence and rational choice theory (Cornish and Clarke, 1987), social learning theory (Akers, 1973), social bonding and control theories (Hirschi, 1969; Agnew, 1985), labelling theory (Becker, 1963), social disorganisation theory (Shaw and McKay, 1942), and anomie and strain theory (Durkheim, 1951; Merton, 1938; Messner and Rosenfeld, 1994) are abstract (Sullivan and McGloin, 2014) and tend to have elusive concepts (Farnworth, Thornberry, and Krohn., 1994). Because latent constructs and core variables in criminology have been observed by both alternative and indirect ways (Sullivan McGloin, 2014), criminologists have to deal with the extra challenges in identifying and capturing mechanisms initially discussed in the theories appropriately.

While the previous measurement methods are insufficient to address the research questions being asked (Bernard, 1987), with innovative measurement techniques in collecting and computing datasets recently developed, crime scientists can explore and understand crime in new ways. Because this change provides abundant opportunities to develop and expand criminology, reviewing and applying newly available datasets are incredibly important. As one example of the use of innovative data in crime studies, this thesis advocates the use of big data - especially georeferenced social media data - as a source of measuring qualitative and personal level information.

While big data shows great promise in criminology, its use has been under-explored. Previous studies have shown that criminological theories have evident gaps in both conceptualisation and the measurement of key concepts, which in turn can challenge the validity of those theories. Weisburd and Piquero (2008) reviewed the explanatory powers of core criminogical theories using 169 empirical papers published in Criminology from 1968 to 2005. According to them, social disorganisation theories yield an average R^2 of 0.444 and routine activity theories generates 0.464 comparing with total average R^2 including all theories of 0.389 (Weisburd and Piquero, 2008). They argue that average of R^2 values are higher in crime-oriented theories such as social disorganisation theory or routine activity theory than individual-oriented theory such as self-control theory or life-course theory and, that the overall explained level is constantly very low over time having only 10 to 20 per cent explanatory power (Weisburd and Piquero, 2008). This low explanatory power of theories could originate from inadequate theory including the articulation of causal mechanisms, poor prediction ability, or inadequate explanation of the crime process. More importantly, these crucial limitations could result from the poor measurement of conceptual constructs. One reason for this is that there are inherent difficulties in measuring many aspects of human behaviour, and consequently approaches to data collection often involve limited and indirect approaches that may poorly reflect the constructs they seek to estimate. Collectively, measurement issues have impeded progress in criminology and hindered understanding of if and how theoretical mechanisms work in practice.

There is therefore much scope for the use of big data in criminology, particularly with respect to understanding spatial and temporal patterns of crime. There is a role that each of the 3Vs can play in advancing such understanding. Firstly, overwhelmingly massive datasets from a volume perspective could be farmed for large-scale correlations of the factors that are related to crime patterns. Of course, it would be important to recognise that data mining alone would not reveal much about causality or the mechanisms through which crime patterns might form. In terms of velocity, this has the potential to facilitate the dynamic detection and understanding of movement patterns and how these might influence criminogenic opportunities both spatially and temporally. Data variety might help in capturing the precise context of criminal opportunities and their surroundings and is a desirable research pursuit in place-based crime studies. When this is recorded, big data offers the real time records of people not only in terms of their location but also their personal information, giving crucial insight into those present in criminal situations. The data are not flawless but when harnessed appropriately may provide local, timely and qualitative information about crime situations.

Historically, the role of the police has often been viewed as solving crime easily and quickly. Nowadays, the public increasingly expect the police to prevent crime before it happens (Dai and Jiang, 2016). The rise of big data opens up possibilities for the police to identify threats to society in advance by adding an extra layer of knowledge. The big data revolution is creating new methods of collecting, analysing, and translating data, which the police should embrace to help inform the ways in which they target limited resources and ultimately, to fight crime.

3.5 BIG DATA IN THE CURRENT STUDY

As Ginni Rometty, IBM chief executive officer said, big data will be the next 'natural resources' for the industrial revolution, there has been a substantial amount of effort in academia to understand the chemical components of the data and to utilise the resource as a proper tool and make it useful (cited from Ginni Rometty's presentation at Council on Foreign Relations in 2014). Recently, a substantial amount of research has been focused on how to measure online-traits from social media platforms in particular Facebook and Twitter.

Note: each empirical chapter discusses relevant concepts and works in more details

Population movement patterns

The research on crime patterns and mobile populations has been conducted in recent years by Malleson and Andresen (see Malleson and Andresen, 2015a; 2015b; 2016). Malleson and Andresen (2015b) measured an ambient population using geolocated Twitter in Leeds, UK with two million Tweets collected from June 2011 to April 2013. They examined the relationship between the Twitter population and crime using spatial scan statistic methods, and found that the ambient population patterns are not directly related to crime volumes (Malleson and Andresen, 2015b). In their 2016 research in London using local spatial statistics (Getis-Ord), they found the workday census data is more in line with the ambient population as measured by Twitter. However, when targeting only crime concentrated areas, a statistically significant relationship between crime and mobile population was not observed (Malleson and Andresen, 2016). Although Malleson and Andresen's (2015a; 2015b; 2016) research could not find robust correlations between mobile population and crime at any place and any time, they found that when these data are applied on carefully selected locations and time frames considering characteristics of crime patterns of the target areas, they indicate more interesting and significant relationships.

Linguistic patterns

Previous research has shown that users leave personality and social identity related "behavioral residue" in their virtual environments (Marwick and Boyd, 2010; Thelwall and Wilkinson, 2010; Back et al., 2010; Yee et al., 2011) and that social identity and personality are expressed in microblogging activities, including texts on Facebook and Twitter, since they are based on daily conversations (Mehl et al., 2006; Back et al., 2010; Yee et al., 2011; Qiu et al., 2012). Meanwhile, research on sociolinguistics has shown that there is an association between self-reported personality traits and linguistic patterns in microblogs (Pennebaker and King, 1999; Qui et al., 2012) and that a person's language used in SNS uncovers their social identities, revealing the communities or groups they belong to (Nguyen et al. 2013; Ritter et al., 2013). Some studies assessed the degree to which demographic characteristics of users from social media posts such as age, gender, income and education can be inferred using Latent Semantic Analysis (LSA). Bokanyi et al. (2016) explored the possibility of inferring the demographic information from social media using pattern analysis. Using LSA methodology on 335 million geotagged Tweets in the U.S., they found that users' word use including slang use, urbanisation, travel, religion and ethnicity are highly correlated with demographic information from census data (Bokanyi et al., 2016). This shows the potential of Twitter in revealing information concerning their users.

Personality patterns

A variety of approaches has been recently proposed to infer a user's personality from social media platforms such as Facebook or Twitter. In many studies, Big Five model (openness, conscientiousness, extroversion, agreeableness, and emotional stability) is a widely accepted framework used to compare between personality in the real world and personality in social network environments. Farnadi et al. (2013) found that females tend to be characterised through openness and extroversion in Facebook use patterns (Farnadi et al., 2013), and Hughes et al. (2012) found people who use Twitter actively show significant correlations with a high level of openness and a low degree of neuroticism. Hughes et al. (2012) also argue Twitter is not used for a tool to mitigate loneliness not like Facebook, and mainly used by people who want to enjoy socialising (Hughes et al, 2012).

Cohesion patterns

Based on the assumption that people use online social media platforms to communicate with others or broadcast information (Preotiuc Pietro et al, 2015), social connectivity between users and others/followers has been studied to understand underlying similarities and differences between online and real-life social networks (Salvatore et al., 2011). In Facebook research, the number of linked friends (Lampe, Ellison and Steinfeld, 2007; Tong et al., 2008), time spent on social media (Burke, Marlow Lento, 2010; Hughes and Batey, 2012), and a number of posting messages or photos (Pempek, Yermolayeva, Calvert, 2009) have been commonly used to measure a degree of social network relations. In many studies, strong ties in online environments are associated with greater feelings of social cohesion and bonding, but do not directly related to actual network size in real life (Burke, Marlow Lento, 2010).

With the evolution of computer technologies, more in-depth information about populations can now be collected, numeralised, stored, possessed, and analysed. This new resource fits the needs of crime analysis and offers great promise as it provides information on the population that traditional sociodemographic and population data could not show (Solymosi and Bowers, 2017).

The aim of this research is to explore the potential utility of social media data and suggest ways to use it to uncover useful details concerning crime situations. Mobile population patterns change dramatically in urban areas spatiotemporally and this change creates criminal situations through the convergence in space and time of offender and victim populations in the absence of guardians. Consequently, estimating the characteristics of population in a way that varies over time and by location and is crucial in advancing our understanding of criminal environment (Solymosi and Bowers, 2017). In the following chapters, an estimation of the characteristics of dynamic population such as locations, feelings, and types of the population are explored, all of which could offer insight into underlying causal mechanisms and interesting features of criminal situations.

4 METHODS

This chapter discusses the research settings shared by all three of the empirical studies that follow. The chapter begins with an overview of the study area selected. In what follows, the data sources used including crime incidents and georeferenced posts from social media are described. Discussion then moves to methods for aggregating data to appropriate spatial and temporal units in preparation for the various analyses. Note that the more specific modelling strategies are explained in depth within each empirical chapter.

4.1 STUDY AREA

This research was conducted using data for New York City USA- technically a borough of Manhattan only- where the area has unique characteristics of population in contrast to other boroughs of the city. Most importantly to the research context, this borough has a substantial amount of mobile and dynamic population which cannot be measured using traditional population census data.

The research area is famous as a major tourist city. According to New York City visitation statistics (NYC Company, 2017), the number of visitors has been increasing every year. In 2013 (the current research employed the datasets collected from 2013), 54.3 million people visited the city, combining 42.8 million visitors from domestic locations and 11.5 million international visitors, and this increased to 65.1 million in 2018 (NYC Company, 2018). The number of the visitors are around three times (2.95 times based on the population of the city in 2013) greater than the total population of the city and are close to 26 times the population of Manhattan only (the population of Manhattan; 1.631 million in 2013 (The U.S. Census Bureau, 2018)).

In addition to visitors, the city also has a substantial subset of people who commute into the city centre every day. Manhattan serves approximately 2.3 million commuting workers and students on weekdays and the volume of commuting population for daily workforce consitutes 52 percent of the daytime population (Moss and Qing, 2012). These statistics indicate that the city has a substantial amount of population flow caused by the non-residential population, both in-flow and out-flow, therefore the official census data from the U.S. Bureau of Census can be seen as inadequate in accurately estimating the mobility of the population.

In criminology, the volume and density of population has been considered as a basic requirement in the analysis of crime events/opportunities (see Chamlin and Cochran, 2004) and the need for alternative data providing a shift of real-time population in lieu of the traditional census data has been raised (Malleson and Andresen, 2015). The NY city setting has the suitable conditions for the current research as it has a significant contrast in the volume of population flow at different times and the benefit of new data should be particularly evident in providing an alternative population layer.

4.2 DATA

The current research is undertaken using two primary data drawn from two difference sources; social media posts collected from Twitter 2013 calendar year and the number of crimes reported by the New York Police Department (NYPD) in 2013.

4.2.1 *Dynamic population data*

This research uses geotagged Twitter data which was collected from January 1st to December 31st of 2013 in New York City including all five boroughs; Manhattan, Staten Island, the Bronx, Brooklyn, and Queens, which matches with the period that the crime data collected covers also [1].

Twitter is one of the most popular social networking and microblogging platforms. It allows people to broadcast short messages or 'tweets' to 'followers' to whom they are connected on the network (Marsick

1 The data was obtained from the Uncertainty of Identity project at UCL. In practice, considering the volume of the data and the collecting methods, it is not possible for an individual to collect data covering a whole year. Twitter Streaming API only possible to collect real-time data and historical data should be collected using users' identification. The Streaming API allows collecting a one per cent of tweets produced at a given time (see Leetaru et al, 2013; Morstatter et al, 2013; Leak, 2017)

boyd, 2010; Jin, 2013). In its early days and during the development of Twitter, the team had originally designed Twitter as an SMS mobile phone-based communication platform. Having learned of suitable mobile communication tools, and having been successful at creating a simple interface, Twitter became one of the biggest interpersonal communications platforms.

Compared to other social media platforms, Twitter has three unique characteristics; mobility, georeferencing, and censorship. In the U.K., 80 per cent of Tweets sent by 15 million active Twitter users (Curtis, 2013) are created via mobile devices (Warren and Knight, 2016). Twitter automatically records the location information associated with Tweets for those users who agree to this upon registration. Although Leetaru et al. (2013) found only 1.6 per cent of tweets have precise geo-codes, they note that the amount of geo-coded tweets remains high even at this level of representation, which means that it is enough to conduct a meaningful analysis on them (Leetaru et al, 2013). Lastly, Twitter has no censorship so it has been used for crime activities such as advertisements for the porn industry or recruitments for new members by terrorist groups. Downloaded Tweets from the Twitter server have the information of user identifications, message content, geospatial coordinates and times.

After eliminating accounts not representing individual such as bots and spammers, 519,561 messages (out of 1,048,864 messages collected from the total five boroughs, 49.54 per cent of messages were created in Manhattan) were utilised in this research.

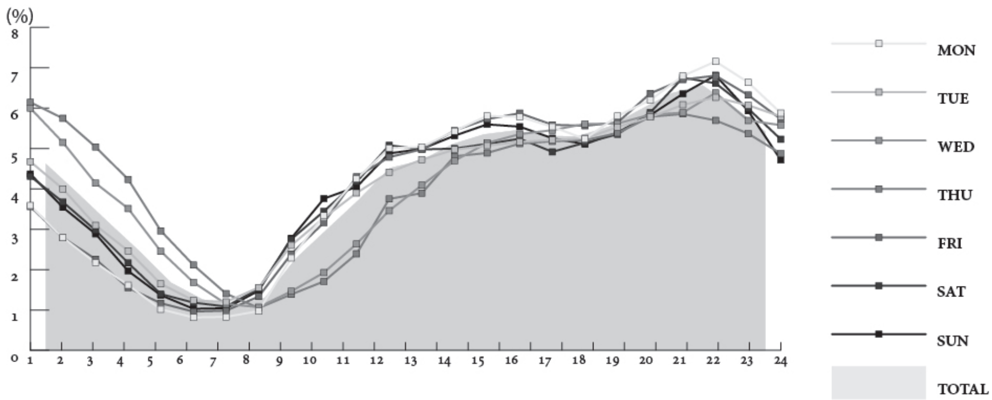
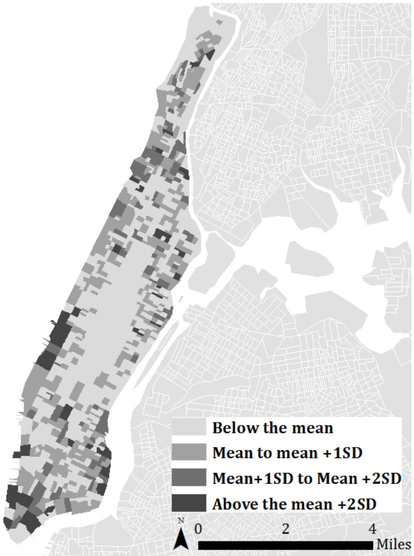
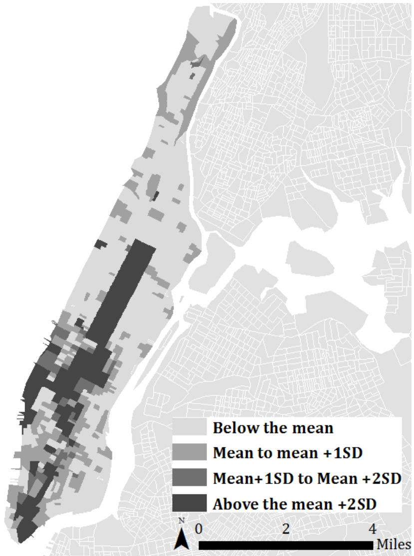


Figure 3: Temporal trends of dynamic population | percentages

Regarding temporal patterns in messaging levels, it can be seen from Table 1 that there is a reasonably equal distribution of tweets across days of the week. In Monday to Sunday time frame, there was no significant difference in volumes observed except for Thursday that generated less messages (60,003 posts, mean=74,223, SD=7,241). Unlike the temporal trends across days of week, the amount of geotagged messages varied very noticeably by time of the day (Figure 3). The temporal patterns indicated that 50 per cent of posts were created during the day (8am to 6pm) and the messages were most likely observed from 7pm to 11pm with this time frame contributing 30.43 per cent of the total messages collected. The off-peak was found from 5am to 9am and only 7.6 per cent of messages were created with this time period. Figure 3 also demonstrates that there is some variation in temporal trends by day of the week, but that in general the patterns are fairly similar.



(a) Census population



(b) Dynamic population

Figure 4: Maps of census and dynamic population

Figure 4 shows the spatial patterns of the census and dynamic populations. While the census population (left) was highly concentrated in the edge of the city, the spatial distribution of the population measured from social media (right) indicated that the population volume was much higher at the city centre (the lower part of a large unit (Central Park) in the middle) that has more attraction for tourists or the shopping/business district and workplaces (see Appendix (p.146) for a daytime population map using a volume of New York City transportation users created by Justin Fung (2018). According to the visual assessment on two daytime maps with different sources, the maps show very similar patterns of the dynamic population distribution)

Table 1: Temporal pattern of the texts by 24 hours timeframe

	Timeframe	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Total
1	0:00-0:59	2,778	3,864	4,346	3,687	2,560	3,343	3,394	23,972
2	1:00-1:59	2,163	3,305	3,730	3,449	2,008	2,847	2,751	20,253
3	2:00-2:59	1,681	2,560	2,999	3,024	1,624	2,306	2,242	16,436

4	3:00-3:59	1,251	2,038	2,545	2,541	1,114	1,679	1,527	12,695
5	4:00-4:59	788	1,362	1,776	1,777	840	1,081	1,061	8,685
6	5:00-5:59	631	1,023	1,220	1,268	690	917	800	6,549
7	6:00-6:59	643	991	832	843	708	842	811	5,670
8	7:00-7:59	753	1,286	775	630	958	1,181	1,148	6,731
9	8:00-8:59	1,776	2,151	1,059	832	1,717	2,134	2,158	11,827
10	9:00-9:59	2,582	2,713	1,396	1,025	2,276	2,671	2,920	15,583
11	10:00-10:59	3,289	3,223	1,910	1,435	3,090	3,237	3,146	19,330
12	11:00-11:59	3,878	3,647	2,507	2,251	3,441	3,936	3,790	23,450
13	12:00-12:59	3,905	3,903	2,962	2,337	3,580	3,857	3,876	24,420
14	13:00-13:59	4,207	4,108	3,401	2,885	3,914	3,873	4,130	26,518
15	14:00-14:59	4,499	4,189	3,722	2,934	4,118	3,963	4,348	27,773
16	15:00-15:59	4,481	4,266	3,879	3,073	4,223	4,063	4,304	28,289
17	16:00-16:59	4,291	4,307	3,950	3,107	4,017	3,811	4,081	27,564
18	17:00-17:59	4,055	4,339	4,062	3,124	3,999	3,986	3,967	27,532
19	18:00-18:59	4,498	4,569	4,069	3,230	4,059	4,148	4,194	28,767
20	19:00-19:59	4,802	4,773	4,230	3,473	4,576	4,565	4,526	30,945
21	20:00-20:59	5,265	5,026	4,264	3,513	4,822	5,234	4,937	33,061
22	21:00-21:59	5,543	5,178	4,619	3,418	4,892	5,128	5,292	34,070
23	22:00-22:59	5,143	5,017	4,124	3,222	4,546	4,641	4,608	31,301
24	23:00-23:59	4,553	4,787	4,040	2,925	4,119	4,052	3,664	28,140
	Total	77,455	82,625	72,417	60,003	71,891	77,495	77,675	519,561

4.2.2 *Crime data*

Types of crime

In total, 114,016 crimes collected from the target area which are suitable for the research purpose were selected considering the characteristics and mechanism of crime. In the first step of selecting process, crimes which are less likely related to/caused by situational or environmental factors were removed; crimes such as bribery, child abandonment, credit card fraud, forgery, gambling, and tampering. In the next step, crimes were categorised by types of targets; property, violent, and drug crime, and then finally sub-categorised into five types of crimes; property, harassment, assault, robbery, and drug crime (Figure 5).

Because robbery is defined as theft by force it means it has a character of both property crime and violent crime (Cook, 1987; Monk, Heinonen, and Eck, 2010), robbery crime was categorised separately. Additionally, violent crime were divided into harassment and assault crime considering intensity of crime i.e. whether it included physical actions or not. Collectively, 74,529 crimes were processed and categorised into property, harassment, assault, robbery, and drug crime. Burglary and theft were not separately considered because the differences between the two types of crime were not accurately distinguishable in the data. For example, dwelling burglary refers

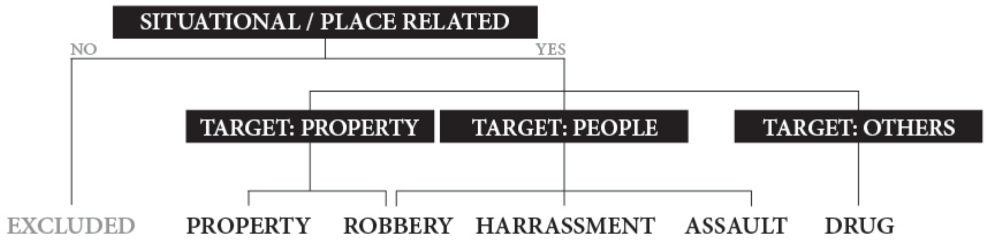


Figure 5: Crime classification trees

to breaking and entering a residence for the intention to steal but dwelling theft is committed by people who have an authority to enter the house. As the police recorded data does not define the differences clearly, they are combined into an overall property crime category.

Types of crime locations

To examine the specific characteristics of the places at which crime events occur, it is necessary to consider the situational patterns of the immediate environments in which crimes take place (Newton and Felson, 2015)[2]. The types of crime locations were categorised by environmental characteristics which could influence routine activities such as levels of guardianship. Influenced by Newman (1973)'s defensible space, the types of place were divided into four categories - private, semiprivate, semipublic, and public[3].

Private places include all types of residential property such as apartments, houses, and public housing. The place is shared by people who have intimate relationships and if there are any unauthorised visitors, it will be easily noticeable. The temporal patterns of the place would vary by their occupations, but in general, the place is more likely to be vacant during the daytime and occupied at night. In terms of property crime, except for crime committed by the member of the household, the place would keep high levels of surveillance 24 hours from visitors of the outside. For violent crime, it is more likely to occur between close relationships, therefore, criminal situations would be clearly different compared with violent crimes at public places, which are more likely to commit targeting random people on the streets.

2 The police recorded data which is utilised in this research contains the information on where each crime occurs such as 'inside of store', 'in the streets', or 'at the station'. Based on the descriptions, crime locations is categorised into 4 categories; private, semi-private, semi-public, and public.

3 Although Newman used the concept to explain residential environments, this research adopted the concept at more expanded level. Newman noted that public spaces can be characterised as anonymous which eventually lead to low levels of ownership and surveillance (Newman, 1973).

Semiprivate places comprise commercial buildings, shops, offices, schools, hospitals, restaurants and so on, and the place will have people (place managers and visitors) when they are open. This type of place has a clear schedule (temporal patterns) compared with other types of places. In this research, semipublic places are defined as the place that opens to the public and has a low level of guardianship. This place includes airport terminals, and public transportation such as inside of bus and subways. Public places consist of the street, park, playground, and any open areas to the public. The characteristics of people who occupy the public space can be defined in one way; passer-by. They may provide informal control (informal control, low responsibility, low wiliness to intervene) but the quality of surveillance provided by them could be less effective compared with people who have responsibility or ownership.

Each location type has a different physical and social setting in supervision. For property crime at private or semiprivate places, the capable guardians are responsible for protecting their property and belongings and would have the highest willingness to take protective action in a hypothetical explanation, thus, criminals are more likely to do their jobs at the highest level from entering the place through the gate to selecting a suitable item to take. Public places have similar temporal patterns in a flow of population with semiprivate places (busy during the day), a level of guardianship would be completely different. Both semipublic and public places have a high level of public accessibility but semipublic places would have some guardianship such as CCTV or ticket sellers unlike public places that rely on informal guardians in the places.

The maps show crime is distributed differently by the types of locations even the same crime type (Figure 6). Crimes at public places are concentrated at the centre of the research area (near/lower the central park that has a larger size spatial unit than all other units) and crimes at private places are more observed from the edge of the city[4].

According to Tables 2 and 3, the study area had property crime at semiprivate places most and had more crimes at public places than other three types of places. Compared with other four boroughs which have more residential areas, crimes at places that allow public access were more likely to be observed and crimes at private places were less occurred.

Table 2: Descriptive Statistics of crime categories

		Manhattan		Others	
Crime type	Place type	Frequency	Percent	Frequency	Percent
Property	Private	5,541	7.34	25,563	11.08
	semiprivate	26,003	34.46	30,091	13.05

4 Maps of property crime included in this chapter as an example (Figure 6. See Appendix (p.138) for the maps of other 4 crime types

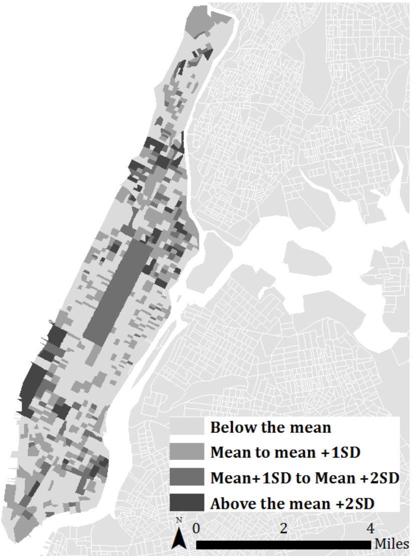
	semipublic	1,715	2.27	4,229	1.83
	public	8,823	11.69	32,082	13.91
Harassment	Private	5,481	7.26	27,956	12.12
	semiprivate	2,439	3.23	6,274	2.72
	semipublic	365	0.48	867	0.38
	public	2,476	3.28	9,085	3.94
Assault	Private	5,382	7.13	33,799	14.66
	semiprivate	2,354	3.12	6,262	2.72
	semipublic	548	0.73	1,138	0.49
	public	4,453	5.9	16,685	7.23
Robbery	Private	737	0.98	2,992	1.30
	semiprivate	636	0.84	1,828	0.79
	semipublic	186	0.25	709	0.31
	public	1,658	2.2	9,649	4.18
Drug	Private	1,822	2.41	6,170	2.68
	semiprivate	364	0.48	742	0.32
	semipublic	325	0.43	451	0.20
	public	4,141	5.49	14,059	6.10
Total		74,529	100	230,621	100

Table 3: Descriptive Statistics of place types

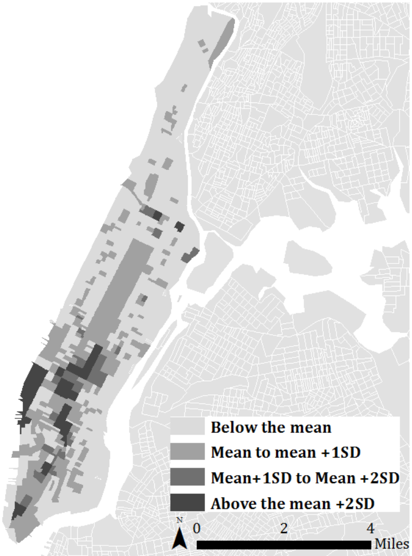
	Property	Harassment	Assault	Robbery	Drug
Private	5,541	5,481	5,382	737	1,822
	(13.17)	(50.98)	(42.25)	(22.91)	(46.42)
Semiprivate	26,003	2,439	2,354	636	364
	(61.79)	(22.68)	(18.48)	(19.77)	(9.27)
Semipublic	1,715	365	548	186	325
	(4.08)	(3.31)	(4.30)	(5.78)	(8.28)
Public	8,823	2,476	4,453	1658	1,414
	(20.97)	(23.03)	(34.96)	(51.54)	(36.03)
Total	42,084	10,752	12,737	3,217	3,925
	(100%)	(100%)	(100%)	(100%)	(100%)

4.3 RESEARCH UNITS

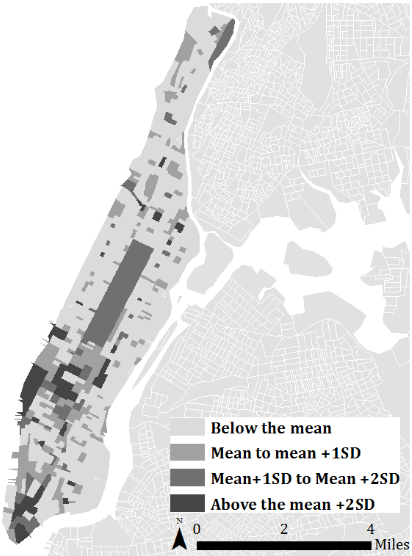
As this thesis primarily focuses on crime situations, a micro level resolution will be used for both the spatial and temporal dimensions.



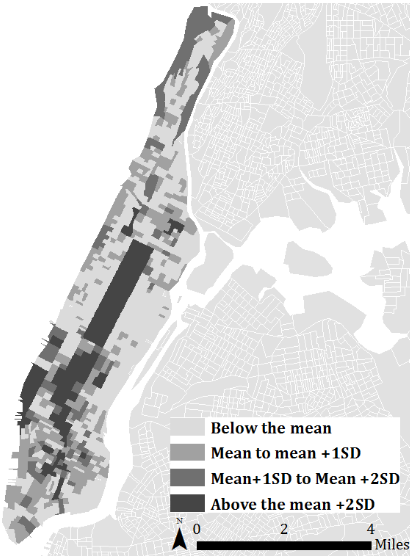
(a) Private places



(b) Semiprivate places



(c) Semipublic places



(d) Public places

Figure 6: Maps of property crime

Although criminology traditionally has focused on macro spatial units, recent studies demonstrate the potential theoretical and empirical advantages of conducting research at micro crime places (Eck and Weisburd, 1995; Sherman, 1995; Taylor, 1997; Weisburd, 2002; Weisburd et al.,2004). Weisburd et al. (2004) emphasised the value of conducting empirical studies of criminology at a micro spatial level because crime is considerably concentrated at micro places. They argue that macro crime hot spots such as socially disadvantaged communities have several crime micro hot spots (such as a street or a spot) within them which means crime is not evenly distributed in disorganised macro units so and hence the latter are inappropriate units at which to measure crime concentration (Weisburd et al., 2004).

4.3.1 *Spatial units*

With the trend moving to micro-space scale, recently several crime patterns research exercises have used street segments but it should be noted that census units would be more useful when estimating neighbourhood effects (see Hipp, 2007). 'Census block groups' is one of the standard units (blocks, block groups, and tracts) for the census survey, which have been most commonly used in previous social science research and each unit is accepted as an area that shares similar demographic and socioeconomic characteristics (Hipp, 2007). Based on the finding from previous studies, the first empirical study reported in this thesis, which aims to compare the census population with dynamic population, employs block groups as the unit of analysis and the second and third empirical studies - which are designed to understand micro-situations of crime - use a micro-level unit: street segments.

Block groups are the second smallest unit of census geography and have been employed in studies on the relationship between socioeconomic status of neighbourhoods and crime (Sampson and Raudenbush, 2004; Hipp, 2007). The research area has 1,071 units and contains between 0 and 8,000 people and the average population is 1,452 (Table 5). The city has 1,091 census blocks groups in the research target area and 1,077 units are selected for the study. The units that have some special facilities (i.e. correction facilities) or are isolated from the main target area and contain no data (census/dynamic population and crime) have been removed to reduce any errors. Whilst census blocks are useful for examining socio-economic influences, in practice, street segments have been used as the basic units for policing, therefore, they have been used in recent crime pattern research substantially (Weisburd et al.,2004; Weisburd, 2018). For the remaining two empirical studies - less concerned with area level demographic trends, 11,836 street segments were employed (the average of segment length=315 feet (approximately 100 meter) (Table 6). As shown in Table 4, among total street segments in the city (N=11,836), 21.05 per cent of streets (N=2,492) does not have any dynamic population values.

Table 4: Descriptive Statistics of dynamic population by spatial units

Units	Mean	Std.Dev.	Min	Max
Block groups				
(N=1,077)	442.25	757.83	3	4,048
Street segments				
(N=11,836)	43.23	121.77	0	2.471

(N=9,344)*Streets without any value	54.77	134.73	1	2,471

Table 5: Descriptive Statistics of crime categories by block groups (N=1,077)

Variable	Mean	Std.Dev.	Min	Max
Property				
Private	5.13	4.44	0	41
Semiprivate	23.83	83.97	0	2,090
Semipublic	1.59	5.10	0	97
Public	8.19	10.77	0	164
Harassment				
Private	5.05	6.91	0	88
Semiprivate	2.25	3.80	0	32
Semipublic	0.34	1.18	0	12
Public	2.30	3.41	0	36
Assault				
Private	4.96	8.18	0	104
Semiprivate	2.18	4.24	0	45
Semipublic	0.51	2.32	0	51
Public	4.13	5.75	0	66
Robbery				
Private	0.68	1.49	0	21
Semiprivate	0.59	1.31	0	17
Semipublic	0.17	0.64	0	7
Public	1.54	2.27	0	26
Drug				
Private	1.68	5.41	0	50
Semiprivate	0.34	3.17	0	98
Semipublic	0.30	1.70	0	30
Public	3.83	9.67	0	134

Table 6: Descriptive Statistics of crime categories by street segments (N=11,836)

Variable	Mean	Std.Dev.	Min	Max

Property				
Private	0.46	1.27	0	24
Semiprivate	2.16	18.23	0	1,599
Semipublic	0.14	1.32	0	91
Public	0.74	1.59	0	24
Harassment				
Private	0.46	1.69	0	49
Semiprivate	0.20	0.76	0	16
Semipublic	0.03	0.32	0	12
Public	0.21	0.66	0	14
Assault				
Private	0.45	1.94	0	69
Semiprivate	0.20	0.89	0	31
Semipublic	0.05	0.68	0	51
Public	0.37	1.06	0	21
Robbery				
Private	0.06	0.39	0	14
Semiprivate	0.05	0.29	0	6
Semipublic	0.02	0.19	0	6
Public	0.14	0.47	0	8
Drug				
Private	0.15	1.31	0	47
Semiprivate	0.03	0.93	0	96
Semipublic	0.03	0.51	0	30
Public	0.35	1.85	0	79

4.3.2 *Temporal units*

As Felson and Poulsen (2003) said "Crime varies greatly more by hour of day than by any other predictor we know" (Felson and Poulsen, 2003, p.595). The temporal characteristics of crime have been treated as a crucial aspect to understand crime behaviours. Previous research has identified that there are differences in temporal patterns of crime by seasons, bank holidays, pay-days, school holidays, day-of- week, and time-of-day and that distinct crime types are uniquely distributed over time (Tompson and Bowers, 2013; Boldt and Born, 2016; Towers et al. 2016). In general, temporal crime patterns have a close relationship with the use rhythm/cycle of places. For example, working hours and non-working hours, or school days and non-school days generate completely different situations at these places. During school days, people would naturally wake-up earlier than non-school days (activity time) and movement patterns (activity location) also would be distinct as well.

To decide upon a temporal unit of analysis, the temporal characteristics of crime such as trend and periodicity have been considered. To produce precise analysis on each crime type, as the true extent and timings of events are sometimes buried in overall patterns, the temporal unit was disaggregated in one-hour time slots for initial exploration.

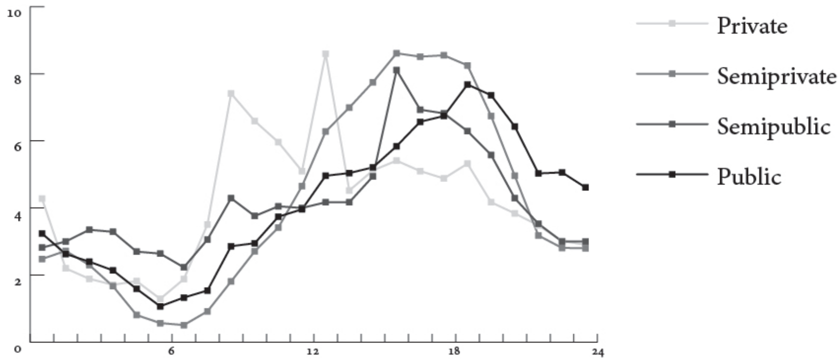


Figure 7: Temporal trends by crime locations: property crime | percentages

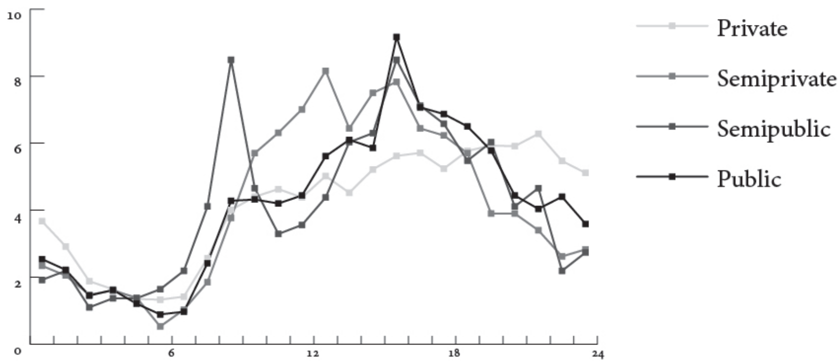


Figure 8: Temporal trends by crime locations: harassment crime | percentages

Figures 7 to 11 show that temporal trends of each crime type used in this thesis at different locations. In all five crimes, temporal patterns vary between crimes at private places and crimes at public places, and semipublic and public places displayed some similar fluctuations by time-of-day.

Temporal trends of each crime by day-of-week showed distinguishable patterns by types of crime but for time-of-day, except for certain time periods, the temporal patterns did not display significant dif-

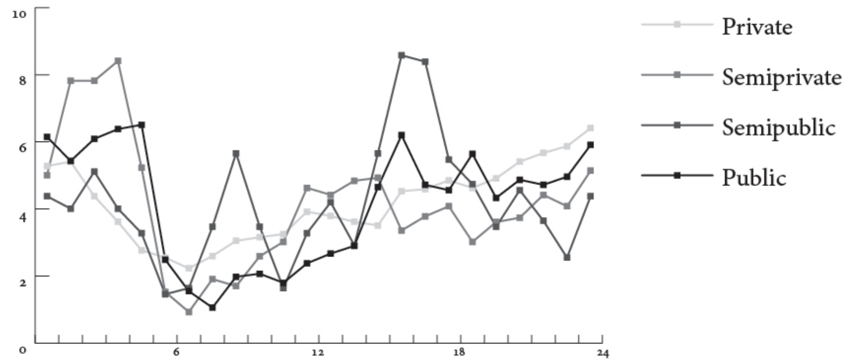


Figure 9: Temporal trends by crime locations: assault crime | percentages

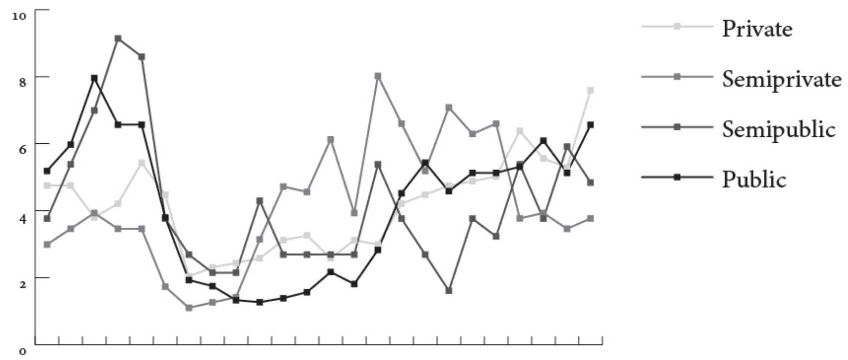


Figure 10: Temporal trends by crime locations: robbery crime | percentages

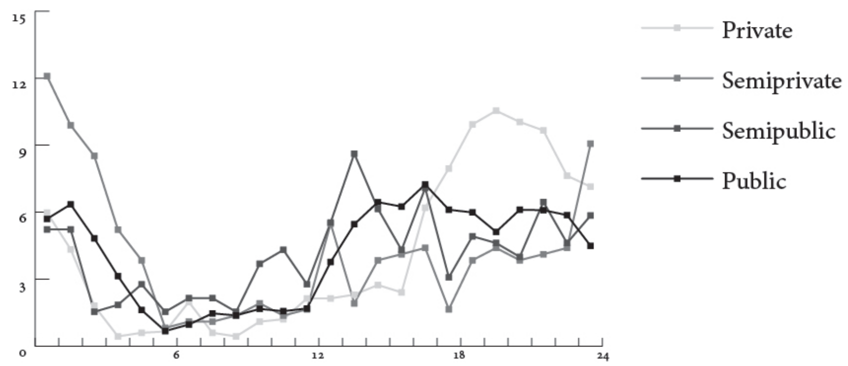


Figure 11: Temporal trends by crime locations: drug crime | percentages

ferences [5]. For example, in a case of property crime at private places (Figure 12), even though subtle daily variations were found at a short period of time-of-day (i.e. midnight to 6 am), there were similar fluctuation patterns of time-of-day overall.

As this research aims to test/maximise the benefits of the new data, 'real-time data', it was very important to employ a temporal unit which allow to capture situational changes according to the time of day. While the temporal patterns indicated a 168 time matrix (24 hours and 7 days)

5 Temporal trends of property crime at private places reported in this chapter as an example (Figure 12). See Appendix (p.141) for the analysis results of other 19 crime types

would be appropriate in a few crime cases, employing the units to the street segment levels generated a considerable amount of non-crime units, which would damage the validity of the analysis. For these reasons, the first study using the block groups which have sufficient volumes of crimes and dynamic population employed 168 time frame and the second and third studies using street segments used 24-hour frame. Collectively, for the first analysis (census block groups and a 168 time matrix) 39.84 per cent of units (72,087 units out of 180,936 units) have no record or crime/dynamic population, and for the second analysis (street segments and 24-hour time matrix), 61 per cent of units (173,259 units out of 284,064 units) have no crime/dynamic population value. For this reason, the model selection was carefully conducted to control/eliminate the issues the analysis model might have, especially in the analysis employing street segments. Future studies using the datasets should carefully select a spatiotemporal unit and the model selection steps are discussed in each empirical chapter in detail.

Recent research has found that a large proportion of crime is concentrated at a limited number of street segments (Weisburd, 2015) and these findings suggest that the model should not ignore the time-invariant latent characteristics of each street. There is also a possibility of time-specific unobservables on crime which could make an omitted variable bias in the model. For the crime data, a cyclic fluctuation in criminal opportunities would generate bias (Raphael and Winter-Ebmer, 2001), hence mitigating unobserved time effects such as time/day/year effects has been considered important to the production of reliable statistical results (Baltagi, 2006).

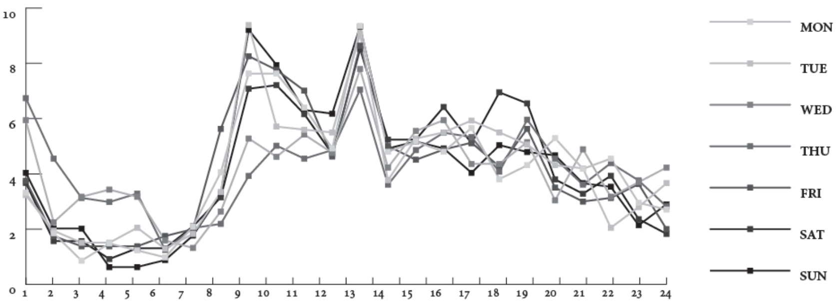


Figure 12: Temporal trends: property crime at private places | percentages

5 MOBILE AMBIENT POPULATION PATTERNS

5.1 INTRODUCTION

For many years, environmental criminologists have been treating the social and physical structure of the neighbourhood as an essential base to explain levels of crime. Further, the majority of the empirical studies exploring crime within this structure have relied on data from the population census to characterise areas.

It is however unsurprising that this traditionally used data has not generally been able to provide sufficient information to allow criminologists to fully operationalise the logic described in the environmental criminology theories. In particular, empirical research on routine activities theory discussing crimes as a by-product of opportunities created by routine activities - generalised social activities undertaken by the basic population - (Cohen and Felson, 1979) has faced criticism on the use of broadly defined, imprecise, and a proxy measurements of the core construct, human activities. Unlike census data, the data collected from social media enables a more nuanced exploration of the role of presence and absence of populations in explaining crime which is only possible using human activity data that is both spatially and temporally referenced.

The main aim of the current study is to examine the utility of social media data in understanding a microenvironment of crime situations in tracking the movements and fluctuations of certain populations. Critically to the argument here, these relationships will examined for a series of distinct behaviour settings which separate between crime types and distinct place contexts.

5.2 Background literature and theoretical rationale

5.2.1 *Crime and the rhythm of places*

Environmental criminology explains the emergence of crime within a spatiotemporal framework. It believes that certain spatiotemporal situations are more likely to generate dangerous outcomes: crimes. In the rationale of the theory, crime occurs when offenders judge a moment where they stand as a crime approachable condition and feel safe enough to execute a crime.

Routine activity theory explains that this moment is not evenly distributed in time and space, and is predictable by understanding the spatiotemporal patterns of routine activities (Clarke and Felson, 1993). A salient characteristic of routine activities is repetition. Each individual has periodic activities within spatiotemporal dimensions and the accumulated activities create concrete situation patterns of places, 'the rhythm of place'.

Note: see the sections about 'Routine activity' in Chapter 2 (p.11)

A specific issue is that the primary concept of guardianship fluctuation has traditionally been tested indirectly by using data which does not contain a sufficient temporal dimension (Sampson and Wooldredge, 1987; Ouimet, 2000). Instead the 'situation' has been conceptually inferred from the average attributes of people in the same geographical area rather than a precise account of the context or moment. In the original research by Cohen and Felson, the absence of guardians against crime is conceptually framed as the consequence of more females in the job market rather than the lack of guardianship in the crime situation (Cohen and Felson, 1979). In subsequent research, the characteristics of people's routines were understood by studying urban forms – land use (Brantingham and Brantingham, 1975; Brantingham and Brantingham, 1981), transportation hubs, street network or urban facilities (Groff, 2011; Bowers, 2013) – implying rather than directly measuring temporal shifts of population.

5.2.2 *Guardianship and the rhythm of places*

Garland (1999) argues that all crime situations are different since they are created in the cultural context of the place. Furthermore, Anderson (1999) notes that places have their own culture and code and cannot be easily measured using traditionally used methods such as census data. To understand the true dynamics between the presence of people and crime, it is necessary to consider how this might vary according to context. For example, Cohen and Felson hypothesised that it was the lack of residents at home during the day that contributed to the rise in property crime post world-war II (Cohen and Felson, 1979). Therefore a lack of population to act as suitable guardians in the private home environment can encourage crime. However, Beavon, Brantingham and Brantingham (1994) found an increased likelihood of burglary on streets that were likely to be busier in terms of their flow and large quantity of population. In this context more people about equates to higher crime.

One of the distinctions here is the degree to which people act as guardians against crime or are, in fact, crime facilitators in certain situations. Indeed, the crime situation is undoubtedly one of the core concepts in environmental criminology (LeClerc et al., 2016). Considering the immediate environment and how the situation shapes the moment is important in understanding when a second or third party might provoke crime (e.g. attempting to split up a fight or by providing a suitable victim) and when they might prevent it (e.g. by acting as a guardian against property crime).

The concept of guardianship is complex and has attracted substantial amounts of research. Miethe and Meier (1994) discuss how the availability of others offer assistance to ward off an attack and this can be in the form of either formal or informal guardianship. Reynald (2009) suggested 'guardianship in action', arguing that the concept of guardianship should be re-assessed to focus on the actual action of guardians and not only their existence. Her findings suggested that actions by the place occupiers are important in assessing whether a situation has or lacks the necessary conditions for crime. The extent to which a population will be capable of acting as a guardian against crime depends on their role in that particular situation (Reynald, 2009). Clarke and Harris (1992) also argued guardianship is not an existence, it is an action, which depends on the varying degrees of responsibility for discouraging crime.

People can have different relationships with the places they routinely visit. At home or a private space, people act as self-motivated guardians of their own territory. In fact, much of the work by Newman (1976) and others discuss as the role of territoriality in encouraging surveillance (Newman, 1976; Taylor, Gottfredson, and Brower, 1984; Reynald and Elffers, 2009). In a work setting, named as semiprivate or semipublic places in this research, it is likely there is some expectation that the employees will assist with guardianship. Eck and Rosenbaum (1994) have referred to this as 'place management' which emphasises the actions by place managers to control criminal activities in the place (Eck and Rosenbaum, 1994; Eck and Weisburd, 1995; Clarke and Eck, 2005).

In some situations the presence of more people can increase the anonymity of an area, weaken the level of informal control and eventually decrease guardianship capabilities. Busy situations might make criminals less likely to be recognised but also provides more eyes as witnesses in the public places (Roncek, 1981; Roncek and Maier, 1991; Nasar, Fisher, and Grannis, 1993). In very public places, where people's roles are generally as 'passers-by' then perhaps the requirement for individuals to act as intervening guardians is less formal and more optional. This is supported by research suggesting that people may feel higher willingness to intervene in situations where they have more knowledge and familiarity with the place (Uittenbogaard, 2014).

The role of the people present in creating opportunities or guarding against crime will also depend upon the type of crime being considered. In terms of property crime, in private homes, the place can at the extreme be guarded 24 hours a day against visitors from the outside. People's daily routines in public space can create crime vulnerability if they lead potential offenders to become aware of opportunities. This can be framed as the development of an offender's mental map (Brantingham and Brantingham, 1988). However, the habitation of a private setting will not protect against crimes committed by members of the household against each other, such as domestic abuse. In the case of violent crime, private and public spaces have clear differences in terms of available third parties who can act as a guardian (Cooney, 1998). Indeed, more generally third parties can shape violent situations in different ways such as taking the role of warriors or peacemakers (Cooney, 1998; Cooney, 2009).

5.2.3 *Measuring population dynamics*

To date, the lack of temporal information concerning routine activities and the use of alternative and proxy measures has made it difficult to develop the theory and to examine its predictive validity. Groff (2006) argues that although routine activity theory has a well-developed framework, its empirical validity is still not confirmed because of the difficulties associated with obtaining individual level data (Groff, 2006). This is backed up by Weisburd and Piquero (2008) who reviewed the explanatory powers of core criminological theories using 169 empirical papers published in Criminology from 1968 to 2005. According to them, in general explanatory power is constantly very low over time. This low explanatory power could originate from inadequate theory including the articulation of causal mechanisms, poor predictive ability, or deficient explanation of the crime process (Weisburd and Piquero, 2008). These are exacerbated by the limited data issue as it deters the development and update of the theories based on adequate empirical examinations.

More importantly, these crucial limitations could result from the poor measurement of conceptual constructs. One reason for this is that there are inherent difficulties in measuring many aspects of human behaviour (Osgood, McMorris, and Potenza, 2002) and consequently approaches to data collection often involve limited and indirect approaches that may poorly reflect the constructs they seek to estimate. Collectively, measurement issues have impeded progress in criminology and hindered understanding of if and how theoretical mechanisms work in practice (Sullivan and McGloin, 2014).

To more directly measure guardianship fluctuation, it is necessary to quantify people flows over the urban area, which are known to be skewed and disproportionate (Felson and Boivin, 2015). Measuring changing populations and populations at risk has been a subject of research in various fields. In geography, a better estimation of the location and size of vulnerable populations in hazard prone areas can be crucial because it guides decision making and policy in national urgent situations (Mondal and Tatem, 2012). Urban planning research has long since evolved from reliance on static population data such as the census. Instead more spatially and temporally specific data such as LandScan or the Global Rural Urban Mapping Project (GRUMP) are used in hazard, public health and economic/business models (Bhaduri et al. 2002; Bhaduri, Bright and Coleman, 2007; Piel et al.2013; Njuguna and McSharry, 2016).

In criminology, the census has been used as a backbone source in measuring the characteristics of population because it gives snapshots of sociodemographic characteristics of where and how people live (Calder and Teague, 2013). From early research such as Burgess (1925) census tracts were seen as the crucial spatial unit of analysis as understanding social features of places was a predominant way of explaining crime location. Whilst this can give a solid understanding of socioeconomic features based on the residential population it accounts poorly for non-residential activities (Felson and Boivin, 2015). Additionally, the increasingly well evidenced consistent finding that crime concentrates in hotspots demonstrates that crime acts are influenced by a flow of population activity rather than static features like census population.

Georeferenced social media data has much scope in the future development of criminology, particularly with respect to understanding spatial and temporal patterns of crime. In particular, it has the potential to facilitate the dynamic detection and understanding of movement patterns. Capturing the precise context of criminal opportunities and their surroundings is a desirable research pursuit in place-based crime studies. When recorded, social media data gives real time records of people not only in terms of their location but also their unobservable information (e.g. emotions and interests), giving crucial insight into those present in criminal situations. Very little has been done to develop methods that study guardianship dynamics using social media data. The only exception is Solymosi, Bowers and Fujiyama (2017) who use crowdsourced data from an online problem reporting system, Fixmystreet [1] , to represent active guardianship in neighbourhood sized areas (Solymosi, Bowers and Fujiyama, 2017). There is therefore much scope in undertaking research that examines the utility of social

1 www.fixmystreet.com

media data in understanding the population microenvironments of crime situations.

5.3 methods

This analysis involves the use of a number of datasets. Two of the core datasets have already been described. To recap, these are the crime data and social media feeds in the region of NYC between the dates of 1st of January, 2013 and 31st of December, 2013. In addition, the research in this chapter also required data that could account for social and economic differences and residential population characteristics in the selected spatiotemporal units of analysis. As explained below these differences will impact on crime and need to be accounted for in the modelling strategy.

Note: see the section about 'Data' in Chapter 4 (p.36)

5.3.1 *Measurements*

Socioeconomic variables

To determine whether the dynamic population affects crimes and whether the effect varies across socioeconomic structures, the analysis model was necessary to include relevant indicators of socioeconomic contexts. The analysis used 2010 census to estimate the socioeconomic and demographic characteristics [2].

In Shaw and McKay's original work, socioeconomic status of a neighbourhood was estimated using census data on Social Economic Status (SES), residential stability, and racial/ethnic heterogeneity (Shaw and McKay, 1942). Other studies have used alternative and expanded measures based on the variables constructed for testing social disorganisation theory. In these neighbourhoods SES was assessed using level of education, economic hardship such as unemployment rates or the percentage of people receiving public assistance/living below the poverty line, or family disruption including the percentage of female-headed/single parents/divorced households (Sampson, 1985; Sampson and Groves, 1989; Bellair, 1997). Residential stability was calculated using the percentage of owner/renter occupied or residents living in the same house over five years (Sampson, 1985; Sampson et al., 1997; Sampson and Raudenbush, 1999; Morenoff et al., 2001).

2 The U.S. Census Bureau collects data about the people living in the country from various sources and publishes the result every 10 years. The main source is the American Community Survey (ACS) which is on-going survey about vital socioeconomic information of neighbourhoods including employment status, educational attainment, whether people own or rent their homes. According to the ACS report, most questionnaires were collected by mail response and online, phone, and in-person interviews were also used.

Racial/ethnic heterogeneity has been estimated using the percentage of Black/Hispanic or foreign-born (Sampson and Raudenbush, 1999; Morenoff et al., 2001) or calculated using Blau's diversity index reflecting how many different types of race/ethnic groups there are in the neighbourhood (Sampson and Groves, 1989).

In this study, the factors that are widely included in the previous research and haven been shown empirically to have a measured effect on crime (in certain neighbourhood conditions) were chosen. Hence, to measure SES, three variables indicating educational attainment (the percentage of people with no schooling/1-4 year of schooling), unemployment (the percentage of unemployed workers in the total labour force), and low-income (the percentage of people living below the poverty line) were selected. Resident mobility was measured by non-ownership of residential properties (the percentage of renters). To capture the range of racial heterogeneity, Blau's diversity index [3] was employed (Blau, 1977; Sampson and Groves, 1989).

Vulnerable population variables

The model also includes the population identified and stereotyped as being at high risk of committing crimes. Official statistics and academic studies reported that the demographic characteristics of criminals are disproportionately clustered in a certain type of population segments (Farrington, 1986; Fabio et al., 2011; Bureau of Justice Statistics, 2014). According to a report titled Crime and Enforcement Activity in New York City (NYPD, 2017), the majority of crimes were committed by Black or Hispanic (Murder: Black 61.7 per cent, Hispanic 27.0 per cent, and White 7.7 per cent, Larceny: Black 58.2 per cent, Hispanic 24.9 per cent, and White 12.4 per cent) in 2013 (NYPD, 2017). The age of offenders was also concentrated in a certain age group. The prevalence of offending across all types of crimes peaked in the teenage years (from 15) and then rapidly declined in the early 20s (up to age 24) (Piquero, Hawkins, and Kazemian, 2012; Ulmer and Steffensmeier, 2014). For these findings, residential population counts of Black and Hispanic young males (the percentage of Black aged 15-24, and the percentage of Hispanic aged 15-24) were included.

In order to examine relationships with crime, the models that follow use two population variables (one being the dynamic Twitter data), five socioeconomic variables and two demographic variables (Figure 13).

5.3.2 *Analysis Methods*

The current study consisted of three different analyses. The first analysis was designed to examine the differences between the temporal

3 Blau's index of heterogeneity $(1 - \sum p_i^2)$, where p_i is the proportion of group members in each of the i categories.

	FROM CENSUS	FROM SOCIAL MEDIA
POPULATION SIZE	RESIDENT POPULATION	DYNAMIC POPULATION
SOCIAL DISORGANISATION	EDUCATION LEVEL EMPLOYMENT STATUS INCOME LEVEL RESIDENTIAL OWNERSHIP RACIAL HETEROGENEITY	
HIGH-RISK POPULATION	BLACK YOUNG MALE HISPANIC YOUNG MALE	

Figure 13: The list of independent variables

patterns of crimes and the fluctuation of the dynamic population visually. For the second analysis, Cross-Correlation Analysis (CCA) was conducted to statistically measure the temporal association between these two variables. Using a spatial panel data model, the last analysis estimated crime rates using the dynamic population and the socioeconomic contexts illustrated in Figure 13.

Chapter 4 gives an overview of the research units used in the first two analyses. For the spatiotemporal units of the Spatial Panel data Model, census block groups and hours (168 hours) were chosen. The final data consists of 1,077 geographic units by 168 temporal units between 1st-time slot (Monday 00:01-01:00 on the 24-hour clock) to 168th time slot (Sunday 23:01-00:00 on the 24-hour clock). The counts of crimes and the values of social disorganisation variables estimated from 2010 census data were appended to each spatiotemporal unit. As a result, the generated data contained the number of crimes, the population from the social media and census, and the values of socioeconomic variables in each census block group over time.

Note: see the section about 'Research Units' in Chapter 4 (p.42)

Cross-correlation analysis

In the relationship between two time-series, the responsiveness of one with another can be lagged in time taking several minutes, hours, or years until one affects the one another (Jenkins and Watt, 1968). Cross-Correlation Analysis (CCA) is a standard method of measuring the level to which two temporal series are correlated and is used to explore time differences and similarities between two discrete-time sequences. The analysis identifies the lags of one variable that might predict another variable. In this study, CCA was carried out to investigate the interactions between the dynamic population and crimes [4].

A solid cross-correlation between the two-time series for different time lags would indicate that the dynamic population influences changes in crime rates after the lag periods. However, importantly, statistically confirming the correlation between the two does not necessarily imply causality. Although correlation analysis is not sufficient to demonstrate causality between the ambient population and crime, it can explore trends in terms of moderate or strong strength correlation between lagged periods.

When there are two-time series X_t and Y_t, the coefficient of cross-correlation between Y_t at the current time point and X_t-_k at k time lags point takes the form

$$\rho_{XY}(k) = \rho(X_{t-k}, Y_t) = \frac{Cov(X_{t-k}, Y_t)}{\sqrt{Var(X_{t-k})}\sqrt{Var(Y_t)}} = \frac{\gamma_{XY}(k)}{\sigma_x \sigma_Y} \quad (1)$$

$$\gamma_{XY}(k) = \frac{1}{n} \sum_{t=k+1}^{n} (x_{t-k} - \bar{x})(y_t - \bar{y})$$

The cross-correlation function, $\rho_{XY}(k)$ refers to the coefficient of cross-correlation between an output series data Y_t and an input series X_{t-k} at a time lag k. The cross-covariance function, $\gamma_{XY}(k)$ is estimated by the sample cross-covariance function, $\bar{x}$ and $\bar{y}$ stand the means of each time-series and σ_x and σ_y are the standard deviations of each time-series respectively.

Spatial panel data model

The crime rates of any types of the spatial unit often depend on the crime rates and the characteristics of neighbouring units (Anselin et al., 2000). To address the inherent autocorrelation issues that the given datasets may have, relevant spatial econometric models have been reviewed with theoretical and practical considerations of spatial dependencies.

Spatial panel models can be used to explain the effects of the interactions of time and space lags of both independent and dependent variables (Elhorst, 2011). The standard model for spatial panel data takes the form

$$Y_{it} = X_{it}\beta + \varepsilon_{it}$$

4 As the analysis requires that all data series involved are stationary, the Augmented Dickey-Fuller (ADF) test was conducted first to check stationarity of the datasets. The null hypothesis for the Augmented Dickey-Fuller (ADF) test is the variable contains a unit root, and the alternative is that the variable was generated from a stationary process. The null hypothesis is rejected in most cases and it indicates that the data does not exhibit a unit root.

where i stands an index for the spatial unit with $i = 1, ..., N$, and t is a temporal index with $t = 1, ..., N$. Y_{it} denotes an observation on the dependent variable at i unit and t time. X_{it} is a vector of the independent variables ($1 \times J$ matrix), β is a matching vector ($J \times 1$ matrix) of the parameter(s) to be estimated, and ε_{it} is an error term.

When regarding interactions between spatial units, the model can be modified into a spatial lag model (spatial autoregressive model) which contain a spatially lagged dependent variables or a spatial error model dealing with a spatial autoregressive process in the error term (Anselin, 2006; Elhorst, 2009; Lee and Yu, 2010; Salima, Julie, and Lionel, 2018).

The spatial lag model controls endogenous interaction effects (δWY) between dependent variables of units (Elhorst, 2011). The model considers the influence of dependent variable y of a unit to the dependent variable y of neighbouring units. In empirical studies on crime, endogenous interaction effects can be explained as situations where crime rates interact with the rates in nearby neighbourhoods. Therefore, Y_{it} is estimated with a spatially lagged dependent variable ($\sum_{j=1}^{N} W_{ij} Y_{jt}$) a spatial autoregression parameter (δ).

$$Y_{it} = \delta \sum_{j=1}^{N} W_{ij} Y_{jt} + X_{it}\beta + \varepsilon_{it}$$

The spatial Lag of X Model, on the other hand, posits that exogenous interactions ($WX\gamma$) where the decision depends on the interactions effect of explanatory variables observed in neighbouring units and on a set of observed local characteristics (Plumpes and Neumaye, 2010). In terms of statistical analysis on crime and neighbourhood structures, the effects of exogenous interaction reflect situations where the crime rate of one neighbourhood is influenced by explanatory variables of adjacent neighbourhoods.

$$Y_{it} = X_{it}\beta + \sum_{j=1}^{N} W_{ij} X_{ijt}\gamma + \varepsilon_{it}$$

The spatial error model reflects the interaction effect (λWu) among the error terms (Elhorst, 2011). In crime analysis, the spatial error model considers a situation where determinants of crime rates omitted from the spatial model are spatially autocorrelated or where unobserved conditions follow a spatial pattern. The model takes the form, and ϕ_{it} denotes the spatially autocorrelated error term and λ is the spatial autocorrelation coefficient.

$$Y_{it} = X_{it}\beta + \phi_{it}$$

$$\phi_{it} = \lambda \sum_{j=1}^{N} W_{ij}\phi_{it} + \varepsilon_{it}$$

The spatial Durbin model (SDM) takes account both endogenous and exogenous effects into the model simultaneously (Lesage et al., 2009). The model contains a spatially lagged dependent variable ($\sum_{j=1}^{N} W_{ij}Y_{jt}$) and spatially lagged independent variables ($\sum_{j=1}^{N} W_{ij}X_{ijt}$).

$$Y_{it} = \delta \sum_{j=1}^{N} W_{ij}Y_{jt} + X_{it}\beta + \sum_{j=1}^{N} W_{ij}X_{ijt}\gamma + \varepsilon_{it}$$

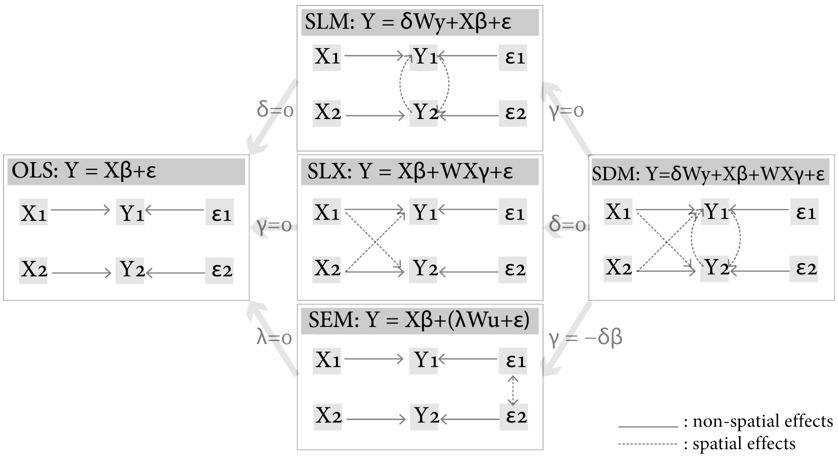


Figure 14: Spatial econometric model specifications

To carry out the analysis, the spatial Durbin model with one-way error component were selected [5]. The main motivation to use the extended model was to include spatial interaction effects which are obviously detected in previous crime spatial analysis (Morenoff and Sampson, 1997; Anselin et al., 2000; Johnson et al., 2007; Andresen and Malleson, 2011) and identified from Moran's I analysis [6] and the values of spatial ρ [7]. If spatial effects exist, statistics models which do not

5 In this research, LeSage-Pace and Elhorst's model was used. See Elhorst(2014)'s *'Spatial Econometrics From Cross-Sectional Data to Spatial Panels'* for more details. To determine which type of model would fit the data, several different spatial models have been tested and SDM generated the highest values in R^2 and log-likelihood values close to zero. After reviewing the results of the statistical investigation and theoretical consideration of the data set and the results of each model, SDM was chosen (see Appendix (p.147) for the results of log-likelihood and R^2 values of SDM, spatial lag and spatial error models (Tables 23).

6 Most variables showed positive clustered patterns. Moran's I analysis is a measure of spatial autocorrelation and the outcome value rages from -1 to 1, with value greater (less) than zero indicates positive (negative) spatial correlation and with zero value means no spatial correlation (see Appendix (p.148) for the results of Moran's I test (Tables 25 and 26)

7 The parameter spatial ρ of SDM was significantly different from zero in most crime types, which indicates that least-squares estimates could be biased and inconsistent. For robbery crimes, the ρ was only significant at public places but in all other crime types, the values were statistically significant which justified using SDM.

consider the effect such as ordinary least square models would lead to biased or unreliable results (Anselin, 1988). To investigate possible spatial autoregressive effects from both explained and explanatory variables, SDM adds average-neighbour values of both explained variables and explanatory variables (Lesage and Pace, 2009). Considering the fact that both the dynamic population and crime have clear temporal patterns, the error terms in the model comprise error terms for i and t (e_{it}) and for time-specific effect (μ_t). μ_t refers to unobserved time heterogeneity. In this model, the time-variant factors control for time-specific shocks that may affect crime such as routine activities and weather which vary by time of the day and day of the week. Regarding the results from a likelihood ratio test [8] and Hausman's specification test [9], the time-period specific effects were treated as fixed effect.

$$Y_{it} = \delta \sum_{j=1}^{N} W_{ij} Y_{jt} + X_{it}\beta + \sum_{j=1}^{N} W_{ij} X_{ijt}\gamma + \mu_t + e_{it} \tag{2}$$

where

- δ a spatial autoregressive coefficient referring the endogenous interaction effect
  - $\delta > 0$: positive spatial dependence
  - $\delta < 0$: negative spatial dependence
  - $\delta = 0$: traditional OLS model
- γ a Vector (J,1) coefficient of spatial lag on independent variables referring the exogenous interaction effect
- μ_t time specific effects

5.4 RESULTS

5.4.1 *Temporal patterns analysis*

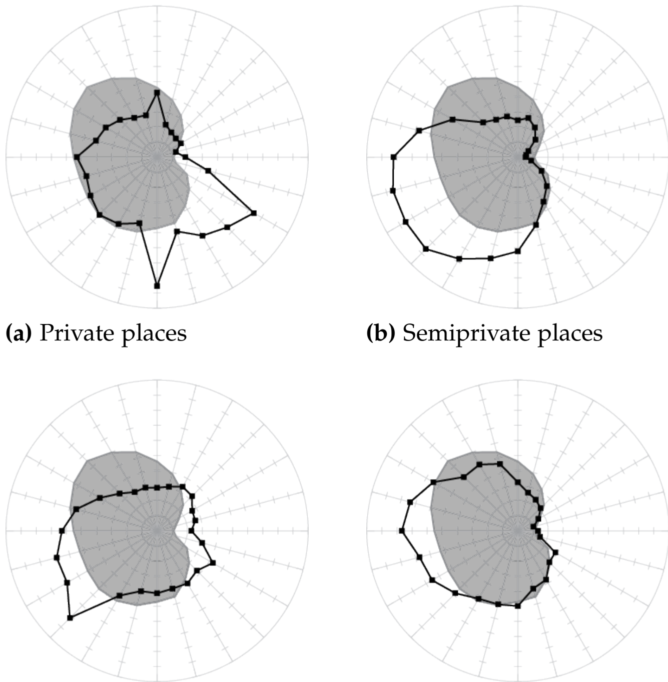
To understand the influence of people's regular activities, quantified by the dynamic population in this research, on crime, visual analysis

8 In all cases, the null hypothesis ($H_0 : v_t = 0$) of the LR was rejected. The test results confirmed the presence of time fixed effects which would bias the estimates of the relationship.

9 The null hypothesis of Hausman test (H_0 :cov(x_{it}, μ_t)=0) was rejected in most cases except for crimes at private places. These results justified the estimation using random effects would include endogenous bias. Although crimes at privates places failed to reject the null hypothesis, intuitively and theoretically the fixed effects model would be suitable as the data covers all populations (all census units in the area) and time which were not randomly selected (see Appendix (p.148) for the results of Hausman's specification test (Table. 27))

was conducted first. The temporal distribution for the two time series are plotted in Table **??** broken down by type of crime and type of space. In each case, the dynamic population is presented in grey and the crime count is shown as a black line. Looking across the row it can be seen that unique interaction patterns exist for the various crime types. The most salient contrast is evident when comparing the patterns of crimes at private places with crimes at public places, and also property crimes with violent crimes.

Note: see the section about 'Research unit; Temporal units' in Chapter 4 (p.46)



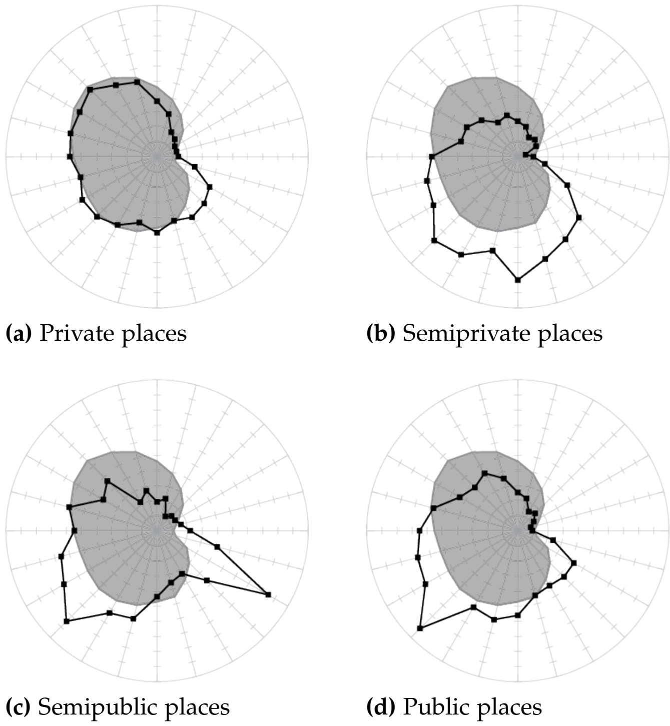
(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 15: Temporal patterns of property crime and dynamic population

As shown in Figure 15, property crimes occurring in residences and on streets showed completely different variations. For example, the temporal variations of property crime and the population appears positively correlated at public places but negatively at private places. Property crime at semiprivate places showed similar patterns with the crime at public places but the peak started a few hours earlier than property crime at public places and lasted only for two to three hours. Unlike property crimes at private or public places, the correlation with the population fluctuations was not obviously observed in property crime at semiprivate and semipublic places.

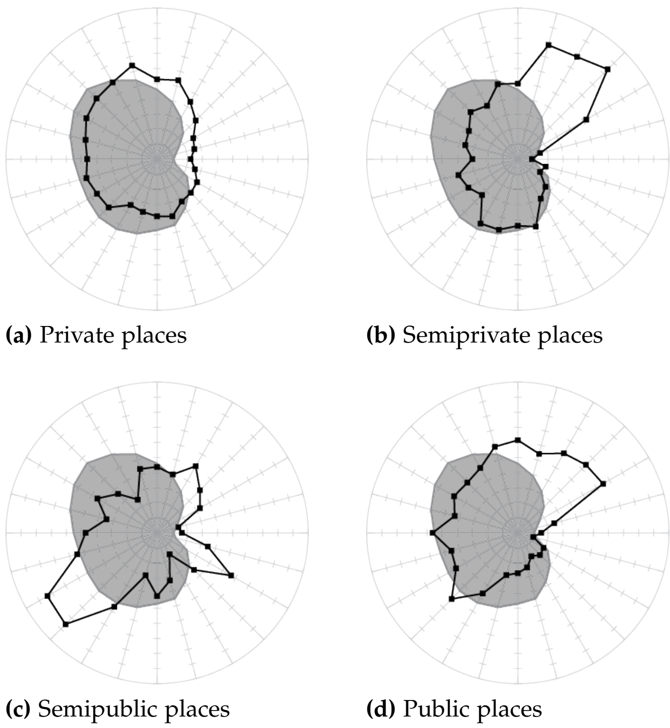
As shown in Figures 16 and 17 , it is clearly evident that crime and population were highly positively correlated in violent crimes, harassment and assault, at private places. These crimes at private places fluctuated almost identically with the population. As shown in the temporal changes of the dynamic population, the largest variation in these crime variables was observed during late evening (7pm to 12am)



(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 16: Temporal patterns of harassment crime and dynamic population

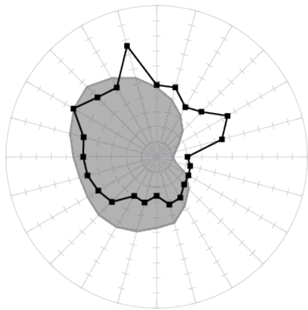


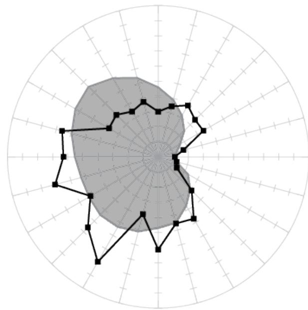
(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

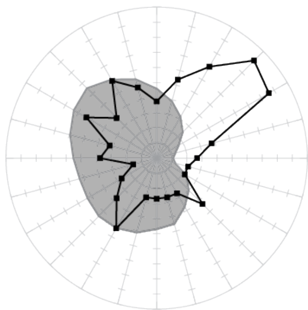
Figure 17: Temporal patterns of assault crime and dynamic population

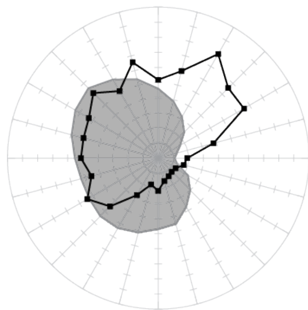
and the variation was sharply dropped after the peak hours around 1am to 2am. For robbery crimes at private places (Figure 18, the peak





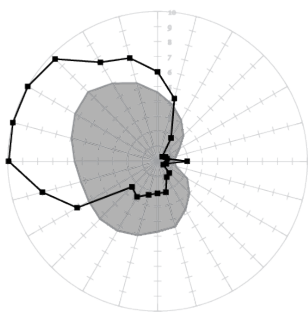
(a) Private places **(b)** Semiprivate places

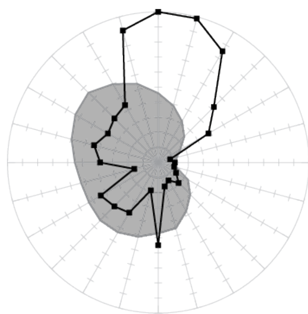




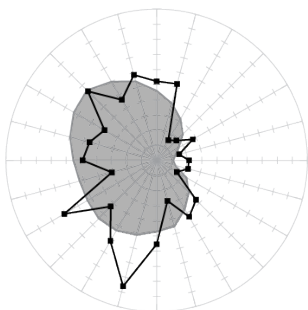
(c) Semipublic places **(d)** Public places

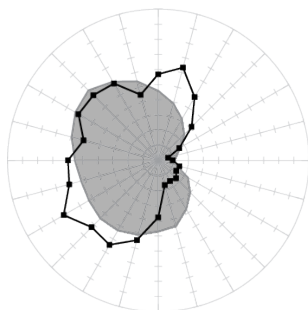
Figure 18: Temporal patterns of robbery crime and dynamic population





(a) Private places **(b)** Semiprivate places





(c) Semipublic places **(d)** Public places

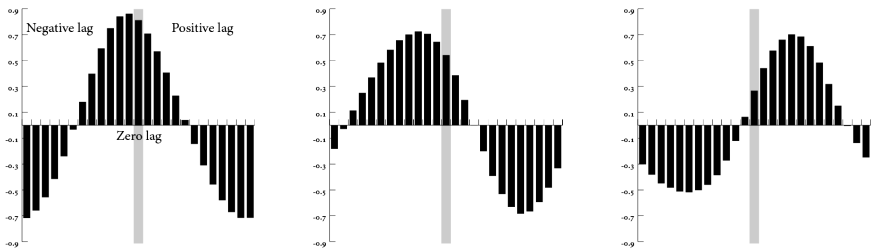
Figure 19: Temporal patterns of drug crime and dynamic population

was found in late evening around 11pm when population activities appear to decrease. Drug crimes (Figure 19 fluctuated intensely in narrow time periods (early evening, 5pm to 9pm) and the pattern did not show any apparent positive correlation with the population changes as was the case with all the other types of crimes.

Overall, observable contrast of temporal interactions was found between crimes at private places and places allowing public access across all crime types, which indicates that even the same type of crime has very distinct interaction patterns with the dynamic population depending on the place context.

5.4.2 *Cross-correlation analysis*

For more formal statistical exploration on whether there is a correlation between two-time series, Cross-correlation analysis (CCA) was employed. The CCA between two series helps to identify the nature of the relationship and how the series are correlated in time.



(a) Type (A) **(b)** Type (B) **(c)** Type (C)

Figure 20: Three patterns of cross correlation

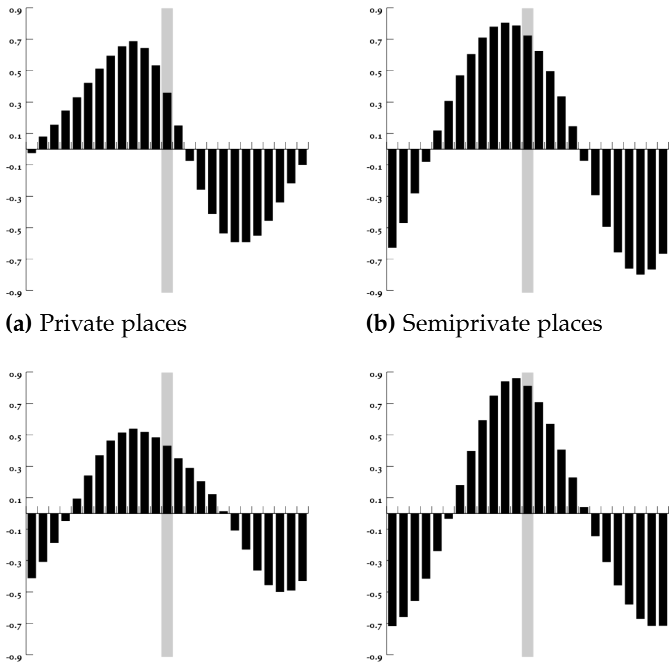
The patterns of CCA results observed were grouped into three types. The most common correlation patterns were Type (A), and Type (B) and Type (C) was occasionally observed. In the graph in Figure 20, zero lag represents immediate correlation of two series, the positive lag side (right side) of the graph shows the cross-correlation coefficient when rises in the dynamic population precedes crime, and the negative lag side (left side) of the graph indicates the value when rises in the dynamic population follows crime.

For example, if the highest positive value is observed at +2 lags, it demonstrates that the population has its greatest positive effect on the future level of crime in two hours.

In Type (A), the dynamic population (T_D) and crime (T_C) are positively correlated in both positive and negative time lags. The highest correlation was mainly found within two leads to two lags periods and the correlation coefficient rapidly declines in ± 2 time periods. The outcome basically refers T_D and T_C instantly respond to each other but the strong correlation only lasts one to two hours. The characteristics of crime having type (A) would be theorised as more people's activities coincide with more crimes, which correspond with natural crime patterns described in environmental criminology. When a place is getting busier it will produce more criminogenic opportunities and build a suitable environment for crime events (Brantingham and Brantingham, 1995). In this analysis, Type (A) was usually found in crimes at places where public access are allowed.

Type (B) occurs when T_D and T_C are negatively correlated in positive time lags. Put another way, this pattern indicates that if T_D increased T_C decreases. This pattern was observed in two cases; property crime at private places and harassment crime at semiprivate places.

Lastly in Type (C), T_D and T_C were positively correlated in positive time lags and negatively correlated in negative time lags. This can be understood when T_D and T_C increases but the population would not cause immediately increase as the highest positive correlation is found in 1 or 2-time lags. This type was observed at assault at semiprivate and public places and robbery at semipublic and public places.



(a) Private places **(b)** Semiprivate places

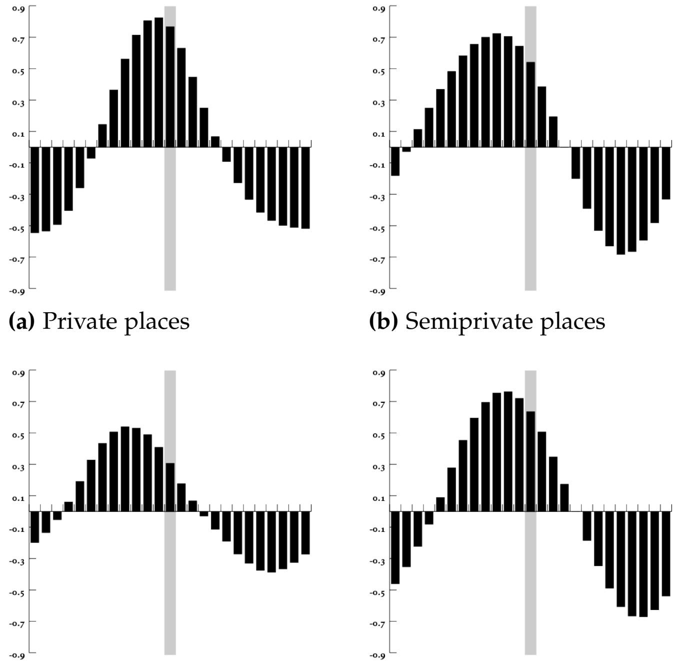
(c) Semipublic places **(d)** Public places

Figure 21: Cross-correlation between $T_{property}$ and $T_{Dynamicpopulation}$

In most types of crime, a strong cross-correlation was found in small lags, within ± 3 [10]. These results were consistently observed across all crime types, which means crimes are an immediate response to the changes in dynamic population, and the effect of the population changes only last 0 to three hours and slowly decreases.

Note:see Appendix for the coefficient values of CCA test of each crime type (p.150)

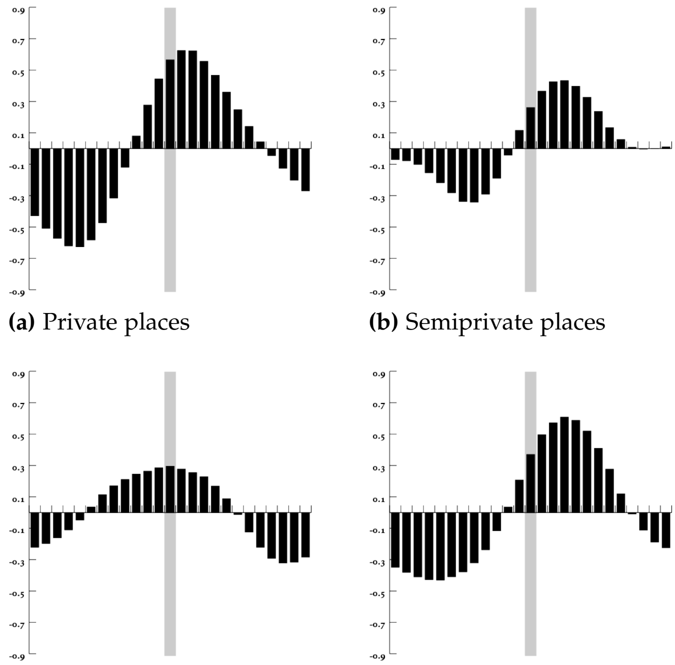
10 Granger causality test was also conducted to confirm the dynamic population leads the crime rates. The results noted that except for one case (property crime at semipublic) most cases rejected the null hypothesis (H_0: x does not Granger-cause y), it, therefore, can be concluded that lagged dynamic population does cause crime. (see Appendix (p.149) for the results of Granger causality test (Tables 28))



(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 22: Cross-correlation between $T_{harassment}$ and $T_{Dynamicpopulation}$

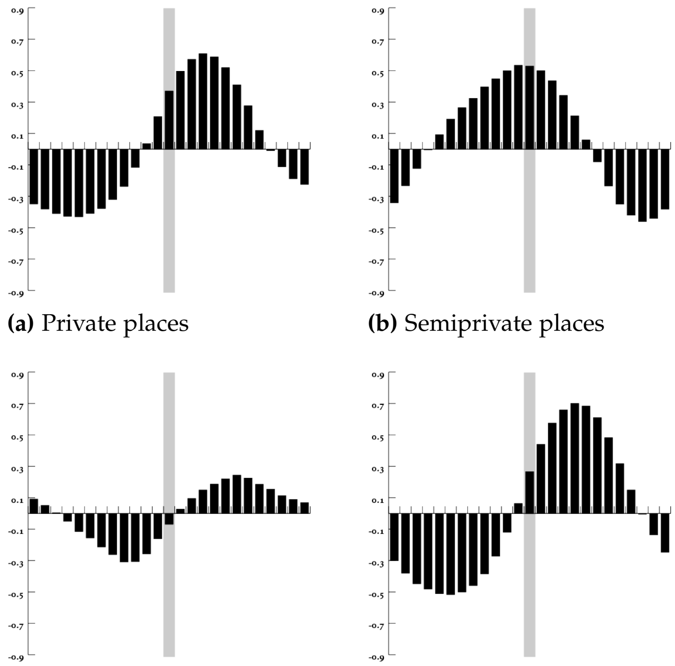


(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 23: Cross-correlation between $T_{assault}$ and $T_{Dynamicpopulation}$

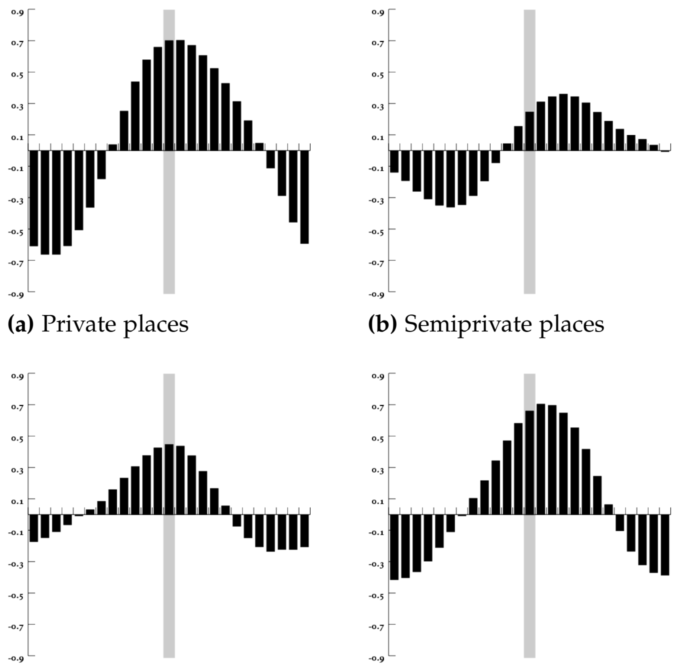
The correlation coefficients were relatively higher in property and harassment crimes than assault, robbery and drug crimes. This outcome indicated that the population was a more significant influence



(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 24: Cross-correlation between $T_{robbery}$ and $T_{Dynamicpopulation}$



(a) Private places **(b)** Semiprivate places

(c) Semipublic places **(d)** Public places

Figure 25: Cross-correlation between T_{drug} and $T_{Dynamicpopulation}$

on property crime and harassment crime than other crimes which can be said to be more serious. In particular, assault and robbery crimes showed narrow and low coefficients, which indicates the crimes are less likely influenced by the flow of the population than property and harassment crime, which either targets goods rather than people or are relatively minor crimes compared with assault and robbery.

For property crime, the results showed apparently different patterns in private and public places. At private places, T_D was negatively correlated with T_C in positive lags and positively correlated with T_C in negative lags, which indicates that when the dynamic population increased, the crime decreases. Unlike the results from crimes at private places, the changes in the population and the crime went hand-in-hand at semi-private and public places. This demonstrates the subtlety of the relationship between busyness and property crime.

Comparing harassment and assault crimes, the population increase led to less harassment crime at semiprivate and public places but more assault crime at semiprivate and public places. These outcomes showed that the dynamic population influences on minor violent and serious violent crimes differently even they occur at the same types of places.

5.4.3 *Spatial panel analysis*

The modeling used the pooled Ordinary Least Squares (OLS) estimation with all explanatory variables as a base model and the estimation by spatial Durbin model (SDM) followed. The second model only comprised the population size variables from the census and the dynamic population data (SDM (A)) and the third model included the social disorganisation variables showing neighbourhoods sociodemographic characteristics (SDM (B)). The last model (SDM (C)), the main model of the current research, included all variables discussed in this research. [11]

Table 7 shows the summarised results of SDM (C) for all crime types, and the SDM (C) results of each crime reported in Table 9 and tables in Appendix (p.154) together with their coefficients and *P*-values. First and of most significance, the results showed that the coefficient of the dynamic population is positive and significant ($p < .1$) for all crimes except for crimes at private places. This outcome was consistent in both SDM (A) and SDM (C) models, which means the effect of the dynamic population was still significant after controlling the effects of social and demographic variables. For crimes at private places, the dynamic population turned out to be insignificant (harassment and robbery) or negatively significant (assault ($\beta = -0.014; p < .1$) and drug ($\beta = -0.009; p < .05$)).

11 To highlight the patterns found in analysis outcomes, the chapter reported 1) summary of findings (SDM(C)) and 2) findings (All four models) of property crime at private places as an example following the analytic process. See Appendix (p.154) for the analysis results of other 19 crime types.

Whereas the coefficient of residential population was positive and significant for crimes at private places (property ($\beta = 0.130; p < .01$), harassment ($\beta = 0.127; p < .01$), assault ($\beta = 0.128; p < .01$), robbery; ($\beta = 0.015; p < .01$), and drug ($\beta = 0.068; p < .01$). Residential population turned to be insignificant or negatively significant for most crimes at places allowing public access (property crime at semiprivate ($\beta = 0.343; p < .01$), harassment crime at semipublic ($\beta = 0.006; p < .01$) and public ($\beta = 0.015; p < .01$), assault crime at semipublic ($\beta = 0.006; p < .01$), drug crime at semipublic ($\beta = 0.003; p < .01$) and public ($\beta = 0.024; p < .01$)). It was clearly observed that the population quantity variables therefore displayed distinct results by crime place context but not necessarily crime types. In contrast with the results of residential population displaying a significant association with crime at private places, dynamic population was found to be more likely associated with crimes at places that may have more substantial in- out- flow of population[12].

The variables of SES and the high-risk population presented a mixture of patterns in particular by crime locations types. For most cases of SES and vulnerable populations, the results showed that effects of the variables were linked to crime rates but positively significant only at crimes at private places, and negatively significant or insignificant at places open to the public. In particular, the level of poverty, Black/Hispanic percentage of the neighbourhoods had a positive and significant in crimes at private places, but the effects of these variables were not shown at public places. The percentage of Hispanic/Black in the community has been used a proxy for social stratification in many studies and statistically, they account for the crime rates of the neighbourhoods (Deaton, 2003). The results shown complied to this argument only at private places but not in crimes at public places.

Consistent with social disorganisation theory, it was expected that all social disorganisation variables included would demonstrate positive correlations, but the results did not conform with expectations. Generally, the effect of these variables was negatively linked to crime in most crime types in the panel model or marginal positive effects were occasionally observed (the latter tended to be in private places). Although the variables generated a negative effect in the SDM models, the pooled OSL estimated for the coefficient in the model were more likely to be positively significant as illustrated in the theory.

For the spatial regression models (Table 8), parameter estimates consist of information about the spatial dependence structure among the observations, a census unit in this research. Because any changes in a single unit with a given explanatory variable will generate the

12 The outcomes from the first two analyses and SDM were not identical. It may be caused by some general reasons such as employing different levels of spatial units for the analysis. It also could be caused by SDM includes all timeframe into the model and generates the average values of it, therefore, the model reflects the times when both were low (most of time) highly.

changes in the unit itself (a direct effect) and potentially affect nearby units (an indirect effect), the model estimates should be interpreted with spatial interactions with given variables (Takagi, Ikeda, and Kawachi, 2012; Tientao, Legros, and Pichery, 2016; Li and Wu, 2017) As regards the average impacts from neighbouring areas, the estimate for the parameters W-dynamic population (the spatially weighted coefficients of the dynamic population) was significantly different from zero except for crimes at private places and robbery crime. In particular, W-dynamic population was positive and significant at all crimes at places accessible by the public, which means the increase of the population can enhance the crime in adjacent neighbourhoods. However, the spatially lagged social variables were not significant in most cases especially in crime at private places, which means those variables do not affect the neighbouring areas.

Table 7: Summary of all types of crime: by independent variables

	Private	Semiprivate	Semipublic	Public
Residential population				
Property	(+)***	(–)***	(+)	(+)
Harassment	(+)***	(–)	(–)***	(–)***
Assault	(+)***	(+)	(–)**	(–)
Robbery	(+)***	(–)	(–)	(–)
Drug	(+)***	(+)***	(–)*	(–)***
Dynamic population				
Property	(+)***	(+)***	(+)***	(+)***
Harassment	(+)	(+)***	(+)***	(+)***
Assault	(–)*	(+)***	(+)***	(+)***
Robbery	(–)	(+)***	(+)***	(+)***
Drug	(–)**	(+)***	(+)***	(+)***
Education level				
Property	(–)	(–)***	(–)**	(–)*
Harassment	(+)	(+)***	(–)	(–)
Assault	(–)	(+)	(–)***	(–)***
Robbery	(+)**	(–)**	(–)*	(–)
Drug	(+)**	(+)	(–)	(–)
Employment status				
Property	(–)***	(–)	(–)***	(–)***
Harassment	(–)***	(–)***	(–)	(–)**
Assault	(–)***	(–)	(–)	(–)**
Robbery	(–)***	(–)	(–)	(–)***

Drug	(–)	(+)***	(–)**	(–)***
Income				
Property	(+)***	(+)***	(+)	(+)***
Harassment	(+)***	(–)**	(+)	(+)
Assault	(+)***	(–)***	(+)	(–)
Robbery	(+)***	(–)	(–)	(–)
Drug	(+)***	(–)***	(–)	(+)
Residential ownership				
Property	(+)	(+)***	(–)***	(–)***
Harassment	(–)	(–)*	(–)**	(–)***
Assault	(+)	(+)**	(–)***	(–)**
Robbery	(+)	(+)	(–)***	(–)***
Drug	(+)***	(–)***	(–)	(–)**
Racial heterogeneity				
Property	(+)***	(+)***	(+)***	(+)*
Harassment	(+)***	(+)***	(+)***	(+)***
Assault	(+)***	(+)***	(+)***	(+)***
Robbery	(+)	(+)*	(+)***	(+)
Drug	(–)	(+)***	(+)***	(+)***
Black young male				
Property	(+)***	(–)***	(–)***	(–)***
Harassment	(+)***	(–)***	(–)	(–)**
Assault	(+)***	(–)***	(–)***	(–)***
Robbery	(+)***	(–)***	(+)	(+)***
Drug	(+)***	(+)	(–)	(+)***
Hispanic young male				
Property	(+)	(–)***	(–)***	(+)**
Harassment	(+)***	(+)	(–)***	(+)
Assault	(+)***	(+)	(–)**	(+)***
Robbery	(+)***	(+)**	(–)***	(+)**
Drug	(+)***	(+)	(–)***	(–)***

** p<.1; ** p<.05; *** p<.01*

Table 8: Summary of all types of crime: by weighted independent variables

	Private	Semiprivate	Semipublic	Public
W-residential population				

Property	(+)	(–)**	(–)	(–)***
Harassment	(+)***	(–)	(+)	(–)***
Assault	(+)	(–)***	(–)*	(–)***
Robbery	(+)	(–)***	(–)***	(–)*
Drug	(–)	(–)*	(–)***	(–)
W-dynamic population				
Property	(–)	(+)***	(+)***	(+)***
Harassment	(–)**	(+)***	(+)***	(+)***
Assault	(–)	(+)***	(+)***	(+)***
Robbery	(–)	(+)***	(+)	(+)
Drug	(–)***	(–)***	(+)***	(–)***
W-education level				
Property	(–)*	(–)***	(–)***	(–)***
Harassment	(+)	(–)	(–)***	(–)**
Assault	(+)	(–)	(–)***	(–)
Robbery	(–)	(–)	(–)	(+)
Drug	(–)***	(–)*	(–)***	(–)***
W-employment status				
Property	(–)*	(–)***	(–)***	(–)***
Harassment	(–)	(+)	(–)	(–)
Assault	(+)	(–)***	(+)	(–)***
Robbery	(+)*	(–)	(–)	(–)**
Drug	(+)***	(+)	(+)	(+)**
W-income				
Property	(+)	(+)***	(+)***	(+)***
Harassment	(–)	(+)	(+)***	(+)***
Assault	(+)**	(+)***	(+)***	(+)***
Robbery	(+)	(+)***	(+)***	(+)***
Drug	(+)***	(+)**	(+)***	(+)***
W-residential ownership				
Property	(+)	(+)***	(+)*	(+)***
Harassment	(–)	(+)	(+)	(+)***
Assault	(–)	(+)***	(+)***	(+)***
Robbery	(+)	(+)	(+)	(+)
Drug	(–)	(–)	(+)***	(+)***
W-racial heterogeneity				
Property	(–)*	(–)***	(–)	(+)**

Harassment	(–)	(–)***	(–)*	(–)*
Assault	(–)**	(–)	(–)***	(–)***
Robbery	(–)*	(+)	(–)***	(+)
Drug	(–)**	(–)***	(–)	(–)*
W-black young male				
Property	(+)***	(+)***	(+)*	(+)
Harassment	(+)***	(+)***	(+)	(+)***
Assault	(+)***	(+)***	(+)***	(+)***
Robbery	(+)***	(+)***	(+)*	(+)***
Drug	(+)	(–)	(+)	(+)***
W-hispanic young male				
Property	(–)	(+)***	(–)	(–)***
Harassment	(–)***	(–)***	(+)*	(–)**
Assault	(–)***	(–)***	(+)	(–)***
Robbery	(–)	(–)***	(+)**	(–)
Drug	(–)***	(–)	(–)	(+)***

** p<.1; ** p<.05; *** p<.01*

Table 9: Property crime at private places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.132***	0.144***		0.130***
Dynamic population	0.061***	0.046***		0.046***
Education level	-0.024		-0.019	-0.011
Employment status	-0.021***		-0.017***	-0.019***
Income	0.019***		0.018***	0.019***
Residential ownership	0.002		0.003	0.000
Racial heterogeneity	0.010***		0.027***	0.018***
Black young male	0.316***		0.173***	0.192***
Hispanic young male	0.028		0.048	0.018
Constant	0.003			
Wx				
Residential population		0.063***		0.018
Dynamic population		-0.032***		-0.014
Education level			-0.033	-0.074*
Employment status			-0.014*	-0.012
Income			-0.005	0.006

Residential ownership	0.009**	0.004
Racial heterogeneity	-0.010	-0.013*
Black young male	0.208***	0.173***
Hispanic young male	-0.052	-0.003

** p<.1; ** p<.05; *** p<.01*

5.5 DISCUSSION

This research was designed to explore whether the dynamic population data collected from social media – not fixed values of the place like traditional census data – would be a useful new source of information for future crime research. The current chapter utilised Twitter data to examine the relationship between fluctuations in available population and crime and, critically, is the first research that disaggregated this for different place contexts and crime types.

The rhythm of places is shaped by the occupiers of the places. Houses are often unoccupied during day and streets are sparsely populated at night-time. As each individual has a regular routine, each place has a routine and crime occurring at a place is affected by the routine of that place. Therefore, measuring and understanding the routine of places - the fluctuations of their population - allows crime scientists to explore the complex relationships between crime and the rhythm of the population.

From the three analyses presented above, a noticeable contrast was found between crimes at private places and crimes at places where there is public access, and between property crimes and violent crime including robbery and assault. For property crime, a strong positive relationship between dynamic population and crime was found when it occurs at public places, but property crime at private and semiprivate places which indicates a higher level of personal guardianship showed a different outcome. For the first two analyses, at private places, property crime decreased when the population increased. Past crime pattern research argues residential burglars are more likely to operate in the morning hours when the places are unoccupied, the children are at school and the parents at work (Walsh, 1980; Lawrence and Cantor, 1982). We can assume theoretically the increased population would be a guardian in the crime situation and eventually reduce criminal opportunities.

According to the first two analyses, unlike property crime, violent crime at private places increased when the population increases. In the previous research on violent crime, researchers have argued that violent crime is context-sensitive crime and has a strong correlation with subculture or local-culture, which means the violent crime should be understood in the cultural context of people who reside in the area (Wolfgang and Ferracuti, 1967; Daly and Wilson, 1989; Buss and Dantely, 2007). As statistics also suggest, serious violent crime is more likely to occur between those in intimate relationships than property crime and the degree of intimacy between offenders and victims is an important variable in violent crime situations (Dawson, 2006; National Crime Victimisation Survey (NCVS), 2019). For violent crime at private places, people who reside there can potentially be a criminal, a victim and even take a role as a guardian. Violent crime, however, produced a different pattern at places where the public can access depending on the degree of violence. Dynamic population increase in these places leads to less harassment crime but more assault, which can be interpreted that the flow of the population might act as a guardian in harassment crime but as more offender-target convergence in assault crime.

For drug crime, the dynamic population only had a marginal or negatively significant effect on crime. Although the crime is concentrated at micro places and specific situations (Weisburd and Mazerolle, 2000), markets can operate independent of the general population. Drug crime is defined as drug possession or use, violent behaviour resulting from drug effect, violence against drug dealers or rival drug dealers and so on (Bureau of Justice Statistic, 2007). Technically, it is crime committed by a specific group of people who sell/buy and use the illegal item (Curtis, 2016) therefore the crime is less likely to be influenced by the dynamic population who represent the regular movements of the majority of people.

The result of the SDMs both supports and conflicts with some of the findings of the earlier analyses including a visual assessment and CCA analysis. However, it does demonstrate a differential effect of the dynamic population on crime in private spaces compared to other types of crime. In other places, it is positively correlated with crime across the board. This is a more consistent result than for the census residential population and demonstrates that the dynamic population appears to have particularly good discriminatory power at explaining the relationship between crime and population within different types of places.

Future research may consider other spatiotemporal units which may generate different outcomes. As with many of the initial studies on the relationship between dynamic population and crime (see Malleson and Anderson,2015a; 2015b; 2016), the research here has used the census unit that has been used fairly frequently in the crime literature for ease of modeling and comparison. The fact that private spaces differ in crime profiles over time from other spaces demonstrates that the unit of analysis selected is likely to have a large impact on results and this could usefully be investigated as a follow-up study.

6 MOBILE POPULATION EMOTION PATTERNS

6.1 INTRODUCTION

Emotions are a snapshot of human behaviours. We all experience a range of emotions as an automatic appraisal of stimuli that we are facing in our daily lives, and emotions take an inherent part in courses of action in particular situation recognition, interactions with others, and decision making processes (Smith and Ellsworth, 1985; De Rivera, 1992; Watson and Spence, 2006). In that sense, crime scientists have come to a consensus that emotions are intrinsically related to behavioural and cognitive process of reaching a criminal choice or creating a criminogenic environment, and some effort has been made to integrate emotional aspects into criminology to fill the holes that traditional theories leaves (Van Gelder et al, 2014; Wortley, 2001). Up until recently, the effort has been focused on theorising conceptual frameworks generally produced by qualitatively-oriented criminologists with narrative and ethnographic approaches, and the empirical studies have not kept pace with the conceptual development (Van Gelder et al, 2014; Walters, 2015). This delay is attributed to a lack of data which can be quantified in a way that reveals emotions in a moment of crime, and also a difficulty in producing simple and valid measurement methodologies.

Unlike previous data collected using qualitative methods, social media data can capture emotions, which are exhibited in day-to-day online communications, of a large population spatio-temporally. The data is not limited to representing criminals' emotions but also captures emotions of other actors who are involved in immediate situations of crime. Analysing these emotions, therefore, can provide a more holistic view of the effect of emotional settings on crime which has not been fully analysed in spite of a well-established theoretical basis on the emotion-behaviour interaction (which is detailed below in 'Background literature and theoretical rationale' of the current research). The research described in this chapter has two main aims. The first is to understand micro-situations of crime, particularly employing crime behaviour settings which is repeatedly observed in a spatiotemporal unit, in relation to emotions exhibited in social media posts. The second is to propose a conceptual framework on how such data should be understood and utilised within criminological approaches as a potential source of future crime prediction research.

6.2 Background literature and theoretical rationale

6.2.1 *Emotions as a mirror of crime situation*

In spite of the evidence that emotional processes play a fundamental role in crime and crime situations, these processes have not received enough attention in criminology (Van Gelder, 2014; Van Gelder, 2016). Because the role of emotions in criminal behaviours - either indirectly mediated by emotions or directly caused by emotions - has been discussed broadly (but often briefly) in criminological theories, there was a need to be selective in this research when building a conceptual framework. The following review covers concepts which discuss emotional states as a vital element of criminal behaviours at two distinguishable levels; individual emotions and collective emotions, and how the emotional status of actors - criminals, victims, as well as third parties who are obliquely involved in the event as a background - influence crime situations.

Emotions as mediators or precipitators of criminal choice making

Unlike deterministic perspectives, rational choice perspectives explain crimes as a result of choices made by rational decision makers. As an economic approach, 'rational choice' was built upon the utilitarian foundation which perceives a human decision as the all-things-considered achievement of goals (Baron, 1999, as cited in Van Gelder et al.2014). Rational choice perspectives describe crime behaviours as an outcome of reasoned decisions made by would-be offenders who seek to benefit themselves through committing crime (Clarke and Cornish, 1985; Leclerc and Wortley, 2013; Van Gelder et al., 2014).

Clarke and Cornish (1985) claimed that the offenders are active in the crime decision making process and are not passive actors whose behaviours are purely and easily predictable based on a given dispositional condition (Leclerc and Wortley, 2013). Hence, offenders rationally weigh up the risks and benefits before making a decision to commit an offence. As stated, Cornish and Clarke do however acknowledge that rationality is endangered by situational contexts such as emotions and cognition (they refer to 'bounded rationality'), but the original work does not integrate emotions more deeply into the theory (Walters, 2015; Wortley, 2015).

According to the studies inspired by rational choice approaches[1], rationality can be compromised by negative emotions such as severe anger and frustration, and constant stress, which eventually prevents rational and deliberate choice and leads to impulsively and risky decision making (Piquero and Tibbetts, 2002; Zafirovski, 2013). Topalli and Wright (2014) also reported that predatory street offenders who are exposed to constant stress due to a rough lifestyle are more likely to be involved in criminal behaviours as their emotional states drive them to be irrational in decision making process easily. In a study based on self-reported data from 76 robbery offenders, Lindergaard et al. (2014) found that the offenders are influenced by various emotions throughout criminal choice processes - in particular happiness, challenge, shame, anger and fear - and, criminals experience fear (before robbery) and happiness (after robbery) intensely while committing crimes (Lindegaard et al., 2014). Research with ordinary students as participants also suggests similar emotional influences - here that anger leads to unethical behaviours (Shalvi, Van Gelder, and Van Der Schalk, 2013). The study found that because aggressive emotions justify unethical behaviours, in this state people weigh the rewards and satisfaction from the behaviours heavily (Shalvi, Van Gelder and Van Der Schalk, 2013).

Richard Wortley (2001; 2006) argued that the person-situation interaction is more complex than rational choice approaches suggest and proposed the concept of situational precipitators of crime (Wortley, 2001; Wortley, 2006). Whereas the rational choice model explains the situational contexts such as emotional provocations and precipitators as elements of a 'cold' decision making process, Wortley argued that situational contexts can encourage certain internal states/dispositions of offenders and called for a modification of rational choice theory (Wortley, 2006; Newman and Freilich, 2013).

In Wortley's model of causation, immediate situational precipitators such as a series of stresses, frustrations and aggression were contemplated as a catalyst intensifying the motivation to commit crime

1 The role of emotions in decision making processes has been also studied in the field of cognitive neuroscience, explaining the differences in the decision caused by the neurocognitive capacities of individuals (Blakemore and Robbins, 2012; Van Gelder et al., 2014). A study on the relationship between emotional states and decision making found that when adolescents experience high arousal emotions and are under pressure, they are more likely to make more risky decisions than adults (Blakemore and Robbins, 2012). Although the concept researched within psychological perspectives are very important for studying crimes of certain groups (i.e. juvenile, the mentally disabled), it was not reviewed in the current study which was conducted in the criminological perspectives and not targeting a relevant group.

(Wortley, 2006). This idea has been supported by a substantial amount of empirical studies. According to a series of Homel's studies, situational and emotional stimuli at night clubs including overcrowding, interactions with rude patrons, and intoxication are strongly linked with a high risk of aggression and subsequent violent crimes (Homel et al., 1992; Macintyre and Homel, 1997; Graham and Homel, 2012). The following studies on bars/pubs or sport-related violence also found that immediate environmental conditions such as rowdiness or a permissive environment (Graham et al., 2006); intoxication (Wells and Graham, 2009); frustration caused by a bad sports performance (Graham et al., 1997; Priks, 2010); and high levels of arousal due to alcohol consumption (Fortes, 2013; Fitzpatrick, 2015) raise severe aggression and violent behaviours (Graham et al., 1997; Graham et al, 2006; Wells and Graham, 2009; Priks, 2010; Fortes, 2013; Fitzpatrick, 2015).

Emotions as a prerequisite of criminal behaviours

While in rational choice approaches, emotions act as mediators or precipitators in the process of criminal decisions, other scholars perceived emotions, in particular, aversive and aggressive emotions, as an essence forming individual disposition. Stated differently, the former states that emotions create immediate criminogenic situations which trigger criminal motivation, but the latter propose that emotions shape distinct criminality, which in turn, would make some individuals more likely to become aggressive and engage a negative action.

General strain theory (GST) says strain such as stress, anger, frustration, and depression motivates criminal actions as one method of coping with these negative emotions (Agnew, 1992). Agnew (2006) explained the mechanism that such strains tend to create low self/social control, increasing the incentive for coping through criminality and leading to the alleviation of strain through the commission of criminal behaviours (Agnew, 1992).

GST also incorporates ideas from Merton's (1938) strain theory which argues that strain can be caused by structural constraints that exist in our society. Robert Merton (1983) stated that when individuals cannot achieve a cultural goal that is socially accepted (i.e. American dream, success and material wealth) by socially-acceptable means, the pressure they feel can lead to deviant behaviours. Merton (1938) briefly and indirectly mentioned how emotions act when individuals respond to this strain but did not precisely explain why and how emotion is associated with the behaviours. According to Merton (1938), greater stress and strain in achieving desired goals vitiates emotional commitment to institutional rules, which eventually leads to deviant behaviours (1938).

Agnew (1992) focused on the strain itself. Rather than discussing frustration of not being able to achieve a goal in the societal structure, he suggested a source of frustration from individual level circumstances such as relationships with others, especially family and friends (Agnew, 1992).

In follow-up research, Agnew (2012) introduced a concept, *criminogenic strains* and argued that a certain type of strain contributes to criminogenic traits more than others. Depending on sources and magnitude (intensity, frequency and continuity) of negative emotions, certain types of emotions would be perceived of as more severe and unjust and generate much anger (Agnew, 2001; Agnew, 2015), which eventually increase likelihood of coping with the emotions through criminal behaviours (Agnew, 2017). He also noted that some types of aversive emotions can reduce crime if they are not overly harsh and are raised in close and trusting relationships (Agnew, 2015).

Additionally, Agnew differentiated between short- and long-lived emotions. Whereas Agnew focused on *state anger* (the frequency of angry episodes) which refers to a transitory emotional condition in response to a frustrating situation, *trait anger* (long-term or chronic anger) representing a personal disposition toward strain, was also discussed in the theoretical accounts (Agnew, 1995; Jang and Agnew, 2015). Jang and Agnew (2015) noted a correlation between trait anger and state anger as individuals with different levels of trait anger would react differently in frustrating situations. For example, individuals who have high degrees of trait anger easily and intensively exhibit state anger more frequently than individuals with lower levels of trait anger (Jang and Agnew, 2015).

Emotions as an element shaping criminal situations

A further theoretical approach focuses on understanding collective emotions toward crime and their effect on neighbourhood crime rates. Unlike the individual-level perspectives, which research the effects of emotions on a specific criminal act, collective level approaches are mainly interested in how emotions shape collective behaviours of a group of people and how these influence crime situations. At the neighbourhood level, a specific emotion - fear of becoming a victim - has been most intensively studied in the past decades. This approach suggests that emotions such as a high level of worry and fear reduce the capacity for informal social control in a neighbourhood, which is more likely to lead to higher crime rates (Garofalo, 1981; Wilson and Kelling, 1982; Baumer, 1985; Markowitz et al., 2001; Taylor, 2001; Robinson et al., 2003).

This framework was built on social disorganisation theory by subdividing a path linking community social/physical disorder and crime. In the rationale negative emotions among residents created from neighbourhood disorder make the residents not want to live the neighbourhood (and physically withdraw) and less likely to intervene in antisocial behaviour (hence socially withdraw). In turn this weakens the capacity of the neighbourhood to deliver social control and collective efficacy, which eventually escalates neighbourhood crime rates as deviancy goes unchallenged (Skogan, 1990; Sampson et al., 1997; Sampson and Raudenbush, 1999; Morenoff et al., 2001; Hipp, 2007).

Although these theories emphasise the role of emotions creating criminal environments, most empirical studies have focused on the causes of fear or its effects on social integration and not necessarily examined the emotion itself in relation to actual crime levels. For example, Jackson and Gray (2009) noted people's behavioural patterns such as locking up the entrance well, Bellair (2000) observed an increased awareness about victimisation, and Melde, Berg, and Esbensen (2016) found constraining people's routines such as avoid socialising with delinquent friends or exposure to risky social settings, but these assume rather than measure the fear that has caused these behavioural changes.

There are some studies that have conceptualised the influence of fear on a decrease in crime rates in a certain type of neighbourhoods (Hipp, 2007; St. Jean, 2007). Hipp (2007) argued that the residents of socially disadvantaged neighbourhoods cannot physically withdraw from where they live, therefore they are forced to maintain informal social control even with high negative emotions toward the neighbourhood and are thus more likely to intervene criminal behaviours. He, however, underlines that the assumed mechanism should carefully consider the physical and social structure of neighbourhoods. By contrast, Agnew (2009) asserts that collective negative emotions will increase crime. He argued that collective strains instigate more negative emotions among individuals in group settings which reduces social control, which is a necessary feature in keeping individuals in legal channels (Agnew et al., 2009)

To sum up, the theories reviewed in the current research argue that emotions are involved in different stages of crime behaviours including developing criminal inclination (Agnew, 1992), precipitating criminal motivation (Wortley, 2001), mediating criminal decisions (Clarke and Cornish, 1985) and lastly creating criminal environments by collective behaviours (Sampson et al., 1997; Hipp, 2007). The theories also highlight that emotions have distinct relationships with crime depending on types and magnitude of emotions. For instance, some type/level of emotions encourage criminal actions because the emotions lower concern over the consequences of the actions and the ability to legally cope (Katz, 1988; Collins, 2008; Agnew, 1992; Agnew, 2015). Others develop crime preventive coping strategies which minimise the possibility of becoming a victim of crime (Swaray, 2006; Hipp, 2007).

6.2.2 *Emotional analysis for crime prediction*

A number of empirical studies analysing emotions in social media texts have been published, the majority dealing with public feedback toward societal topics or products. Examples are emotions in response to political issues such as elections and political orientation (O'Connor et al., 2010; Tumasjan et al.,2011; Preotiuc-Pietro et al., 2017), financial issues such as the stock market (Bollen et al., 2011; Sul, Dennis, and Yuan, 2017; Chen and Chen, 2019), breaking news or popular events -such as the death of Michael Jackson, Oscars, and Tiger Woods' confessions (Kim et al., 2009; Thelwall, Buckley, and Paltoglou, 2010; Lee and Goh. 2013)- or health-related issues such as suicide and vaccination (Ueda et al., 2017; Blankenship et al., 2018). In general, detecting emotions hidden in the texts focuses on exploring the data as a possible supplemental tool to survey emotional opinions toward such topics.

Some studies simply classify emotions into two sentiment groups, positive and negative to measure public sentiment toward popular events (Thelwall, Buckley, and Paltoglou, 2010). Bollen, Mao, and Pepe (2011) used six-dimensional mood including tension, depression, anger, vigor, fatigue, confusion to capture collective emotional trends on socio-economic events such as the U.S. presidential campaign and election, the failure of several large international banks and the Dow Jones significant dropping (Bollen, Mao, and Pepe, 2011). Bermingham and Smeaton (2010) categorised three sentiment groups (positive, negative, mixed) and a non-sentiment group (neutral) to monitor public sentiment on political campaigns and elections. Kim and her colleagues used the Affective Norms for English Words (ANEW) database which consist of 1,034 words representing three emotional dimensions; valence (pleasure vs. displeasure), arousal (excitement vs. calmness), and dominance (strength vs. weakness) to measure public emotions on breaking news (Kim et al., 2009). While such studies employed a simplified classification to capture the contrast between polarised emotions such as like or dislike, more recent studies have focused on developing algorithms reflecting basic emotions such as anger, disgust, fear, happiness, sadness, and surprise and explored their validity in capturing emotions on social media texts (Roberts et al., 2012). These studies demonstrate that there is a precedent for using social media data to represent human emotions.

6.3 methods

The two core datasets have already been described in Chapter 4. To recap, these are the crime data and social media feeds in the region of

Note: see the section about 'Data' in Chapter 4 (p.36)

NYC between the dates of 1st of January, 2013 and 31st of December, 2013. The research in this chapter required production of new datasets that can measure emotions using social media data as the source. This involves the extraction of these emotions from each post and then the aggregation of the emotions into the selected spatiotemporal unit.

6.3.1 *Measurement*

Emotion variables

The emotion variables are measured by IBM Watson Tone Analyser[2] which extracts probable emotions from text fields. The program returns the results in two different categories, an emotional tone and a language tone [3] and the current research only utilised the emotion tone. The emotional tone consists of four emotions, *Anger, Fear, Joy* and *Sadness* and the definition of each emotion is given below.

NOTE
Some of the language which is used in the following research is offensive. Given the informal nature of messages of social media, the messages include swearwords in the expression of emotions, it was necessary for clarity, to refer to these in their raw form.

Anger Anger is evoked due to injustice, conflict, humiliation, negligence, or betrayal. If anger is active, the individual attacks the target, verbally or physically. If anger is passive, the person silently sulks and feels tension and hostility.

2 IBM Watson Tone Analyser (www.ibm.com/watson/services/tone-analyzer/) is the psycholinguistic analysis tool pioneered by IMB and designed for detecting and interpreting emotional cues observed in texts including punctuation and emoticons. The program was originally proposed for a commercial purpose such as enhancing customer service or understanding people's feedback on brands and products newly realised.

3 Language tones in IBM Watson Tone Analyser is not included in the research. Although emotions and feelings arise in momentary response to stimuli (Brown, 2003), tones are close to attitude toward a subject displaying a writer's intentions (Sinha, 2006) so that the examples of the tone would be polite, sarcastic, arrogant, kind and so on. According the program description provided IBM, the language tone provides how analytical, confident and tentative one's post is (Sinha, 2006).

Analytical An analytical tone indicates a person's reasoning and analytical attitude about things. An analytical person might be perceived as intellectual, rational, systematic, emotionless, or impersonal

Confident A confident tone indicates a person's degree of certainty. A confident person might be perceived as assured, collected, hopeful, or egotistical

Tentative A tentative tone indicates a person's degree of inhibition. A tentative person might be perceived as questionable, doubtful, or debatable

Examples: *"I'm so fucking stressed out omg", "It's so annoying how the school doesn't let us in until 8:30 on the dot Because of that we are all technically late when we sign in stupid", "So pissed off This line for the bus is ridiculous"*

Fear Fear is a response to impending danger. It is a survival mechanism that is triggered as a reaction to some negative stimulus. Fear can be a mild caution or an extreme phobia.

Examples: *"Scary streets", "I know that wasn't an earthquake. Why this building shaking?", "At this point I'm scared"*

Joy (Happiness) Joy has shades of enjoyment, satisfaction, and pleasure. Joy brings a sense of well-being, inner peace, love, safety, and contentment.

Examples: *"Watching friends still laying on the couch lol", "That nap was what I needed Feel good now", "Finally got a new phone charger"*

Sadness Sadness indicates a feeling of loss and disadvantage. When a person is quiet, less energetic, and withdrawn, it can be inferred that they feel sadness.

Examples: *"Late work night cancelled my social plans but at least I can get a workout in", "do i even exist anymore", "People have insecurities that others won't ever be able to fix :("*

Swearing Variable

The lists of swear words were imported from Thelwall's 'swear word selection' (Table 10) (Thelwall, 2008). The word was composed of nearly 100 swear words and classified by strength of the swearing words guided by the British Board of Film Classification (Thelwall, 2008; quoted in McEnery, 2005). In this research, the strength scale was not considered as all swearing words were simply merged as an indicator of aggressive/negative emotions.

Table 10: A list of swear words

Swear words	Strength
Cunt, jew, motherfuckin, motherfucking, muthafucker, muthafuckin, mutherfucker, nigga, niggah, niggas, niggaz, nigger, nigguh, paki	Very strong
Fuck, fucked, fucken, fucker, fuckin, fucking, fuckstick, spastic	Strong
Arsehole, asshole, bastard, bollock, cock, dick, gay, piss, pissin, pissing, poof, poofter, poofy, prick, pussy, queer, shag, shagged, shagging, twat, wank, wanker, wanking, whore	Moderate

Arse, arsed, ass, bitch, bugger, butthole, christ, cow, dickhead, dipshit, fanny, fart, jesus, moron, pissed, retard, screw, screwed, screwing, shit, shite, shithead, shittin, shitty, slag, slagged, slut, tit, titties, tosser	Mild
Bap, bimbo, bird, bloody, bonk, bonking, boob, bullshit, butt, butthead, crap, damn, dork, dorky, git, god, hell, hussy, idiot, jerk, jug, knocker, pig, pillock, pimp, sod, tart, tarty, turd, wuss	Very mild

Emotional atmosphere variable

An emotional atmosphere variable was calculated using the four emotions (anger, fear, joy and sadness) using Shannon's equitability index (E_H), which is widely used to measure species richness and diversity in a given community (Shannon and Weaver, 1949; Spellerberg and Fedor, 2003). The index takes the number of emotions present (richness) and the number of individuals per emotion (abundance) into account and provides a mathematical value of abundance and evenness of the emotions present (Shannon and Weaver, 1949; Sheldon, 1969). The equitability index is calculated by dividing H (the Shannon's diversity index) by H_{max} (H_{max}=lnS). The proportion of individual emotion ($p_i = n/N$) is calculated by the number of one particular emotion found (n) and the total number of individuals found in a given spatiotemporal unit (N). The resulting value is between 0 and 1 and higher values indicate that emotions are more evenly distributed and that a unit has more diverse emotions.

$$E_H = H/H_{max} = H/lnS \tag{3}$$

$$H = -\sum_{i=1}^{s} p_i ln p_i$$

6.3.2 *Analysis methods*

For the units of analysis, street segments and hour (24 hours) were chosen. The final data consisted of 11,836 geographic units by 24 temporal units between the 1st time slot (00:00-00:59) to 24th time slot (23:00-23:59). The analysis began with the identification of emotions using IBM Watson Tone analyser. Each message was recorded against the corresponding emotion if the program detected it from the text and the emotion variables were coded dichotomously. If a text obtained multiple emotions, all identified emotions by the program were recorded. The number of incidents of each emotion were then aggregated to street segments for each of the temporal units, along

Note: see the section about 'Research Units' in Chapter 4 (p.42)

with a simple count of the total number of tweets. After completing the coding process, the sum of each emotion and the equitability index of each spatiotemporal unit was calculated. Collectively, the final data generated included the number of crimes and three formats for each emotion in each street over time; 1) the sum of each emotion (Anger, Fear, Joy(Happiness) and Sadness), 2) the sum of swearing and 3) emotional evenness (Figure 26). A panel data model (two-way fixed effects negative binomial model) was chosen to estimate the effects of the emotion variables on crime.

Figure 26: The list of independent variables

Panel data model

The count of the number of crimes that occurred in each street segment in each time unit was the dependent variables (five crimes at four different places; property, harassment, assault, robbery, and drug crime at private, semiprivate, public and semipublic) in the models. As crimes are rare events and the data is counts, the data contains considerable amounts of zero which create over-dispersion in a Poisson model. Therefore, negative binomial modelling [4] is selected.

The model considers omitted variables which are constant over time and varies between streets but not within streets, and those which can possibly vary over time but are constant across all street segments (Charbonneau, 2012; Jorgenson, Rice and Clark, 2012). In considering the spatial patterns of crimes, there is a strong assumption that crime would be linked to place-specific attributes which do not vary over time. Recent research has found that a large proportion of crime is concentrated at a limited number of street segments (Weisburd, 2015) and these findings suggest that the model should not ignore the time-invariant latent characteristics of each street. There is also a possibility of time-specific unobservables on crime which could make

4 A negative binomial distribution of panel data with expected value and variance are given by $E(Y_{it}) = \lambda_{it}$, $var(Y_{it}) = \lambda_{it}(1 + \theta_i)$. Unlike the mean and the variance are both equal (λ_{it}) in the Poisson distribution, the variance of the negative binomial distribution exceeds the mean, over-dispersed. From the over-dispersion parameter test results, the null hypothesis (H_0:α=0) was rejected in all cases indicating that the model should use negative binomial distribution. When comparing the results with Poisson and negative binomial model has higher the values of log-likelihood

an omitted variable bias in the model. For the crime data, a cyclic fluctuation in criminal opportunities would generate bias (Raphael and Winter-Ebmer, 2001), hence mitigating unobserved time effects such as time/day/year effects has been considered important to the production of reliable statistical results (Baltagi, 2006).

Both error terms were considered as fixed effects. Intuitively, it is hard to assume the error terms would be drawn independently from probability distributions because the data represents all individuals in the target area not a randomly selected sample. The F-test[5] and Hausman's specification test results also showed that fixed effects modelling is suitable for in this case rather than random effects [6]. These results justified the extension of the model to incorporate geographical and time-period fixed effects.

After all, two-way fixed effects negative binomial model [7] was chosen as appropriate and is expressed as follows (Baltagi, 2008).

$$Y_{it} = \beta X_{it} + \varepsilon_{it} \tag{4}$$

$$\varepsilon_{it} = \mu_i + v_t + e_{it}$$

where $_i$ denotes an index of the spatial units with i=1,..., N, and t is a temporal index with t=1, .., N. Y_{it} denotes the dependent variable, X_{it} represents each explanatory variable, and β is the coefficient for that independent variables. For the error terms, μ_i is unobserved individual effects, v_t controls unobserved time effects, and e_{it} is an error term which varies by both time and individual.

6.4 RESULTS

6.4.1 *Descriptive statistics*

Among the total messages collected in the target area, any emotions were observed for 49.51 per cent of the total posts (257,216 out of 519,561 posts). The four emotions - anger, fear, joy and sadness - were collected from 36.93 per cent of the total and 866 posts have more than two more emotions (Table 11). Joy was the most prevalent (25.58

5 In all cases, the null hypothesis ($H_0 : \mu_i = 0$ and $H_0 : v_t = 0$) of the F-test was rejected. It confirms the presence of group and time fixed effects which would bias the estimates of the relationship.

6 The null hypothesis of Hausman test (H_0 :cov(x_{it}, μ_t)=0) was rejected in all cases. The results indicate that the estimation using random effects would include endogenous bias (see Appendix (p.165) for the results of Hausman's specification test (Table. 53)).

7 The time specific effects were included into the model as dummy variables and the total social media posts created each analysis unit was included as a control variable. The model also controls/drops the streets with no crime or dynamic population in the analysis as it considers them as an time-invariant values/stable characteristics.

percent of the total posts) followed by sadness (7.04 percent), anger (3.42 percent), and fear (0.89 percent). Swearing was observed in 26.1 percent of the total (Figure 27). Table 12 contains the descriptive statistics of emotional atmosphere, calculated by Shannon's equitability index.

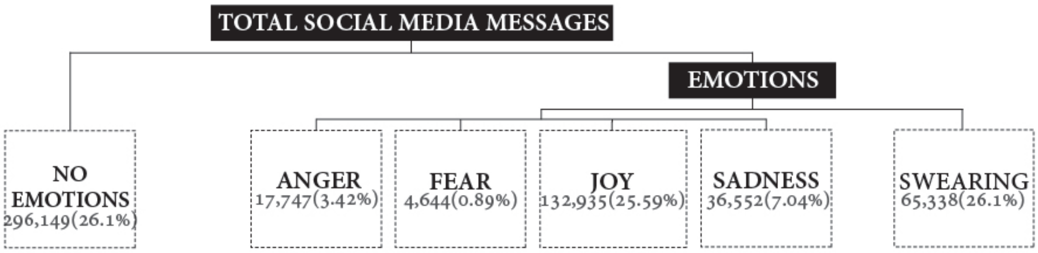


Figure 27: The list of emotion variables

Table 11: Descriptive Statistics: overlap of emotion variables

	Anger	Fear	Joy	Sadnesss
Anger	-			
Fear	93	-		
Joy	18	2	-	
Sadness	350	259	0	-

Table 12: Descriptive Statistics of emotional equitability index by crime types

Variable	obs	Mean	Std.Dev.	Min	Max
Property					
Private	2,047	0.28	0.23	0	1
Semiprivate	3,191	0.27	0.20	0	1
Semipublic	612	0.26	0.21	0	1
Public	3,260	0.28	0.22	0	1
Harassment					
Private	1,616	0.30	0.24	0	1
Semiprivate	1,261	0.26	0.20	0	1
Semipublic	183	0.27	0.23	0	1
Public	1,400	0.27	0.22	0	1

Assault					
Private	1,324	0.30	0.25	0	1
Semiprivate	1,034	0.27	0.21	0	1
Semipublic	193	0.30	0.25	0	1
Public	1,928	0.27	0.23	0	1
Robbery					
Private	395	0.30	0.26	0	1
Semiprivate	459	0.28	0.21	0	1
Semipublic	98	0.27	0.24	0	1
Public	1,027	0.28	0.24	0	1
Drug					
Private	360	0.30	0.28	0	1
Semiprivate	164	0.28	0.20	0	1
Semipublic	86	0.24	0.24	0	1
Public	1,314	0.28	0.25	0	1

6.4.2 *Panel analysis*

The pooled Ordinary Least Squares (OLS) estimation with all explanatory variables was used as a base model and a series of estimations using two-way fixed effects negative binomial (FENB) models followed. The second to fourth models comprised each base emotion (anger, fear, joy (happiness) and sadness) (FENB (A)), swearing (FENB (B)), and emotional equitability (FENB (C)). The last model (FENB (D)), the primary model of the current study, included all explanatory variables discussed in this research [8].

Table 13 shows the summarised results of FENB (D) for all crime types and the findings from OLS models were reported in Table 14). Most relevant theories make a distinction between transitory and stable emotions and argue that both emotions are profoundly interrelated to each other in crime situations. For example, when individuals are exposed to the same level of provocative stimuli, we cannot assume that they would respond in the same way because individuals accept/judge the stimuli differently. For these reasons, to compare the results of the relationship between emotions and crime with and without temporal effects, both OLS and FENB summary results were reported. Applying Agnew (1995)'s concepts, the results

8 To highlight the patterns found in analysis outcomes, the chapter reported 1) summary of findings (OLS and FENB (D)) and 2) findings (All five models) of property crime at private places as an example following the analytic process. See Appendix (p.165) for the analysis results of other 19 crime types.

of FENB and OLS were labelled as emotional atmosphere which refers to transitory emotional condition on crime (FENB in Table 13), and emotional traits which indicates general tendency (OLS in Table 14), the relationship between emotions and crime rates of places with no temporal perspectives. The OLS and all FENB (A, B, C, and D) results of each crime reported in Table 15 and in Tables 54 to 72 in Appendix (p.165) together with coefficients, incidence rate ratios (IRR) and *P*-values.

Emotional states and crime

The FENB (D) results (Table 13) showed anger has positive and significant effects on harassment crime at public places (IRR=1.072; p<.1) and assault crime at private (IRR=1.092; p<.05) and semiprivate places (IRR=1.116; p<.01), but negative and significant effects on property crime at semiprivate places (IRR=0.972; p<.1) and public places (IRR=0.926; p<.01). It also can be read that for every unit increase in anger, there were 9.2 per cent increase in the incidence of assault crime at private places, 11.6 per cent increase in the incidence of assault crime at semiprivate places and 2.8 per cent and 7.4 per cent decrease in the incidence of property crime at semiprivate and public places respectively. The results indicated that anger is more likely to be related to violent crimes but less likely to be correlated with property crimes.

For fear, negative and significant effects were found at assault crimes at semiprivate places (IRR=0.721; p<.01), robbery crime at private places (IRR=0.491; p<0.1), and drug crime at semipublic places (IRR=0.610; p<0.1). Joy displayed positive and significant effects at property crimes at private (IRR=1.022; p<.05), semiprivate (IRR=1.026; p<.01), and public places (IRR=1.017; p<.01), which indicates that property crimes would be highly increased by joy, a positive emotion. Assault crime at semipublic places (IRR=1.049; p<.1) and drug crime at public places (IRR=1.024; p<.01) were also related positively with joy. Lastly, sadness had negative and significant effects on property crime at semiprivate places (IRR=0.982; p<.1) and positive effects on assault crime at public places (IRR=1.055; p<.01) and drug crime at semiprivate places (IRR=1.346; p<.01). The outcome showed fear tends to be associated with less crime, unlike sadness even though both emotions were categorised negative emotions.

Compared with the four base emotions, swearing and emotional evenness showed significant and consistent effects on crime especially with crimes at semiprivate places. Swearing variables were positive and significant at all crimes at semiprivate places (property (IRR=1.047; p<.01), harassment (IRR=1.034; p<.05), assault (IRR=1.041; p<.01), robbery (IRR=1.145; p<.01), and drug (IRR=1.085; p<.01)). The emotional equitableness was positively related with crimes at semiprivate places except for drug crime (property (b=1.540; p<.01), harassment (b=1.603; p<.05), assault (b=1.646; p<.01), and robbery (b=1.522; $p<.01$))[9], showing that evenness of emotions were found to increase crimes at semiprivate.

Because emotion variables were not statistically significant in most cases, only a few recognisable patterns were observed from the FENB (D) model. Crimes at semiprivate, a place type that attracts many visitors, increased with more swearing and emotional evenness, which indicates the moment when a place has diverse emotions, not a single emotion, the likelihood of crime increases. Most importantly, the findings from FENB suggested that groups of negative emotions do not have a consistent impact on crime and generate distinct influences on crime. For example, assault crime at semiprivate places increased by 11.6 per cent in anger but decreased by 27.9 per cent in fear. Anger was also related to violent crime and property crime differently. The effects of anger were positive at assault crime but negative with the number of property crime.

Emotional traits and crime

Table 14 shows the summary results from the OLS analysis. Without taking into account the temporal effects, all emotion variables generated more significant results in general.

For every unit increase in Anger, property crime decreased by 11.3 per cent at semiprivate (IRR=0.887; $p<.01$), 20.9 per cent at semipublic (IRR=0.791; $p<.01$), and 14 per cent at public places (IRR=0.860; $p<.01$) but assault crime at private places and drug crime at private places increased by 25.9 per cent (IRR=1.259; $p<.01$) and 44.9 per cent (IRR=1.449; $p<.01$) respectively.

Fear had negative and significant effects on property crime at public places (IRR=0.901; $p<.1$), violent crimes, harassment and assault at semiprivate places (harassment (IRR=0.826; $p<.05$) and assault (IRR=0.618; $p<.01$)) and robbery crime at private places (IRR=0.377; $p<.01$) but had positive effects on robbery at semipublic places (IRR=1.739; $p<.1$). These results are supported by previous research that has demonstrated the positive effects of fear on crime which leads to a constraining risk behaviours and strengthening social cohesion at the collective level (Bellair, 2000; Hipp, 2007; Melde, Berg, and Esbensen, 2016). In particular, the positive role of fear was found in violent crimes at semiprivate places.

In contrast with the results of anger and fear, joy and sadness were found to be positively associated with crimes. Joy increases property, harassment, and drug crimes at all types of places, and crimes at places allowing public access. Sadness was found to increase crimes at private places except for drug crime, which means that private places with a sad climate are more likely to experience crimes. From the

9 The IRR values of the equitability index are not reported. Since the index is a ratio not an absolute amount like other dependent and independent variables, IRR values which represent the effect of every unit increase on dependent variables did not generate meaningful outcomes.

OLS results, swearing appears to have positive effects on crimes at semiprivate places, displaying the same patterns found in the FENB results. The equitability of emotions was positively related at most crime cases, which indicates that the balance of emotions is interrelated with crimes positively even when time effects were not included. This results generally suggest higher emotional equitability of places can lead to more crime.

Table 13: Summary of all types of crime: by the types of emotion (FENB (D))

	Private	**Semiprivate**	**Semipublic**	**Public**
Anger				
Property	(–)	(–)*	(+)	(–)***
Harassment	(+)	(+)	(–)	(+)*
Assault	(+)**	(+)***	(–)	(–)
Robbery	(+)	(+)	(–)	(+)
Drug	(+)	(+)	(–)	(+)
Fear				
Property	(–)	(–)	(–)	(+)
Harassment	(+)	(–)	(+)	(–)
Assault	(–)	(–)***	(–)	(+)
Robbery	(–)*	(+)	(+)	(+)
Drug	(+)	(+)	(–)*	(–)
Joy				
Property	(+)**	(+)***	(+)	(+)***
Harassment	(+)	(+)	(+)	(+)
Assault	(–)	(+)	(+)*	(+)
Robbery	(–)	(–)	(–)	(+)
Drug	(–)	(–)	(+)	(+)***
Sadness				
Property	(+)	(–)*	(–)	(–)
Harassment	(+)	(+)	(–)	(–)
Assault	(–)	(+)	(–)	(+)***
Robbery	(+)	(–)	(+)	(+)
Drug	(–)	(+)***	(+)	(+)
Swearing				
Property	(+)	(+)***	(+)	(+)***
Harassment	(–)	(+)**	(+)	(+)
Assault	(–)	(+)***	(–)	(+)

Robbery	(+)	(+)***	(+)*	(–)*
Drug	(+)	(+)**	(+)	(+)
Equitability				
Property	(+)	(+)***	(+)	(+)***
Harassment	(+)	(+)***	(+)	(+)
Assault	(+)	(+)***	(+)	(+)
Robbery	(+)	(+)**	(+)	(–)
Drug	(–)	(–)	(+)*	(+)

** p<.1; ** p<.05; *** p<.01*

Table 14: Summary of all types of crime: by the types of emotion (OLS)

	Private	Semiprivate	Semipublic	Public
Anger				
Property	(–)	(–)***	(–)***	(–)***
Harassment	(+)	(–)**	(–)	(–)
Assault	(+)***	(+)	(–)*	(–)
Robbery	(+)	(–)	(–)	(+)
Drug	(+)***	(+)	(–)	(+)
Fear				
Property	(–)	(+)	(+)	(–)*
Harassment	(–)	(–)**	(+)	(–)
Assault	(–)	(–)***	(+)	(–)
Robbery	(–)**	(–)	(+)*	(–)
Drug	(+)	(–)	(–)	(–)
Joy				
Property	(+)***	(+)***	(+)***	(+)***
Harassment	(+)***	(+)***	(+)***	(+)***
Assault	(–)	(+)***	(+)***	(+)***
Robbery	(–)	(+)***	(–)	(+)**
Drug	(–)***	(+)***	(+)***	(+)***
Sadness				
Property	(+)*	(+)	(+)	(–)
Harassment	(+)*	(+)	(–)	(–)
Assault	(+)*	(–)	(+)	(+)*
Robbery	(+)**	(–)**	(+)	(+)
Drug	(+)	(+)	(–)	(+)

Swearing				
Property	(+)*	(+)***	(+)	(+)***
Harassment	(–)	(+)***	(+)	(+)**
Assault	(+)	(+)***	(+)	(+)
Robbery	(+)	(+)**	(+)	(–)
Drug	(+)	(+)	(–)	(+)
Equitability				
Property	(+)***	(+)***	(+)***	(+)***
Harassment	(+)***	(+)***	(+)***	(+)***
Assault	(+)*	(+)***	(+)***	(+)***
Robbery	(+)	(+)***	(+)***	(+)***
Drug	(–)***	(+)***	(+)***	(+)***

** p<.1; ** p<.05; *** p<.01*

Table 15: Property crime at Private places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.047 (0.954)	-0.006 (0.994)			-0.013 (0.987)
Fear	-0.088 (0.916)	0.004 (1.004)			-0.007 (0.993)
Joy	0.059*** (1.061)	0.022** (1.023)			0.022** (1.022)
Sadness	0.044* (1.045)	0.011 (1.011)			0.006 (1.006)
Swearing	0.027* (1.027)		0.021** (1.021)		0.005 (1.005)
Equitability	2.687***			0.285	0.159

** p<.1; ** p<.05; *** p<.01*

6.5 DISCUSSION

Because emotions, as immediate reactions to given external stimuli, are constantly changing (Ekman, 1992; Plutchik, 2001), mining emotions at the moment of crime situations was not an easy task before social media data became available. The current research aimed to explore micro-crime situations, specially using spatiotemporal units, with real-time emotions collected from Twitter which would be expected to provide a more unique and in-depth understanding of crime situations. Such text expressions may be used to identify emotional atmosphere of crime situations created by people who would be more likely to make an irrational criminal decision, become a victim of crime and/or be willing to take a role as a potential guardian against crime.

As demonstrated by the work of Agnew (2012) and Wortley (2016), certain types of negative emotion such as anger and frustration might be more strongly related to crime more than other emotions with a relatively low magnitude of intensity and frequency. The results from this empirical research found that, in contrast to other negative emotion variables, fear was more likely to be correlated with crime. This finding is supported by some of the research cited in the literature review. These past studies explain that emotions related to fear and worry, adjusts people's preventive behaviours such as always being aware their surroundings and taking preventive actions to avoid being victimised (Bellair, 2000; Jackson and Gray, 2009).

It was also interesting to observe how anger works differentially in different crime situations. Anger was only positively associated with violent crimes and not property crime. Many studies have highlighted the role of anger in violent crime and the emotion as a crucial motivation to violent offenders (Agnew, 2012; Felman, 2016; Wortley, 2015) and the results from the current research are generally consistent with these findings. In particular, assault crime at private places was increased by anger in both emotional states (i.e. temporary emotions at crime situation) and emotional traits (i.e. emotional tendency of crime place). Considering the fact that assault crime at private places presumably occur between those who share more intimate relationships at the places where they reside, the results can be interpreted the overall aggressive atmosphere of private places may be a significant predictor of temporal aggressive atmosphere, which encourages serious violent behaviours such as higher-harm violent offences inflicting physical harm. However, anger was associated with a low level of property crime in which emotional states of offenders are less considered as a criminogenic motivation.

Interestingly, the results found that joy increased crime after controlling all potential influential effects. One possible explanation might be that, with respect to the assumption when people express joy, joy would be highly related with facilities or events attracting the influx of people, in particular, in this case, a massive influx of visitors (see Ristea et al., 2018). Put differently, joy might correlate with entertainment routine activities which would create more crime opportunities,

Except for joy, the four base emotions displayed less consistent outcomes than other emotional variables including swearing and equitability. In previous research, emotions have been utilised to gauge feedback in terms of reactions to specific social or political topics. Put differently, such emotions displayed could be not temporary emotional states but emotional tones (attitudes) toward a certain topic which is less related to moment-to-moment behaviours. Such emotions are closer to an opinion/attitude rather than may be less related with instant reactions about situations that people face. However, swearing, a signal of the speaker's state of mind (Byrne, 2018), was highly correlated with an increase in crime, especially crimes at semiprivate places. This observation is not to say that criminal situations would be caused by basic emotions people express naturally, rather, that the act of criminal behaviour may be associated with the use of swear words more.

This study also found consistent significant effects of equitability of emotions on crime. In particular, based on the results shown, it was evident that evenness of emotions is highly associated with crime in the OLS model, which to recap does not take into account the temporal effects. This sheds light on a role of collective emotions in crime situations in contrast to individual emotions. Emotions are contagious thus people sharing the same space catch and assimilate with the mood of the space quickly (Wild, Erb, and Bartels, 2001). A substantial amount of research suggested that one's personal emotion is not independent and is shareable, and therefore has a significant influence on others who are make contact with that person in daily life (Grandey, 2000; Bono et al., 2007; Pekrun, 2009). It may be assumed that a group of individuals who are making consistent and often contact would share and transfer similar emotions which would lower the evenness of emotions (recall that 'evenness' here means equal shares of different emotions and no one dominant emotion).

However, this research cannot avoid a common criticism of social media based analysis. The samples filtered out the whole population by users' decisions such as choices of Twitter, and sharing locations. In future study, as a new form of future crime data, more advanced understanding on how the users are filtered and selected and what the sample population represents is necessary to increase the validity of any results derived from the data.

In seeking to understand this outcome better, more follow-up studies would be needed to determine the association between collective emotions and crime in social disorganisation approaches. This initial research however might lead to the hypothesis that there is a strong correlation between sharing similar emotions of people in a neighbourhood and a level of cohesion or collective concern.

Further, the research outcomes from this study suggest that more static emotional tendency (or climate) of the places is important in addition to temporal emotions of crime situations. Overall, the OLS model showed a more significant and consistent association between emotions and crime, than the models that included temporal fluctuation. This confirms that crimes can be driven by temporal trends of emotions but also be influenced by emotional inclination of the areas. In particular, in contrast to the results from the panel model, a general inclination for sadness displayed consistent and positive correlations with crimes at private places. It indicates that while sadness states do not drive crime, sadness traits (general tendency) of places would be a significant variable to be linked with crimes especially at private places which are more likely to be driven by the people who spend most of their time in those place.

According to Agnew (1995), emotional traits have a significant impact on emotional states, the latter being the actual experience of emotional response to particular events. In other words, differences in a basic emotional level of individuals would lead to distinct behavioural responses at the time of an incident because interactional style and degree would be elicited by emotional traits. Despite the importance of the relationship between emotional traits and states, the current research did not fully explore this mechanism - explaining the effects of underlying emotions on temporary emotions - and this is thus left as an avenue for future research. Future research might also employ the data into the past key works in criminology such as non-recursive modelling on the links between crime, emotion, and cohesion suggested by Markowitz et al. (2001) or Tylor's work (2001) studying effects of emotions on crime at two levels; individual emotions and collective emotions.

In the past, it was not an easy task to capture emotions of a significant population, which can provide important cues to understand criminal situations. Notwithstanding its critical role as in building criminogenic atmosphere, quantifying a level of emotions in the moment of crime wasn't possible up until recently. The findings confirmed that the data that the current research utilised has the potential to extend the relevant theories and research on emotions and crime in various directions.

7 MOBILE POPULATION ATTRIBUTE PATTERNS

7.1 INTRODUCTION

Social media platforms, like Twitter, Facebook and Instagram, are an emerging means of self-expression. The users can release their inner thoughts and broadcast daily activities by short messages at anywhere and anytime to anyone. As an important means of self-expression that enables people to broadcast their thoughts on topics that they are interested in, it is considered as a new source of in-depth personal information which indicates 'behaviour signals' (Minamikawa et al.,2012; Lai and To, 2015). While traditional environmental criminology primarily relies on the census population data to capture social-behaviour indications such routine activities or a level of actions as guardians, this research explores signals of human behaviours by analysing online traits hidden in the texts people post on social media. In other words, this research aims to move beyond the traditional practice of using non-opinion characteristics of individuals such as colour, income, and education in the exploration of crime trends.

The main objective of the research is to analyse the context of crime situations in terms of online-traits (online-activities and especially shown interests) collected from social media posts. In particular, the current study aims to assess whether (and if possible, why) some traits are related to crime and whether these findings can be used to disclose criminal behaviour situations/settings which is repeatedly observed in a spatiotemporal unit.

7.2 BACKGROUND LITERATURE AND THEORETICAL RATIONALE

7.2.1 *Online-traits and offline-traits*

Like a cartoon published in 1993 says 'On the internet, nobody knows you're a dog' (The New Yorker, 1993). Online anonymity is a matter of great controversy that has been debated from the very early days of the internet. Since the internet gained an indisputable position and became a necessary tool for communication, we have spent more and more time in the online world reflecting our real-world selves in it. Indeed, today it is hard to fake or hide who you are online because our real lives cannot be separable from online activities (Buchanan, 2001; Marriott and Buchanan, 2014).

Multiple recent studies also support the premise that personality can be inferred by online traits such as social medias, blogs and forums. Based on the comparison of survey-based personality estimation with online text-based personality, the studies found that personal traits from the survey showed correlation with certain online behaviours such as people's decisions, actions, and tastes (Golbeck and Norris, 2013) and attitudes toward social networking (Gangadharbatla, 2008; Moore and McElroy, 2012).

Even though a fair number of studies have focused on how to extract personal attributes accurately from social medias (Correa, Hinsley and De Zuniga, 2010; Golbeck, Robles and Turner, 2011; Hughes et al.,2012; Park et al.,2015), the majority of them have examined online personal attributes relying exclusively on 'Big Five' personality traits[1]

Further, considering the increase in evidence of the significant linkage between online behaviours and personality (Bachrach et al., 2012; Lee et al., 2014; Zhao et al, 2015), studies on how offline behaviours relate to online personal attributes are very limited. Although some research has shown that certain types of behaviours in real life may be predicted by online traits such as political engagement (Di Gennaro and Dutton, 2006) and sexual behaviours (Boies, 2002), further studies are required to enhance understanding of online traits as a potential indicator of real world behaviours.

Online-traits as behaviour signals

In order to maximise the utility of Twitter in capturing personal attributes, the user characteristics of the platform require further examination. Twitter has different usage patterns from other popular social medias such as Facebook, Youtube and Instagram. Unlike other social networking platforms initially designed for sharing daily activities via

Note: the section titled 'What is Twitter' in Chapter 3 (p.28)

1 The Big Five personality so-called Five-Factor Model or OCEAN was developed to measure personality and five factors are openness, conscientiousness, extraversion, agreeableness, and neuroticism (see McCrae and Costa (1987))

various formats such as visual or audio media, Twitter is optimised for informative social networking. An analysis of Twitter traffic disclosed that a large number of messages on the platform are text-based and the posts about breaking news, how-to or social activities tend to get more attention than daily chats. For these reasons, Twitter has been considered as a barometer monitoring trending topics and there is an ample amount of research analysing Tweets to gauge people's predominant interests which are likely to be those frequently being mentioned (Cataldi, Di Caro and Schifanella, 2010; Lau, Collier and Baldwin, 2012; Xie et al., 2016). Considering these recognised features of the platform, analysing topics/interests and contents of the posts was seen as the best approach to investigating online-traits utilising Twitter.

There have not been enough studies to analyse the relationship of types of topics publicly posted on social medias and criminal behaviours within the perspectives and methodologies drawn from Criminology. Although there were some efforts to detect anti-social behaviours using micro-blogging messages (Wang, Brown, and Gerber, 2012; Gerber, 2014), these were limited to classifying contents related to crime events, suggesting computational methods or finding statistical evidence of the relationship, rather than building a theoretical framework. The latter could answer questions concerning why these contents are related to crime and how we should investigate them as a potential source for future crime studies.

The relationship between topics we broadcast and the types of people we are (or our behavioural traits) can be conceptually understood by theories of interests. These theories were employed in the fields of Psychology and Pedagogy to examine the effect of interests on educational or vocational achievement. According to John Dewey (1913)'s definition, interest is "being engaged, engrossed, or entirely taken up with some activities" or topics (Dewey, 1913, p.17 as cited in Harackiewicz and Hulleman, 2010). Intuitively, we can assume being interested in any subject implies it is important to one's daily life and that it is a personal concern (Harackiewicz and Hulleman, 2010).

Dewey (1913) notes that being interested in something also means positive feeling toward the subjects and therefore a willingness to put energy to the activities or topics. The majority of the past empirical research using Dewey's theory has examined whether interest relates to people's action regarding the subjects of interests. Previous studies on the relationship between interest and performance found that the status of interest was strongly correlated with the levels of performance. Interest was found to promote attention, recall, and task persistence (Su, Rounds, and Armstrong, 2009). Since past studies have mainly been conducted in Psychology, they have focused on how personal traits and interests are correlated with performance and achievement (McCown and Johnson, 1991; McKenzie, 1989, Schiefele, Krapp, and Winteler, 1992), educational and occupational choices (Hansen and Sackett, 1993; Lapan, Shaughnessy and Boggs, 1996; Fouda, 1999) and job satisfaction (Morris, 2003).

The theories and empirical research studies clearly show the connection between 'interests' and 'actions' and this logic can be applied to some possible scenarios. For example, a person who always shows interest in community news – demonstrated through posting about or adding comments on community activities such as events held by local churches or schools- or communicates with other people constantly will be more likely to participate in such events than people who never post about the local activities. Likewise, a person who intensively publishes his opinions on social and political issues on his social medias would have a higher chance of acting out for the social issues and getting involved in political actions than people who do not post on social topics at all, which can be possibly explained by social cohesion/collective efficacy emphasising the influence of sharing the same values on collective actions at the macro-level.

Note: see the section about 'Collective Efficacy and Social cohesion' in Chapter2 (p.8)

7.2.2 *Contents-based Analysis for capturing Crime Situations*

Content-based analysis refers to the process of converting artifacts such as written documents, verbal discourse, and visual representations into a relevant and manageable format of data (Krippendorff, 1980; Weber, 1990; Titscher et al.,2000). Basically, this analysis allows researchers to extract valuable information from unstructured data and to draw valid inferences based on the information (Bengtsson, 2016).

Measurements of online-traits from texts

To decide on appropriate methods for the current research, the content-based analysis methods which have been widely used to analyse online-traits were reviewed. Manual-content analysis, hand-coded by human coders, has a strong advantage in capturing the nuances and complexities hidden in texts. However, processing the current research data which is in the form of so-called 'Big Data' is beyond the capabilities of individual coders (Guo et al., 2016). This limitation of manual analysis has dramatically accelerated the need of methodological innovations which can process substantial amounts of data automatically and in a systemic manner (Lewis, Zamith, and Hermida, 2013). Indeed, many researchers have become actively involved in the development of computational techniques and recently these methods have been widely used for processing big data (Guo et al., 2016).

Within computer-assisted analysis, it is useful to distinguish between lexicon-based approaches and Machine learning approaches. Early computer-assisted analysis employed lexicon-based approaches, particularly dictionary-based methods. The methods include simple automated tasks such as counting the occurrence of predefined key words in given texts (Riffe et al.2014; Guo et al.,2016). It entails a process of developing a predetermined dictionary that researchers want to isolate. For this reason, establishing the appropriate dictionary is crucial as it corresponds directly with a research purpose and producing it remains a manual task (Guo, Shi, and Tu, 2017). This method has definite advantages for investigating specific problems, if the dictionary adequately reflects the attribute and context of the text to be analysed (Dalianis, 2018). However, in cases where it doesn't, it will achieve a poor performance (Guo, Shi, and Tu, 2017).

Machine learning approaches, in contrast to lexicon-based approaches, minimise intervention by researchers in the analysis process. Basically, the approaches involve a set of statistical methods that use 'learning' from previous patterns in texts to establish a set of rules/filters to gauge the text (Guo, Shi, and Tu, 2017; Dalianis, 2018). Within these approaches it is possible to distinguish between supervised methods and unsupervised methods- both are widely used methods in recent text analysis and are deployed depending on purposes of the analysis or types of texts.

Among many different unsupervised methods, topic modelling using the Latent Dirichlet Allocation (LDA) is very commonly applied to analyse text documents including the text collected from social media. The effectiveness of LDA on micro-blogging texts has been demonstrated, notwithstanding the problems in these texts such as their relatively shorter length than other text-based statements or informality/noisiness (Weng, 2010; Naveed, Gottron, and Kunegis, 2011; Guo et al.2016; Risch, 2016). While the lexicon-based approaches use deductive coding techniques which begin with a predetermined set of dictionaries, unsupervised machine learning adopts inductive coding techniques that are also known as data-driven methods (Eickhoff and Wieneke, 2018). Therefore, unsupervised machine learning would be more advantageous to explore unexpected outcomes that researchers initially cannot capture such as extracting trending topics from massive amounts of data.

To determine the analysis method for this research, any risks which could reduce the validity of the measurements were considered. The LDA algorithm needs reasonable large size of texts to produce meaningful topics (Tang et al., 2014). Past studies have, therefore, combined a number of texts based on analysis criteria such as time or location of posting or authorship (Hong and Davison, 2010; Zhao et al.,2011). A study using similar computational methods for personality analysis also suggested that at least 25 messages are required to fully measure language patterns consistently found in the texts (Xu et al., 2015). In addition, the model can produce different outcomes depending on how texts are merged, and the same texts will not show the same results when the different criteria are applied to them (Guo et al.,2016). Additionally, LDA merges individual texts into a bulk data set and extracts the most discussed topics. It appears that even though the LDA is useful to extract topics from texts with a given criteria such as time and location, the process of aggregation could eliminate the individuality of each text that should be included and analysed in this research. As in previous studies, LDA would be suitable for capturing overall tendency from massively aggregated messages but, in this case, (some spatiotemporal units only have under ten texts) may not be a well-matched method as it would not have statistically enough number of messages to conduct the inductive computation method. To avoid this problem, a dictionary-based method was chosen, which is performed on each text not the information in bulk. In summary, the analysis in this chapter is done using a lexicon-based non supervised method, more detail of which is given below.

Measurements of online-traits for crime analysis

Robert Weber (1990) underlines that the data-reduction processes should be carefully designed for conducting a content-based analysis. Technically, the analysis refers to a process of cutting and simplifying many words of texts into selected content categories decided by explicit rules of coding (Weber, 1990; Stemler, 2001). Therefore, designing a valid content classification based on the characteristics of datasets and the context and objective of studies is very important for content-based analysis.

For the main analysis two content analysis measurements were employed; online-activity and topic/interest-based, these suited the characteristics of the data and the objectives and the theoretical framework of the current research. For example, it distinguishes between messages that are conversations and those that are location broadcasts. Here topic-based refers to the subject of the communication. For example, it might be about body and appearance, business or religion and politics. Online-activity refers to the style or type of the tweet.

Measurement of the online activity type is based on Java et al.(2007) and Dann (2010) which are Twitter content categorisation studies. Java et al.(2007) categorise Twitter by users' intentions and find the main reasons for using Twitter can be classified into four types; Daily chatter, Conversations, Sharing information/URLs and Reporting news. Besides categorising the messages by how and why people use the social media, they also categorise them by types of users. According to Java et al.(2007), there are three types of users; Information Sources, who functions as a hub and have a large number of followers, Friends, who are the most common types of users and communicate with other users in online-community actively, and Information seekers who rarely creates a post but regularly follows other users (Java et al., 2007).

In contrast, Dann's (2010) classification only focuses on the types of content. After reviewing 16 studies on Twitter categorisation, he suggests six broad categories (with 23 detailed subcategories); Conversational, Pass Along, News, Status, Phatic communication and Spam. He defines Conversational as the posts created for the social interaction between users, Pass Along for re-tweeting other users comments or showing an act of endorsement, and News covering mainstream media issues such as sports, events, and weather news. Status refers the posts providing the update of activities the user is currently undertaking, the posts of Phatic communication means meaningless exchanges such as simple daily notes and unclassifiable contents.

Combining the two Twitter classifications from previous research, five categories of online-activity are used in the current research; Conversation, Informal Broadcasting, Informal Broadcasting with location status, location status, and Informative Broadcasting. Figure 28) indicates the relationship between the various concepts mentioned above. The added emphasis on activities related to location status is deliberate and is not obvious in the previous research classifications.

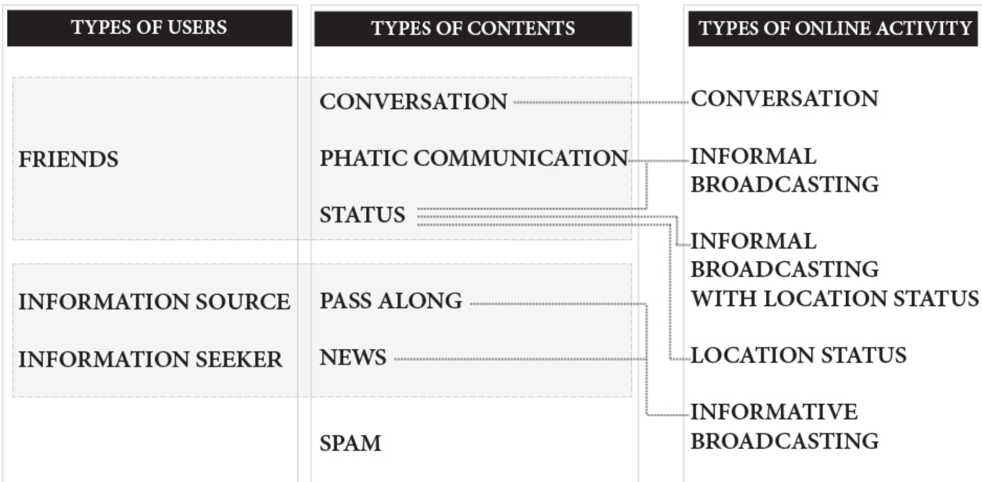


Figure 28: Online activity categories

In previous research exploring topic-based analysis which infers personality traits and interests from texts, one of the most popularly used dictionary-based methods is the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker and King, 1999; Hirsh and Peterson, 2009; Tausczik and Pennebaker, 2010; Tumasjan, Sprenger, and Sandner, 2010). Instead of simply classifying key words into relevant topics, the LIWC groups words by psychology-relevant connotation. Therefore, it sees, for example, pronouns such as 'we' and 'you' as an important source for capturing a quality of a close relationship/social coordination and the words related to 'Family' and 'Friends' as the key words for measuring 'Social Processes'. It has topic/interest related categories but only consists of seven topics including Work, Achievement, Leisure, Home, Money, Religion, and Death under 'Personal Concerns' (Tausczik and Pennebaker, 2010).

Although the LIWC generated some significant and valid findings in studies about inferring personality traits using social media, showed better outcomes with writing samples (Pennebaker and King, 1999) or spoken dialogue (Mehl, Gosling, and Pennebaker, 2006) and some limitations with short texts from social media. These have a number of challenges since they are often not well-constructed texts. In many cases, the texts are not written in full sentences: instead, they are composed with some key words and contain abbreviations, truncated words, and poor grammar (Guo et al., 2016). To overcome identified limitations in the previous studies, the current work employs Oxford Topic Dictionaries to explore the relationship of topics of micro-blogging texts and crime. The topic dictionaries are selected because they are composed from a broader range of topics and from common subjects in everyday life using basic key words.

To summarise, online activity categorisation is important as it reflects specific actions by users. In other words, it answers the question; what sort of behaviour happened? An information broadcast has a different purpose to a conversation for example. In addition, the topic or subject of social media texts likely to reflect the interests and concerns of those tweeting them. In the next section, there is an account of how online-activity based and topic-based measurements were operationalised for the analysis. This is followed by a description of the analysis.

7.3 METHODS

The two core datasets have already been described in Chapter 4. To recap, these are the crime data and social media messages in the region of NYC between the dates of 1st of January, 2013 and 31st of December, 2013. The research in this chapter required production of new datasets that can measure online traits using social media data as the source. This involves the extraction of online-activity variables and topic/interest variables from each post and then the aggregation of them into the selected spatiotemporal unit.

Note: see the section about 'Data' in Chapter 4 (p.36)

7.3.1 *Measurement*

Categorisation of Online activity

There is a range of purposes for posts for example, formal business posts are very different to those that are informal conversations. As illustrated in Figure 28, this research uses five online-activity categories which exhibit distinct behaviours on social media. This section describes the process by which posts were sorted into these categories.

Conversation These posts are created for communicating and trying to engage conversations across all subjects. The messages go back and forth between more than two users. If the messages start with "@", they were deemed as conversation regardless of the contents of the messages and the receivers of the messages. For instance, if a user sends news to other users and adds comments about it, the messages are coded into conversation as they are created to communicate about the news with the specific receivers not to spread out the news across followers or general users.

Informal broadcasting This category covers the use of the social media to broadcast personal thoughts, opinions, statements, and emotional status. This includes general posting such as daily activities, any personal statements, and emotional status.

Location status These messages only have information about the locations where the messages are created. The messages in this category only contain the location status and have no extra text with them. They commonly start "I am at-" or just tag the place where they are when they create the post. It also could be automated location tags by mobile applications such as Foursquare or Yelp check-ins.

Informal broadcasting with location status These messages have the location status with additional text-based comments about the location or the status such as emotions or opinions. In many cases, the posts are created at tourist attractions or restaurants where users want to leave a note about the locations.

Informative broadcasting These messages are created to deliver information to other users and do not contain any comments revealing the users' thoughts or emotions. The users purely use their account to pass along the messages, (called re-tweets), and perform a role as messengers or distributors. If users add any comments revealing their own thoughts or opinion with it, the messages are not counted as informative post as the users' intention is not delivering but commenting on the messages. The posts also incorporate the main stream media news including breaking news, sports, weather, and events and information such as job postings.

Categorisation of Topic/interest

The study uses the Oxford topic dictionaries generated using words we use on a daily-basis. The dictionaries have 24 sets of topics related to common subject areas and each topic category consists of groups of words and word-stems. There are several detailed subcategories under each topic group. For instance, *Family and Life Stages* has five 1st-level subcategories including Age, Children, Family, and Love. These subcategories are composed of 2nd-level subcategories such as Death, Middle Age, Old Ages, and Youth in the subcategory 'Age' and Babies, Birth, Pregnancy, and Raising Children for in the subcategory 'Children'. Table 1 gives full details of the topic classification.

Topic	**1st-level**	**2nd-level**
Animal	Features	Animal homes, Groups of animals, Parts of animals
	Types	Amphibians and reptiles, Birds, Domesticated mammals, Fish, Invertebrates, Wild mammals
Body and appearance	Appearance	Attractiveness, Body shape, Facial expressions, Female attractiveness, Male attractiveness, Position and movement
	Body	Body parts, Face, Hands and nails, Internal anatomy, Mouth and teeth, Skeleton and muscles, Skin
	Hair	Describing hair, Facial hair, Hair colour, Hair products and accessories, Styling hair
Business	Companies	Business deals, Business people, Running a business
	Economics	Banking, Economy, Trends
	Products	Manufacturing, Marketing
Clothes and fashion	Clothing	Accessories, Clothes, Describing clothes, Footwear, Jewellery, Parts of clothing
	Fashion	Beauty products, The fashion world
Crime and law	Crime	Committing crime, Criminals, Prison, Solving crime, The police, Types of crime, Types of punishment
	Law	Justice, Legal documents, Legal processes, People in law
Culture	Art	Art equipment, Artwork and techniques, Describing art, Styles of art, The art world

	Film	Film people, Film plots, Film reviews and promotion, Film-making equipment, Making films, Showing films, Types of film
	Literature	Characters in a story, Describing a story, Elements of a story, Parts of a book, People in publishing, Poetry, Types of story, Types of text, Writing and publishing
	Music	Describing music, Listening to music, Live music, Musical instruments, People in the music world, Pieces of music, Producing music, Reading music, Styles of music
	Theatre	Elements of a play, In the theatre, People in theatre, Producing a play, Types of play
Education	School	Access to education, Exams and assessment, In school, People in schools, School life, Subjects and courses, Teaching and learning, Types of school
	University	Exams and degrees, Higher education institutions, Study routes, University life, University people
Family and life stages	Age	Death, Middle age, Old age, Youth
	Children	Babies, Birth, Pregnancy, Raising children
	Family	Family background, Friends, Names, Pets, Relations
	Love	Marriage, Romance, Separation
Food and drink	Cooking	In the kitchen, Preparing food, Ways of cooking
	Describing food	Taste of food, Texture of food
	Meals	At the dining table, Savoury dishes, Sweets and desserts, Types of meal
	Restaurants	Dining out, Restaurant people, Types of restaurant
	Types of food	Alcoholic drinks, Carbohydrates, Fruit, Herb and spices, Meat, Sauces, Soft drinks, Vegetables
Health	Diet	Diet, Healthy eating habits, Unhealthy eating habits
	Fitness	Exercise, Good health, Poor health
	Illness	Ailments and diseases, Being ill, Injuries, Recovering from illness

	Medicine	Complementary medicine, Conventional medicine, Hospitals, Medical equipment, Medical examinations, Medical staff, Medication, Operations
	Mental health	Addition, Mental and emotional problems, Mental health care
Houses and buildings	Architecture	Architectural features, Architectural styles, Describing architecture
	Buildings	Buildings, Construction, Historic buildings, People in construction, Public spaces, Religious buildings
	Houses	Furniture, Gardening, House equipment, House location, How a building looks, In the garden, Interior decor, Parts of a house, Rooms in a house, Types of home
	Owning a home	Buying a home, Renting a home
Language	Language learning	Language skills, Languages
	Linguistics	Grammar, Linguistic devices, Phonetics, Punctuation
Leisure	Ball sports	American football, Baseball, Basketball, Cricket, Golf, Rugby, Soccer, Tennis
	Games	Board games, Card games, Children's games, Games, Pool and snooker
	Hobbies	Crafts and skills, Hobbies
	Other activities	Athletics, Boating, Combat sports, Cycling, Diving, Equine sports, Extreme sports, Swimming, Water sports, Winter sports
Nature	Agriculture	Animal farming, Crops, Farm animals, Farm people, Growing crops, On the farm
	Environment	Conservation, Natural disasters, The power industry, Waste and pollution
	Landscape	Coastlines and the sea, Describing geographic regions, Mountains and valleys, Other geographic regions, Plants, River and lakes
	The Earth and space	Space travel, The Earth and the atmosphere, The sun and the moon, The universe
	Weather	Rain, Sky, Snow and ice, Wind

Personality and emotions	Emotions	Anger, Boredom, Disappointment, Disgust, Embarrassment, Excitement, Fear, Happiness, Loneliness, Love, Showing interest, Surprise, Unhappiness
	Feelings	Hunger, Thirst, Tiredness
	Personality	Brave, Clever, Confident, Describing annoying traits, Describing strange traits, Describing unpleasant traits, Dishonest, Energetic, Friendly, Honest, Immoral, Kind, Lazy, Moral, Nervous, Proud, Selfish, Stupid
Religion and politics	Politics	Elections, International relations, Parliament, Political views and systems
	Religion	Religious ceremonies, Religious holidays and festivals, Religious items, Religious people, Religious places, Religious texts, Types of belief
Retail	Shopping	Cost and payment, Online shopping, Shopping in stores
	Stores	Departments in stores, In the store, Types of stores
Science	Engineering	Devices, How machines work, Materials and properties, Structures
	Mathematics	Mathematical terminology, Types of mathematics
	The natural sciences	Cell biology, Electronics, Energy and physical forces, Experiments and research, Genetic engineering, Molecules and matter
Social issues	People in need	Helping others, Poverty and famine
	Social groups	Belonging to society, Immigration, Race, Social justice
Technology	Computers	Computer hardware, computer problems, Computer programming, Computer software, Mobile devices, Using a computer, Video games
	Telephones	Communication devices, Making calls, Phone services, Text messages
	The Internet	Email, Social networking, Using the Internet, Websites
The media	News media	Celebrity news, Journalism, People in journalism, The press
	Radio	People in radio, Radio broadcasting, Radio technology

	Television	Producing TV shows, TV people, TV shows, Watching TV
Travel and tourism	Air travel	Aircraft, Airports, Parts of a plane, People in air travel, Plane travel
	Public transport	Railway tracks and stations, Train and bus travel, Trains
	Road travel	Controlling traffic, Driving, Features of roads, Motoring problems and accidents, Parts of a car, People in motoring, The car industry, Types of road, Types of vehicle
	Sea travel	Parts of boats and ships, People in sea travel, Travelling by boat or ship, Types of boats and ships
	Tourism	Camping, Countries and continents, Holiday accommodation, Hotel people, Staying in a hotel, The tourist industry, Types of holiday/vacation
War and conflict	Conflict	Protest, Terrorism
	The armed forces	Conflict, Peacekeeping, The air force, The army, The navy, Weapons
Work	Jobs	Describing jobs, Job titles, Professions
	The job market	Job interviews, Job skills and personal qualities, Unemployment
	Working life	Business meetings, Describing work, In the office, Office life, Pay and conditions at work

Table 16: The list of topic/interest categories

7.3.2 *Analysis methods*

For the analysis units, street segments and day/hour (24 hours and 7 days) were chosen. The analysis began with a coding process. Each text was coded by the presence/absence of corresponding online-activity categories and specific topic related lexicon. When the dictionaries indicated a topic appeared at least once in a post, the post was recorded as falling into that topic category. If a text obtained multiple topics, all detected topics were recorded, regardless of how many times the words appeared. The data was cleaned to remove any errors which might affect the coding process including checking for spelling errors and converting words into the simple/singular form before being analysed. After finishing the coding process, the number of online-activities/topics present in each spatiotemporal unit was counted.

Note: see the section about 'Research Units' in Chapter4 (p.42)

This was added to the crime information for these units. As a result, the generated data contained the number of crimes and a online-activities/topics occurred/mentioned in each street segment over time.

The analysis was performed using two different levels of comparison (Figure. 29). The first level of analysis was designed to compare the differences between the streets with crime records (Group A) and the streets no recorded crime (Group B). For the second analysis, Group A was divided into two groups, Group A1 and Group A2, and Group B was excluded from the analysis. This second level of analysis examined the differences between time units with crimes and time units with no-crime records. If the time unit had any recorded crime, it was assigned to Group A1 and the remaining units were categorised into Group A2. The same process of classification applied to all crime types (e.g. assault or property crime) and locations type (e.g. public or private) and consequently created four groups (Group A, Group B, Group A1 and Group A2) for 20 types of crime. Differences in proportions of online traits across each group were analysed using chi-square test. Fisher's exact test was also used but results are not included here as the two tests produced the same results.

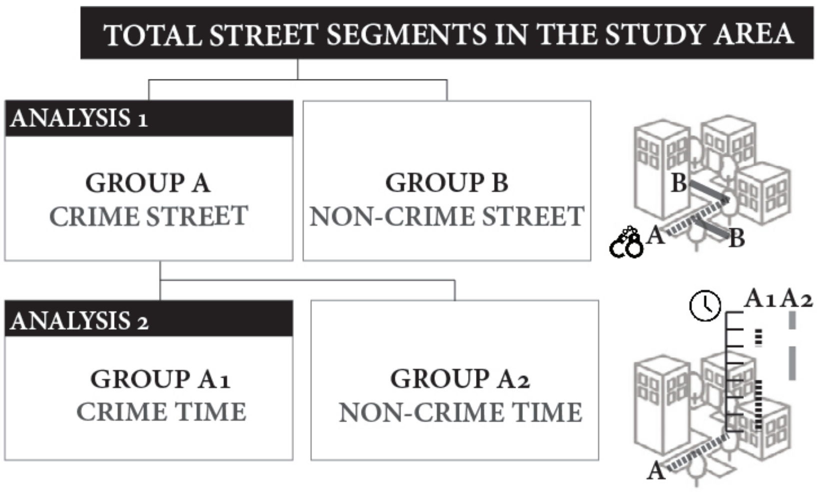


Figure 29: Study flowchart (image @shutterstock)

Before moving to the results, it is worth noting that the analysis does not track the feelings, actions or behaviours of specific individuals over time and/or space. Rather, it captures the collective current 'traits', 'behaviours' or 'concerns' of those tweeting in a spatiotemporal unit. It can be seen in some senses as the population-driven 'social atmosphere' of that area at that time. Strengths and weaknesses of undertaking analysis using this method are discussed later in this chapter.

7.4 RESULTS

7.4.1 *Descriptive statistics*

For online activity-based variables (Table 17), *Informal broadcasting* was generated the most. Over one-third of posts were generated for a conversational purpose. *Conversation* and *Informal broadcasting with location status* were observed nearly 20 per cent respectively, showing that the social media is more likely being used for social networking via posting their own thoughts and interests to communicate with other users. The data analysis showed that a small percentage of the posts were produced to share information (2.9 per cent) such as breaking news, sports, weather, and events.

Table 18 shows that the frequency distribution of topic-based variables. For topic/interest-based variables, *Personality and emotions*(40.31 per cent) was detected the most. A substantial number of texts were identified as tourism-related (i.e. *Leisure* (27.91 per cent), *Culture* (21.59 per cent) and *Travels and tourism* (19.93 per cent)) but political and social topics were observed less (i.e. *Religion* (4.7 per cent), *Social issues* (4.43 per cent), and *War* (3.15 per cent)).

Table 17: Descriptive Statistics of online activity categories

Online activity	Frequency	Percent
Conversation	119,740	23.05
Informal broadcasting	220,608	42.46
Location Status	57,569	11.08
Informal with location status	106,237	20.45
Informative broadcasting	15,407	2.9

Table 18: Descriptive Statistics of topics/interests categories

Topic/interest	Frequency	Percent	Topic/interest	Frequency	Percent
Animals	31,119	5.99	*Leisure*	114,986	27.91
Body and appearance	55,032	10.59	*Nature*	79,781	15.36
Business	52,559	10.12	*Personality and emotions*	209,429	40.31
Clothes and fashion	77,924	15.00	*Religion and politics*	24,437	4.70
Crime and law	56,150	10.81	*Retail*	39,405	7.58
Culture	112,190	21.59	*Science*	42,912	8.26

Education	60,496	11.64	***Social issues***	23,005	4.43
Family and life	82,878	15.95	***Technology***	58,409	11.24
Food and drink	76,538	14.73	***The media***	42,485	8.18
Health	105,944	20.39	***Travel and tourism***	103,552	19.93
Houses and buildings	95,624	18.40	***War and conflict***	16,359	3.15
Language	114,769	22.09	***Work***	49,439	9.52

7.4.2 *Chi-squared analysis*

Table 19 shows the summarised results of online-activities analysis for all crime types, and the findings from topic/interest-based were reported in Table 21. In the summary tables (Tables 19 and 21), '(+)' indicates that the topics/online activities are over-represented in crime units (Group A or Group A1) and '(–)' refers to those are under-represented in the units [2].

Online activity-based analysis

The online activity-based analysis results of each crime are reported for the example of property crime in Table 20 and the remainder can be found in Tables 73 to 91 in Appendix (p.173).

Comparison between crime and non-crime streets (segments comparison)

The analysis of Group A against Group B indicated that online-activity types showed statistically significant associations (p < .01) for all crime types and crime locations. For crime at private places, three online activities, *Conversation*,(Property: $x^2 = 1.00, p < .01$, Harassment: $x^2 = 560.37, p < .01$, Assault: $x^2 = 177.46, p < .01$, Robbery: $x^2 = 748.22, p < .01$, Drug: $x^2 = 629.30, p < .01$), *Informal broadcasting* (Property: $x^2 = 615.08, p < .01$, Harassment: $x^2 = 2.20, p < .01$, Assault: $x^2 = 1.40, p < .01$, Robbery: $x^2 = 235.90, p < .01$, Drug: $x^2 = 939.72, p < .01$), and *Informative broadcasting* (Property: $x^2 = 24.90, p < .01$, Assault: $x^2 = 1.00, p < .01$, Robbery: $x^2 = 22.82, p < .01$) were more detected in Group A segments (crime present on segments).

In contrast to *Informal broadcasting with location status* (Property: $x^2 = 1.50, p < .01$, Harassment: $x^2 = 1.90, p < .01$, Assault: $x^2 = 2.20, p < .01$, Robbery: $x^2 = 1.10, p < .01$, Drug: $x^2 = 1.60, p < .01$) and *Location status* (Property: $x^2 = 1.20, p < .01$, Harassment: $x^2 = 1.10, p < .01$, Assault: $x^2 = 1.20, p < .01$, Robbery: $x^2 = 473.76, p < .01$, Drug: $x^2 = 952.6, p < .01$)

2 To highlight the patterns found in analysis outcomes, the chapter reports 1) summary of findings of online activity-based analysis (Table 19) and the findings of property crime at private places as an example of each crime type following the analytic process (Table 20). (See Appendix (p.173) for the analysis results of other 19 crime types) and 2) summary of findings of topic-based analysis (Table 21) and the findings of property crime at private places as an example following the analytic process (Table 22). (See Appendix (p.182) for the analysis results of other 19 crime types)

were less identified from crime at private places in Group A segments. The findings also showed a salient difference by types of crime locations, especially segments with crimes at private places in comparison to the other crime location segments, but there appeared to be no noticeable difference by types of crimes. The results indicate that crimes at private places have more online activities such as Conversation, Informal broadcasting and Informative broadcasting and posts with location information are under-represented in Group A segments for crimes at private places.

Comparison between crime and non-crime times (units comparison)

For the comparison of online activity types between Group A1 (crime time) and Group A2 (non-crime time) units, significant differences were found in *Conversation* and *Informative Broadcasting*. Both of these online activities were less identified in Group A1, in particular property and assault crimes, and more in Group A2 units.

There is a significant difference by crimes at different locations in *Informative broadcasting* and *Informal broadcasting with location status*. For crimes at private places, *Informative broadcasting* (Property: $x^2 = 7.48, p < .01$, Harassment: $x^2 = 4.08, p < .05$, Robbery: $x^2 = 71.30, p < .01$, Drug: $x^2 = 33.71, p < .01$) was more detected at time of crimes (group A1 units), and *Informal broadcasting with location status* (Property: $x^2 = 4.89, p < .05$, Harassment: $x^2 = 4.11, p < .05$, Assault: $x^2 = 9.73, p < .01$) was less identified at the time. Both online activities showed different distributions for group A1 and A2 units for crimes at the other types of places. Lastly, *Location status* (Semiprivate: $x^2 = 112.23, p < .01$, Semipublic: $x^2 = 19.28, p < .01$, Public: $x^2 = 13.97, p < .01$) was more highly observed at crime times (in group A1 units) in property crime than other types of crimes.

Table 19: Summary of all types of crime: frequency of online activity by each group

Crime locations: a = private, b = semiprivate, c = semipublic, and d = public

(+): over-represented, (-): under-represented in crime segments or units

	Group A and B				Group A1 and A2			
	a	b	c	d	a	b	c	d
Conversation								
Property	(+)***	(–)***	(–)***	(–)***	(+)	(–)***	(–)***	(–)***
Harassment	(+)***	(–)***	(–)***	(–)***	(+)	(+)**	(+)	(+)
Assault	(+)***	(–)***	(–)***	(–)***	(+)***	(–)***	(–)**	(–)**
Robbery	(+)***	(–)***	(–)***	(+)***	(–)***	(–)	(–)	(+)
Drug	(+)***	(–)***	(–)***	(–)***	(–)***	(–)*	(–)	(–)***
Informal broadcasting								
Property	(+)***	(–)***	(–)***	(–)*	(+)	(–)***	(+)	(–)***
Harassment	(+)***	(–)***	(–)***	(–)**	(+)*	(+)	(–)	(–)**
Assault	(+)***	(–)***	(–)***	(–)***	(+)*	(+)	(+)	(+)
Robbery	(+)***	(–)***	(–)***	(+)***	(–)	(+)***	(+)**	(+)

Drug	(+)***	(–)***	(–)***	(–)	(–)	(+)**	(+)	(–)
Location Status								
Property	(–)***	(+)***	(+)***	(+)***	(–)	(+)***	(+)***	(+)***
Harassment	(–)***	(+)***	(+)***	(+)***	(–)*	(–)	(+)***	(+)
Assault	(–)***	(+)***	(+)***	(+)***	(–)	(–)	(+)	(–)
Robbery	(–)***	(+)***	(+)***	(–)***	(–)	(–)	(–)	(–)
Drug	(–)***	(+)***	(+)***	(+)***	(+)**	(–)***	(–)	(+)
Informal with location status								
Property	(–)***	(+)***	(+)	(+)***	(–)**	(+)***	(+)	(+)***
Harassment	(–)***	(+)***	(+)***	(+)***	(–)**	(–)	(+)	(+)
Assault	(–)***	(+)***	(+)***	(+)***	(–)***	(+)***	(+)	(+)***
Robbery	(–)***	(+)***	(+)***	(–)***	(–)	(–)**	(–)	(–)
Drug	(–)***	(+)***	(+)**	(+)***	(+)	(+)***	(+)	(+)***
Informative broadcasting								
Property	(+)***	(+)***	(+)***	(+)***	(+)***	(–)***	(–)***	(–)***
Harassment	(–)***	(+)***	(+)***	(–)***	(+)**	(–)*	(–)***	(–)
Assault	(+)***	(–)***	(–)***	(–)*	(–)***	(–)***	(–)	(–)***
Robbery	(+)***	(–)***	(+)**	(–)***	(+)***	(+)***	(+)	(–)*
Drug	(–)	(–)***	(–)***	(–)***	(+)***	(–)***	(–)	(–)***

** p<.1; ** p<.05; *** p<.01*

Table 20: Property at private places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	50,185 (25.44)***	69,555 (21.58)	859 (26.23)	49,326 (25.42)
Informal broadcasting	88,065 (44.63)***	132,543 (41.13)	1,480 (45.19)	86,585 (44.62)
Location Status	18,032 (9.14)***	39,537 (12.27)	276 (8.43)	17,756 (9.15)
Informal with location status	34,876 (17.68)***	71,361 (22.14)	531 (16.21)**	34,345 (17.7)
Informative broadcasting	6,147 (3.12)***	9,260 (2.87)	129 (3.94)***	6,018 (3.10)

** p<.1; ** p<.05; *** p<.01*

Topic-based Analysis

The topic-based analysis results of each crime are reported for the example of property crime in Table 22 and the remainder can be found in Tables 92 to 110 in Appendix (p.182).

Comparison between crime and non-crime streets (segments comparison)

Most topics were found to be significantly different (p<.01) between Group A and Group B in most crime types and locations (Table 21). Seven topics were found more often in Group A (crime present) including *Culture, Food/Drink, Houses/Buildings, Language, Personality/Emotions, Retail* and *Science*. The Remaining seventeen topics were less captured in Group A than Group B units. Among the less observed topics in Group A, notable differences were found in *Business, Clothes, Education, Health, Leisure, Nature*, and *Religion*, which indicates that these topics which are more related to everyday living were less generated at high crime places.

Inspection of the results revealed that there is more variation in topics associated with Group A segments (crime streets) across the different types of crime locations than between crime types. In particular, segments with crimes reported at private places were associated with topics that were not significant for segments with the other crime location types (semiprivate, semipublic, and public places). For example, the eight topics identified more in segments containing crimes at private places; *Animal, Body/Appearance, Clothes/Fashion, Culture, Family/Life stages, Health, Personality/Emotions* and *Social issue* were less present at the other three sets of crime location segments.

Among these, four topics; *Body/Appearance* (Property: $x^2 = 46.16, p < .01$, Harassment: $x^2 = 89.35, p < .01$, Assault: $x^2 = 29.35, p < .01$, Robbery: $x^2 = 10.50, p < .01$), *Family/Life stages* (Property: $x^2 = 87.15, p < .01$, Harassment: $x^2 = 83.67, p < .01$, Assault: $x^2 = 75.67, p < .01$, Robbery: $x^2 = 24.017, p < .01$), *Health* (Property: $x^2 = 8.52, p < .01$, Harassment: $x^2 = 103.88, p < .01$, Assault: $x^2 = 77.010, p < .01$, Robbery: $x^2 = 38.39, p < .01$) and *Social issue* (Property: $x^2 = 60.50, p < .01$, Harassment: $x^2 = 25.84, p < .01$, Assault: $x^2 = 174.30, p < .01$, Robbery: $x^2 = 6.89, p < .01$) displayed a statistically significant difference between segments with crimes at private places and those with crime at other places for most types of crimes.

There were also some topics less observed in segments with crimes at private places and more at those with crimes at semiprivate, semipublic and public places. Topics including *Business, Housing/Buildings, Language, Nature, Retail, Science, The media, Travel/Tourism*, and *War/Conflict* were less identified in all five crime types for segments with crime in private places and the outcomes are statistically significant. *Food/Drink* was more captured from all crimes at semiprivate places, and the largest disparity between Group A and Group B was found for property crimes at semiprivate places.

The findings also presented some general patterns by crime types. Although there are a few exceptions, for property crimes (Table 22 for the example of private places and Table 21 for a summary), a significant difference was found between Group A and B segments in topics including *Business* (semiprivate: $x^2 = 27.15, p < .01$, semipublic: $x^2 = 277.22, p < .01$, Public: $x^2 = 143.97, p < .01$), *Crime/Law* (semiprivate: $x^2 = 26.010, p < .01$, semipublic: $x^2 = 36.94, p < .01$, Public: $x^2 = 70.22, p < .01$), *Culture* (Private: $x^2 = 71.18, p < .01$, semiprivate: $x^2 = 132.35, p < .01$, semipublic: $x^2 = 39.50, p < .01$, Public: $x^2 = 138.94, p < .01$), *Education* (Private: $x^2 =$

6.24, $p < .05$, semiprivate: $x^2 = 97.67, p < .01$, Public: $x^2 = 58.37, p < .01$), *Food/Drink* (Private: $x^2 = 77.016, p < .01$, semiprivate: $x^2 = 685.78, p < .01$, semipublic: $x^2 = 16.64, p < .01$, Public: $x^2 = 58.88, p < .01$), *The media* (semiprivate: $x^2 = 216.965, p < .01$, semipublic: $x^2 = 361.58, p < .01$, Public: $x^2 = 188.67, p < .01$), *War/Conflict* (semiprivate: $x^2 = 2.10, p < .01$, semipublic: $x^2 = 277.92, p < .01$, Public: $x^2 = 126.29, p < .01$), and *Work* (semiprivate: $x^2 = 163.13, p < .01$, semipublic: $x^2 = 586.22, p < .01$, Public: $x^2 = 384.79, p < .01$).

For a general trend of harassment and assault crimes, a significant difference between crime and non-crime segments was observed in *Personality/Emotions* (Harassment at Private: $x^2 = 128.39, p < .01$, semipublic: $x^2 = 20.67, p < .01$, Public: $x^2 = 75.52, p < .01$) / Assault at Private: $x^2 = 24.98, p < .01$, semiprivate: $x^2 = 6.53, p < .05$, semipublic: $x^2 = 112.32, p < .01$). This result fits to the findings from previous studies that insist the important role of emotions in violent crime situations (Agnew, 1992; Wortley, 2001).

Contrary to most topics that showed different associations based on the types of crime and locations, *Religion/Politics* were less identified in Group A segments in all 20 cases (crime types and locations) and differences were significant in all segments where crimes occurred at semiprivate and public places. The largest differences of *Religion/Politics* were found in segments with assault and robbery crimes.

Comparison between crime and non-crime time (units comparison)

In the results from the analysis between time of crime and non-crime, significant differences in topics are conspicuously found between group A1 (crime time) and A2 (non-crime time) units for crimes in semiprivate places. Most topics did not show a statistical difference between group A1 and group A2 units for crimes at private places unlike the results found in the analysis between Group A and Group B. In other words, for the location category crimes at private places, there was a significant difference in topics at crime places and non-crime places but there is no difference between topics at times when crime is present and those when there is no crime.

Across all types of crimes and crime locations, a substantial difference between Group A1 and Group A2 was observed in eight topics including *Business, Food/Drink, Health, Nature, Retail, Technology, The media* and *War/Conflict*. All those topics are less detected in Group A1 (crime times) than Group A2 (non-crime times) units.

Considering crime types, consistent patterns were observed in topics for group A1 and group A2 for property crimes and assault crimes. For property crimes, topics such as *Houses/Buildings* (semiprivate: $x^2 = 26.06, p < .01$, semipublic: $x^2 = 42.99, p < .01$, Public: $x^2 = 11.90, p < .01$), *Language* (semiprivate: $x^2 = 159.59, p < .01$, semipublic: $x^2 = 12.82, p < .01$, Public: $x^2 = 30.14, p < .01$), and *Science* (Private: $x^2 = 3.07, p < .1$, semiprivate: $x^2 = 17.41, p < .01$, Public: $x^2 = 7.48, p < .01$) were seen in higher proportions in Group A1 (crime times). For assault crimes, topics including *Nature* (semiprivate: $x^2 = 24.14, p < .01$, semipublic: $x^2 = 5.12, p < .01$), *The media* (semiprivate: $x^2 = 24.14, p < .01$, semipublic: $x^2 = 5.12, p < .01$), and *War/Conflict* (semiprivate: $x^2 = 4.05, p < .05$, semipublic: $x^2 = 17.57, p < .01$) were less detected in Group A1. For both crime types, the largest differences were found in crimes at semipublic places.

Table 21: Summary of all types of crime: frequency of topic by each group

Crime locations: a = private, b = semiprivate, c = semipublic, and d = public
(+): over-represented, (-): under-represented in crime segments or units

	Group A and Group B				Group A1 and Group A2			
	a	b	c	d	a	b	c	d
Animals								
Property	(+)***	(–)***	(–)	(+)	(+)**	(–)	(–)	(+)
Harassment	(+)***	(–)***	(–)***	(–)	(–)	(+)	(–)	(–)
Assault	(+)**	(–)*	(–)***	(–)***	(–)	(–)*	(–)	(–)***
Robbery	(–)	(–)***	(+)***	(–)***	(+)	(+)**	(–)	(–)
Drug	(–)**	(–)**	(–)***	(–)**	(+)	(–)	(–)	(–)
Body and appearance								
Property	(+)***	(–)	(–)***	(+)***	(+)***	(–)***	(–)	(–)
Harassment	(+)***	(–)***	(–)***	(–)	(+)	(+)	(+)	(–)
Assault	(+)***	(–)	(–)***	(–)***	(+)*	(–)**	(–)	(–)
Robbery	(+)***	(+)	(–)**	(–)***	(+)***	(+)	(–)	(+)
Drug	(–)**	(–)***	(–)***	(–)	(+)	(–)**	(–)	(–)*
Business								
Property	(–)***	(+)***	(+)***	(+)***	(+)***	(–)	(–)***	(–)
Harassment	(–)***	(+)***	(+)***	(–)*	(–)	(–)	(–)***	(+)
Assault	(–)***	(–)***	(–)***	(–)***	(–)	(–)***	(–)**	(–)***
Robbery	(–)***	(–)***	(–)	(–)***	(–)	(–)	(–)	(–)
Drug	(–)***	(–)***	(–)***	(–)***	(+)	(–)***	(+)	(–)
Clothes and fashion								
Property	(+)***	(–)**	(–)***	(+)	(–)*	(–)	(+)	(+)
Harassment	(+)***	(–)***	(–)***	(+)*	(+)	(+)	(+)	(–)
Assault	(+)*	(–)***	(–)***	(–)***	(–)	(–)	(–)	(–)
Robbery	(+)	(+)	(+)	(–)***	(–)	(+)	(–)	(–)
Drug	(–)***	(–)***	(–)***	(–)***	(–)	(–)*	(+)	(–)
Crime and law								
Property	(+)	(+)***	(+)***	(+)***	(–)	(–)	(–)	(–)
Harassment	(–)	(+)***	(–)***	(–)***	(+)	(–)	(–)	(–)
Assault	(+)	(–)	(–)***	(–)***	(–)	(–)***	(–)	(–)
Robbery	(–)	(–)***	(–)***	(–)***	(–)	(+)	(–)	(+)

Drug	(–)***	(–)***	(–)***	(–)	(–)	(+)	(+)	(–)**
Culture								
Property	(+)***	(+)***	(+)***	(+)***	(+)	(–)***	(–)*	(+)
Harassment	(+)***	(+)***	(–)***	(–)	(–)	(+)	(–)	(–)
Assault	(+)***	(+)***	(–)***	(+)***	(–)	(–)	(–)	(–)
Robbery	(–)	(–)	(–)***	(–)***	(–)	(–)	(–)	(–)
Drug	(–)***	(+)	(–)***	(–)	(–)	(–)**	(+)	(–)**
Education								
Property	(+)**	(+)***	(–)***	(+)***	(–)	(–)	(–)	(+)
Harassment	(–)***	(+)*	(–)***	(–)	(+)	(–)	(+)	(–)
Assault	(–)	(–)*	(–)***	(–)***	(–)	(–)*	(+)	(–)
Robbery	(–)***	(–)	(–)***	(–)***	(+)	(+)	(–)**	(–)**
Drug	(–)***	(–)***	(–)***	(–)***	(–)	(+)***	(–)	(–)
Family and life stages								
Property	(+)***	(–)***	(–)	(–)	(–)	(–)*	(–)	(+)
Harassment	(+)***	(–)***	(–)***	(+)***	(+)	(–)	(–)	(–)
Assault	(+)***	(–)***	(–)	(–)***	(+)	(–)**	(–)	(+)
Robbery	(+)***	(–)***	(–)	(+)	(+)	(+)	(+)	(+)*
Drug	(–)	(–)***	(+)	(–)	(+)	(–)	(+)	(–)*
Food and drink								
Property	(+)***	(+)***	(+)***	(+)***	(+)	(–)**	(–)**	(–)
Harassment	(+)	(+)***	(–)***	(+)	(+)	(+)	(–)	(+)
Assault	(+)	(+)***	(–)***	(+)***	(–)***	(–)***	(+)	(–)
Robbery	(–)***	(+)***	(+)***	(–)***	(–)	(+)***	(+)	(–)
Drug	(–)***	(–)*	(–)***	(–)***	(+)	(–)***	(+)	(–)***
Health								
Property	(+)***	(–)***	(–)***	(–)***	(+)	(–)***	(–)	(–)***
Harassment	(+)***	(–)***	(–)***	(+)***	(+)**	(–)	(–)	(+)
Assault	(+)***	(–)***	(–)***	(–)***	(+)	(–)***	(–)	(–)**
Robbery	(+)***	(–)***	(–)***	(–)***	(–)	(+)	(–)	(–)
Drug	(–)	(–)***	(–)***	(+)	(–)	(–)***	(+)	(–)***
Houses and buildings								
Property	(–)***	(+)***	(+)***	(+)***	(+)	(+)***	(+)***	(+)***
Harassment	(–)***	(+)***	(+)***	(+)***	(–)	(–)***	(+)	(+)

Assault	(–)***	(+)***	(+)***	(+)***	(–)	(+)	(–)	(+)
Robbery	(–)***	(–)*	(+)***	(–)***	(+)*	(–)	(–)*	(–)
Drug	(–)***	(+)***	(+)***	(+)***	(–)	(+)***	(–)	(+)**
Language								
Property	(–)***	(+)***	(+)***	(+)***	(+)	(+)***	(+)***	(+)***
Harassment	(–)***	(+)***	(+)***	(+)***	(–)**	(+)	(+)**	(–)
Assault	(–)***	(+)***	(+)***	(+)***	(–)	(–)***	(+)	(–)**
Robbery	(–)***	(+)***	(+)***	(–)***	(+)	(–)	(–)	(–)
Drug	(–)***	(+)***	(+)***	(+)***	(+)	(–)	(–)	(+)
Leisure								
Property	(–)***	(–)***	(+)***	(–)	(–)*	(–)*	(+)	(+)
Harassment	(–)***	(+)***	(–)***	(+)***	(+)	(–)	(–)***	(+)
Assault	(–)***	(–)*	(–)***	(–)***	(+)	(–)	(–)**	(+)
Robbery	(–)	(–)***	(–)*	(–)***	(+)	(+)	(–)	(–)***
Drug	(–)***	(+)***	(–)	(–)	(+)	(–)	(–)	(+)
Nature								
Property	(–)***	(–)***	(+)***	(–)***	(+)	(–)***	(+)***	(–)
Harassment	(–)***	(–)***	(+)***	(–)***	(+)	(–)	(–)**	(+)
Assault	(–)***	(–)***	(+)***	(–)**	(–)	(–)***	(–)**	(–)***
Robbery	(–)***	(–)***	(–)***	(–)***	(–)	(+)	(+)	(+)
Drug	(–)***	(–)***	(+)***	(–)**	(+)	(–)***	(–)***	(–)***
Personality and emotions								
Property	(+)	(–)***	(+)	(–)	(+)	(–)	(+)	(–)
Harassment	(+)***	(–)	(+)***	(+)***	(+)	(–)**	(–)	(+)
Assault	(+)***	(+)**	(+)***	(–)***	(+)	(–)*	(+)	(–)
Robbery	(+)***	(–)***	(–)	(+)*	(–)	(+)	(+)	(–)
Drug	(–)***	(+)**	(–)***	(+)***	(–)	(–)***	(–)	(–)
Religion and politics								
Property	(–)***	(–)***	(–)***	(–)***	(+)	(–)*	(–)	(–)
Harassment	(–)**	(–)***	(–)	(–)***	(+)	(–)	(–)	(–)
Assault	(+)	(–)***	(–)***	(–)***	(–)	(–)**	(–)	(–)
Robbery	(–)	(–)***	(–)***	(–)***	(+)	(–)	(+)	(–)
Drug	(–)	(–)***	(–)***	(–)***	(–)	(–)	(–)	(–)***
Retail								

Property	(–)***	(+)***	(+)***	(+)***	(–)	(+)***	(–)	(–)
Harassment	(–)***	(+)***	(+)***	(+)***	(–)	(+)	(–)***	(+)
Assault	(–)***	(+)***	(–)**	(–)***	(–)	(–)**	(–)*	(+)
Robbery	(–)***	(+)***	(+)***	(–)***	(–)	(+)	(–)	(–)
Drug	(–)***	(+)***	(–)	(+)***	(–)**	(–)***	(–)	(+)
Science								
Property	(–)***	(+)***	(+)***	(+)***	(+)*	(+)***	(+)	(+)***
Harassment	(–)***	(+)***	(+)***	(+)***	(–)	(–)**	(–)	(+)**
Assault	(–)***	(+)***	(+)***	(+)***	(–)	(–)***	(+)***	(+)
Robbery	(–)***	(–)***	(+)	(–)	(+)	(–)	(–)**	(–)
Drug	(–)***	(+)***	(+)***	(+)***	(–)*	(–)***	(+)	(+)***
Social issues								
Property	(+)***	(–)	(–)***	(+)**	(+)	(–)***	(–)	(–)**
Harassment	(+)***	(–)***	(–)	(–)***	(–)	(–)	(–)	(–)
Assault	(+)***	(–)***	(–)***	(–)***	(–)	(+)	(+)	(–)
Robbery	(+)***	(–)	(–)**	(–)	(+)	(+)	(–)	(–)
Drug	(–)***	(–)***	(–)***	(–)***	(–)	(+)	(+)	(–)
Technology								
Property	(+)	(–)***	(+)***	(–)	(–)	(–)***	(+)**	(–)
Harassment	(–)	(–)***	(+)*	(+)***	(–)	(–)	(–)	(+)**
Assault	(+)	(–)***	(–)***	(–)	(+)	(–)***	(–)	(–)**
Robbery	(+)**	(–)***	(–)***	(–)***	(+)	(+)**	(–)*	(–)**
Drug	(–)**	(–)***	(+)	(+)	(–)	(–)***	(–)	(–)***
The media								
Property	(–)**	(+)***	(+)***	(+)***	(–)	(–)***	(+)***	(–)
Harassment	(–)***	(–)***	(+)***	(–)***	(–)**	(+)	(–)	(–)
Assault	(–)***	(–)***	(+)***	(+)***	(–)	(–)**	(–)***	(–)**
Robbery	(–)***	(–)	(–)***	(–)***	(–)	(+)	(–)	(–)
Drug	(–)***	(–)***	(+)***	(+)***	(–)	(+)	(–)**	(–)***
Travel and tourism								
Property	(–)***	(–)	(+)***	(–)	(–)	(–)***	(+)***	(+)
Harassment	(–)***	(–)***	(+)***	(–)	(+)***	(–)	(–)**	(–)
Assault	(–)***	(+)	(+)***	(–)***	(–)	(–)***	(–)	(–)
Robbery	(–)***	(–)***	(+)***	(–)***	(+)	(–)	(–)	(+)
Drug	(–)***	(–)***	(+)***	(+)***	(–)	(–)***	(–)	(–)*

War and conflict								
Property	(–)***	(+)***	(+)***	(+)***	(+)	(–)**	(+)***	(–)
Harassment	(–)***	(+)	(+)***	(–)***	(+)	(+)	(–)	(–)
Assault	(–)***	(–)***	(+)***	(+)***	(+)	(–)**	(–)***	(–)**
Robbery	(–)***	(–)***	(–)***	(–)***	(–)	(+)	(+)	(–)
Drug	(–)***	(–)***	(+)***	(+)***	(+)	(–)	(–)***	(–)***
Work								
Property	(–)***	(+)***	(+)***	(+)***	(+)**	(–)	(–)***	(–)
Harassment	(–)***	(+)***	(+)***	(–)***	(+)	(+)	(–)**	(–)***
Assault	(+)***	(–)***	(–)***	(–)	(–)	(–)*	(+)	(–)
Robbery	(–)***	(–)***	(–)***	(–)***	(+)*	(+)***	(–)	(–)
Drug	(–)***	(–)***	(+)***	(–)***	(+)	(–)***	(–)	(–)***

** p<.1; ** p<.05; *** p<.01*

Table 22: Property at private places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	12,088 (6.13)***	19,031 (5.91)	230 (7.02)**	11,858 (6.11)
Body and appearance	21,630 (10.96)***	33,402 (10.37)	408 (12.46)***	21,222 (10.94)
Business	19,208 (9.74)***	33,351 (10.35)	378 (11.54)***	18,830 (9.70)
Clothes and fashion	30,102 (15.26)***	47,822 (14.84)	464 (14.17)*	29,638 (15.27)
Crime and law	21,338 (10.81)	34,812 (10.80)	330 (10.08)	21,008 (10.83)
Culture	43,819 (22.21)***	68,371 (21.22)	733 (22.38)	43,086 (22.21)
Education	23,254 (11.79)**	37,242 (11.56)	381 (11.63)	22,873 (11.79)
Family and life stages	32,669 (16.56)***	50,209 (15.58)	526 (16.06)	32,143 (16.57)
Food and drink	30,154 (15.28)***	46,384 (14.39)	518 (15.82)	29,636 (15.27)
Health	40,644 (20.60)***	65,300 (20.26)	712 (21.74)	39,932 (20.58)

Houses and buildings	31,170 (15.80)***	64,454 (20.00)	524 (16.00)	30,646 (15.79)
Language	38,999 (19.77)***	75,770 (23.51)	648 (19.79)	38,351 (19.77)
Leisure	52,364 (26.54)***	92,622 (28.74)	826 (25.22)*	51,538 (26.56)
Nature	28,045 (14.21)***	51,736 (16.05)	494 (15.08)	27,551 (14.20)
Personality and emotions	79,775 (40.43)	129,654 (40.23)	1,336 (40.79)	78,439 (40.43)
Religion and politics	9,059 (4.59)***	15,378 (4.77)	167 (5.10)	8,892 (4.58)
Retail	14,434 (7.32)***	24,971 (7.75)	234 (7.15)	14,200 (7.32)
Science	13,602 (6.89)***	29,310 (9.10)	251 (7.66)*	13,351 (6.88)
Social issues	9,296 (4.71)***	13,709 (4.25)	167 (5.10)	9,129 (4.70)
Technology	22,190 (11.25)	36,219 (11.24)	368 (11.24)	21,822 (11.25)
The media	15,937 (8.08)**	26,548 (8.24)	263 (8.03)	15,674 (8.08)
Travel and tourism	36,831 (18.67)***	66,721 (20.70)	592 (18.08)	36,239 (18.68)
War and conflict	5,229 (2.65)***	11,130 (3.45)	90 (2.75)	5,139 (2.65)
Work	18,415 (9.33)***	31,024 (9.63)	346 (10.56)**	18,069 (9.31)

* *p<.1;* ** *p<.05;* *** *p<.01*

7.5 Discussion

This research starts with a basic concept; potential behaviours and actions can be captured by the traits of an individual. Among various approaches to quantify such traits, this study uses digital traits in particular those 'stamped' on social media posts. In traditional criminology, potential traits that might explain criminal behaviours have often been defined by a feature of an individual such as gender, level of education, or ethnicity. We cannot completely deny the leverage of these static features on our potential behaviours, however, we cannot completely define ourselves by them either, because what we believe and what does or does not interest us may be a more reasonable and logical way to define us in terms of our likely actions than static characteristics.

In this research, online traits were measured by a content-based analysis of social media posts using topic-based and online activity-based classifications. The results of this study demonstrated that online traits at places are substantially related to crimes and therefore meaningful indicators of criminal situations. Results revealed that the posts about tourist attractions were generated in higher levels at places with property crime records. Further, locations with any type of crime at semiprivate places had more posts referring to commercial and retail topics. In particular, these posts containing topics of Culture, Food/Drinks, Houses/Buildings (i.e. Historic and Religious building, Public places, etc), and Retail were highly captured at the places with property crime or crimes at semiprivate places. Unlike property crimes, places with violent crimes such as harassment and assault crimes produced more posts about Personality/Emotions. This is interesting, as it appears to be representing the instrumental nature of property crime and the more expressive nature of violent crimes.

Results also demonstrated that some topics are highly generated in places without any recorded crime. The posts showing interests on topics which we frequently use in our daily conversation such as Clothes/Fashion, Health (i.e. Diet, Fitness, etc), Leisure (i.e. Hobbies, Other activities, etc), and Nature (i.e. Weather), and subjects with respect to current affairs or issues such as Business, Crime/Law, Education, Social issues and Religion/Politics were less found in crime places. Notably, messages about Religion/Politics were consistently less found in crime places. It seems that where we discuss our passions and interests at deeper levels of interaction less crime tends to happen. The analysis comparing topics between crime and non-crime times demonstrated similar findings. Results indicated that the posts to discuss serious subjects such as Business, Retail, Technology and The media, or to have casual conversation including Food/Drink, Health and Nature were less observed at time of crimes in all crime types at all location types.

The results of online activity analysis indicated that the posts for Conversation, Informal broadcasting and Informative broadcasting were found less at most crime locations by crime types. The posts with location information such as Location status and Informal broadcasting with location status were found substantially more at crime places. For analysis on differences between time of crime and non-crime, the online activities including Conversation and Informative broadcasting were less found at the time of crimes and the activities such as revealing location information were more detected at crime time.

It is interesting to report that people who post Conversation, or Informal broadcasting are more likely to be associated with less crime. This might be understood in the concept of social cohesion, which is known as a key feature of deciding a level of informal guardianship of the area. In previous studies, social cohesion is estimated by various social and physical factors; a survey (e.g. a level of neighbours chat (Yang et al.,2002; Cramm et al, 2013)), physical factors (e.g. gated communities (Manzi and Smith-Bowers, 2006), schools (Witten et al., 2001), and greenspaces (Farahani et al., 2018). Like those indicators, a level of social cohesion can be measured by social behaviours, here, sociable online-activities such as communicating with other uses, sharing their thought via posts, and showing interests about social issues. These findings are consistently observed at crime places that have in- and out-flow of dynamic populations.

Overall, this study suggests that online traits generated some meaningful and valuable insight into people in crime situations and how they take their roles in the situations. According to the results from the topic-based and online activity-based analysis, a place and time that has a considerable amount of people who post topics related to daily conversation and who use the social media to communicate with others are more likely to be related to lower crime rates. The outcome also indicates that people who discuss social and religious issues on social media are less likely to be found in a place and time where and when crimes occurred.

Equally importantly, some online traits are more likely to contribute to criminal situations. People who create social media texts about any types of touristic attractions and who share the locations where they visit are more likely to be discovered at property crime places and time of crime only at the places allowing public access. Lastly, people who post comments about their emotions are more likely to be exposed to violent crimes.

Based upon the theoretical framework of the current research, these findings can be interpreted as the people showing a sociable disposition and having interests in social issues or religious belief could potentially be guardians of the places where they are or potentially indicate community spirit that helps inoculate against crime. The findings also can be read as the places having a substantial amount of posts about touristic attractions such as a place to enjoy, go, and eat are more likely to generate high crime opportunities.

While the analysis produces meaningful outputs for future crime science research, some significant limitations of the utilised data still remain. Although comparing with other social media platforms Twitter is optimised for sharing the topics that users are interested in rather than sharing self-structured personal data (Hughes et al., 2012), the ability to harvest the information cannot fully justify the validity of the data. Therefore, the self-selecting nature of the sample still needs to be consider to interpret the outputs of the current research and future research needs to understand there might be a significant gap between the self-selecting samples and the total population.

Note: over and above data representativeness which is covered in chapters 3 (p.3.1.3)

The diversity, intensity or interrelation of topics was not considered in the analysis. According to the studies on the diversity of interests, people with less constrained interests have greater flexibility to tailor themselves to others adopting and communicating with the other people. Taking a closer look at the intensity of some online traits such as political orientation may generate interesting findings related to a level of bonding with people who share the traits. More importantly, topics are not independent. Examining the interrelation of different topics would create how they correlate and act in crime situations.

As a primary study attempting to understand the correlation between online-traits and crime, the research tried to employ a reasonably simple and straightforward method that would be easily modified or updated in future studies on crime patterns. For instance, instead of using the existence of the topic words, unsupervised computer assisted methods or developing more complicated coding rules could suggest more nuanced findings.

Further issues are the relationship between individual posts and the collective 'concerns' of those in a particular spatiotemporal unit. As mentioned above the measures discussed here can be seen in some senses as population-driven 'social atmospheres' of areas at certain times. The degree to which the social atmosphere of a street segment is the correct unit of analysis for understanding crime situations at a specific place is an open debate and should be explored in further research.

It might also be interesting to explore the interaction between the concepts measured by the posts and other measures or predictors of crime at place. For instance, in some cases the posts could be seen to be partly representing land use rather than 'social atmospheres'. This is evident for some of the property crime analysis. However, it is rare to be able to measure concepts such as personalities and emotions at the small area level and therefore posts certainly add an extra layer to crime analysis.

8 DISCUSSION AND CONCLUSION

8.1 SUMMARY OF THESIS

The purpose of this thesis has been to examine the potential utility of using geocoded social network data to test criminological theory concerned with crime patterns. While theories of environmental criminology concern the dynamic interactions between offenders, victims, and guardians in crime situations, traditional methods of data collection are ill-equipped to capture the dynamic patterns associated with people's daily activities. Moreover, the approaches to data collection adopted in the field rarely offer insight into the attributes of population, focusing on numbers rather than characteristics. Neither do they examine social cohesion in a dynamic way, thereby ignoring the potential for collective action in specific settings.

These shortcomings of the data used in previous research cannot be addressed using traditional methods of data collection, which impedes progress both conceptually and empirically. Traditional methods have tended to rely on large-scale census data or small-scale detailed surveys. Research using the former is restricted by the prescribed fields collected for administrative purposes, and that using the latter suffers from inadequately sized population samples. Both these sources are also insufficiently dynamic - representing only one or several points in time. Geolocated social network data, however, provides a potential opportunity to advance research conducted in the discipline, providing as it does information on people's patterns of movement as well as semantic content.

Based on the findings and limitations identified from previous studies, there was a need to signpost for crime scientists the various directions they can explore in future research on the relationship between crime and dynamic population, specifically computed from georeferenced social media messages. Considering that the previous studies did not successfully produce robust correlations between crime and the population and only focused on the volume of them (see Malleson and Andresen, 2015b; 2016), the current study suggested to explore the relationship employing 1) further categorisations of crime types and locations and 2) in-depth information about users - in this study emotions and tendencies, which can be extracted from the messages.

In the current research, two activities were necessary to pursue the aim. First was the development of a theoretical frame more for exploring the relationship between identified online-population and crime. This was necessary because the majority of previous research tends to suggest advanced-computational methods or demonstrate statistical significance between social media data and crime rather than providing an in-depth criminological framework, which therefore fills a gap in current research in the field. The second was an examination of big data analysis methods previously used in related fields, which enabled the effective identification of methods suitable for spatiotemporal crime pattern analysis. Building on these activities, this research analysed the relationship between mobile population as measured by social media data and crime with different crime types (property, harassment, assault, robbery, and drug) at different locations (private, semiprivate, semipublic, and public). Whilst categorising by both crime type and place have increased the number of analyses necessary in exploring relationships, this can be considered a strength of the research approach as it is evident from the results that they are important contextual characteristics in understanding the crime-population relationship.

Summary of results

This section briefly summarises the thesis findings from the three empirical chapters. In each case the research questions are recapped and the main results from the analysis are reported.

Mobile ambient population patterns

What are the spatial and temporal patterns of mobile population? How can these patterns of movement and fluctuation be used in efforts to test theories of crime pattern formation?

To answer these questions, the movement and fluctuation of estimated ambient population derived from Twitter data was analysed at the spatiotemporal level. The research found that Twitter appeared to be a particularly appropriate denominator for crimes at places where there is public access. In most cases, crime increased when the floating population increases, but for minor violent crimes (harassment), the high flow of population tends to reduce crime. Additionally, violent crimes at private places, which might be understood in the neighbourhood cultural context of residents, were less associated with mobile population, in other words, were less likely to be affected by the existence of potential guardians. This research indicates that (1) as shown in other research studies, it is helpful to use ambient populations to study crime and (2) that this relationship is heavily mediated by the type of crime and the type of location considered.

Mobile population emotion patterns

What "emotions" of people of the mobile population are associated with patterns of crime concentration at place and time? What types of crime are significantly correlated with particular mobile user emotional patterns?

In answer to these questions, the second analysis suggested that certain types of negative emotions such as anger and swearing might be related to crime more than other negative emotions. In particular, the positive role of anger and swearing on violent crimes was observed at private places and semiprivate places. As highlighted in previous studies that discuss the positive effects of fear on encouraging crime preventive behaviours, fear as measured in the ambient Twitter data is more likely to be associated with a lower volume of crime. The results also underlined the correlation between emotional traits (emotional tendency) of the places and emotional states of crime situations (temporal emotion changed by stimuli).

Mobile population attribute patterns

How social cohesion can be measured in the microblogging messages? What types of online-traits of the present population (topic/interest or online-activity) are associated with crime patterns positively or negatively? Do different online-traits correlate with certain crime types and crime situations?

These questions were answered by exploring online traits (topic-based and online activity-based) analysis. The results revealed that posts about tourist attractions were highly detected at places with a high level of property crimes. At spatiotemporal units that did not experience a crime (non-crime units), posts containing topics which we frequently use in our daily conversations (i.e. fitness, hobbies, and weather) were more frequently identified. For online activities-based analysis, social online behaviours such as having a conversation or posting thoughts were found more in non-crime places and situations . These findings offer empirical support towards the mechanism of social cohesion/collective efficacy; that areas with 'social atmospheres' are less likely to have high crime levels.

8.2 research implications

There are a number of general implications that arise from this research. These relate to (1) the development of more theoretical big data approaches; (2) the use of social media data in understanding social connections that influence crime; (3) the role of social media approaches in measuring near crime causes and reducing bias against vulnerable population groups. These are discussed in turn in this section.

Stories behind big data

Nate Silver, a statistician and a developer of prediction and forecasting programmes expressed concern in an attitude that sees big data as a cure-all and identified issues that could be caused by data-driven predictions. In his book *The Signal and the Noise: The Art and Science of Prediction*, he said "The numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning...Before we demand more of our data, we need to demand more of ourselves" (Silver, 2012, p.7). Like Silver pointed out, this data (such as the social media data used in this thesis) is the sheer information and does not speak meaningful things before being transformed from information into knowledge.

Although it is hard to deny that recent studies on big data, which use advanced computational methods, increase efficiency and have been brought an excellent new perspective to the crime science field, analysis built upon weak theoretical understanding clearly has a limitation since this generates facts rather than knowledge. A theory-led approach is more likely to generate knowledge by revealing underlying stories that help explain why the patterns have been observed and generating future direction on how to utilise the data to develop crime prevention strategies.

In other fields, there have been numerous attempts to develop forecasting models using big data such as medical, natural disasters, politics, economy via innovative and highly mathematical approaches but purely naïve models have limited success in generating successful predictions (Lazer et al., 2014). Taking one example, in the early stages of big data analysis, Google introduced 'Google Flu Trends' on Nature which expected to provide an accurate and inexpensive way to track the outbreak of flu symptoms (Ginsberg et al., 2009). The engineers simply set up two key search words, 'flu symptoms' or 'pharmacies near me' in belief that these would be directly linked with the outbreaks but the this not enough to produce the knowledge that Google sought. The keywords were selected by simple intuition and did not incorporate careful hypothesis (Harford, 2014); people without the flu would google for the information. As a result, it missed the 2009 swine flu pandemic and over-predicted, nearly double, flu outbreaks in 2013 (Salzberg, 2014). Of course, whilst reasons for model failure might have come from other factors such as technical errors or inappropriate analysis designs, an absence of theoretical rationale is a threshold issue for big data analysis.

With this in mind, producing a theoretical structure was deemed the first step that we, crime scientists, need to take in turning big data sources into a worthwhile asset for studying crime and constructing crime prevention strategies. For this reason, as a study at the beginning stage, the data was construed within an environmental criminology framework, using particular theories of neighbourhood effects and opportunity perspectives that has been studied in the past decades using various approaches and methods. Thus, the research attempted to examine the data utilising the fundamental concepts and factors adopted from previous studies, to best highlight the usefulness of this data in future crime research.

A new wave of criminology

Although the validity of using social connections identified from online platforms as an indicator of real-world interpersonal relation has not yet been firmly established, recent studies have found reciprocal interactions between users in on- and offline social networks (Baym, 2002; Chan and Cheng, 2004). Theories of social cohesion and disorganisation hypothesise that socially well-organised neighbourhoods suppress crime through increased levels of collective efficacy, which enhance a willingness to intervene for the common good (see Sampson and Grove, 1989; Sampson, 1991; Sampson and Raudenbush, 1999).

In previous empirical studies, social cohesion has been explored with various approaches, but ways to analyse collective efficacy have been limited by past measurement. For example, when Shaw and McKay published their initial ideas, the mechanism they wanted to discuss was not the effects of three features of neighbourhoods; mobility, racial heterogeneity, socioeconomic status on crime, but the informal mechanism of social control which would be damaged by poor social cohesion. However, at the time there was no data available to measure the latter directly. It is not easy to deny that social and demographic characteristics such as race, the period people have been resident in an area, or the levels of income and education, describe types of individuals who share similar characteristics, but such characteristics are also not a direct way to measure how the individuals in a group effectively function. As the theory states that shared sense of identity, shared norms, and shared values are the core indicator of the willingness to enforce these norm and values (Sampson et al., 1997), the different ability of neighbourhoods to act collectively should be measured based on how they define their identity themselves not through the use of proxies.

In the pre-internet era, people were only connected in the physical environment and the boundary of people's lives were very limited. This meant that people who shared physical boundaries were those making real contact and hence interpersonal relationships. In the internet-age the physical boundary has been collapsed and social interaction limits are widened. As we continue to build up relationships with diverse people who have different backgrounds and opinions, the concept, collective efficacy, should be measured via the data fit for representing the conditions we live in today. The dynamic status such as what people think, and how people feel, as measured by their attitudes, communications and emotions, therefore, would be a more direct/accurate/correct indicator to predict their behaviours about how they would deal/react with disorders or crimes rather than static information/tangible features such as what they look like or where they come from. In modern society, sharing the same tangible features might be not enough to understand how much people feel a sense of belonging to the group they are in or the community they belong to. In this thesis, therefore, the construction of new direct and indirect indicators of social connectivity are explored in the scientific faith that social media can be accepted as a crucial channel of communication and can provide a new perspective for future crime research.

The data for ruling out bias

To build an easy and quick crime prevention strategy, it is very tempting to start it with oversimplified demographic characteristics of criminals; young, poor, coloured, and male. However, this biased prevention approach has been damaging to society and this discrimination on the 'prime suspects', who equally can be called the minority, has been aggravating social divisions between them and others (see *'Why people obey the law'*(2006) and *Trust and legitimacy* (2011) by Tom Tyler).

Not only can it lead to a negative preconception against a certain population segments, causing disproportionate law enforcement and unbalanced judicial decisions, but it also shortens the sight of crime researchers blocking creativity and innovation when exploring the mechanisms of criminal behaviours. Because the abundance of statistical studies representing 'crime causes' using demographic characteristics, these studies self-perpetuate and limit innovation. Although the demographic characteristics could statistically reveal a group of individuals more likely to commit law-violating behaviours than others, putting too much weight on their characteristics overestimates distal-causes, (e.g. socioeconomic status), and underestimate near-causes, (e.g. immediate criminogenic situations), which can create more effective, efficient and fair solutions for crime prevention.

The official statistics of criminal demography show that 33 per cent of people who were sentenced were black in the US (According to 2017 census, black represented 12 per cent of the U.S. population) (Pew Research Centre, 2019). Although alternative statistics such as criminal conviction, stop and search, or victimisation persistently show that a certain group of people is more likely to fall into a problematic category, this does not justify an oversimplified selection of the 'usual suspects'. As the Criminal Justice System exists to punishing the guilty and protect the innocent, crime prevention policies should not accuse the people who just belong in the vulnerable group or simply categorise them as potential criminals.

The empirical observation on crime patterns has consistently generated the same outcome; crime is concentrated at certain and a few places and time (see Johnson, 2010). As this finding is an accepted axiom, it has taken an important role in policy implications and has been led to significant changes in policing strategies. However, place-based approaches are inclined to treat the reasons why crime are concentrated in the places less importantly (Weisburd and White, 2019). The policing strategies could be effective but could be questionable in the longer term because preventive actions are developed based on the results of crime - spatial patterns - not the reasons that lead to crime manifestation in the particular situation. Recent research has begun to expand crime hot spot research with this in mind, such as looking at crime places in relation to the level of mental and physical illness of the place (Weisburd and White, 2019). This is in its infancy and more studies are needed to offer unique and different ways of framing crime places.

Measuring demographic and socioeconomic trends have to date been treated as the fundamental method of understanding crime behaviours. Factors such age, race, sex, levels of education and income, marital status, occupation, and religion, are typically used to identify groups and community features and predict the behaviours of those groups (Lavrakas, 2008).

However, the framework is incomplete according to one of the findings from this research- that 'stable' demographic and socioeconomic characteristics are only effective at explaining crimes at private places, and a complementary approach is needed . An approach that employs big data as has been described in this thesis is ageless, financialess, raceless, and genderless and represents opinions and emotions rather than characteristics has promise and could combat some of the social problems the current policing strategies create. As emphasised in the findings of three empirical studies, mobile population would be a more precise indicator of the characteristics of people and explain crime at public places that allow public accessibility and have a high level of population flow. It follows that utilising social media data, which has real-time information of people with their location, would offer more benefits for building crime prevention strategies targeting crimes at public places rather than those at private places.

8.3 Research limitations and future research

This research has only focused on text-based analysis which is the most basic and simplest way to explore trends in social media data, which might well not be the optimum approach. This research focused on understanding 'placenality' which refers to the combination of characteristics forming a distinctive character of a place as measured by online-population. In other words, the research tried to catch the pattern of the places by the accumulated patterns of dynamic population but did not consider characteristics of individuals directly impacted by the event which may bring more interesting findings and more extensively utilise the capacity of the new data source. So, this research did not study the interactions between users at the micro-level based on individual histories and activities.

One of the most notable characteristics of social networking services are networking. In the case of Twitter, it has a networking system which is called "following", in which each individual can decide who to 'follow' or 'communicate'. According to a recent study, nearly 70 per cent of the users of Twitter follows more than 80 per cent of the followers meaning that a group of people who follow each other would share some similarity in activities or subjects (Weng et al., 2010). For future research, adding reciprocity in the following relationships such as network between users and the characteristics of the followers could generate more interesting results.

There are many diverse platforms of social media. I utilised text-based Twitter data for this research but an analysis with Facebook or Instagram that includes pictures and videos might have produced different results. For future research, using a variety of social media platforms would draw more solid and concrete conclusions, particularly if the analysis triangulated findings from multiple data sources.

The use of new types of data such as social media text faces several critiques that are worth considering here. First of all, the data may not be sufficient to represent the entire population, which means the online-population data could be vulnerable to selection bias. The most significant doubt comes from the distrust that the data would not be enough to represent the target population in the real-world. Furthermore, the geotagged messages are a self-selecting sample collected from Twitter users who made the active decision to publish their locations from the total population in the area. Unlike a random sample selection, this decision made by the users could generate a bias future researchers should carefully consider. This concern can be alleviated by analysing explicit demographic characteristics of the users (e.g. last names, gender, age, and languages) (Sloan and Morgan, 2015; Leak, 2017) beyond their user names. In looking to future crime research using the data, while the data offer great opportunities to explore the information which has not been captured and tested in the field of crime science, a criticism about how representative the data are needs to be addressed for its broader applicability.

Note: see the section 'Big Data hubris' in Chapter 3 (p.26)

However, unless it is a complete enumeration, the data that we collect using more traditional ways such observation, any types of survey or interview has the same potential risk even if it is randomly selected. Indeed, even though over 70 per cent of the population is using the social media daily basis, the risk that specific groups in the population could be over-represented cannot be removed completely. This limitation of the data could be mediated by conducting studies on various online platforms and accumulating the data of the characteristics of population measured from each platform.

Second, there is a critique on the relationship between online-personality and off-line personality. The data may not represent the real face of the people and some people might not be genuine in the online environment. Some people could hide a genuine side of their personality from online-environments. Put a different way, people could intentionally or unintentionally create a different identity or emotion, meaning how they behave in the virtual community could not be reflective of their behaviours in the off-line world. In defense of this assumption however, research recently published found a strong correlation between online traits and offline traits (Chamorro-

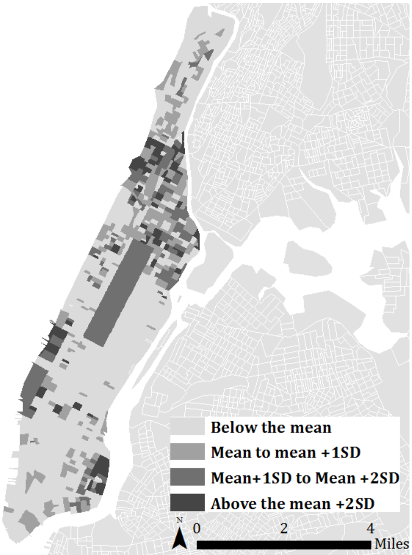
Note: see the section 'online-traits and offline-traits' in Chapter 7 (p.101)

Premuzic, 2015) as the time we spend on online-world are increasing so that it is hard to fake who we are on online.

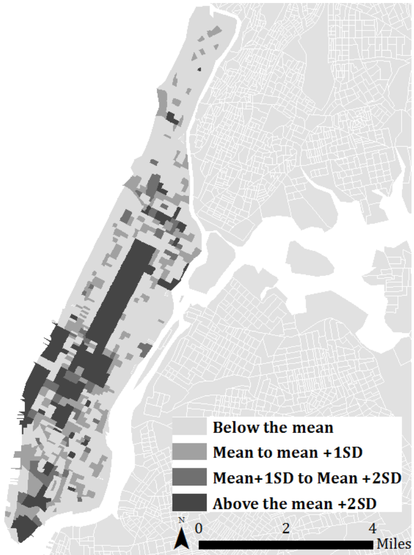
Perhaps the best way of integrating social media data in a field such as crime science is to understand the data as an another population layer; the online-census population. As criminologists have been adding a layer that could have a correlated with crime one by one in the past, the data from social media is one more data set to assimilate - such as risky or safety facilities, street network, and so on (Johnson and Bowers, 2013; Bowers, 2014). Suffice to say all datasets have their known limitations and these have not restricted empirical exercises using other data sources in the past. Social media data have limitations but also bring new possibilities. With technological development, an incomparable amount of data can be collected with a low cost and level of effort. Over 70 per cent of people use social media - such a level of information is impossible to collect in the off-line world, and the allows crime scientists to scrutinise some of the hard-to-reach populations. This advantage can help expand criminological enquiry and will contribute to building solid policy implications. The data also combines temporal data with geospatial information, which reveal real-time crime atmospheres which have been shown in this thesis to be related to criminal behaviours and situations. This information that people are creating every moment should help to tell the true story of crime.

A
APPENDIX

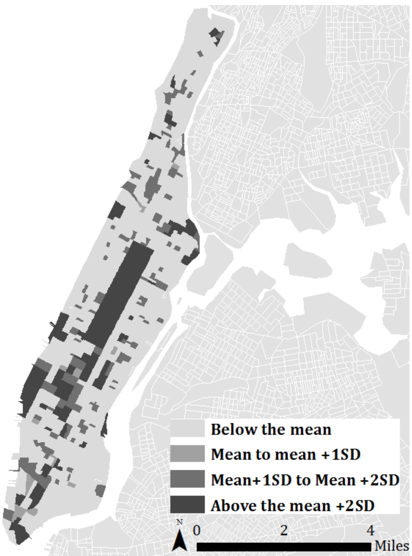
A.1 CHAPTER 4: METHODS



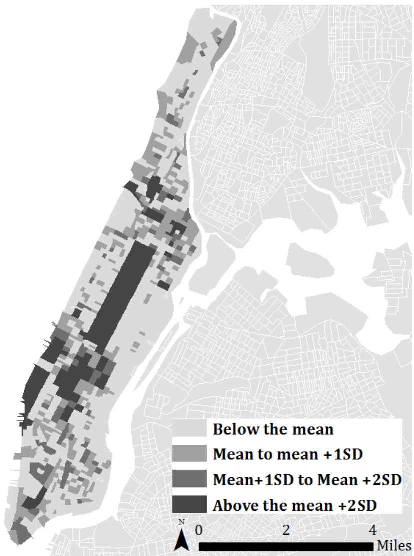
(a) Private places



(b) Semiprivate places

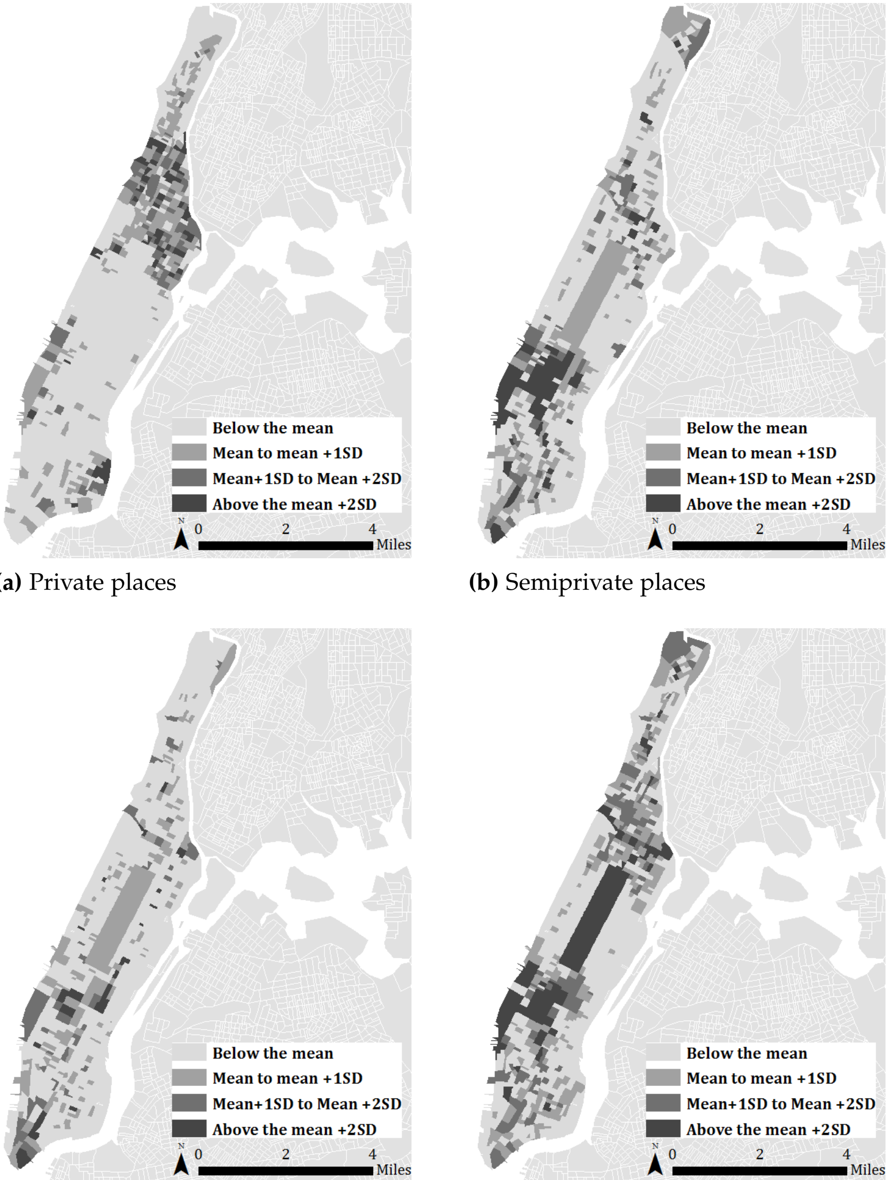


(c) Semipublic places



(d) Public places

Figure 30: Maps of harassment crime



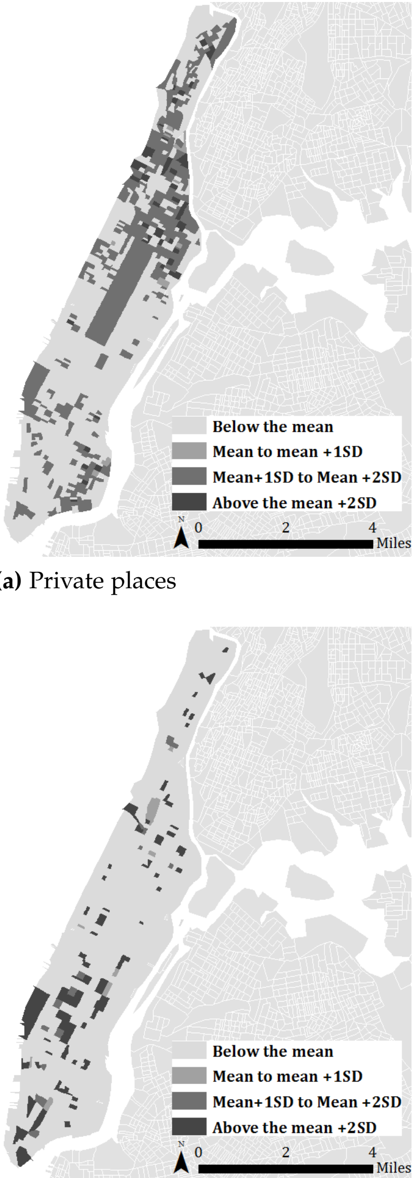
(a) Private places

(b) Semiprivate places

(c) Semipublic places

(d) Public places

Figure 31: Maps of assault crime



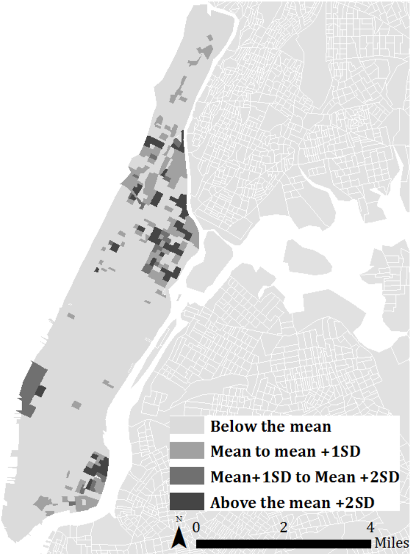
(a) Private places

(b) Semiprivate places

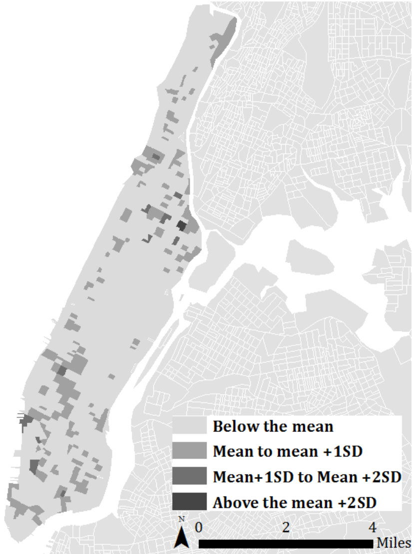
(c) Semipublic places

(d) Public places

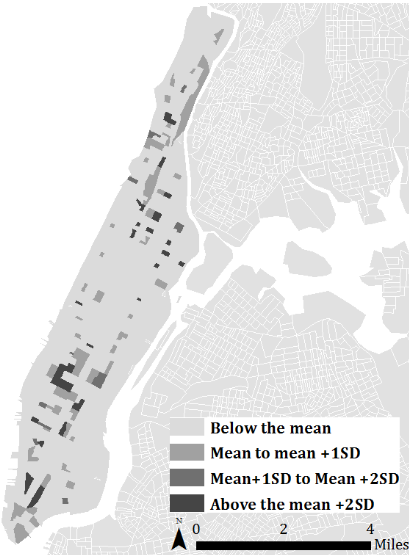
Figure 32: Maps of robbery crime



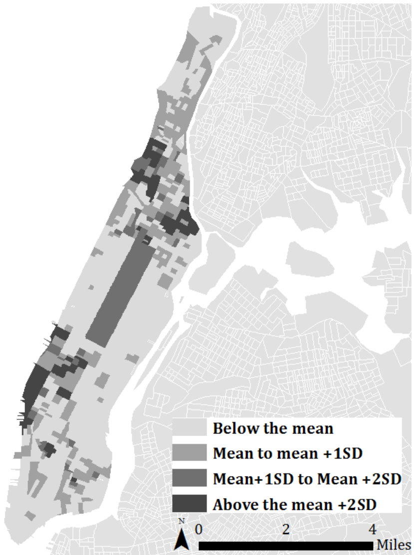
(a) Private places



(b) Semiprivate places



(c) Semipublic places



(d) Public places

Figure 33: Maps of drug crime

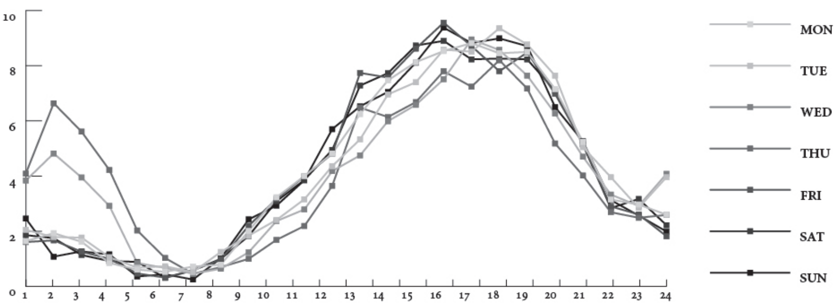


Figure 34: Temporal trends: property crime at semiprivate places | percentages

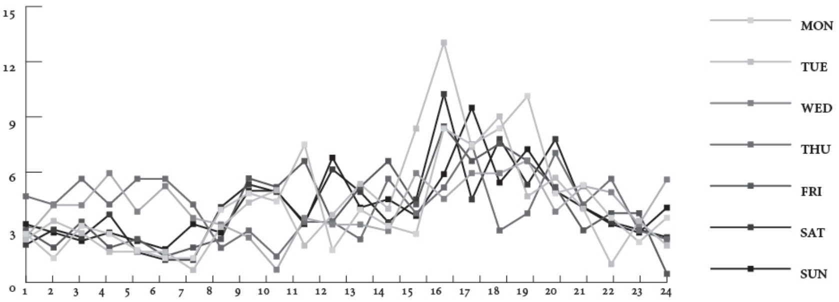


Figure 35: Temporal trends: property crime at semipublic places | percentages

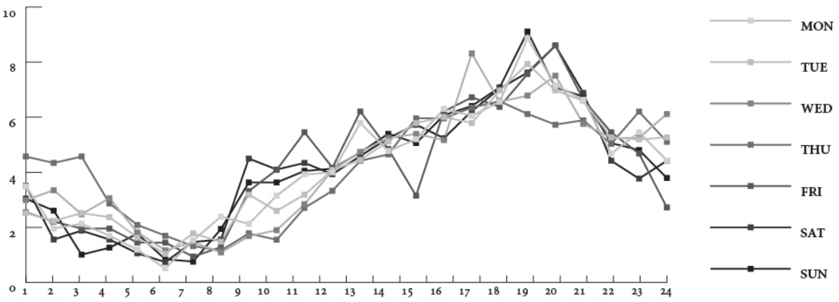


Figure 36: Temporal trends: property crime at public places | percentages

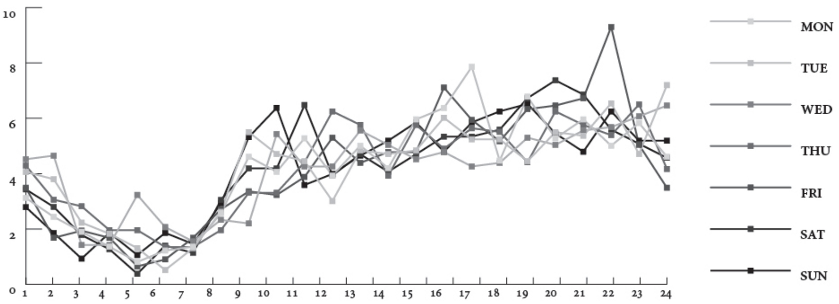


Figure 37: Temporal trends: harassment crime at private places | percentages

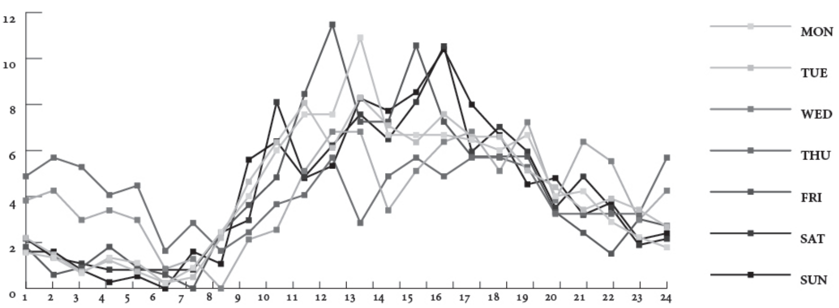


Figure 38: Temporal trends: harassment crime at semiprivate places | percentages

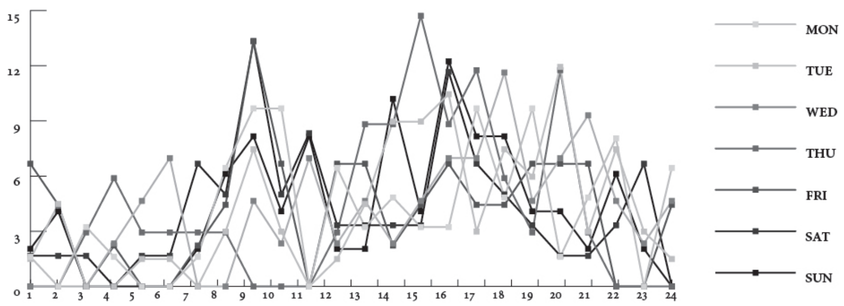


Figure 39: Temporal trends: harassment crime at semipublic places | percentages

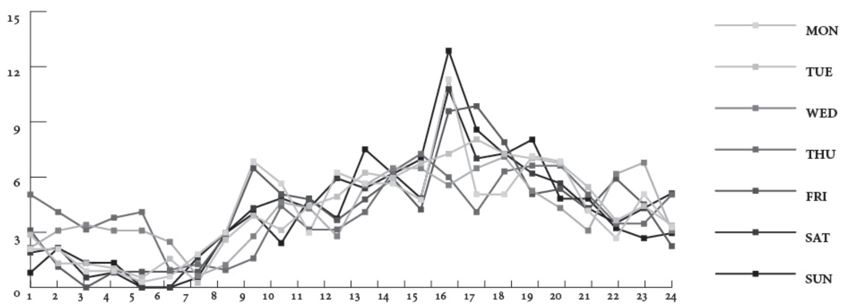


Figure 40: Temporal trends: harassment crime at public places | percentages

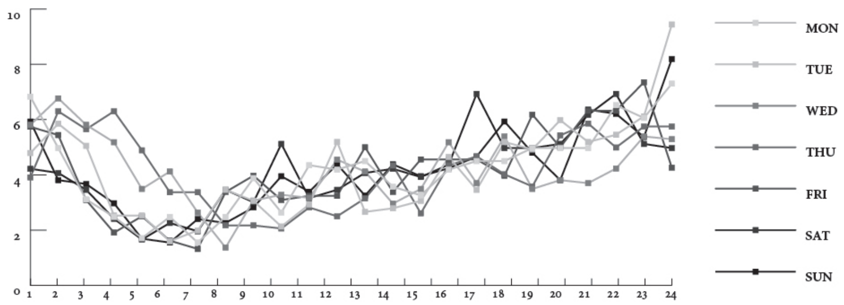


Figure 41: Temporal trends: assault crime at private places | percentages

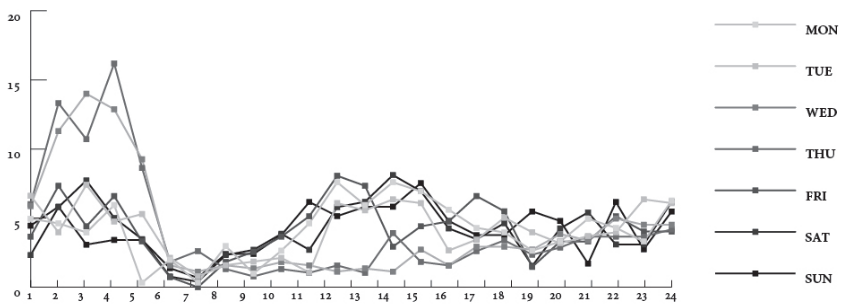


Figure 42: Temporal trends: assault crime at semiprivate places | percentages

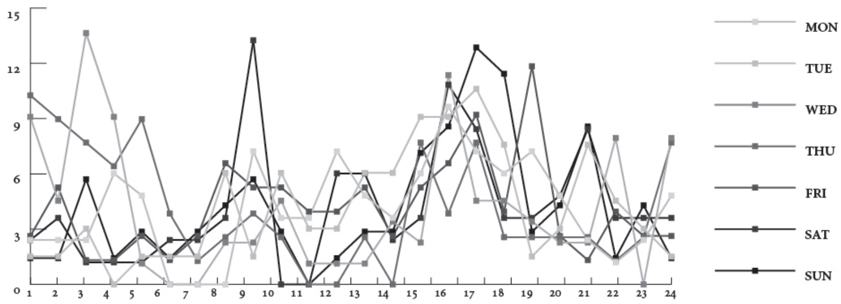


Figure 43: Temporal trends: assault crime at semipublic places | percentages

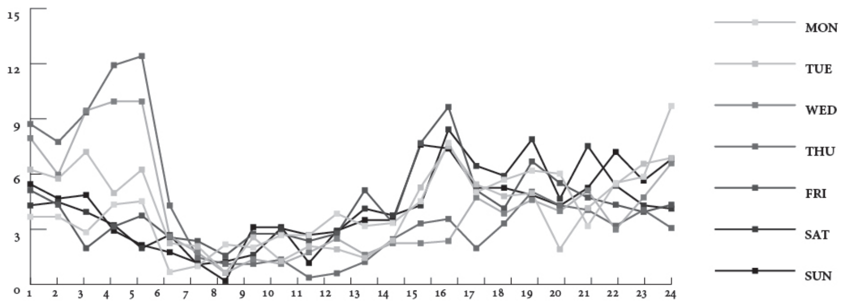


Figure 44: Temporal trends: assault crime at public places | percentages

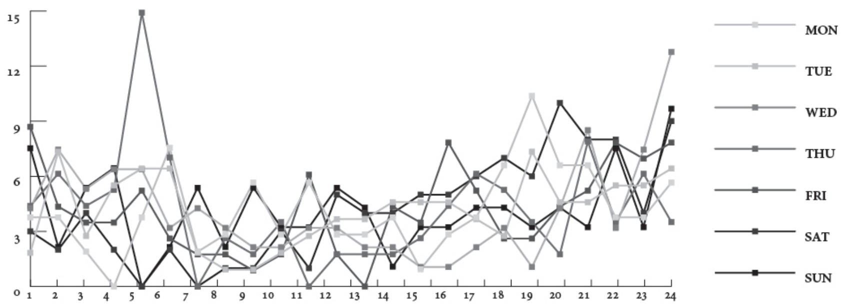


Figure 45: Temporal trends: robbery crime at private places | percentages

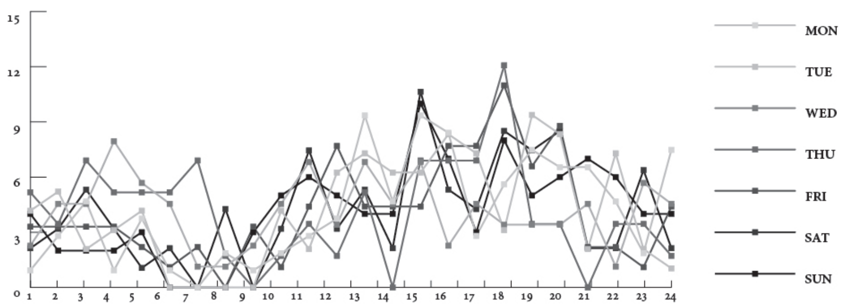


Figure 46: Temporal trends: robbery crime at semiprivate places | percentages

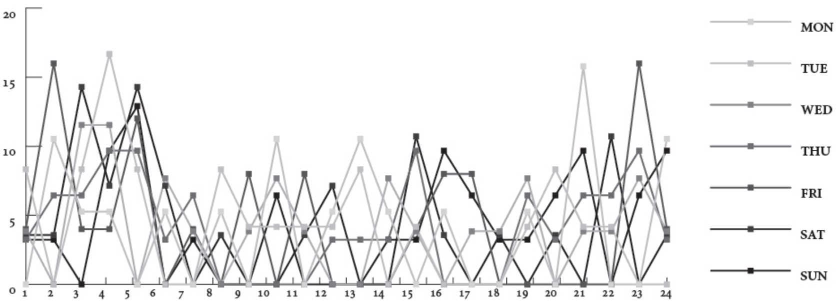


Figure 47: Temporal trends: robbery crime at semipublic places | percentages

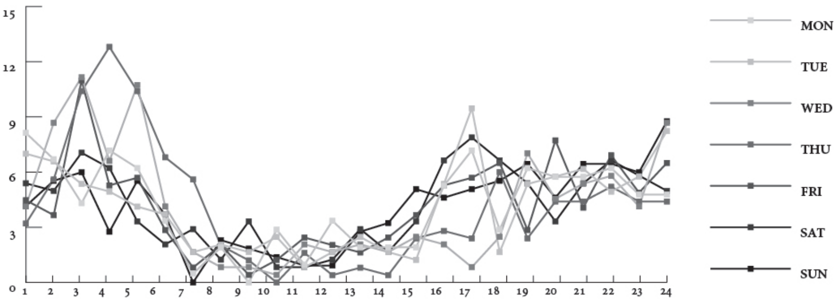


Figure 48: Temporal trends: robbery crime at public places | percentages

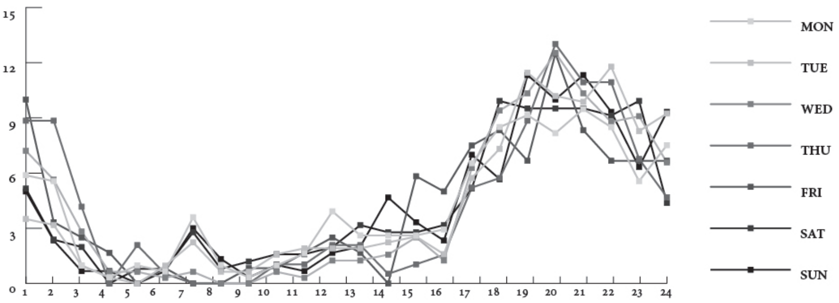


Figure 49: Temporal trends: drug crime at private places | percentages

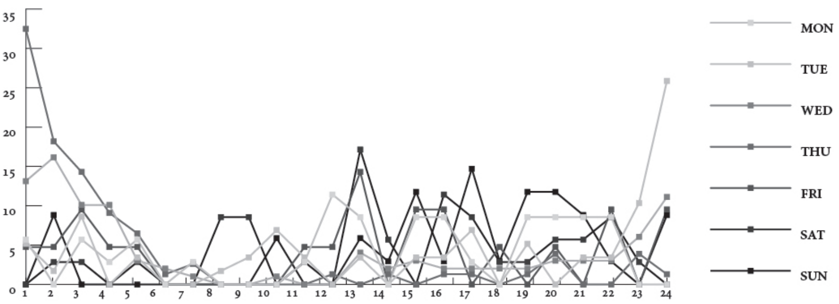


Figure 50: Temporal trends: drug crime at semiprivate places | percentages

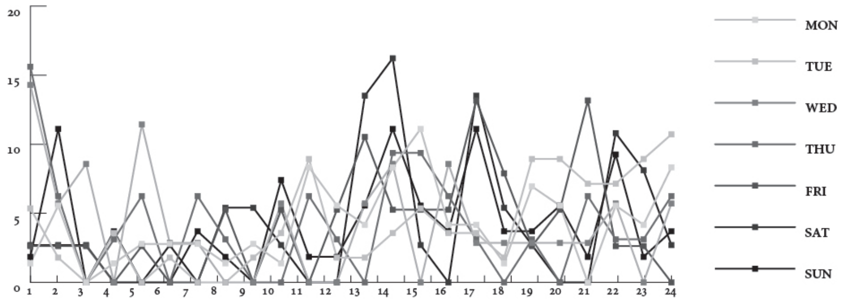


Figure 51: Temporal trends: drug crime at semipublic places | percentages

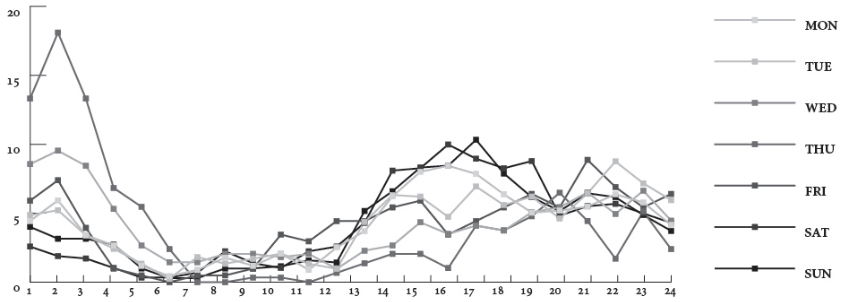
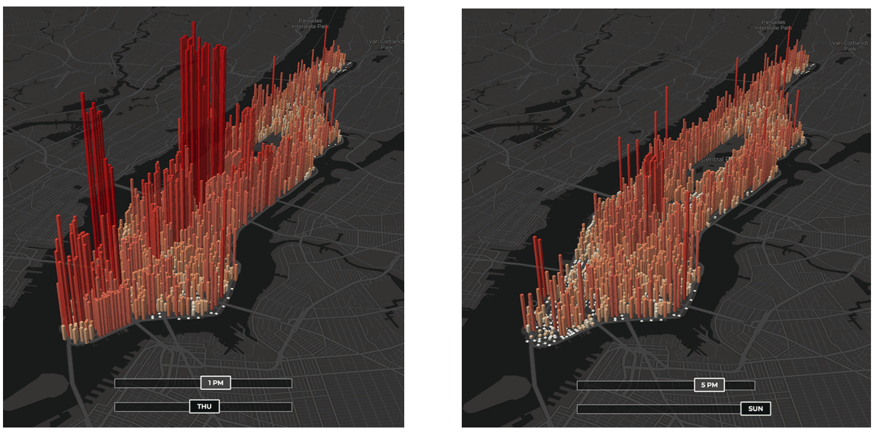


Figure 52: Temporal trends: drug crime at public places | percentages



(a) Weekday **(b)** Weekend

Figure 53: The Population Pulse of a Manhattan Workday by Fung (2018)

A.2 chapter 5: mobile population attribute patterns

A.2.1 *Footnote tables*

Table 23: Results of log-likelihood of SDM, spatial lag and spatial error models

		Private	Semiprivate	Semipublic	Public
Property	SDM	56845.591	-207167.87	155240.59	9020.0794
	SAR	56845.595	-206198.39	155242.45	9028.2695
	SEM	56859.667	-205236.17	155330.99	9189.639
Harassment	SDM	52514.994	125535.68	299887.22	124560.74
	SAR	52516.729	125544.65	299887.11	124563.53
	SEM	52579.483	125721.28	299901.67	124639.35
Assault	SDM	40223.074	112882.13	249050.25	55714.334
	SAR	40224.638	112892.02	249052.57	55729.361
	SEM	40293.283	112979.92	249136.5	55947.101
Robbery	SDM	241605.54	251073.32	364996.08	166461.61
	SAR	241605.56	251073.36	364996.01	166463.59
	SEM	241627.34	251140.1	365030.56	166519.43
Drug	SDM	139535.8	186219.56	303507.14	28892.962
	SAR	139553.33	186219.74	303509.53	28848.273
	SEM	139674.24	186252.63	303634.78	29019.466

Table 24: Results of R^2 of SDM, spatial lag and spatial error models

		Private	Semiprivate	Semipublic	Public
Property	SDM	0.01	0.17	0.03	0.05
	SAR	0.01	0.16	0.02	0.05
	SEM	0.01	0.15	0.02	0.05
Harassment	SDM	0.02	0.02	0.01	0.01
	SAR	0.02	0.02	0.01	0.01
	SEM	0.02	0.02	0.01	0.01
Assault	SDM	0.03	0.01	0.01	0.02
	SAR	0.03	0.01	0.01	0.01
	SEM	0.03	0.01	0.01	0.01
Robbery	SDM	0.01	0.01	0.01	0.01
	SAR	0.01	0.01	0.01	0.01
	SEM	0.01	0.01	0.01	0.01
Drug	SDM	0.02	0.01	0.01	0.03
	SAR	0.02	0.01	0.01	0.02

SEM	0.02	0.01	0.01	0.02

Table 25: Moran's I statistics by crimes

	Private		Semiprivate		Semipublic		Public	
	I	p	I	p	I	p	I	p
Property	0.155	0.000	0.238	0.000	0.211	0.000	0.289	0.000
Harassment	0.309	0.000	0.305	0.000	0.119	0.000	0.256	0.000
Assault	0.357	0.000	0.298	0.000	0.146	0.000	0.378	0.000
Robbery	0.237	0.000	0.189	0.000	0.021	0.191*	0.287	0.000
Drug	0.307	0.000	0.011	0.132*	0.140	0.000	0.363	0.000

** denotes Random. Otherwise, Clustered*

Table 26: Moran's I statistics by independent variables

	Moran's I	p -value	Type
Population size			
Residential population	0.181	0.000	Clustered
Dynamic population	0.450	0.000	Clustered
Social disorganisation			
Education level	0.342	0.000	Clustered
Employment status	0.245	0.000	Clustered
Income	0.449	0.000	Clustered
Residential ownership	0.434	0.000	Clustered
Racial heterogeneity	0.748	0.000	Clustered
High-risk population			
Black young male	0.677	0.000	Clustered
Hispanic young male	0.760	0.000	Clustered

Table 27: Hausman's specification test by crime

	Private		Semiprivate		Semipublic		Public	
	x^2	p	x^2	p	x^2	p	x^2	p
Property	0.13	0.714	8.43	0.037	217.21	0.000	143.25	0.000
Harassment	0.04	0.851	25.44	0.000	45.99	0.000	84.22	0.000
Assault	0.54	0.462	88.42	0.000	33.05	0.000	107.56	0.000
Robbery	1.42	0.233	17.16	0.000	14.31	0.000	57.23	0.000
Drug	0.48	0.488	38.25	0.000	7.63	0.050	1.23	0.267

Table 28: Results of unit root test

Lags	0	3	6	9	12
TW	-2.060	-5.731 ***	-5.865 ***	-6.047 ***	-5.984 ***
Property					
Private	-6.075 ***	-4.267 ***	-5.418 ***	-5.472 ***	-6.011 ***
semiprivate	-2.308	-6.214 ***	-6.543 ***	-6.260 ***	-5.283 ***
semipublic	-6.987 ***	-5.384 ***	-4.410 ***	-4.821 ***	-4.410 ***
public	-3.300 **	-5.078 ***	-6.810 ***	-7.225 ***	-6.989 ***
Harassment					
Private	-4.637 ***	-4.899 ***	-6.249 ***	-5.390 ***	-5.960 ***
semiprivate	-3.817 ***	-4.754 ***	-5.722 ***	-5.826 ***	-4.962 ***
semipublic	-8.926 ***	-5.386 ***	-4.020 ***	-4.872 ***	-5.320 ***
public	-5.084 ***	-4.120 ***	-5.777 ***	-7.578 ***	-7.708 ***
Assault					
Private	-5.021 ***	-4.135 ***	-4.319 ***	-3.775 ***	-4.089 ***
semiprivate	-5.263 ***	-5.614 ***	-5.179 ***	-4.044 ***	-3.402 ***
semipublic	-9.504 ***	-5.646 ***	-4.733 ***	-5.042 ***	-4.455 ***
public	-4.759 ***	-4.474 ***	-4.970 ***	-4.144 ***	-4.355 ***
Robbery					
Private	-8.709 ***	-4.846 ***	-4.891 ***	-5.333 ***	-5.354 ***
semiprivate	-8.312 ***	-4.717 ***	-4.462 ***	-4.480 ***	-4.716 ***
semipublic	-11.938 ***	-5.785 ***	-4.332 ***	-4.833 ***	-4.406 ***
public	-6.227 ***	-4.701 ***	-4.778 ***	-5.684 ***	-5.756 ***
Drug					
Private	-3.199 *	-5.101 ***	-5.409 ***	-4.582 ***	-5.487 ***
semiprivate	-5.363 ***	-4.811 ***	-4.002 ***	-3.740 ***	-3.135 *
semipublic	-9.713 ***	-3.870 ***	-4.515 ***	-3.268 *	-3.353 **
public	-3.073 *	-3.556 **	-4.163 ***	-3.448 **	-2.988 *

** p<.1; ** p<.05; *** p<.01*

A.2.2 *Results: cross-correlation analysis*

Table 29: CCA results of property crime

Lag	Private	Semiprivate	Semipublic	Public
-12	-0.024	-0.627	-0.412	-0.718
-11	0.079	-0.471	-0.308	-0.659
-10	0.156	-0.282	-0.187	-0.557
-9	0.245	-0.081	-0.048	-0.414
-8	0.329	0.119	0.094	-0.239
-7	0.421	0.306	0.240	-0.033

-6	0.512	0.469	0.369	0.181
-5	0.594	0.604	0.463	0.398
-4	0.654	0.710	0.514	0.593
-3	0.687	0.779	0.539	0.749
-2	0.643	0.805	0.518	0.839
-1	0.533	0.786	0.483	0.860
0	0.358	0.723	0.430	0.810
1	0.149	0.624	0.350	0.707
2	-0.073	0.496	0.288	0.571
3	-0.257	0.335	0.205	0.406
4	-0.413	0.145	0.122	0.227
5	-0.537	-0.073	0.013	0.041
6	-0.591	-0.292	-0.108	-0.144
7	-0.591	-0.494	-0.229	-0.309
8	-0.550	-0.656	-0.363	-0.458
9	-0.455	-0.760	-0.456	-0.579
10	-0.339	-0.797	-0.499	-0.671
11	-0.217	-0.765	-0.490	-0.715
12	-0.101	-0.665	-0.430	-0.715

Table 30: CCA results of harassment crime

Lag	Private	Semiprivate	Semipublic	Public
-12	-0.545	-0.181	-0.199	-0.461
-11	-0.535	-0.028	-0.135	-0.353
-10	-0.493	0.113	-0.053	-0.224
-9	-0.405	0.250	0.061	-0.082
-8	-0.260	0.369	0.191	0.090
-7	-0.072	0.482	0.328	0.279
-6	0.145	0.583	0.434	0.453
-5	0.365	0.656	0.506	0.595
-4	0.562	0.700	0.539	0.695
-3	0.715	0.724	0.530	0.754
-2	0.806	0.706	0.490	0.762
-1	0.824	0.644	0.408	0.720
0	0.767	0.540	0.306	0.635
1	0.630	0.386	0.177	0.506
2	0.447	0.194	0.067	0.348
3	0.250	-0.002	-0.030	0.173
4	0.067	-0.201	-0.114	-0.001

5	-0.093	-0.392	-0.191	-0.185
6	-0.228	-0.532	-0.271	-0.348
7	-0.334	-0.631	-0.331	-0.489
8	-0.415	-0.683	-0.377	-0.607
9	-0.468	-0.665	-0.388	-0.666
10	-0.499	-0.594	-0.367	-0.672
11	-0.512	-0.482	-0.326	-0.627
12	-0.519	-0.332	-0.273	-0.539

Table 31: CCA results of assault crime

Lag	Private	Semiprivate	Semipublic	Public
-12	-0.430	-0.071	-0.222	-0.349
-11	-0.509	-0.078	-0.198	-0.382
-10	-0.574	-0.102	-0.162	-0.411
-9	-0.621	-0.155	-0.111	-0.428
-8	-0.627	-0.218	-0.048	-0.431
-7	-0.583	-0.283	0.037	-0.410
-6	-0.474	-0.338	0.115	-0.378
-5	-0.316	-0.342	0.171	-0.322
-4	-0.120	-0.292	0.211	-0.238
-3	0.081	-0.190	0.246	-0.117
-2	0.278	-0.043	0.264	0.036
-1	0.444	0.116	0.286	0.208
0	0.565	0.262	0.296	0.370
1	0.625	0.366	0.278	0.497
2	0.623	0.426	0.255	0.572
3	0.556	0.433	0.229	0.608
4	0.468	0.398	0.169	0.588
5	0.361	0.327	0.089	0.519
6	0.248	0.238	-0.013	0.410
7	0.141	0.134	-0.126	0.277
8	0.043	0.058	-0.222	0.120
9	-0.045	0.009	-0.293	-0.011
10	-0.126	-0.004	-0.323	-0.114
11	-0.203	-0.001	-0.316	-0.189
12	-0.270	0.011	-0.284	-0.225

Table 32: CCA results of robbery crime

Lag	Private	Semiprivate	Semipublic	Public
-12	-0.290	-0.342	0.091	-0.302
-11	-0.367	-0.233	0.051	-0.381
-10	-0.407	-0.123	0.005	-0.448
-9	-0.435	-0.005	-0.052	-0.483
-8	-0.446	0.093	-0.116	-0.512
-7	-0.402	0.192	-0.156	-0.517
-6	-0.322	0.264	-0.214	-0.501
-5	-0.235	0.325	-0.262	-0.460
-4	-0.119	0.396	-0.310	-0.385
-3	-0.014	0.448	-0.307	-0.273
-2	0.096	0.500	-0.258	-0.121
-1	0.208	0.535	-0.162	0.064
0	0.313	0.529	-0.069	0.266
1	0.383	0.501	0.027	0.440
2	0.437	0.436	0.096	0.576
3	0.438	0.342	0.149	0.660
4	0.422	0.213	0.187	0.701
5	0.385	0.059	0.221	0.684
6	0.334	-0.081	0.244	0.611
7	0.267	-0.235	0.226	0.483
8	0.167	-0.351	0.186	0.318
9	0.062	-0.422	0.154	0.149
10	-0.054	-0.463	0.113	-0.006
11	-0.164	-0.442	0.090	-0.138
12	-0.245	-0.382	0.070	-0.248

Table 33: CCA results of drug crime

Lag	Private	Semiprivate	Semipublic	Public
-12	-0.609	-0.139	-0.173	-0.416
-11	-0.663	-0.193	-0.148	-0.403
-10	-0.663	-0.261	-0.110	-0.365
-9	-0.608	-0.310	-0.066	-0.298
-8	-0.508	-0.349	-0.009	-0.212
-7	-0.363	-0.361	0.032	-0.111
-6	-0.181	-0.346	0.085	-0.009
-5	0.038	-0.288	0.160	0.104
-4	0.252	-0.196	0.234	0.218

-3	0.439	-0.079	0.307	0.343
-2	0.579	0.045	0.377	0.470
-1	0.659	0.154	0.425	0.581
0	0.701	0.246	0.447	0.660
1	0.703	0.310	0.436	0.704
2	0.670	0.344	0.376	0.696
3	0.606	0.361	0.276	0.649
4	0.523	0.344	0.166	0.554
5	0.428	0.304	0.055	0.417
6	0.313	0.244	-0.076	0.244
7	0.191	0.188	-0.150	0.064
8	0.047	0.136	-0.206	-0.104
9	-0.113	0.099	-0.237	-0.235
10	-0.288	0.072	-0.224	-0.322
11	-0.457	0.035	-0.224	-0.370
12	-0.594	-0.008	-0.208	-0.387

A.2.3 *Results: spatial Durbin model analysis*

Table 34: Property crime at semiprivate places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.468***	-0.299***		-0.343***
Dynamic population	4.135***	3.091***		3.083***
Education level	-0.724***		-0.856***	-0.636***
Employment status	-0.006		-0.015	-0.014
Income	0.107***		0.105***	0.092***
Residential ownership	0.040***		-0.019*	0.032***
Racial heterogeneity	0.204***		0.291***	0.172***
Black young male	-0.804***		-2.181***	-1.475***
Hispanic young male	-2.039***		-1.606***	-1.453***
Constant	0.014			
Wx				
Residential population		-0.102*		-0.121**
Dynamic population		1.660***		1.725***
Education level			-1.852***	-1.190***
Employment status			-0.134***	0.183***
Income			0.486***	0.364***
Residential ownership			-0.066***	0.066***

Racial heterogeneity			0.235***	-0.235***
Black young male			-1.130***	0.650***
Hispanic young male			-2.866***	0.644***

* p<.1; ** p<.05; *** p<.01

Table 35: Property crime at semipublic places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.002	0.002		0.001
Dynamic population	0.203***	0.189***		0.186***
Education level	-0.024**		-0.045***	-0.027**
Employment status	-0.010***		-0.009***	-0.009***
Income	0.005*		0.004	0.003
Residential ownership	-0.003***		-0.008***	-0.005***
Racial heterogeneity	0.023***		0.027***	0.018***
Black young male	-0.015		-0.169***	-0.109***
Hispanic young male	-0.134***		-0.114***	-0.129***
Constant	0.001			
Wx				
Residential population		0.002		-0.008
Dynamic population		0.053***		0.052***
Education level			-0.154***	-0.101***
Employment status			-0.022***	-0.005
Income			0.051***	0.042***
Residential ownership			-0.000	0.004*
Racial heterogeneity			0.023***	-0.004
Black young male			-0.030	0.061*
Hispanic young male			-0.271***	-0.017

* p<.1; ** p<.05; *** p<.01

Table 36: Property crime at public places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.022**	0.003		0.003
Dynamic population	0.672***	0.621***		0.601***
Education level	-0.006		-0.108***	-0.046*
Employment status	-0.022***		-0.023***	-0.022***
Income	0.029***		0.022***	0.016***
Residential ownership	-0.019***		-0.037***	-0.026***
Racial heterogeneity	0.042***		0.041***	0.013*
Black young male	-0.033		-0.514***	-0.316***

Hispanic young male	-0.149***		0.240***	0.132**
Constant	0.032***			
Wx				
Residential population		-0.041**		-0.104***
Dynamic population		0.062***		0.086***
Education level			-0.463***	-0.285***
Employment status			-0.044***	-0.000
Income			0.175***	0.149***
Residential ownership			0.019***	0.025***
Racial heterogeneity			0.093***	0.024**
Black young male			-0.224***	0.030
Hispanic young male			-1.442***	-0.611***

** p<.1; ** p<.05; *** p<.01*

Table 37: Harassment crime at private places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.135***	0.167***		0.127***
Dynamic population	0.019***	0.002		0.002
Education level	0.021		0.024	0.024
Employment status	-0.031***		-0.029***	-0.031***
Income	0.043***		0.038***	0.040***
Residential ownership	-0.000		0.003	-0.000
Racial heterogeneity	0.035***		0.050***	0.044***
Black young male	0.821***		0.509***	0.513***
Hispanic young male	0.034		0.214***	0.180***
Constant	-0.013***			
Wx				
Residential population		0.181***		0.053***
Dynamic population		-0.071***		-0.024**
Education level			0.103**	0.040
Employment status			0.001	-0.002
Income			-0.025**	-0.009
Residential ownership			0.003	-0.005
Racial heterogeneity			-0.011	-0.009
Black young male			0.547***	0.485***
Hispanic young male			-0.222***	-0.225***

** p<.1; ** p<.05; *** p<.01*

Table 38: Harassment crime at semiprivate places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.007	0.004		-0.002
Dynamic population	0.212***	0.168***		0.164***
Education level	0.032**		0.022	0.038***
Employment status	-0.010***		-0.011***	-0.011***
Income	-0.004		-0.005*	-0.006**
Residential ownership	-0.002		-0.005***	-0.003*
Racial heterogeneity	0.023***		0.034***	0.025***
Black young male	0.083***		-0.155***	-0.105***
Hispanic young male	-0.107***		0.003	0.020
Constant	0.004***			
Wx				
Residential population		0.016*		-0.008
Dynamic population		0.099***		0.108***
Education level			-0.077***	-0.019
Employment status			-0.017***	0.005
Income			0.022***	0.012
Residential ownership			-0.008**	0.002
Racial heterogeneity			0.018***	-0.014***
Black young male			0.176***	0.299***
Hispanic young male			-0.353***	-0.112***

** p<.1; ** p<.05; *** p<.01*

Table 39: Harassment crime at semipublic places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.006***	-0.006***		-0.006***
Dynamic population	0.040***	0.038***		0.038***
Education level	-0.011**		-0.012**	-0.009
Employment status	-0.001		-0.001	-0.001
Income	0.002		0.002*	0.002
Residential ownership	-0.001**		-0.002***	-0.002**
Racial heterogeneity	0.005***		0.008***	0.006***
Black young male	0.007		-0.026**	-0.014
Hispanic young male	-0.028***		-0.038***	-0.044***
Constant	0.001*			
Wx				
Residential population		0.004		0.004
Dynamic population		0.006**		0.005**
Education level			-0.040***	-0.029***

	OLS	SDM (A)	SDM (B)	SDM (C)
Employment status			-0.006***	-0.003
Income			0.010***	0.008***
Residential ownership			-0.000	0.000
Racial heterogeneity			0.002	-0.004*
Black young male			0.002	0.019
Hispanic young male			-0.030**	0.024*

** p<.1; ** p<.05; *** p<.01*

Table 40: Harassment crime at public places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.022***	-0.005		-0.015***
Dynamic population	0.169***	0.151***		0.147***
Education level	-0.016		-0.036**	-0.020
Employment status	-0.008***		-0.007***	-0.007**
Income	0.008***		0.006*	0.004
Residential ownership	-0.004**		-0.009***	-0.006***
Racial heterogeneity	0.020***		0.026***	0.019***
Black young male	0.139***		-0.109***	-0.062**
Hispanic young male	-0.029		0.023	0.016
Constant	0.006***			
Wx				
Residential population		0.004		-0.038***
Dynamic population		0.016***		0.034***
Education level			-0.127***	-0.071**
Employment status			-0.018***	-0.005
Income			0.049***	0.038***
Residential ownership			0.004	0.009***
Racial heterogeneity			0.010*	-0.009*
Black young male			0.143***	0.228***
Hispanic young male			-0.287***	-0.087**

** p<.1; ** p<.05; *** p<.01*

Table 41: Assault crime at private places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.131***	0.183***		0.128***
Dynamic population	-0.006	-0.019***		-0.014*
Education level	0.000		-0.014	-0.013
Employment status	-0.022***		-0.022***	-0.024***
Income	0.048***		0.039***	0.041***

Residential ownership	0.003		0.007***	0.004
Racial heterogeneity	0.019***		0.035***	0.028***
Black young male	1.155***		0.760***	0.759***
Hispanic young male	0.238***		0.365***	0.352***
Constant	-0.018***			
Wx				
Residential population		0.184***		0.021
Dynamic population		-0.084***		-0.005
Education level			0.079*	0.025
Employment status			0.009	0.007
Income			0.015	0.028**
Residential ownership			-0.002	-0.006
Racial heterogeneity			-0.024***	-0.020**
Black young male			0.610***	0.561***
Hispanic young male			-0.149**	-0.178***

** p<.1; ** p<.05; *** p<.01*

Table 42: Assault crime at semiprivate places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.009*	0.012**		0.001
Dynamic population	0.193***	0.177***		0.172***
Education level	0.033**		0.001	0.021
Employment status	-0.005*		-0.002	-0.002
Income	-0.004		-0.006**	-0.008***
Residential ownership	0.007***		0.001	0.004**
Racial heterogeneity	0.023***		0.029***	0.020***
Black young male	0.035		-0.150***	-0.095***
Hispanic young male	-0.075***		0.028	0.030
Constant	-0.004***			
Wx				
Residential population		-0.006		-0.054***
Dynamic population		0.040***		0.053***
Education level			-0.091***	-0.026
Employment status			-0.038***	-0.020***
Income			0.060***	0.048***
Residential ownership			0.001	0.009***
Racial heterogeneity			0.016***	-0.007
Black young male			0.030	0.131***
Hispanic young male			-0.382***	-0.157***

** p<.1; ** p<.05; *** p<.01*

Table 43: Assault crime at semipublic places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.008***	-0.004*		-0.006**
Dynamic population	0.058***	0.052***		0.050***
Education level	-0.020***		-0.026***	-0.021***
Employment status	0.000		-0.000	-0.000
Income	0.003*		0.002	0.001
Residential ownership	-0.004***		-0.006***	-0.005***
Racial heterogeneity	0.010***		0.015***	0.012***
Black young male	-0.006		-0.103***	-0.088***
Hispanic young male	-0.038***		-0.042***	-0.038**
Constant	0.002**			
Wx				
Residential population		-0.006		-0.009*
Dynamic population		0.021***		0.026***
Education level			-0.081***	-0.061***
Employment status			-0.002	0.004
Income			0.024***	0.021***
Residential ownership			0.003*	0.005***
Racial heterogeneity			-0.004*	-0.013***
Black young male			0.046**	0.082***
Hispanic young male			-0.067***	0.005

* *p<.1; ** p<.05; *** p<.01*

Table 44: Assault crime at public places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.023***	0.014**		-0.011
Dynamic population	0.236***	0.210***		0.202***
Education level	-0.022		-0.078***	-0.055***
Employment status	-0.014***		-0.010**	-0.009**
Income	0.011***		0.001	-0.002
Residential ownership	0.003		-0.009***	-0.005**
Racial heterogeneity	0.034***		0.038***	0.028***
Black young male	0.206***		-0.361***	-0.297***
Hispanic young male	0.071***		0.166***	0.178***
Constant	0.003			
Wx				
Residential population		0.043***		-0.064***
Dynamic population		0.032***		0.083***
Education level			-0.133***	-0.051

Employment status			-0.047***	-0.024***
Income			0.109***	0.093***
Residential ownership			0.011**	0.022***
Racial heterogeneity			0.008	-0.021***
Black young male			0.460***	0.600***
Hispanic young male			-0.478***	-0.202***

** p<.1; ** p<.05; *** p<.01*

Table 45: Robbery crime at private places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.016***	0.023***		0.015***
Dynamic population	-0.002	-0.002		-0.001
Education level	0.019***		0.016**	0.016**
Employment status	-0.004***		-0.005***	-0.005***
Income	0.010***		0.009***	0.009***
Residential ownership	0.000		0.000	0.000
Racial heterogeneity	0.001		0.004**	0.003
Black young male	0.135***		0.071***	0.071***
Hispanic young male	0.052***		0.067***	0.062***
Constant	-0.002***			
Wx				
Residential population		0.034***		0.006
Dynamic population		-0.019***		-0.005
Education level			0.005	-0.003
Employment status			0.006**	0.005*
Income			0.001	0.003
Residential ownership			0.002	0.000
Racial heterogeneity			-0.005**	-0.004*
Black young male			0.097***	0.088***
Hispanic young male			-0.017	-0.019

** p<.1; ** p<.05; *** p<.01*

Table 46: Robbery crime at semiprivate places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.005**	0.001		-0.001
Dynamic population	0.049***	0.039***		0.037***
Education level	-0.013*		-0.019***	-0.014**
Employment status	-0.002		-0.002	-0.002
Income	0.000		-0.001	-0.001

Residential ownership	0.001		-0.000	0.000
Racial heterogeneity	0.005***		0.005***	0.003*
Black young male	0.004		-0.064***	-0.053***
Hispanic young male	-0.011		0.021	0.032**
Constant	0.001			
Wx				
Residential population		-0.010**		-0.021***
Dynamic population		0.024***		0.028***
Education level			-0.024	-0.005
Employment status			-0.007***	-0.001
Income			0.015***	0.011***
Residential ownership			-0.003*	0.001
Racial heterogeneity			0.007***	0.000
Black young male			0.040**	0.075***
Hispanic young male			-0.108***	-0.057***

** p<.1; ** p<.05; *** p<.01*

Table 47: Robbery crime at semipublic places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.003**	-0.001		-0.002
Dynamic population	0.010***	0.011***		0.011***
Education level	-0.004		-0.008**	-0.007*
Employment status	-0.001		-0.001	-0.001
Income	-0.000		-0.001	-0.001
Residential ownership	-0.001**		-0.002***	-0.001***
Racial heterogeneity	0.002***		0.003***	0.003***
Black young male	0.024***		-0.002	0.002
Hispanic young male	-0.006		-0.023***	-0.022***
Constant	0.001***			
Wx				
Residential population		-0.005**		-0.007***
Dynamic population		0.001		0.003
Education level			-0.016**	-0.010
Employment status			-0.002	-0.001
Income			0.010***	0.009***
Residential ownership			0.000	0.001
Racial heterogeneity			-0.003*	-0.004***
Black young male			0.012	0.019*
Hispanic young male			0.009	0.022**

** p<.1; ** p<.05; *** p<.01*

Table 48: Robbery crime at public places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.008**	0.004		-0.006
Dynamic population	0.032***	0.033***		0.034***
Education level	0.003		-0.010	-0.007
Employment status	-0.012***		-0.010***	-0.010***
Income	0.002		-0.002	-0.002
Residential ownership	-0.002		-0.004***	-0.003***
Racial heterogeneity	0.010***		0.005*	0.004
Black young male	0.214***		0.065***	0.076***
Hispanic young male	0.078***		0.062***	0.058**
Constant	0.005***			
Wx				
Residential population		0.035***		-0.012*
Dynamic population		-0.016***		0.001
Education level			-0.006	0.008
Employment status			-0.012***	-0.010**
Income			0.023***	0.020***
Residential ownership			0.003	0.003
Racial heterogeneity			0.007*	0.004
Black young male			0.160***	0.176***
Hispanic young male			-0.045	-0.003

** p<.1; ** p<.05; *** p<.01*

Table 49: Drug crime at private places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.063***	0.090***		0.068***
Dynamic population	-0.004	-0.013***		-0.009**
Education level	-0.001		0.028**	0.030**
Employment status	0.004*		0.000	-0.001
Income	0.041***		0.038***	0.039***
Residential ownership	0.002*		0.006***	0.004***
Racial heterogeneity	-0.006***		0.003	-0.000
Black young male	0.634***		0.440***	0.441***
Hispanic young male	0.056***		0.115***	0.105***
Constant	-0.013***			
Wx				
Residential population		0.037***		-0.012
Dynamic population		-0.056***		-0.021***
Education level			-0.268***	-0.293***

Employment status			0.032***	0.030***
Income			0.040***	0.045***
Residential ownership			-0.002	-0.005
Racial heterogeneity			-0.017***	-0.011**
Black young male			0.056	0.023
Hispanic young male			-0.102***	-0.132***

** p<.1; ** p<.05; *** p<.01*

Table 50: Drug crime at semiprivate places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	0.038***	0.040***		0.039***
Dynamic population	0.066***	0.077***		0.076***
Education level	0.001		-0.007	0.002
Employment status	0.007***		0.006***	0.006***
Income	-0.006***		-0.005**	-0.006***
Residential ownership	-0.005***		-0.005***	-0.005***
Racial heterogeneity	0.002		0.014***	0.009***
Black young male	0.036**		0.001	0.029
Hispanic young male	0.025*		0.055***	0.025
Constant	-0.005***			
Wx				
Residential population		-0.017***		-0.012*
Dynamic population		-0.029***		-0.027***
Education level			-0.040*	-0.036*
Employment status			-0.002	0.000
Income			0.012**	0.012**
Residential ownership			0.003	-0.000
Racial heterogeneity			-0.007**	-0.011***
Black young male			-0.022	-0.022
Hispanic young male			-0.101***	-0.014

** p<.1; ** p<.05; *** p<.01*

Table 51: Drug crime at semipublic places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.007***	-0.002		-0.003*
Dynamic population	0.028***	0.020***		0.019***
Education level	-0.009*		-0.012**	-0.009
Employment status	-0.002**		-0.002**	-0.002**
Income	-0.000		-0.001	-0.002

Residential ownership	-0.000		-0.001**	-0.001
Racial heterogeneity	0.007***		0.006***	0.005***
Black young male	0.027***		-0.018	-0.013
Hispanic young male	-0.034***		-0.051***	-0.038***
Constant	0.000			
Wx				
Residential population		-0.011***		-0.019***
Dynamic population		0.020***		0.023***
Education level			-0.078***	-0.063***
Employment status			-0.002	0.003
Income			0.023***	0.020***
Residential ownership			0.002	0.005***
Racial heterogeneity			0.001	-0.003
Black young male			-0.013	0.013
Hispanic young male			-0.029**	-0.002

** p<.1; ** p<.05; *** p<.01*

Table 52: Drug crime at public places: OLS and SDM

	OLS	SDM (A)	SDM (B)	SDM (C)
Residential population	-0.031***	0.002		-0.024***
Dynamic population	0.372***	0.390***		0.395***
Education level	-0.026		-0.062**	-0.025
Employment status	-0.031***		-0.031***	-0.031***
Income	0.016***		0.008	0.005
Residential ownership	-0.002		-0.016***	-0.007**
Racial heterogeneity	0.038***		0.044***	0.031***
Black young male	0.763***		0.247***	0.383***
Hispanic young male	0.079**		-0.026	-0.176***
Constant	-0.001			
Wx				
Residential population		0.081***		-0.020
Dynamic population		-0.174***		-0.134***
Education level			-0.311***	-0.222***
Employment status			0.009	0.020**
Income			0.075***	0.061***
Residential ownership			0.027***	0.017***
Racial heterogeneity			0.011	-0.015*
Black young male			0.113*	0.178***
Hispanic young male			-0.272***	0.222***

** p<.1; ** p<.05; *** p<.01*

A.3 CHAPTER 6: MOBILE POPULATION EMOTION PATTERNS

A.3.1 *Footnote tables*

Table 53: Hausman's specification test by crime

	Private		Semiprivate		Semipublic		Public	
	x^2	p	x^2	p	x^2	p	x^2	p
Property	800.38	0.000	-31617.76	n/a	165.64	0.000	980.43	0.000
Harassment	1180.76	0.000	449.71	0.000	49.04	0.000	248.22	0.000
Assault	32.59	0.000	415.42	0.000	69.15	0.000	343.11	0.000
Robbery	41.11	0.000	103.18	0.000	2068.28	0.000	110.70	0.000
Drug	0.04	1.000	53.26	0.000	15.01	0.036	127.76	0.000

A.3.2 *Results: FENB model analysis*

Table 54: Property crime at semiprivate places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.120*** (0.887)	-0.488*** (0.614)			-0.028* (0.972)
Fear	0.005 (1.005)	-11.381 (0.000)			-0.027 (0.974)
Joy	0.349*** (1.417)	0.089*** (1.093)			0.026*** (1.026)
Sadness	0.026 (1.026)	-0.608*** (0.544)			-0.018* (0.982)
Swearing	0.205*** (1.228)		0.078*** (1.081)		0.046*** (1.047)
Equitability	4.286***			1.661***	1.540***

** p<.1; ** p<.05; *** p<.01*

Table 55: Property crime at semipublic places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.234***	0.031			0.004
	(0.791)	(1.032)			(1.004)
Fear	0.027	-0.034			-0.055
	(1.028)	(0.967)			(0.946)
Joy	0.183***	0.007			0.005
	(1.201)	(1.007)			(1.005)
Sadness	0.028	0.011			-0.005
	(1.029)	(1.011)			(0.995)
Swearing	0.051		0.032***		0.027
	(1.052)		(1.033)		(1.027)
Equitability	3.652***			0.562	0.482

** p<.1; ** p<.05; *** p<.01*

Table 56: Property crime at public places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.151***	-0.030			-0.077***
	(0.860)	(0.970)			(0.926)
Fear	-0.105*	0.055			0.014
	(0.901)	(1.056)			(1.014)
Joy	0.138***	0.020***			0.017***
	(1.148)	(1.020)			(1.017)
Sadness	-0.023	0.027*			-0.001
	(0.977)	(1.028)			(0.999)
Swearing	0.038***		0.039***		0.030***
	(1.039)		(1.040)		(1.031)
Equitability	3.842***			0.708***	0.706***

** p<.1; ** p<.05; *** p<.01*

Table 57: Harassment crime at private places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.074	0.015			0.051
	(1.077)	(1.015)			(1.053)
Fear	-0.019	0.010			0.005
	(0.981)	(1.010)			(1.005)
Joy	0.041***	-0.006			0.001
	(1.042)	(0.994)			(1.001)
Sadness	0.055*	0.011			0.019
	(1.057)	(1.011)			(1.020)

Swearing	-0.012		-0.008		-0.035
	(0.988)		(0.992)		(0.965)
Equitability	1.999***			0.193	0.096

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 58: Harassment crime at semiprivate places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.140**	0.130***			0.047
	(0.869)	(1.138)			(1.048)
Fear	-0.191**	0.044			-0.028
	(0.826)	(1.045)			(0.972)
Joy	0.143***	0.011			0.006
	(1.153)	(1.011)			(1.006)
Sadness	0.033	0.059***			0.018
	(1.033)	(1.061)			(1.018)
Swearing	0.078***		0.059***		0.033**
	(1.081)		(1.061)		(1.034)
Equitability	4.417***			1.942***	1.603***

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 59: Harassment crime at semipublic places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.227	-0.032			-0.078
	(0.797)	(0.969)			(0.925)
Fear	0.208	0.299			0.248
	(1.231)	(1.348)			(1.281)
Joy	0.135***	-0.002			0.001
	(1.145)	(0.998)			(1.001)
Sadness	-0.034	-0.002			-0.035
	(0.966)	(0.998)			(0.965)
Swearing	0.067		0.010		0.013
	(1.069)		(1.010)		(1.013)
Equitability	4.326***			1.096	1.179

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 60: Harassment crime at public places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)

Anger	-0.036	0.095***			0.070*
	(0.965)	(1.100)			(1.072)
Fear	-0.141	0.006			-0.011
	(0.868)	(1.006)			(0.989)
Joy	0.109***	0.005			0.003
	(1.115)	(1.005)			(1.003)
Sadness	-0.045	-0.004			-0.017
	(0.956)	(0.996)			(0.983)
Swearing	0.053**		0.046***		0.023
	(1.055)		(1.047)		(1.023)
Equitability	3.047***			0.407	0.190

** p<.1; ** p<.05; *** p<.01*

Table 61: Assault crime at private places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.231***	0.071***			0.088**
	(1.259)	(1.074)			(1.092)
Fear	-0.097	-0.087			-0.087
	(0.908)	(0.916)			(0.916)
Joy	-0.003	-0.018			-0.015
	(0.997)	(0.982)			(0.985)
Sadness	0.073*	-0.033			-0.027
	(1.076)	(0.968)			(0.973)
Swearing	0.000		0.003		-0.015
	(1.000)		(1.003)		(0.985)
Equitability	0.711*			-0.069	0.009

** p<.1; ** p<.05; *** p<.01*

Table 62: Assault crime at semiprivate places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.005	0.180***			0.110***
	(1.005)	(1.197)			(1.116)
Fear	-0.481***	-0.250***			-0.328***
	(0.618)	(0.779)			(0.721)
Joy	0.184***	0.011*			0.007
	(1.201)	(1.011)			(1.007)
Sadness	-0.051	0.050**			0.01
	(0.950)	(1.052)			(1.010)

Swearing	0.135***		0.081***		0.040***
	(1.144)		(1.084)		(1.041)
Equitability	3.603***			2.013***	1.646***

** p<.1; ** p<.05; *** p<.01*

Table 63: Assault crime at semipublic places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.291*	-0.044			-0.057
	(0.748)	(0.957)			(0.945)
Fear	0.114	-0.278			-0.297
	(1.121)	(0.757)			(0.743)
Joy	0.174***	0.047*			0.048*
	(1.190)	(1.048)			(1.049)
Sadness	0.068	-0.006			-0.014
	(1.070)	(0.994)			(0.986)
Swearing	0.072		0.017		-0.002
	(1.075)		(1.017)		(0.998)
Equitability	3.148***			0.086	0.395

** p<.1; ** p<.05; *** p<.01*

Table 64: Assault crime at public places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.029	-0.017			-0.027
	(0.971)	(0.983)			(0.973)
Fear	-0.067	0.04			0.027
	(0.935)	(1.041)			(1.028)
Joy	0.111***	0.013			0.012
	(1.117)	(1.013)			(1.012)
Sadness	0.057*	0.059***			0.053***
	(1.058)	(1.060)			(1.055)
Swearing	0.031		0.032***		0.005
	(1.031)		(1.032)		(1.005)
Equitability	2.378***			0.615**	0.284

** p<.1; ** p<.05; *** p<.01*

Table 65: Robbery crime at private places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)

Anger	0.112	0.016			0.004
	(1.118)	(1.016)			(1.004)
Fear	-0.974**	-0.682*			-0.711*
	(0.377)	(0.506)			(0.491)
Joy	-0.028	-0.018			-0.022
	(0.972)	(0.982)			(0.979)
Sadness	0.152**	0.071			0.063
	(1.164)	(1.074)			(1.065)
Swearing	0.012		0.011		0.009
	(1.012)		(1.011)		(1.009)
Equitability	0.986			0.278	0.476

** p<.1; ** p<.05; *** p<.01*

Table 66: Robbery crime at semiprivate places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.026	0.235***			0.061
	(0.974)	(1.265)			(1.063)
Fear	-0.034	0.222			0.104
	(0.967)	(1.249)			(1.110)
Joy	0.085***	0.018			-0.007
	(1.089)	(1.018)			(0.993)
Sadness	-0.151**	0.016			-0.088
	(0.860)	(1.016)			(0.916)
Swearing	0.080**		0.154***		0.135***
	(1.083)		(1.167)		(1.145)
Equitability	4.808***			1.870***	1.522**

** p<.1; ** p<.05; *** p<.01*

Table 67: Robbery crime at semipublic places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.109	0.056			-0.183
	(0.897)	(1.058)			(0.833)
Fear	0.554*	0.625**			0.352
	(1.739)	(1.868)			(1.421)
Joy	-0.017	-0.061			-0.097
	(0.983)	(0.941)			(0.908)
Sadness	0.025	0.143			0.014
	(1.025)	(1.154)			(1.015)

Swearing	0.102		0.132*		0.182*
	(1.108)		(1.141)		(1.200)
Equitability	2.979***			2.664**	2.246

** p<.1; ** p<.05; *** p<.01*

Table 68: Robbery crime at public places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.076	0.027			0.105
	(1.079)	(1.027)			(1.111)
Fear	-0.048	-0.01			0.03
	(0.953)	(0.990)			(1.030)
Joy	0.024**	-0.007			0.007
	(1.024)	(0.993)			(1.007)
Sadness	0.056	0.016			0.047
	(1.058)	(1.016)			(1.048)
Swearing	-0.019		-0.013		-0.070*
	(0.981)		(0.988)		(0.933)
Equitability	1.851***			-0.098	-0.358

** p<.1; ** p<.05; *** p<.01*

Table 69: Drug crime at private places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.371***	0.063			0.033
	(1.449)	(1.066)			(1.034)
Fear	0.209	0.207			0.238
	(1.233)	(1.230)			(1.269)
Joy	-0.109***	-0.032			-0.039
	(0.897)	(0.968)			(0.962)
Sadness	0.129	-0.042			-0.055
	(1.138)	(0.959)			(0.946)
Swearing	0.052		-0.006		0.033
	(1.054)		(0.994)		(1.034)
Equitability	-2.483***			-0.608	-0.543

** p<.1; ** p<.05; *** p<.01*

Table 70: Drug crime at semiprivate places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)

Anger	0.188	0.164***			0.072
	(1.207)	(1.178)			(1.074)
Fear	-0.042	0.109			0.133
	(0.959)	(1.115)			(1.142)
Joy	0.117***	-0.003			-0.01
	(1.124)	(0.997)			(0.990)
Sadness	0.045	0.107			0.297***
	(1.046)	(1.113)			(1.346)
Swearing	0.128		0.112***		0.081**
	(1.136)		(1.119)		(1.085)
Equitability	2.652***			1.327	-0.584

** p<.1; ** p<.05; *** p<.01*

Table 71: Drug crime at semipublic places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	-0.129	0.015			-0.05
	(0.879)	(1.015)			(0.951)
Fear	-0.45	-0.391			-0.495*
	(0.638)	(0.676)			(0.610)
Joy	0.244***	0.008			0.015
	(1.276)	(1.008)			(1.015)
Sadness	-0.051	0.082			0.036
	(0.950)	(1.086)			(1.036)
Swearing	-0.052		0.008		0.003
	(0.950)		(1.008)		(1.003)
Equitability	5.500***			1.253	1.655*

** p<.1; ** p<.05; *** p<.01*

Table 72: Drug crime at public places: OLS and FENB

	OLS	FENB (A)	FENB (B)	FENB (C)	FENB (D)
Anger	0.057	0.032			0.009
	(1.058)	(1.033)			(1.009)
Fear	-0.022	-0.041			-0.062
	(0.978)	(0.960)			(0.940)
Joy	0.165***	0.026***			0.023***
	(1.179)	(1.026)			(1.024)
Sadness	0.023	0.015			0.005
	(1.024)	(1.015)			(1.005)

Swearing	0.033 (1.033)		0.050*** (1.052)		0.019 1.019)
Equitability	1.693***			0.573**	0.356

** p<.1; ** p<.05; *** p<.01*

A.4 CHAPTER 7: MOBILE POPULATION ATTRIBUTE PATTERNS

A.4.1 *Results: chi-squared analysis*

Table 73: Property at semiprivate places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	73,817 (20.88)***	45,923 (27.65)	6,000 (18.8)***	67,817 (21.09)
Informal broadcasting	143,029 (40.46)***	77,579 (46.71)	12,010 (37.63)***	131,019 (40.75)
Location Status	44,902 (12.70)***	12,667 (7.63)	4,655 (14.59)***	40,247 (12.52)
Informal with location status	80,794 (22.86)***	25,443 (15.32)	8,529 (26.73)***	72,265 (22.47)
Informative broadcasting	10,924 (3.09)***	4,483 (2.70)	718 (2.25)***	10,206 (3.17)

** p<.1; ** p<.05; *** p<.01*

Table 74: Property at semipublic places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	19,095 (21.09)***	100,645 (23.46)	414 (18.46)***	18,681 (21.15)
Informal broadcasting	37,409 (41.31)***	183,199 (42.70)	963 (42.93)	36,446 (41.27)

Location Status	10,871 (12.00)***	46,698 (10.89)	336 (14.98)***	10,535 (11.93)
Informal with location status	18,603 (20.54)	87,634 (20.43)	470 (20.95)	18,133 (20.53)
Informative broadcasting	4,582 (5.06)***	10,825 (2.52)	60 (2.67)***	4,522 (5.12)

** p<.1; ** p<.05; *** p<.01*

Table 75: Property at public places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	73,373 (21.68)***	46,367 (25.60)	1,579 (19.54)***	71,794 (21.73)
Informal broadcasting	143,415 (42.38)*	77,193 (42.62)	3,197 (39.57)***	140,218 (42.44)
Location Status	39,399 (11.64)***	18,170 (10.03)	1,047 (12.96)***	38,352 (11.61)
Informal with location status	70,512 (20.83)***	35,725 (19.72)	2,042 (25.28)***	68,470 (20.73)
Informative broadcasting	11,736 (3.47)***	3,671 (2.03)	214 (2.65)***	11,522 (3.49)

** p<.1; ** p<.05; *** p<.01*

Table 76: Harassment at private places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	35,752 (25.31)***	83,988 (22.20)	575 (25.43)	35,177 (25.31)
Informal broadcasting	67,409 (47.72)***	153,199 (40.50)	1,122 (49.62)*	66,287 (47.69)
Location Status	12,247 (8.67)***	45,322 (11.98)	173 (7.65)*	12,074 (8.69)
Informal with location status	23,273 (16.48)***	82,964 (21.93)	337 (14.90)**	22,936 (16.50)
Informative broadcasting	2,577 (1.82)***	12,830 (3.39)	54 (2.39)**	2,523 (1.82)

** p<.1; ** p<.05; *** p<.01*

Table 77: Harassment at semiprivate places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	36,492 (20.43)***	83,248 (24.42)	655 (21.90)**	35,837 (20.41)
Informal broadcasting	70,568 (39.51)***	150,040 (44.00)	1,186 (39.65)	69,382 (39.51)
Location Status	22,589 (12.65)***	34,980 (10.26)	373 (12.47)	22,216 (12.65)
Informal with location status	42,162 (23.61)***	64,075 (18.79)	681 (22.77)	41,481 (23.62)
Informative broadcasting	6,783 (3.80)***	8,624 (2.53)	96 (3.21)*	6,687 (3.81)

** p<.1; ** p<.05; *** p<.01*

Table 78: Harassment at semipublic places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	6,588 (17.21)***	113,152 (23.51)	83 (17.85)	6,505 (17.20)
Informal broadcasting	14,670 (38.33)***	205,938 (42.79)	167 (35.91)	14,503 (38.36)
Location Status	5,441 (14.21)***	52,128 (10.83)	88 (18.92)***	5,353 (14.16)
Informal with location status	8,536 (22.30)***	97,701 (20.30)	107 (23.01)	8,429 (22.29)
Informative broadcasting	3,042 (7.95)***	12,365 (2.57)	20 (4.30)***	3,022 (7.99)

** p<.1; ** p<.05; *** p<.01*

Table 79: Harassment at public places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	33,922 (21.20)***	85,818 (23.87)	454 (22.00)	33,468 (21.19)

Informal broadcasting	67,548 (42.22)**	153,060 (42.57)	824 (39.92)**	66,724 (42.25)
Location Status	18,977 (11.86)***	38,592 (10.73)	265 (12.84)	18,712 (11.85)
Informal with location status	35,988 (22.49)***	70,249 (19.54)	481 (23.30)	35,507 (22.48)
Informative broadcasting	3,561 (2.23)***	11,846 (3.29)	40 (1.94)	3,521 (2.23)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 80: Assault at private places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	26,700 (24.56)***	93,040 (22.65)	510 (27.73)***	26,190 (24.50)
Informal broadcasting	51,632 (47.49)***	168,976 (41.13)	909 (49.43)*	50,723 (47.46)
Location Status	8,923 (8.21)***	48,646 (11.84)	134 (7.29)	8,789 (8.22)
Informal with location status	16,658 (15.32)***	89,579 (21.80)	234 (12.72)***	16,424 (15.37)
Informative broadcasting	4,803 (4.42)***	10,604 (2.58)	52 (2.83)***	4,751 (4.45)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 81: Assault at semiprivate places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	29,863 (19.65)***	89,877 (24.45)	619 (16.46)***	29,244 (19.73)
Informal broadcasting	58,387 (38.41)***	162,221 (44.13)	1,488 (39.56)	56,899 (38.38)
Location Status	20,625 (13.57)***	36,944 (10.05)	477 (12.68)	20,148 (13.59)
Informal with location status	39,330 (25.88)***	66,907 (18.20)	1,134 (30.15)***	38,196 (25.77)
Informative broadcasting	3,791 (2.49)***	11,616 (3.16)	43 (1.14)***	3,748 (2.53)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 82: Assault at semipublic places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	5,727 (19.49)***	114,013 (23.26)	76 (15.57)**	5,651 (19.55)
Informal broadcasting	11,974 (40.74)***	208,634 (42.56)	199 (40.78)	11,775 (40.74)
Location Status	4,606 (15.67)***	52,963 (10.8)	88 (18.03)	4,518 (15.63)
Informal with location status	6,670 (22.70)***	99,567 (20.31)	122 (25.00)	6,548 (22.66)
Informative broadcasting	412 (1.40)***	14,995 (3.06)	3 (0.61)	409 (1.42)

** p<.1; ** p<.05; *** p<.01*

Table 83: Assault at public places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	48,274 (22.13)***	71,466 (23.71)	780 (20.79)**	47,494 (22.16)
Informal broadcasting	90,005 (41.27)***	130,603 (43.32)	1,566 (41.75)	88,439 (41.26)
Location Status	25,916 (11.88)***	31,653 (10.50)	431 (11.49)	25,485 (11.89)
Informal with location status	47,543 (21.80)***	58,694 (19.47)	907 (24.18)***	46,636 (21.76)
Informative broadcasting	6,352 (2.91)*	9,055 (3.00)	67 (1.79)***	6,285 (2.93)

** p<.1; ** p<.05; *** p<.01*

Table 84: Robbery at private places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	7,750 (30.04)***	111,990 (22.68)	42 (20.59)***	7,708 (30.11)
Informal broadcasting	12,144 (47.07)***	208,464 (42.22)	96 (47.06)	12,048 (47.07)

Location Status	1,789 (6.93)***	55,780 (11.30)	14 (6.86)	1,775 (6.93)
Informal with location status	3,226 (12.50)***	103,011 (20.86)	23 (11.27)	3,203 (12.51)
Informative broadcasting	892 (3.46)***	14,515 (2.94)	29 (14.22)***	863 (3.37)

* *p<.1; ** p<.05; *** p<.01*

Table 85: Robbery at semiprivate places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	12,389 (21.83)***	107,351 (23.20)	121 (19.64)	12,268 (21.85)
Informal broadcasting	22,777 (40.13)***	197,831 (42.75)	282 (45.78)***	22,495 (40.07)
Location Status	7,882 (13.89)***	49,687 (10.74)	73 (11.85)	7,809 (13.91)
Informal with location status	13,000 (22.91)***	93,237 (20.15)	116 (18.83)**	12,884 (22.95)
Informative broadcasting	703 (1.24)***	14,704 (3.18)	24 (3.90)***	679 (1.21)

* *p<.1; ** p<.05; *** p<.01*

Table 86: Robbery at semipublic places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	2,831 (21.57)***	116,909 (23.08)	27 (19.85)	2,804 (21.59)
Informal broadcasting	5,190 (39.55)***	215,418 (42.54)	66 (48.53)**	5,124 (39.46)
Location Status	1,749 (13.33)***	55,820 (11.02)	13 (9.56)	1,736 (13.37)
Informal with location status	2,916 (22.22)***	103,321 (20.40)	24 (17.65)	2,892 (22.27)
Informative broadcasting	436 (3.32)**	14,971 (2.96)	6 (4.41)	430 (3.31)

* *p<.1; ** p<.05; *** p<.01*

Table 87: Robbery at public places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	25,896 (26.12)***	93,844 (22.32)	230 (28.19)	25,666 (26.10)
Informal broadcasting	43,161 (43.53)***	177,447 (42.21)	364 (44.61)	42,797 (43.52)
Location Status	9,861 (9.94)***	47,708 (11.35)	71 (8.70)	9,790 (9.95)
Informal with location status	18,333 (18.49)***	87,904 (20.91)	142 (17.40)	18,191 (18.50)
Informative broadcasting	1,908 (1.92)*	13,499 (3.21)	9 (1.10)*	1,899 (1.93)

** p<.1; ** p<.05; *** p<.01*

Table 88: Drug at private places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	6,234 (30.26)***	113,506 (22.75)	94 (23.10)***	6,140 (30.40)
Informal broadcasting	10,879 (52.81)***	209,729 (42.03)	208 (51.11)	10,671 (52.84)
Location Status	920 (4.47)***	56,649 (11.35)	28 (6.88)**	892 (4.42)
Informal with location status	1,979 (9.61)***	104,258 (20.90)	46 (11.30)	1,933 (9.57)
Informative broadcasting	590 (2.86)	14,817 (2.97)	31 (7.62)***	559 (2.77)

** p<.1; ** p<.05; *** p<.01*

Table 89: Drug at semiprivate places: frequency of online activity by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Conversation	6,175 (19.64)***	113,565 (23.27)	200 (17.68)*	113,565 (19.71)
Informal broadcasting	12,148 (38.63)***	208,460 (42.71)	475 (42.00)**	208,460 (38.51)

Location Status	4,061 (12.91)***	53,508 (10.96)	103 (9.11)***	53,508 (13.06)
Informal with location status	8,539 (27.16)***	97,698 (20.02)	348 (30.77)***	97,698 (27.02)
Informative broadcasting	522 (1.66)***	14,885 (3.05)	5 (0.44)***	14,885 (1.71)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 90: Drug at semipublic places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	3,305 (19.39)***	116,435 (23.17)	58 (18.24)	3,247 (19.41)
Informal broadcasting	6,964 (40.86)***	213,644 (42.51)	137 (43.08)	6,827 (40.81)
Location Status	2,934 (17.21)***	54,635 (10.87)	44 (13.84)	2,890 (17.28)
Informal with location status	3,598 (21.11)**	102,639 (20.43)	77 (24.21)	3,521 (21.05)
Informative broadcasting	244 (1.43)***	15,163 (3.02)	2 (0.63)	242 (1.45)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 91: Drug at public places: frequency of online activity by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Conversation	29,537 (20.90)***	90,203 (23.85)	687 (17.99)***	28,850 (20.98)
Informal broadcasting	59,946 (42.42)	160,662 (42.47)	1,587 (41.57)	58,359 (42.45)
Location Status	17,342 (12.27)***	40,227 (10.63)	471 (12.34)	16,871 (12.27)
Informal with location status	31,633 (22.39)***	74,604 (19.72)	1,032 (27.03)***	30,601 (22.26)
Informative broadcasting	2,848 (2.02)***	12,559 (3.32)	41 (1.07)***	2,807 (2.04)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 92: Property at semiprivate places: frequency of topic by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Animals	20,961 (5.93)***	10,158 (6.12)	1,836 (5.75)	19,125 (5.95)
Body and appearance	37,319 (10.56)	17,713 (10.66)	3,203 (10.04)***	34,116 (10.61)
Business	36,285 (10.27)***	16,274 (9.80)	3,264 (10.23)	33,021 (10.27)
Clothes and fashion	52,760 (14.93)**	25,164 (15.15)	4,721 (14.79)	48,039 (14.94)
Crime and law	38,732 (10.96)***	17,418 (10.49)	3,452 (10.82)	35,280 (10.97)
Culture	77,916 (22.04)***	34,274 (20.64)	6,786 (21.26)***	71,130 (22.12)
Education	42,222 (11.95)***	18,274 (11.00)	3,803 (11.92)	38,419 (11.95)
Family and life stages	55,435 (15.68)***	27,443 (16.52)	4,900 (15.35)*	50,535 (15.72)
Food and drink	55,190 (15.61)***	21,348 (12.85)	4,843 (15.18)**	50,347 (15.66)
Health	69,972 (19.80)***	35,972 (21.66)	5,796 (18.16)***	64,176 (19.96)
Houses and buildings	68,400 (19.35)***	27,224 (16.39)	6,519 (20.43)***	61,881 (19.24)
Language	86,054 (24.35)***	28,715 (17.29)	8,693 (27.24)***	77,361 (24.06)
Leisure	96,916 (27.42)***	48,070 (28.94)	8,621 (27.01)*	88,295 (27.46)
Nature	53,244 (15.06)***	26,537 (15.98)	4,467 (14.00)***	48,777 (15.17)
Personality and emotions	141,205 (39.95)***	68,224 (41.08)	12,658 (39.67)	128,547 (39.98)
Religion and politics	15,997 (4.53)***	8,440 (5.08)	1,375 (4.31)*	14,622 (4.55)
Retail	28,418 (8.04)***	10,987 (6.61)	2,811 (8.81)***	25,607 (7.96)
Science	29,993 (8.49)***	12,919 (7.78)	2,906 (9.11)***	27,087 (8.42)
Social issues	15,624 (4.42)	7,381 (4.44)	1,227 (3.84)***	14,397 (4.48)

Technology	38,960 (11.02)***	19,449 (11.71)	3,200 (10.03)***	35,760 (11.12)
The media	30,260 (8.56)***	12,225 (7.36)	2,328 (7.30)***	27,932 (8.69)
Travel and tourism	70,380 (19.91)	33,172 (19.97)	5,888 (18.45)***	64,492 (20.06)
War and conflict	11,789 (3.34)***	4,570 (2.75)	992 (3.11)**	10,797 (3.36)
Work	34,894 (9.87)***	14,545 (8.76)	3,078 (9.65)	31,816 (9.89)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 93: Property at semipublic places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	5,381 (5.94)	25,738 (6.00)	116 (5.17)	5,265 (5.96)
Body and appearance	9,281 (10.25)***	45,751 (10.66)	226 (10.08)	9,055 (10.25)
Business	10,534 (11.63)***	42,025 (9.80)	222 (9.90)***	10,312 (11.68)
Clothes and fashion	13,286 (14.67)***	64,638 (15.07)	330 (14.71)	12,956 (14.67)
Crime and law	10,303 (11.38)***	45,847 (10.69)	233 (10.39)	10,070 (11.40)
Culture	20,262 (22.37)***	91,928 (21.43)	464 (20.69)*	19,798 (22.42)
Education	9,887 (10.92)***	50,609 (11.80)	225 (10.03)	9,662 (10.94)
Family and life stages	14,291 (15.78)	68,587 (15.99)	326 (14.53)	13,965 (15.81)
Food and drink	13,736 (15.17)***	62,802 (14.64)	307 (13.69)**	13,429 (15.21)
Health	17,147 (18.93)***	88,797 (20.70)	403 (17.97)	16,744 (18.96)
Houses and buildings	19,388 (21.41)***	76,236 (17.77)	606 (27.02)***	18,782 (21.27)
Language	21,586 (23.84)***	93,183 (21.72)	606 (27.02)***	20,980 (23.76)
Leisure	26,477 (29.24)***	118,509 (27.62)	690 (30.76)	25,787 (29.20)

Nature	14,736 (16.27)***	65,045 (15.16)	485 (21.62)***	14,251 (16.14)
Personality and emotions	36,570 (40.38)	172,859 (40.29)	916 (40.84)	35,654 (40.37)
Religion and politics	3,979 (4.39)***	20,458 (4.77)	83 (3.70)	3,896 (4.41)
Retail	8,556 (9.45)***	30,849 (7.19)	201 (8.96)	8,355 (9.46)
Science	8,711 (9.62)***	34,201 (7.97)	222 (9.90)	8,489 (9.61)
Social issues	3,434 (3.79)***	19,571 (4.56)	73 (3.25)	3,361 (3.81)
Technology	10,799 (11.92)***	47,610 (11.10)	299 (13.33)**	10,500 (11.89)
The media	8,830 (9.75)***	33,655 (7.84)	340 (15.16)***	8,490 (9.61)
Travel and tourism	19,442 (21.47)***	84,110 (19.61)	621 (27.69)***	18,821 (21.31)
War and conflict	5,031 (5.56)***	11,328 (2.64)	237 (10.57)***	4,794 (5.43)
Work	10,560 (11.66)***	38,879 (9.06)	220 (9.81)***	10,340 (11.71)

** p<.1; ** p<.05; *** p<.01*

Table 94: Property at public places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	20,375 (6.02)	10,744 (5.93)	498 (6.16)	19,877 (6.02)
Body and appearance	36,199 (10.7)***	18,833 (10.40)	856 (10.60)	35,343 (10.70)
Business	35,479 (10.48)***	17,080 (9.43)	830 (10.27)	34,649 (10.49)
Clothes and fashion	50,908 (15.04)	27,016 (14.92)	1,219 (15.09)	49,689 (15.04)
Crime and law	37,469 (11.07)***	18,681 (10.31)	877 (10.86)	36,592 (11.08)
Culture	74,745 (22.09)***	37,445 (20.67)	1,788 (22.13)	72,957 (22.08)
Education	40,248 (11.89)***	20,248 (11.18)	969 (11.99)	39,279 (11.89)

Family and life stages	53,958 (15.94)	28,920 (15.97)	1,337 (16.55)	52,621 (15.93)
Food and drink	50,790 (15.01)***	25,748 (14.22)	1,197 (14.82)	49,593 (15.01)
Health	68,095 (20.12)***	37,849 (20.90)	1,433 (17.74)***	66,662 (20.18)
Houses and buildings	64,416 (19.03)***	31,208 (17.23)	1,658 (20.52)***	62,758 (19.00)
Language	77,543 (22.91)***	37,226 (20.55)	2,056 (25.45)***	75,487 (22.85)
Leisure	94,315 (27.87)	50,671 (27.98)	2,263 (28.01)	92,052 (27.86)
Nature	51,419 (15.19)***	28,362 (15.66)	1,223 (15.14)	50,196 (15.19)
Personality and emotions	136,284 (40.27)	73,145 (40.38)	3,236 (40.05)	133,048 (40.27)
Religion and politics	15,541 (4.59)***	8,896 (4.91)	363 (4.49)	15,178 (4.59)
Retail	27,012 (7.98)***	12,393 (6.84)	644 (7.97)	26,368 (7.98)
Science	29,630 (8.76)***	13,282 (7.33)	776 (9.61)***	28,854 (8.73)
Social issues	15,156 (4.48)**	7,849 (4.33)	325 (4.02)**	14,831 (4.49)
Technology	38,032 (11.24)	20,377 (11.25)	863 (10.68)	37,169 (11.25)
The media	28,967 (8.56)***	13,518 (7.46)	669 (8.28)	28,298 (8.57)
Travel and tourism	67,347 (19.90)	36,205 (19.99)	1,616 (20.00)	65,731 (19.90)
War and conflict	11,656 (3.44)***	4,703 (2.60)	266 (3.29)	11,390 (3.45)
Work	34,181 (10.10)***	15,258 (8.42)	812 (10.05)	33,369 (10.10)

* *p<.1; ** p<.05; *** p<.01*

Table 95: Harassment at private places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	8,709 (6.17)***	22,410 (5.92)	121 (5.35)	8,588 (6.18)

Body and appearance	15,895 (11.25)***	39,137 (10.35)	257 (11.37)	15,638 (11.25)
Business	12,442 (8.81)***	40,117 (10.60)	190 (8.40)	12,252 (8.81)
Clothes and fashion	22,099 (15.64)***	55,825 (14.76)	365 (16.14)	21,734 (15.64)
Crime and law	15,176 (10.74)	40,974 (10.83)	250 (11.06)	14,926 (10.74)
Culture	30,964 (21.92)***	81,226 (21.47)	490 (21.67)	30,474 (21.92)
Education	15,831 (11.21)***	44,665 (11.81)	264 (11.68)	15,567 (11.20)
Family and life stages	23,607 (16.71)***	59,271 (15.67)	385 (17.03)	23,222 (16.71)
Food and drink	20,806 (14.73)	55,732 (14.73)	352 (15.57)	20,454 (14.72)
Health	30,121 (21.32)***	75,823 (20.04)	524 (23.18)**	29,597 (21.29)
Houses and buildings	21,781 (15.42)***	73,843 (19.52)	344 (15.21)	21,437 (15.42)
Language	27,182 (19.24)***	87,587 (23.15)	396 (17.51)**	26,786 (19.27)
Leisure	38,502 (27.26)***	106,484 (28.15)	640 (28.31)	37,862 (27.24)
Nature	19,727 (13.97)***	60,054 (15.87)	330 (14.60)	19,397 (13.95)
Personality and emotions	58,722 (41.57)***	150,707 (39.84)	969 (42.86)	57,753 (41.55)
Religion and politics	6,491 (4.60)**	17,946 (4.74)	111 (4.91)	6,380 (4.59)
Retail	9,450 (6.69)***	29,955 (7.92)	146 (6.46)	9,304 (6.69)
Science	9,523 (6.74)***	33,389 (8.83)	139 (6.15)	9,384 (6.75)
Social issues	6,590 (4.67)***	16,415 (4.34)	96 (4.25)	6,494 (4.67)
Technology	15,739 (11.14)	42,670 (11.28)	250 (11.06)	15,489 (11.14)
The media	10,915 (7.73)***	31,570 (8.35)	148 (6.55)**	10,767 (7.75)
Travel and tourism	26,090 (18.47)***	77,462 (20.48)	469 (20.74)***	25,621 (18.43)

War and conflict	3,585 (2.54)***	12,774 (3.38)	63 (2.79)	3,522 (2.53)
Work	11,988 (8.49)***	37,451 (9.90)	197 (8.71)	11,791 (8.48)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 96: Harassment at semiprivate places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	10,366 (5.80)***	20,753 (6.09)	182 (6.08)	10,184 (5.80)
Body and appearance	18,513 (10.37)***	36,519 (10.71)	314 (10.50)	18,199 (10.36)
Business	18,799 (10.53)***	33,760 (9.90)	313 (10.46)	18,486 (10.53)
Clothes and fashion	26,350 (14.75)***	51,574 (15.13)	461 (15.41)	25,889 (14.74)
Crime and law	20,984 (11.75)***	35,166 (10.31)	335 (11.20)	20,649 (11.76)
Culture	40,344 (22.59)***	71,846 (21.07)	710 (23.74)	39,634 (22.57)
Education	20,981 (11.75)*	39,515 (11.59)	341 (11.40)	20,640 (11.75)
Family and life stages	27,464 (15.38)***	55,414 (16.25)	440 (14.71)	27,024 (15.39)
Food and drink	27,316 (15.30)***	49,222 (14.44)	469 (15.68)	26,847 (15.29)
Health	35,053 (19.63)***	70,891 (20.79)	569 (19.02)	34,484 (19.64)
Houses and buildings	36,367 (20.36)***	59,257 (17.38)	537 (17.95)***	35,830 (20.40)
Language	44,363 (24.84)***	70,406 (20.65)	763 (25.51)	43,600 (24.83)
Leisure	50,297 (28.16)***	94,689 (27.77)	821 (27.45)	49,476 (28.17)
Nature	26,137 (14.63)***	53,644 (15.73)	433 (14.48)	25,704 (14.64)
Personality and emotions	71,952 (40.29)	137,477 (40.32)	1,144 (38.25)**	70,808 (40.32)
Religion and politics	7,961 (4.46)***	16,476 (4.83)	128 (4.28)	7,833 (4.46)

Retail	14,766 (8.27)***	24,639 (7.23)	252 (8.43)	14,514 (8.27)
Science	17,226 (9.65)***	25,686 (7.53)	251 (8.39)**	16,975 (9.67)
Social issues	7,486 (4.19)***	15,519 (4.55)	120 (4.01)	7,366 (4.19)
Technology	19,228 (10.77)***	39,181 (11.49)	316 (10.57)	18,912 (10.77)
The media	14,236 (7.97)***	28,249 (8.28)	254 (8.49)	13,982 (7.96)
Travel and tourism	34,953 (19.57)***	68,599 (20.12)	571 (19.09)	34,382 (19.58)
War and conflict	5,711 (3.20)	10,648 (3.12)	101 (3.38)	5,610 (3.19)
Work	18,831 (10.54)***	30,608 (8.98)	331 (11.07)	18,500 (10.54)

* *p<.1; ** p<.05; *** p<.01*

Table 97: Harassment at semipublic places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	1,973 (5.15)***	29,146 (6.06)	19 (4.09)	1,954 (5.17)
Body and appearance	3,544 (9.26)***	51,488 (10.70)	47 (10.11)	3,497 (9.25)
Business	4,947 (12.92)***	47,612 (9.89)	40 (8.60)***	4,907 (12.98)
Clothes and fashion	5,064 (13.23)***	72,860 (15.14)	62 (13.33)	5,002 (13.23)
Crime and law	3,421 (8.94)***	52,729 (10.96)	39 (8.39)	3,382 (8.94)
Culture	7,629 (19.93)***	104,561 (21.73)	83 (17.85)	7,546 (19.96)
Education	4,252 (11.11)***	56,244 (11.69)	59 (12.69)	4,193 (11.09)
Family and life stages	5,303 (13.85)***	77,575 (16.12)	64 (13.76)	5,239 (13.86)
Food and drink	5,408 (14.13)***	71,130 (14.78)	56 (12.04)	5,352 (14.15)
Health	6,992 (18.27)***	98,952 (20.56)	76 (16.34)	6,916 (18.29)

Houses and buildings	10,505 (27.44)***	85,119 (17.69)	130 (27.96)	10,375 (27.44)
Language	9,138 (23.87)***	105,631 (21.95)	134 (28.82)**	9,004 (23.81)
Leisure	10,410 (27.20)***	134,576 (27.96)	98 (21.08)***	10,312 (27.27)
Nature	6,373 (16.65)***	73,408 (15.25)	59 (12.69)**	6,314 (16.70)
Personality and emotions	15,849 (41.41)***	193,580 (40.22)	184 (39.57)	15,665 (41.43)
Religion and politics	1,795 (4.69)	22,642 (4.70)	18 (3.87)	1,777 (4.70)
Retail	4,725 (12.34)***	34,680 (7.21)	33 (7.10)***	4,692 (12.41)
Science	5,481 (14.32)***	37,431 (7.78)	60 (12.90)	5,421 (14.34)
Social issues	1,637 (4.28)	21,368 (4.44)	16 (3.44)	1,621 (4.29)
Technology	4,418 (11.54)*	53,991 (11.22)	49 (10.54)	4,369 (11.55)
The media	4,820 (12.59)***	37,665 (7.83)	53 (11.40)	4,767 (12.61)
Travel and tourism	8,973 (23.44)***	94,579 (19.65)	91 (19.57)**	8,882 (23.49)
War and conflict	2,811 (7.34)***	13,548 (2.81)	27 (5.81)	2,784 (7.36)
Work	4,641 (12.12)***	44,798 (9.31)	40 (8.60)**	4,601 (12.17)

** p<.1; ** p<.05; *** p<.01*

Table 98: Harassment at public places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	9,531 (5.96)	21,588 (6.00)	119 (5.77)	9,412 (5.96)
Body and appearance	16,933 (10.58)	38,099 (10.60)	205 (9.93)	16,728 (10.59)
Business	15,282 (9.55)*	37,277 (10.37)	198 (9.59)	15,084 (9.55)
Clothes and fashion	24,206 (15.13)*	53,718 (14.94)	301 (14.58)	23,905 (15.14)

Crime and law	16,483 (10.30)***	39,667 (11.03)	202 (9.79)	16,281 (10.31)
Culture	34,370 (21.48)	77,820 (21.64)	415 (20.11)	33,955 (21.50)
Education	18,507 (11.57)	41,989 (11.68)	225 (10.90)	18,282 (11.58)
Family and life stages	25,842 (16.15)***	57,036 (15.86)	331 (16.04)	25,511 (16.15)
Food and drink	23,596 (14.75)	52,942 (14.72)	327 (15.84)	23,269 (14.73)
Health	32,977 (20.61)***	72,967 (20.29)	438 (21.22)	32,539 (20.60)
Houses and buildings	31,911 (19.94)***	63,713 (17.72)	430 (20.83)	31,481 (19.93)
Language	36,926 (23.08)***	77,843 (21.65)	472 (22.87)	36,454 (23.08)
Leisure	45,920 (28.70)***	99,066 (27.55)	593 (28.73)	45,327 (28.70)
Nature	23,157 (14.47)***	56,624 (15.75)	311 (15.07)	22,846 (14.47)
Personality and emotions	65,911 (41.20)***	143,518 (39.91)	879 (42.59)	65,032 (41.18)
Religion and politics	7,275 (4.55)***	17,162 (4.77)	85 (4.12)	7,190 (4.55)
Retail	13,319 (8.32)***	26,086 (7.25)	181 (8.77)	13,138 (8.32)
Science	14,572 (9.11)***	28,340 (7.88)	221 (10.71)**	14,351 (9.09)
Social issues	6,828 (4.27)***	16,177 (4.50)	85 (4.12)	6,743 (4.27)
Technology	18,584 (11.62)***	39,825 (11.08)	269 (13.03)**	18,315 (11.60)
The media	12,390 (7.74)***	30,095 (8.37)	158 (7.66)	12,232 (7.75)
Travel and tourism	31,692 (19.81)	71,860 (19.99)	388 (18.80)	31,304 (19.82)
War and conflict	4,163 (2.60)***	12,196 (3.39)	51 (2.47)	4,112 (2.60)
Work	14,336 (8.96)***	35,103 (9.76)	151 (7.32)***	14,185 (8.98)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 99: Assault at private places: frequency of topic by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Animals	6,683 (6.15)**	24,436 (5.95)	113 (6.14)	6,570 (6.15)
Body and appearance	12,006 (11.04)***	43,026 (10.47)	229 (12.45)*	11,777 (11.02)
Business	10,747 (9.89)***	41,812 (10.18)	174 (9.46)	10,573 (9.89)
Clothes and fashion	16,502 (15.18)*	61,422 (14.95)	277 (15.06)	16,225 (15.18)
Crime and law	11,883 (10.93)	44,267 (10.77)	198 (10.77)	11,685 (10.93)
Culture	24,038 (22.11)***	88,152 (21.46)	402 (21.86)	23,636 (22.12)
Education	12,506 (11.50)	47,990 (11.68)	193 (10.49)	12,313 (11.52)
Family and life stages	18,274 (16.81)***	64,604 (15.72)	320 (17.40)	17,954 (16.80)
Food and drink	16,111 (14.82)	60,427 (14.71)	233 (12.67)***	15,878 (14.86)
Health	23,205 (21.34)***	82,739 (20.14)	421 (22.89)	22,784 (21.32)
Houses and buildings	16,318 (15.01)***	79,306 (19.30)	273 (14.85)	16,045 (15.01)
Language	19,706 (18.13)***	95,063 (23.14)	318 (17.29)	19,388 (18.14)
Leisure	29,069 (26.74)***	115,917 (28.21)	513 (27.90)	28,556 (26.72)
Nature	15,284 (14.06)***	64,497 (15.70)	251 (13.65)	15,033 (14.07)
Personality and emotions	44,541 (40.97)***	164,888 (40.13)	782 (42.52)	43,759 (40.94)
Religion and politics	5,163 (4.75)	19,274 (4.69)	79 (4.30)	5,084 (4.76)
Retail	7,783 (7.16)***	31,622 (7.70)	123 (6.69)	7,660 (7.17)
Science	7,436 (6.84)***	35,476 (8.63)	121 (6.58)	7,315 (6.84)
Social issues	5,610 (5.16)***	17,395 (4.23)	90 (4.89)	5,520 (5.16)

Technology	12,249 (11.27)	46,160 (11.24)	214 (11.64)	12,035 (11.26)
The media	8,461 (7.78)***	34,024 (8.28)	126 (6.85)	8,335 (7.80)
Travel and tourism	20,397 (18.76)***	83,155 (20.24)	329 (17.89)	20,068 (18.78)
War and conflict	2,768 (2.55)***	13,591 (3.31)	55 (2.99)	2,713 (2.54)
Work	11,107 (10.22)***	38,332 (9.33)	168 (9.14)	10,939 (10.24)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 100: Assault at semiprivate places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	8,951 (5.89)*	22,168 (6.03)	196 (5.21)*	8,755 (5.91)
Body and appearance	16,071 (10.57)	38,961 (10.60)	359 (9.55)**	15,712 (10.60)
Business	14,589 (9.60)***	37,970 (10.33)	303 (8.06)***	14,286 (9.64)
Clothes and fashion	22,051 (14.51)***	55,873 (15.20)	513 (13.64)	21,538 (14.53)
Crime and law	16,375 (10.77)	39,775 (10.82)	347 (9.23)***	16,028 (10.81)
Culture	33,385 (21.96)***	78,805 (21.44)	819 (21.78)	32,566 (21.97)
Education	17,515 (11.52)*	42,981 (11.69)	400 (10.64)*	17,115 (11.55)
Family and life stages	23,473 (15.44)***	59,405 (16.16)	532 (14.15)**	22,941 (15.48)
Food and drink	23,685 (15.58)***	52,853 (14.38)	490 (13.03)***	23,195 (15.65)
Health	29,804 (19.61)***	76,140 (20.71)	659 (17.52)***	29,145 (19.66)
Houses and buildings	32,650 (21.48)***	62,974 (17.13)	828 (22.02)	31,822 (21.47)
Language	38,620 (25.41)***	76,149 (20.72)	883 (23.48)***	37,737 (25.46)
Leisure	42,129 (27.72)*	102,857 (27.98)	998 (26.54)	41,131 (27.75)

Nature	21,751 (14.31)***	58,030 (15.79)	434 (11.54)***	21,317 (14.38)
Personality and emotions	61,679 (40.58)**	147,750 (40.20)	1,471 (39.11)*	60,208 (40.62)
Religion and politics	6,395 (4.21)***	18,042 (4.91)	134 (3.56)**	6,261 (4.22)
Retail	12,230 (8.05)***	27,175 (7.39)	264 (7.02)**	11,966 (8.07)
Science	14,361 (9.45)***	28,551 (7.77)	306 (8.14)***	14,055 (9.48)
Social issues	6,156 (4.05)***	16,849 (4.58)	170 (4.52)	5,986 (4.04)
Technology	16,055 (10.56)***	42,354 (11.52)	348 (9.25)***	15,707 (10.60)
The media	12,096 (7.96)***	30,389 (8.27)	258 (6.86)**	11,838 (7.99)
Travel and tourism	30,401 (20.00)	73,151 (19.90)	666 (17.71)***	29,735 (20.06)
War and conflict	3,666 (2.41)***	12,693 (3.45)	72 (1.91)**	3,594 (2.42)
Work	14,007 (9.22)***	35,432 (9.64)	315 (8.38)*	13,692 (9.24)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 101: Assault at semipublic places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	1,476 (5.02)***	29,643 (6.05)	24 (4.92)	1,452 (5.02)
Body and appearance	2,976 (10.13)***	52,056 (10.62)	40 (8.20)	2,936 (10.16)
Business	2,833 (9.64)***	49,726 (10.14)	32 (6.56)**	2,801 (9.69)
Clothes and fashion	4,167 (14.18)***	73,757 (15.05)	67 (13.73)	4,100 (14.19)
Crime and law	2,557 (8.70)***	53,593 (10.93)	42 (8.61)	2,515 (8.70)
Culture	5,457 (18.57)***	106,733 (21.77)	80 (16.39)	5,377 (18.60)
Education	2,766 (9.41)***	57,730 (11.78)	47 (9.63)	2,719 (9.41)

Family and life stages	4,607 (15.68)	78,271 (15.97)	76 (15.57)	4,531 (15.68)
Food and drink	3,953 (13.45)***	72,585 (14.81)	66 (13.52)	3,887 (13.45)
Health	5,573 (18.96)***	100,371 (20.48)	82 (16.80)	5,491 (19.00)
Houses and buildings	8,500 (28.92)***	87,124 (17.77)	138 (28.28)	8,362 (28.93)
Language	7,581 (25.80)***	107,188 (21.87)	141 (28.89)	7,440 (25.74)
Leisure	7,983 (27.16)***	137,003 (27.95)	112 (22.95)**	7,871 (27.23)
Nature	5,435 (18.49)***	74,346 (15.17)	71 (14.55)**	5,364 (18.56)
Personality and emotions	12,712 (43.25)***	196,717 (40.13)	213 (43.65)	12,499 (43.25)
Religion and politics	1,177 (4.00)***	23,260 (4.75)	16 (3.28)	1,161 (4.02)
Retail	2,136 (7.27)**	37,269 (7.60)	25 (5.12)*	2,111 (7.30)
Science	3,915 (13.32)***	38,997 (7.96)	89 (18.24)***	3,826 (13.24)
Social issues	1,130 (3.84)***	21,875 (4.46)	19 (3.89)	1,111 (3.84)
Technology	3,131 (10.65)***	55,278 (11.28)	42 (8.61)	3,089 (10.69)
The media	3,972 (13.52)***	38,513 (7.86)	44 (9.02)***	3,928 (13.59)
Travel and tourism	7,257 (24.69)***	96,295 (19.65)	112 (22.95)	7,145 (24.72)
War and conflict	2,592 (8.82)***	13,767 (2.81)	17 (3.48)***	2,575 (8.91)
Work	2,581 (8.78)***	46,858 (9.56)	44 (9.02)	2,537 (8.78)

** p<.1; ** p<.05; *** p<.01*

Table 102: Assault at public places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	12,520 (5.74)***	18,599 (6.17)	172 (4.59)***	12,348 (5.76)

Body and appearance	22,784 (10.45)***	32,248 (10.70)	376 (10.02)	22,408 (10.45)
Business	21,495 (9.86)***	31,064 (10.30)	315 (8.40)***	21,180 (9.88)
Clothes and fashion	31,885 (14.62)***	46,039 (15.27)	525 (14.00)	31,360 (14.63)
Crime and law	23,224 (10.65)***	32,926 (10.92)	389 (10.37)	22,835 (10.65)
Culture	47,677 (21.86)***	64,513 (21.40)	791 (21.09)	46,886 (21.87)
Education	24,673 (11.31)***	35,823 (11.88)	421 (11.22)	24,252 (11.31)
Family and life stages	34,380 (15.76)***	48,498 (16.09)	610 (16.26)	33,770 (15.76)
Food and drink	32,780 (15.03)***	43,758 (14.51)	531 (14.16)	32,249 (15.05)
Health	41,616 (19.08)***	64,328 (21.34)	662 (17.65)**	40,954 (19.11)
Houses and buildings	42,185 (19.34)***	53,439 (17.73)	749 (19.97)	41,436 (19.33)
Language	50,651 (23.22)***	64,118 (21.27)	818 (21.81)**	49,833 (23.25)
Leisure	59,236 (27.16)***	85,750 (28.44)	1,050 (27.99)	58,186 (27.15)
Nature	33,159 (15.20)**	46,622 (15.46)	487 (12.98)***	32,672 (15.24)
Personality and emotions	86,481 (39.65)***	122,948 (40.78)	1,478 (39.40)	85,003 (39.66)
Religion and politics	9,524 (4.37)***	14,913 (4.95)	162 (4.32)	9,362 (4.37)
Retail	16,101 (7.38)***	23,304 (7.73)	289 (7.7)	15,812 (7.38)
Science	18,662 (8.56)***	24,250 (8.04)	322 (8.58)	18,340 (8.56)
Social issues	9,058 (4.15)***	13,947 (4.63)	155 (4.13)	8,903 (4.15)
Technology	24,346 (11.16)	34,063 (11.30)	373 (9.94)**	23,973 (11.18)
The media	18,728 (8.59)***	23,757 (7.88)	282 (7.52)**	18,446 (8.61)
Travel and tourism	42,706 (19.58)***	60,846 (20.18)	715 (19.06)	41,991 (19.59)

War and conflict	8,084 (3.71)***	8,275 (2.74)	116 (3.09)**	7,968 (3.72)
Work	20,588 (9.44)	28,851 (9.57)	337 (8.98)	20,251 (9.45)

** p<.1; ** p<.05; *** p<.01*

Table 103: Robbery at private places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	1,514 (5.87)	29,605 (6.00)	14 (6.86)	1,500 (5.86)
Body and appearance	2,889 (11.20)***	52,143 (10.56)	38 (18.63)***	2,851 (11.14)
Business	2,297 (8.90)***	50,262 (10.18)	15 (7.35)	2,282 (8.92)
Clothes and fashion	3,927 (15.22)	73,997 (14.99)	28 (13.73)	3,899 (15.23)
Crime and law	2,718 (10.53)	53,432 (10.82)	19 (9.31)	2,699 (10.54)
Culture	5,491 (21.28)	106,699 (21.61)	36 (17.65)	5,455 (21.31)
Education	2,566 (9.95)***	57,930 (11.73)	25 (12.25)	2,541 (9.93)
Family and life stages	4,397 (17.04)***	78,481 (15.89)	36 (17.65)	4,361 (17.04)
Food and drink	3,535 (13.70)***	73,003 (14.79)	27 (13.24)	3,508 (13.70)
Health	5,652 (21.91)***	100,292 (20.31)	42 (20.59)	5,610 (21.92)
Houses and buildings	3,533 (13.69)***	92,091 (18.65)	37 (18.14)*	3,496 (13.66)
Language	4,262 (16.52)***	110,507 (22.38)	37 (18.14)	4,225 (16.51)
Leisure	7,130 (27.63)	137,856 (27.92)	63 (30.88)	7,067 (27.61)
Nature	3,735 (14.48)***	76,046 (15.40)	27 (13.24)	3,708 (14.49)
Personality and emotions	11,031 (42.75)***	198,398 (40.18)	83 (40.69)	10,948 (42.77)
Religion and politics	1,210 (4.69)	23,227 (4.70)	12 (5.88)	1,198 (4.68)

Retail	1,696 (6.57)***	37,709 (7.64)	11 (5.39)	1,685 (6.58)
Science	1,596 (6.19)***	41,316 (8.37)	14 (6.86)	1,582 (6.18)
Social issues	1,227 (4.76)***	21,778 (4.41)	12 (5.88)	1,215 (4.75)
Technology	3,008 (11.66)**	55,401 (11.22)	29 (14.22)	2,979 (11.64)
The media	1,813 (7.03)***	40,672 (8.24)	12 (5.88)	1,801 (7.04)
Travel and tourism	4,580 (17.75)***	98,972 (20.04)	41 (20.10)	4,539 (17.73)
War and conflict	694 (2.69)***	15,665 (3.17)	2 (0.98)	692 (2.70)
Work	2,085 (8.08)***	47,354 (9.59)	24 (11.76)*	2,061 (8.05)

* *p<.1; ** p<.05; *** p<.01*

Table 104: Robbery at semiprivate places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	3,208 (5.65)***	27,911 (6.03)	49 (7.95)**	3,159 (5.63)
Body and appearance	6,044 (10.65)	48,988 (10.58)	73 (11.85)	5,971 (10.64)
Business	5,312 (9.36)***	47,247 (10.21)	53 (8.60)	5,259 (9.37)
Clothes and fashion	8,580 (15.12)	69,344 (14.98)	100 (16.23)	8,480 (15.11)
Crime and law	5,488 (9.67)***	50,662 (10.95)	64 (10.39)	5,424 (9.66)
Culture	12,118 (21.35)	100,072 (21.62)	127 (20.62)	11,991 (21.36)
Education	6,580 (11.59)	53,916 (11.65)	79 (12.82)	6,501 (11.58)
Family and life stages	8,514 (15.00)***	74,364 (16.07)	95 (15.42)	8,419 (15.00)
Food and drink	8,882 (15.65)***	67,656 (14.62)	124 (20.13)***	8,758 (15.60)
Health	11,005 (19.39)***	94,939 (20.51)	130 (21.10)	10,875 (19.37)

Houses and buildings	10,281 (18.12)*	85,343 (18.44)	109 (17.69)	10,172 (18.12)
Language	14,649 (25.81)***	100,120 (21.63)	158 (25.65)	14,491 (25.81)
Leisure	15,106 (26.62)***	129,880 (28.06)	172 (27.92)	14,934 (26.60)
Nature	8,132 (14.33)***	71,649 (15.48)	89 (14.45)	8,043 (14.33)
Personality and emotions	22,490 (39.63)***	186,939 (40.39)	256 (41.56)	22,234 (39.61)
Religion and politics	2,320 (4.09)***	22,117 (4.78)	25 (4.06)	2,295 (4.09)
Retail	4,600 (8.11)***	34,805 (7.52)	57 (9.25)	4,543 (8.09)
Science	4,321 (7.61)***	38,591 (8.34)	40 (6.49)	4,281 (7.63)
Social issues	2,510 (4.42)	20,495 (4.43)	35 (5.68)	2,475 (4.41)
Technology	5,895 (10.39)***	52,514 (11.35)	79 (12.82)**	5,816 (10.36)
The media	4,633 (8.16)	37,852 (8.18)	59 (9.58)	4,574 (8.15)
Travel and tourism	10,585 (18.65)***	92,967 (20.09)	107 (17.37)	10,478 (18.67)
War and conflict	1,459 (2.57)***	14,900 (3.22)	19 (3.08)	1,440 (2.57)
Work	4,721 (8.32)***	44,718 (9.66)	71 (11.53)***	4,650 (8.28)

** p<.1; ** p<.05; *** p<.01*

Table 105: Robbery at semipublic places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	882 (6.72)***	30,237 (5.97)	7 (5.15)	875 (6.74)
Body and appearance	1,307 (9.96)**	53,725 (10.61)	11 (8.09)	1,296 (9.98)
Business	1,324 (10.09)	51,235 (10.12)	7 (5.15)	1,317 (10.14)
Clothes and fashion	2,000 (15.24)	75,924 (14.99)	20 (14.71)	1,980 (15.25)

Crime and law	1,226 (9.34)***	54,924 (10.85)	10 (7.35)	1,216 (9.36)
Culture	2,600 (19.81)***	109,590 (21.64)	24 (17.65)	2,576 (19.84)
Education	1,342 (10.23)***	59,154 (11.68)	7 (5.15)**	1,335 (10.28)
Family and life stages	2,029 (15.46)	80,849 (15.96)	22 (16.18)	2,007 (15.46)
Food and drink	2,046 (15.59)***	74,492 (14.71)	22 (16.18)	2,024 (15.59)
Health	2,542 (19.37)***	103,402 (20.42)	22 (16.18)	2,520 (19.41)
Houses and buildings	2,580 (19.66)***	93,044 (18.37)	18 (13.24)*	2,562 (19.73)
Language	3,271 (24.93)***	111,498 (22.02)	29 (21.32)	3,242 (24.97)
Leisure	3,572 (27.22)*	141,414 (27.92)	37 (27.21)	3,535 (27.22)
Nature	1,771 (13.50)***	78,010 (15.40)	22 (16.18)	1,749 (13.47)
Personality and emotions	5,229 (39.85)	204,200 (40.32)	56 (41.18)	5,173 (39.84)
Religion and politics	545 (4.15)***	23,892 (4.72)	7 (5.15)	538 (4.14)
Retail	1,112 (8.47)***	38,293 (7.56)	10 (7.35)	1,102 (8.49)
Science	1,112 (8.47)	41,800 (8.25)	5 (3.68)**	1,107 (8.52)
Social issues	530 (4.04)**	22,475 (4.44)	4 (2.94)	526 (4.05)
Technology	1,359 (10.36)***	57,050 (11.26)	8 (5.88)*	1,351 (10.4)
The media	988 (7.53)***	41,497 (8.19)	7 (5.15)	981 (7.55)
Travel and tourism	2,747 (20.93)***	100,805 (19.90)	24 (17.65)	2,723 (20.97)
War and conflict	310 (2.36)***	16,049 (3.17)	6 (4.41)	304 (2.34)
Work	1,139 (8.68)***	48,300 (9.54)	8 (5.88)	1,131 (8.71)

** p<.1; ** p<.05; *** p<.01*

Table 106: Robbery at public places: frequency of topic by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Animals	5,695 (5.74)***	25,424 (6.05)	41 (5.02)	5,654 (5.75)
Body and appearance	10,138 (10.22)***	44,894 (10.68)	90 (11.03)	10,048 (10.22)
Business	8,306 (8.38)***	44,253 (10.53)	62 (7.60)	8,244 (8.38)
Clothes and fashion	14,543 (14.67)***	63,381 (15.08)	105 (12.87)	14,438 (14.68)
Crime and law	9,818 (9.90)***	46,332 (11.02)	86 (10.54)	9,732 (9.90)
Culture	20,834 (21.01)***	91,356 (21.73)	169 (20.71)	20,665 (21.01)
Education	10,039 (10.12)***	50,457 (12.00)	63 (7.72)**	9,976 (10.14)
Family and life stages	15,819 (15.95)	67,059 (15.95)	149 (18.26)*	15,670 (15.93)
Food and drink	13,946 (14.06)***	62,592 (14.89)	113 (13.85)	13,833 (14.07)
Health	19,540 (19.71)***	86,404 (20.55)	155 (19.00)	19,385 (19.71)
Houses and buildings	16,612 (16.75)***	79,012 (18.79)	120 (14.71)	16,492 (16.77)
Language	19,854 (20.02)***	94,915 (22.58)	159 (19.49)	19,695 (20.03)
Leisure	25,985 (26.21)***	119,001 (28.31)	180 (22.06)***	25,805 (26.24)
Nature	13,574 (13.69)***	66,207 (15.75)	112 (13.73)	13,462 (13.69)
Personality and emotions	40,217 (40.56)*	169,212 (40.25)	328 (40.20)	39,889 (40.56)
Religion and politics	4,253 (4.29)***	20,184 (4.8)	29 (3.55)	4,224 (4.30)
Retail	6,606 (6.66)***	32,799 (7.80)	47 (5.76)	6,559 (6.67)
Science	8,114 (8.18)	34,798 (8.28)	55 (6.74)	8,059 (8.19)
Social issues	4,383 (4.42)	18,622 (4.43)	31 (3.80)	4,352 (4.43)

Technology	10,768 (10.86)***	47,641 (11.33)	68 (8.33)**	10,700 (10.88)
The media	7,221 (7.28)***	35,264 (8.39)	55 (6.74)	7,166 (7.29)
Travel and tourism	17,550 (17.70)***	86,002 (20.46)	153 (18.75)	17,397 (17.69)
War and conflict	2,544 (2.57)***	13,815 (3.29)	20 (2.45)	2,524 (2.57)
Work	7,894 (7.96)***	41,545 (9.88)	57 (6.99)	7,837 (7.97)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 107: Drug at private places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	1,149 (5.58)**	29,970 (6.01)	26 (6.39)	1,123 (5.56)
Body and appearance	2,092 (10.15)**	52,940 (10.61)	43 (10.57)	2,049 (10.15)
Business	1,570 (7.62)***	50,989 (10.22)	37 (9.09)	1,533 (7.59)
Clothes and fashion	2,846 (13.81)***	75,078 (15.05)	49 (12.04)	2,797 (13.85)
Crime and law	1,918 (9.31)***	54,232 (10.87)	32 (7.86)	1,886 (9.34)
Culture	4,002 (19.43)***	108,188 (21.68)	77 (18.92)	3,925 (19.44)
Education	1,850 (8.98)***	58,646 (11.75)	30 (7.37)	1,820 (9.01)
Family and life stages	3,273 (15.89)	79,605 (15.95)	69 (16.95)	3,204 (15.87)
Food and drink	2,439 (11.84)***	74,099 (14.85)	50 (12.29)	2,389 (11.83)
Health	4,129 (20.04)	101,815 (20.41)	78 (19.16)	4,051 (20.06)
Houses and buildings	2,594 (12.59)***	93,030 (18.64)	46 (11.30)	2,548 (12.62)
Language	2,562 (12.44)***	112,207 (22.49)	58 (14.25)	2,504 (12.4)
Leisure	5,033 (24.43)***	139,953 (28.05)	101 (24.82)	4,932 (24.42)

Nature	2,747 (13.33)***	77,034 (15.44)	65 (15.97)	2,682 (13.28)
Personality and emotions	8,040 (39.03)***	201,389 (40.36)	149 (36.61)	7,891 (39.07)
Religion and politics	949 (4.61)	23,488 (4.71)	12 (2.95)	937 (4.64)
Retail	1,225 (5.95)***	38,180 (7.65)	14 (3.44)**	1,211 (6.00)
Science	1,172 (5.69)***	41,740 (8.37)	15 (3.69)*	1,157 (5.73)
Social issues	837 (4.06)***	22,168 (4.44)	13 (3.19)	824 (4.08)
Technology	2,203 (10.69)**	56,206 (11.26)	36 (8.85)	2,167 (10.73)
The media	1,263 (6.13)***	41,222 (8.26)	24 (5.90)	1,239 (6.14)
Travel and tourism	3,439 (16.69)***	100,113 (20.06)	65 (15.97)	3,374 (16.71)
War and conflict	456 (2.21)***	15,903 (3.19)	9 (2.21)	447 (2.21)
Work	1,505 (7.31)***	47,934 (9.61)	32 (7.86)	1,473 (7.29)

* *p<.1;* ** *p<.05;* *** *p<.01*

Table 108: Drug at semiprivate places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	1,796 (5.71)**	29,323 (6.01)	61 (5.39)	1,735 (5.72)
Body and appearance	3,086 (9.81)***	51,946 (10.64)	87 (7.69)**	2,999 (9.89)
Business	2,776 (8.83)***	49,783 (10.20)	58 (5.13)***	2,718 (8.97)
Clothes and fashion	4,409 (14.02)***	73,515 (15.06)	138 (12.20)*	4,271 (14.09)
Crime and law	2,898 (9.22)***	53,252 (10.91)	107 (9.46)	2,791 (9.21)
Culture	6,892 (21.92)	105,298 (21.57)	216 (19.10)**	6,676 (22.02)
Education	3,386 (10.77)***	57,110 (11.70)	280 (24.76)***	3,106 (10.25)

Family and life stages	4,790 (15.23)***	78,088 (16.00)	159 (14.06)	4,631 (15.28)
Food and drink	4,526 (14.39)*	72,012 (14.75)	92 (8.13)***	4,434 (14.63)
Health	5,584 (17.76)***	100,360 (20.56)	163 (14.41)***	5,421 (17.88)
Houses and buildings	7,322 (23.29)***	88,302 (18.09)	351 (31.03)***	6,971 (23.00)
Language	7,857 (24.99)***	106,912 (21.90)	264 (23.34)	7,593 (25.05)
Leisure	9,293 (29.55)***	135,693 (27.80)	318 (28.12)	8,975 (29.61)
Nature	4,131 (13.14)***	75,650 (15.50)	108 (9.55)***	4,023 (13.27)
Personality and emotions	12,848 (40.86)**	196,581 (40.27)	389 (34.39)***	12,459 (41.10)
Religion and politics	1,179 (3.75)***	23,258 (4.76)	37 (3.27)	1,142 (3.77)
Retail	3,389 (10.78)***	36,016 (7.38)	92 (8.13)***	3,297 (10.88)
Science	3,566 (11.34)***	39,346 (8.06)	80 (7.07)***	3,486 (11.50)
Social issues	1,185 (3.77)***	21,820 (4.47)	49 (4.33)	1,136 (3.75)
Technology	3,234 (10.28)***	55,175 (11.30)	83 (7.34)***	3,151 (10.39)
The media	2,449 (7.79)***	40,036 (8.20)	90 (7.96)	2,359 (7.78)
Travel and tourism	5,562 (17.69)***	97,990 (20.08)	155 (13.70)***	5,407 (17.84)
War and conflict	746 (2.37)***	15,613 (3.20)	19 (1.68)	727 (2.40)
Work	2,681 (8.53)***	46,758 (9.58)	56 (4.95)***	2,625 (8.66)

** p<.1; ** p<.05; *** p<.01*

Table 109: Drug at semipublic places: frequency of topic by each group

	Group A N (%)	**Group B** N (%)	**Group A1** N (%)	**Group A2** N (%)
Animals	856 (5.02)***	30,263 (6.02)	13 (4.09)	843 (5.04)

Body and appearance	1,546 (9.07)***	53,486 (10.64)	21 (6.60)	1,525 (9.12)
Business	1,606 (9.42)***	50,953 (10.14)	34 (10.69)	1,572 (9.40)
Clothes and fashion	2,381 (13.97)***	75,543 (15.03)	46 (14.47)	2,335 (13.96)
Crime and law	1,399 (8.21)***	54,751 (10.90)	33 (10.38)	1,366 (8.17)
Culture	2,999 (17.59)***	109,191 (21.73)	65 (20.44)	2,934 (17.54)
Education	1,517 (8.90)***	58,979 (11.74)	21 (6.60)	1,496 (8.94)
Family and life stages	2,739 (16.07)	80,139 (15.95)	55 (17.30)	2,684 (16.05)
Food and drink	2,243 (13.16)***	74,295 (14.78)	45 (14.15)	2,198 (13.14)
Health	3,097 (18.17)***	102,847 (20.47)	60 (18.87)	3,037 (18.16)
Houses and buildings	4,923 (28.88)***	90,701 (18.05)	82 (25.79)	4,841 (28.94)
Language	4,590 (26.93)***	110,179 (21.93)	84 (26.42)	4,506 (26.94)
Leisure	4,667 (27.38)	140,319 (27.92)	79 (24.84)	4,588 (27.43)
Nature	3,855 (22.62)***	75,926 (15.11)	50 (15.72)***	3,805 (22.75)
Personality and emotions	6,677 (39.17)***	202,752 (40.35)	120 (37.74)	6,557 (39.20)
Religion and politics	617 (3.62)***	23,820 (4.74)	8 (2.52)	609 (3.64)
Retail	1,245 (7.30)	38,160 (7.59)	22 (6.92)	1,223 (7.31)
Science	1,655 (9.71)***	41,257 (8.21)	33 (10.38)	1,622 (9.70)
Social issues	595 (3.49)***	22,410 (4.46)	12 (3.77)	583 (3.49)
Technology	1,920 (11.26)	56,489 (11.24)	32 (10.06)	1,888 (11.29)
The media	3,089 (18.12)***	39,396 (7.84)	41 (12.89)**	3,048 (18.22)
Travel and tourism	4,950 (29.04)***	98,602 (19.62)	80 (25.16)	4,870 (29.11)

War and conflict	2,254 (13.22)***	14,105 (2.81)	21 (6.60)***	2,233 (13.35)
Work	1,726 (10.13)***	47,713 (9.49)	27 (8.49)	1,699 (10.16)

** p<.1; ** p<.05; *** p<.01*

Table 110: Drug at public places: frequency of topic by each group

	Group A N (%)	Group B N (%)	Group A1 N (%)	Group A2 N (%)
Animals	8,303 (5.88)**	22,816 (6.03)	222 (5.81)	8,081 (5.88)
Body and appearance	14,823 (10.49)	40,209 (10.63)	364 (9.53)*	14,459 (10.52)
Business	12,414 (8.79)***	40,145 (10.61)	310 (8.12)	12,104 (8.80)
Clothes and fashion	20,424 (14.45)***	57,500 (15.20)	518 (13.57)	19,906 (14.48)
Crime and law	15,177 (10.74)	40,973 (10.83)	367 (9.61)**	14,810 (10.77)
Culture	30,379 (21.50)	81,811 (21.63)	766 (20.06)**	29,613 (21.54)
Education	15,672 (11.09)***	44,824 (11.85)	408 (10.69)	15,264 (11.10)
Family and life stages	22,348 (15.82)	60,530 (16.00)	566 (14.82)*	21,782 (15.84)
Food and drink	20,425 (14.45)***	56,113 (14.83)	490 (12.83)***	19,935 (14.50)
Health	28,813 (20.39)	77,131 (20.39)	688 (18.02)***	28,125 (20.46)
Houses and buildings	29,398 (20.80)***	66,226 (17.51)	854 (22.37)**	28,544 (20.76)
Language	33,008 (23.36)***	81,761 (21.62)	934 (24.46)	32,074 (23.33)
Leisure	39,296 (27.81)	105,690 (27.94)	1,084 (28.39)	38,212 (27.79)
Nature	21,411 (15.15)**	58,370 (15.43)	512 (13.41)***	20,899 (15.20)
Personality and emotions	57,990 (41.04)***	151,439 (40.04)	1,535 (40.20)	56,455 (41.06)
Religion and politics	6,073 (4.30)***	18,364 (4.85)	125 (3.27)***	5,948 (4.33)

Retail	11,026 (7.80)***	28,379 (7.50)	310 (8.12)	10,716 (7.79)
Science	13,137 (9.30)***	29,775 (7.87)	406 (10.63)***	12,731 (9.26)
Social issues	5,742 (4.06)***	17,263 (4.56)	153 (4.01)	5,589 (4.07)
Technology	15,904 (11.26)	42,505 (11.24)	376 (9.85)***	15,528 (11.29)
The media	13,010 (9.21)***	29,475 (7.79)	284 (7.44)***	12,726 (9.26)
Travel and tourism	29,378 (20.79)***	74,174 (19.61)	748 (19.59)*	28,630 (20.82)
War and conflict	5,360 (3.79)***	10,999 (2.91)	104 (2.72)***	5,256 (3.82)
Work	12,358 (8.75)***	37,081 (9.80)	269 (7.05)***	12,089 (8.79)

* *p<.1;* ** *p<.05;* *** *p<.01*

BIBLIOGRAPHY

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, *30*(1), 47–88.

Agnew, R. (1995). Testing the leading crime theories: An alternative strategy focusing on motivational processes. *Journal of Research in Crime and Delinquency*, *32*(4), 363–398.

Agnew, R. (2001). Building on the foundation of general strain theory: Specifying the types of strain most likely to lead to crime and delinquency. *Journal of Research in Crime and Delinquency*, *38*(4), 319–361.

Agnew, R. (2006). General strain theory: Current status and directions for further research. *Taking Stock: The Status of Criminological Theory*, *15*, 101–123.

Agnew, R. (2009). The contribution of "mainstream" theories to the explanation of female delinquency. *The Delinquent Girl*, 7–29.

Agnew, R. (2012). Reflection on "A Revised Strain Theory of Delinquency". *Social Forces*, *91*(1), 33–38. https://doi.org/10.1093/sf/sos117

Agnew, R. (2015). Using general strain theory to explain crime in Asian societies. *Asian Journal of Criminology*, *10*(2), 131–147.

Agnew, R. (2017). Revitalizing merton: General strain theory. In *The origins of American criminology* (pp. 137–158). Routledge.

Agnew, R., Piquero, N. L., Cullen, F. T. (2009). General strain theory and white-collar crime. In *The criminology of white-collar crime* (pp. 35–60). Springer.

Ahmed, W., Bath, P., Demartini, G. (2017). Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges. In *TheEthics of Online Research. Advances in Research Ethics and Integrity* (pp. 79–107).

Akers, R. L. (1973). *Deviant behavior;: A social learning approach*.

Akers, R. L. (2013). *Criminological theories: Introduction and evaluation*.

Anderson, E. (1999). *Code of the Street*. New York: Norton.

Bachrach, Y., Graepel, T., Minka, T., Guiver, J. (2012). How to grade a test without knowing the answers: A Bayesian graphical model for adaptive crowdsourcing and aptitude testing. *ArXiv Preprint ArXiv:1206.6386*.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., Gosling, S. D. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science*, *21*(3), 372–374.

Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., Hughes, M. (2013). *Sentiment analysis of political tweets: Towards an accurate classifier*.

Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley Sons.

Baltagi, B. H. (2006). Estimating an economic model of crime using panel data from North Carolina. *Journal of Applied Econometrics*, *21*(4), 543–547.

Baumer, T. L. (1985). Testing a general model of fear of crime: Data from a national sample. *Journal of Research in Crime and Delinquency*, *22*(3), 239–255.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

Baym, N. (2002). Interpersonal life online. *The Handbook of New Media*, *4*, 62–76.

Beavon, D. J., Brantingham, P. L., Brantingham, P. J. (1994). The influence of street networks on the patterning of property offenses. *Crime Prevention Studies*, *2*, 115–148.

Becker, H. S. (1963). *Outsiders*. Simon and Schuster.

Bellair, P. E. (1997). Social interaction and community crime: Examining the importance of neighbor networks. *Criminology*, *35*(4), 677–704.

Bellair, P. E. (2000). Informal Surveillance and Street Crime: A Complex Relationship*. *Criminology*, *38*(1), 137–170.

Bendler, J., Ratku, A., Neumann, D. (2014). Crime Mapping through Geo-Spatial Social Media Activity. *ICIS 2014 Proceedings*.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8–14.

Benson, M. L., Livelsberger, T. (2012). Emotions, Choice, and Crime. *The Oxford Handbook of Criminological Theory*.

Bermingham, A., Smeaton, A. (2011). On using Twitter to monitor political sentiment and predict election results. *Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011)*, 2–10.

Bermingham, A., Smeaton, A. F. (2010). Classifying sentiment in microblogs: Is brevity an advantage" *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1833–1836. ACM.

Blakemore, S.-J., Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184.

Boetig, B. P., King, M. J. (2006). The routine activity theory. *FBI Law Enforcement Bulletin*, *75*(6), 12.

Boies, S. C. (2002). University students' uses of and reactions to online sexual information and entertainment: links to online and offline sexual behaviour. *Canadian Journal of Human Sexuality*, *11*(2).

Bokanyi, E., Kondor, D., Dobos, L., Sebők, T., St"ger, J., Csabai, I., Vattay, G. (2016). Race, religion and the city: Twitter word frequency patterns reveal dominant demographic dimensions in the United States. *Palgrave Communications*, *2*(1), 1–9. https://doi.org/10.1057/palcomms.2016.10

Boldt, M., Borg, A. (2016). Evaluating Temporal Analysis Methods Using Residential Burglary Data. *ISPRS International Journal of Geo-Information*, *5*(9), 148. https://doi.org/10.3390/ijgi5090148

Bollen, J., Mao, H., Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Fifth International AAAI Conference on Weblogs and Social Media*.

Bollen, J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bowers, K. J. (2013). *Situational crime prevention: What we know and what we are starting to understand. Keynote speech.*

Bowers, Kate J., Johnson, S. D., Pease, K. (2004). Prospective Hot-SpottingThe Future of Crime Mapping" *The British Journal of Criminology*, *44*(5), 641–658. https://doi.org/10.1093/bjc/azh036

Bowers, Kate J., Johnson, S. D., Pease, K. (2005). Victimisation and re-victimisation risk, housing type and area: A study of interactions. *Crime Prevention and Community Safety*, *7*(1), 7–17.

Bowling, B., Phillips, C. (2003). *Policing ethnic minority communities*.

Brantingham, P. L., Brantingham, P. J. (1975). Residential burglary and urban form. *Urban Studies*, *12*(3), 273–284.

Brantingham, P. J., Brantingham, P. L. (1978). A theoretical model of crime site selection. *Crime, Law and Sanctions.* Sage Publications Inc.

Brantingham, P. J., Brantingham, P. L. (1981). *Environmental criminology*. Sage Publications Beverly Hills, CA.

Brantingham, P. L., Brantingham, P. J. (1988). Situational crime prevention in British Columbia. *Journal of Security Administration*, *11*(2), 17–27.

Brantingham, P. J., Brantingham, P. J., Wong, P. S. (1991). How public transit feeds private crime: Notes on the Vancouver 'Skytrain'experience. *Security Journal*, *2*(2), 91–95.

Brantingham, P. J., Brantingham, P. L. (1993). Environment, routine and situation: Toward a pattern theory of crime. *Advances in Criminological Theory*, *5*(2), 259–94.

Brantingham, P. L., Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28.

Brass, D. J. (1984). Being in the Right Place: A Structural Analysis of Individual Influence in an Organization. *Administrative Science Quarterly*, *29*(4), 518–539.

Breul, J. D. (2017). *Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation*. Transaction Publishers.

Broidy, L. M., Daday, J. K., Crandall, C. S., Sklar, D. P., Jost, P. F. (2006). Exploring Demographic, Structural, and Behavioral Overlap Among Homicide Offenders and Victims. *Homicide Studies*, *10*(3), 155–180.

Browning, C. R., Feinberg, S. L., Dietz, R. D. (2004). The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. *Social Forces*, *83*(2), 503–534.

Bruinsma, G., Johnson, S. D. (2018). *The Oxford Handbook of Environmental Criminology*. Oxford University Press.

Buchanan, T. (2001). *Online personality assessment*. Pabst Science Publishers.

Burgess, E. W. (1928). Residential segregation in American cities. *The Annals of the American Academy of Political and Social Science*, *140*(1), 105–115.

Bursik, Grasmick, H. G. (2012). *Neighborhoods Crime*. Lexington Books.

Bursik Jr, R. J., Grasmick, H. G. (1993). Economic deprivation and neighborhood crime rates, 1960-1980. *Law Soc'y Rev.*, *27*, 263.

Bursik, R. J. (1988). Social Disorganization and Theories of Crime and Delinquency: Problems and Prospects*. *Criminology*, *26*(4), 519–552.

Carr, P. J. (2005). *Clean Streets: Controlling Crime, Maintaining Order, and Building Community Activism*. NYU Press.

Cataldi, M., Di Caro, L., Schifanella, C. (2010). Emerging topic detection on twitter based on temporal and social terms evaluation. *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 4. ACM.

Chan, C.-K., Cheng, L.-M. (2004). Hiding data in images by simple LSB substitution. *Pattern Recognition*, *37*(3), 469–474.

Charbonneau, K. (2012). Multiple fixed effects in nonlinear panel data models. *Unpublished Manuscript*.

Chavis, D. M., Hogge, J. H., McMillan, D. W., Wandersman, A. (1986). Sense of community through Brunswik's lens: A first look. *Journal of Community Psychology*, *14*(1), 24–40.

Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, *27*(2), 755–762.

Chen, M.-Y., Chen, T.-H. (2019). Modeling public mood and emotion: Blog and news sentiment and socio-economic phenomena. *Future Generation Computer Systems*, *96*, 692–699. https://doi.org/10.1016/j.future.2017.10.028

Chen, X., Cho, Y., Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. *2015 Systems and Information Engineering Design Symposium*, 63–68. https://doi.org/10.1109/SIEDS.2015.7117012

Clarke, R. V., Eck, J. E. (2005). Crime Analysis for Problem Solvers in 60 Small Steps-Step 35. *Center for Problem Oriented Policing. Washington, DC: US Department of Justice Office of Community Oriented Policing Services*, 68.

Clarke, R. V. G. (1997). *Situational Crime Prevention: Successful Case Studies*. Lynne Rienner Publishers.

Clarke, R. V. G., Felson, M. (1993). *Routine Activity and Rational Choice*. Transaction Publishers.

Clarke, Ronald V., Cornish, D. B. (1985). Modeling Offenders' Decisions: A Framework for Research and Policy. *Crime and Justice*, *6*, 147–185.

Clarke, Ronald V., Harris, P. M. (1992). Auto theft and its prevention. *Crime and Justice*, *16*, 1–54.

Clarke, Ronald VG. (1980). ' situational' Crime Prevention: Theory and Practice. *The British Journal of Criminology*, *20*(2), 136–147.

Cohen, A. K. (1955). *Delinquent boys; The culture of the gang*. New York, NY, US: Free Press.

Cohen, L. E., Cantor, D. (1981). Residential Burglary in the United States: Life-Style and Demographic Factors Associated With the Probability of Victimization. *Journal of Research in Crime and Delinquency*, *18*(1), 113–127.

Cohen, L. E., Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 588–608.

Cohen, L. E., Kluegel, J. R., Land, K. C. (1981). Social Inequality and Predatory Criminal Victimization: An Exposition and Test of a Formal Theory. *American Sociological Review*, *46*(5), 505–524. https://doi.org/10.2307/2094935

Cooney, M. (1998). *Warriors and peacemakers: How third parties shape violence*. NYU Press.

Cooney, M. (2009). The scientific significance of Collins's Violence 1. *The British Journal of Sociology*, *60*(3), 586–594.

Cornish, D. B., Clarke, R. V. (1987). Understanding Crime Displacement: An Application of Rational Choice Theory. *Criminology*, *25*(4), 933–948.

Cornish, D., Clarke, R. (1986). Situational prevention, displacement of crime and rational choice theory. *Situational Crime Prevention: From Theory into Practice*, 1–16.

Correa, T., Hinsley, A. W., De Zuniga, H. G. (2010). Who interacts on the Web": The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2), 247–253.

Curtis, S. (2013, June 9). *Twitter claims 15m active users in the UK*. Retrieved from http://www.telegraph.co.uk/technology/twitter/10291360/Twitter-claims-15m-active-users-in-the-UK.html

Daday, J. K., Broidy, L. M., Crandall, C. S., Sklar, D. P. (2005). Individual, Neighborhood, and Situational Factors Associated with Violent Victimization and Offending. *Criminal Justice Studies*, *18*(3), 215–235.

Dai, M., Jiang, X. (2016). A comparative study of satisfaction with the police in the United States and Australia. *Australian New Zealand Journal of Criminology*, *49*(1), 30–52.

Dalianis, H. (2018). *Clinical text mining: Secondary use of electronic patient records*. Springer.

Dann, S. (2010). Redefining social marketing with contemporary commercial marketing definitions. *Journal of Business Research*, *63*(2), 147–153.

De Rivera, J. (1992). Emotional climate: Social structure and emotional dynamics. *A Preliminary Draft of This Chapter Was Discussed at a Workshop on Emotional Climate Sponsored by the Clark European Center in Luxembourg on Jul 12–14, 1991.* John Wiley Sons.

Degarmo, M. (2011). Understanding the Comparisons of Routine Activities and Contagious Distributions of Victimization: Forming a Mixed Model of Confluence and Transmission. *International Journal of Criminology and Sociological Theory*, *4*(1).

Delivanis, C. (n.d.). Why Big Data Fails Entrepreneurs. Retrieved 2 March 2017, Retrieved 2 March 2017, from http://www.forbes.com/sites/groupthink/2014/06/27/why-big-data-fails-entrepreneurs/

Dewey, J. (1913). *Interest and effort in education*. Houghton Mifflin.

Di Gennaro, C., Dutton, W. (2006). The Internet and the public: Online and offline political participation in the United Kingdom. *Parliamentary Affairs*, *59*(2), 299–313.

Durkheim. (1951). *Suicide: A study in sociology*. Glencoe, IL: Free Press.

Eck, J. E. (2015). Who should prevent crime at places" The advantages of regulating place managers and challenges to police services. *Policing: A Journal of Policy and Practice*, *9*(3), 223–233.

Eck, J. E., Rosenbaum, D. (1994). The new police order: Effectiveness, equity, and efficiency in community policing. *The Challenge of Community Policing: Testing the Promises*, 3–23.

Eck, J. E., Weisburd, D. (1995). *Crime and place* (Vol. 4). Criminal Justice Press. Monsey, NY.

Eickhoff, M., Wieneke, R. (2018). Understanding Topic Models in Context: A Mixed-Methods Approach to the Meaningful Analysis of Large Document Collections. *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200.

Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. (n.d.). Retrieved 30 September 2019, from https://journals.plos.org/plosone/article"id=10.1371/journal.pone.0205151

Farahani, L. M., Maller, C., Phelan, K. (2018). Private Gardens as Urban Greenspaces: Can They Compensate for Poor Greenspace Access in Lower Socioeconomic Neighbourhoods?. **Landscape Online**, **59**, 1-18.

Farnadi, G., Zoghbi, S., Moens, M.-F., Cock, M. D. (2013). Recognising Personality Traits Using Facebook Status Updates. *Seventh International AAAI Conference on Weblogs and Social Media*. Presented at the Seventh International AAAI Conference on Weblogs and Social Media.

Farnworth, M., Thornberry, T. P., Krohn, M. D., Lizotte, A. J. (1994a). Measurement in the Study of Class and Delinquency: Integrating Theory and Research. *Journal of Research in Crime and Delinquency*, *31*, 32–61.

Felson, Marcus. (1987). ROUTINE ACTIVITIES AND CRIME PREVENTION IN THE DEVELOPINGMETROPOLIS. *Criminology*, *25*, 911–932.

Felson, Marcus. (1995). Those who discourage crime. *Crime and Place*, *4*, 53–66.

Felson, MARCUS. (1998). Routine activities and crime prevention in the developing metropolis. *Criminology Theory Reader. Stuart Henry and Werner Einstadter.(Eds.) NY University Press: NY*.

Felson, Marcus. (2002). *Crime and everyday life Thousand Oaks: Sage*.

Felson, Marcus. (2006). *Crime and Nature*. SAGE.

Felson, Marcus. (2013). Routine activity approach. In *Environmental criminology and crime analysis* (pp. 92–99). Willan.

Felson, Marcus, Cohen, L. E. (1980). Human ecology and crime: A routine activity approach. *Human Ecology*, *8*, 389–406.

Felson, Marcus, Poulsen, E. (2003a). Simple indicators of crime by time of day. *International Journal of Forecasting*, *19*, 595–601.

Felson, Marcus, Wortley, R. (2008). Routine activity approach. *Environmental Criminology and Crime Analysis*, 70–77.

Ferrara, E., Yang, Z. (2015). Measuring Emotional Contagion in Social Media. *PLOS ONE*, *10*(11), e0142390.

Ferraro, K. F., Grange, R. L. (1987). The Measurement of Fear of Crime*. *Sociological Inquiry*, *57*, 70–97.

Finkelhor, D., Asdigian, N. L. (1996). Risk Factors for Youth Victimization: Beyond a Lifestyles/Routine Activities Theory Approach. *Violence and Victims*, *11*, 3–19.

Fisher, B. S., Sloan, J. J., Cullen, F. T., Lu, C. (1998). Crime in the Ivory Tower: The Level and Sources of Student Victimization*. *Criminology*, *36*, 671–710.

Fortes, P. R. (2013). The law relating to Brazilian sports fans: An introduction for a British audience. *ESLJ*, *11*, 1.

Fox, K. A., Lane, J., Akers, R. L. (2010). Do perceptions of neighborhood disorganization predict crime or victimization" An examination of gang member versus non-gang member jail inmates. *Journal of Criminal Justice*, *38*, 720–729.

Fraser, G., Kilbride, P. L. (1980). Child abuse and neglect—Rare, but perhaps increasing, phenomena among the Samia of Kenya. *Child Abuse Neglect*, *4*, 227–232.

Friedkin, N. E. (1993). Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study. *American Sociological Review*, *58*(6), 861–872.

Frijters, P., Johnston, D. W., Shields, M. A. (2014). THE EFFECT OF MENTAL HEALTH ON EMPLOYMENT: EVIDENCE FROM AUSTRALIAN PANEL DATA. *Health Economics*, *23*(9), 1058–1071.

Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*, 137–144.

Gangadharbatla, H. (2008). Facebook me: Collective self-esteem, need to belong, and internet self-efficacy as predictors of the iGeneration's attitudes toward social networking sites. *Journal of Interactive Advertising*, *8*, 5–15.

Garland, D. (1999). The Commonplace and the Catastrophic: Interpretations of Crime in Late Modernity. *Theoretical Criminology*, *3*, 353–364.

Garofalo, J. (1981). The fear of crime: Causes and consequences. *J. Crim. L. Criminology*, *72*, 839.

Gartner, I. T. (2014). Glossary.(2013). *Big Data*.

Gerhardt, B., Griffin, K., Klemann, R. (2012a). Unlocking value in the fragmented world of big data analytics. *Cisco Internet Business Solutions Group*.

Gerhardt, B., Griffin, K., Klemann, R. (2012b). Unlocking value in the fragmented world of big data analytics. *Cisco Internet Business Solutions Group*, *7*.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014.

Glynn, T. J. (1981). Psychological Sense of Community: Measurement and Application. *Human Relations*, *34*(9), 789–818.

Golbeck, J., Robles, C., Turner, K. (2011). Predicting personality with social media. *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 253–262. ACM.

Gottfredson, M. R. (1981). On the etiology of criminal victimization. *J. Crim. L. Criminology*, *72*, 714.

Graham, K., Homel, R. (2008). Raising the bar: Preventing aggression in and around bars, clubs and pubs. *UK: Willan Publishing, London*.

Graham, K., Homel, R. (2012). *Raising the bar*. Willan.

Graham, K., Tremblay, P. F., Wells, S., Pernanen, K., Purcell, J., Jelley, J. (2006). Harm, intent, and the nature of aggressive behavior: Measuring naturally occurring aggression in barroom settings. *Assessment*, *13*, 280–296.

Grandey, A. A. (2000). Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, *5*, 95.

Gray, E., Jackson, J., Farrall, S. (2010). Feelings and functions in the fear of crime: Applying a new approach to victimisation insecurity. *The British Journal of Criminology*, *51*, 75–94.

Greenberg, S. W., Rohe, W. M., Williams, J. R. (1985). *Informal citizen action and crime prevention at the neighborhood level: Synthesis and assessment of the research*. US Department of Justice, National Institute of Justice.

Groff, E. (2006). *Exploring the Geography of Routine Activity Theory: A Spatio-Temporal Test Using Street Robbery*.

Groff, E. (2011). Exploring 'near': Characterizing the spatial extent of drinking place influence on crime. *Australian New Zealand Journal of Criminology*, *44*, 156–179.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49.

Hansen, J.-I. C., Sackett, S. A. (1993). Agreement between college major and vocational interests for female athlete and non-athlete college students. *Journal of Vocational Behavior*, *43*, 298–309.

Harackiewicz, J. M., Hulleman, C. S. (2010). The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass*, *4*, 42–52.

Harcourt, B. E., Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city social experiment. *The University of Chicago Law Review*, 271–320.

Harford, T. (2014, March 28). Big data: Are we making a big mistake" Retrieved 17 September 2019, from Financial Times website

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*, 260–271.

Hawley, A. H. (1950). *Ecology.* Retrieved from http://www.popline.org

Hindelang, M. J., Gottfredson, M. R., Garofalo, J. (1978). *Victims of personal crime: An empirical foundation for a theory of personal victimization*.

Hipp, J. R. (2007). Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point. *American Sociological Review*, *72*(5), 659–680. https://doi.org/10.1177/000312240707200501

Hirsh, J. B., Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, *43*, 524–527.

Homel, R., Tomsen, S., Thommeny, J. (1992). Public drinking and violence: Not just an alcohol problem. *Journal of Drug Issues*, *22*, 679–697.

Hong, L., Davison, B. D. (2010). Empirical study of topic modeling in twitter. *Proceedings of the First Workshop on Social Media Analytics*, 80–88. acm.

Hughes, D. J., Rowe, M., Batey, M., Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2), 561-569.

IBM CEO Ginni Rometty believes big data and social will change everything – how about other CEOs" - Top-Rated Futurist Keynote Speaker - Andrew Grill. (2013, March 11). https://andrew.london/2013/03/ibm-ceo-ginni-rometty-believes-big-data-and-social-will-change-everything-how-about-other-ceos/

III, F. P. W., McShane, M. D. (2015). *Criminology Theory: Selected Classic Readings*. Routledge.

Jackson, J., Gray, E. (2009). Functional fear and public insecurities about crime. *The British Journal of Criminology*, *50*, 1–22.

Jacobs, J. (1961). *The death and life of American cities*.

Jang, S. J., Agnew, R. (2015). Strain Theories and Crime. In *International Encyclopedia of the Social Behavioral Sciences* (pp. 495–500).

Jansen, B. J., Zhang, M., Sobel, K., Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169–2188.

Java, A., Song, X., Finin, T., Tseng, B. (2007). Why we twitter: Understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, 56–65. ACM.

Jay, T. (1999). *Why We Curse: A neuro-psycho-social theory of speech*. John Benjamins Publishing.

Jay, T. (2009). The Utility and Ubiquity of Taboo Words. *Perspectives on Psychological Science*, *4*, 153–161. https://doi.org/10.1111/j.1745-6924.2009.01115.x

Jean, P. K. B. S. (2007). *Pockets of Crime: Broken Windows, Collective Efficacy, and the Criminal Point of View*. University of Chicago Press.

Jeffery, C. R. (1971). *Crime Prevention Through Environmental Design*. SAGE Publications.

Jensen, G. F., Brownfield, D. (1986). Gender, Lifestyles, and Victimization: Beyond Routine Activity. *Violence and Victims*, *1*, 85–99.

Jin, S.-A. A. (2013). Peeling back the multiple layers of Twitter's private disclosure onion: The roles of virtual identity discrepancy and personality traits in communication privacy management on Twitter. *New Media Society*, *15*(6), 813–833.

John, O. P., Donahue, E. M., Kentle, R. L. (1991). *The big five inventory—Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Johnson, S. D. (2010). A brief history of the analysis of crime concentration. *European Journal of Applied Mathematics*, *21*(4–5), 349–370.

Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., Townsley, M. (2007). Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization. *Journal of Quantitative Criminology*, *23*, 201–219. https://doi.org/10.1007/s10940-007-9025-3

Johnson, S. D., Bowers, K. J. (2004). The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting. *European Journal of Criminology*, *1*, 237–255. https://doi.org/10.1177/1477370804041252

Jorgenson, A., Rice, J., Clark, B. (2012). Assessing the temporal and regional differences in the relationships between infant and child mortality and urban slum prevalence in less developed countries, 1990–2005. *Urban Studies*, *49*(16), 3495–3512.

Kaisler, S., Armour, F., Espinosa, J. A., Money, W. (2013). Big Data: Issues and Challenges Moving Forward. *2013 46th Hawaii International Conference on System Sciences*, 995–1004. https://doi.org/10.1109/HICSS.2013.645

Katz, J. (1988). *Seductions of crime: Moral and sensual attractions in doing evil*.

Kennedy, L. W., Forde, D. R. (1990). Routine Activities and Crime: An Analysis of Victimization in Canada *. *Criminology*, *28*, 137–152.

Krippendorff, K. (1980). *Validity in content analysis*.

Kubrin, C. E., Weitzer, R. (2003). New directions in social disorganization theory. *Journal of Research in Crime and Delinquency*, *40*, 374–402.

Land, K. C., McCall, P. L., Cohen, L. E. (1990). Structural covariates of homicide rates: Are there any invariances across time and social space" *American Journal of Sociology*, *95*, 922–963.

Laney, D. (2011a). Infonomics: The economics of information and principles of information asset management. *The Fifth MIT Information Quality Industry Symposium. Cambridge*, *9*, 2.

Laney, D. (2011b). The Economics of Information Assets. *The Center for Infonomics.*

Lapan, R. T., Shaughnessy, P., Boggs, K. (1996). Efficacy expectations and vocational interests as mediators between sex and choice of math/science college majors: A longitudinal study. *Journal of Vocational Behavior*, *49*, 277–291.

Larsen-Freeman, D. (2000). *Techniques and Principles in Language Teaching*. Oxford University Press.

Lau, J. H., Collier, N., Baldwin, T. (2012). On-line trend analysis with topic models:# twitter trends detection topic model online. *Proceedings of COLING 2012*, 1519–1534.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publications.

Lazer, D. M., Kennedy, R., King, G., Vespignani, A. (2014). *The Parable of Google Flu: Traps in Big Data Analysis*.

Leak, A. B. (2017). *Applications of new forms of data to demographics* (PhD Thesis). UCL (University College London).

Leavitt Harold, J. (1951). *Some effects of communication patterns on group performance. Journal of abnormal and social psychology vol. 46*.

Leclerc, B., Chiu, Y.-N., Cale, J., Cook, A. (2016). Sexual violence against women through the lens of environmental criminology: Toward the accumulation of evidence-based knowledge and crime prevention. *European Journal on Criminal Policy and Research*, *22*, 593–617.

LeClerc, B., Wortley, R. (2013). *Cognition and crime: Offender decision making and script analyses*. Routledge.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, *18*(5). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/4366

Lewis, S. C., Zamith, R., Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting Electronic Media*, *57*, 34–52.

Lindegaard, M. R., Jacques, S. (2014). Agency as a cause of crime. *Deviant Behavior*, *35*, 85–100.

Liska, A. E., Sanchirico, A., Reed, M. D. (1988). Fear of crime and constrained behavior specifying and estimating a reciprocal effects model. *Social Forces*, *66*, 827–837.

Lohr, S. (2012). The age of big data. *New York Times*, *11*(2012).

Lynch, J. P. (1987). Routine activity and victimization at work. *Journal of Quantitative Criminology*, *3*, 283–300.

Macintyre, S., Homel, R. (1997). Danger on the dance floor: A study of interior design, crowding and aggression in nightclubs. *Policing for Prevention: Reducing Crime, Public Intoxication and Injury*, *7*, 91–113.

Madensen, T. D. (2010). Eck, John E.: Places and the Crime Triangle. In *Encyclopedia of Criminological Theory* (Vols 1–2, pp. 281–284).

Madero-Hernandez, A., Fisher, B. S. (2012). *Routine activity theory*.

Malleson, N., Andresen, M. A. (2015a). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, *4*, 10. https://doi.org/10.1186/s40163-015-0023-8

Malleson, N., Andresen, M. A. (2015b). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, *42*, 112–121. https://doi.org/10.1080/15230406.2014.905756

Malleson, N., Andresen, M. A. (2016). Exploring the impact of ambient population measures on London crime hotspots. *Journal of Criminal Justice*, *46*, 52–63.

Manzi, T., Smith-Bowers, B. (2005). Gated communities as club goods: segregation or social cohesion?. *Housing Studies*, *20*(2), 345-359

Markowitz, F. E., Bellair, P. E., Liska, A. E., Liu, J. (2001). Extending Social Disorganization Theory: Modeling the Relationships Between Cohesion, Disorder, and Fear*. *Criminology*, *39*, 293–319.

Marriott, T. C., Buchanan, T. (2014). The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline. *Computers in Human Behavior*, *32*, 171–177.

Martinez, R., Rosenfeld, R., Mares, D. (2008). Social Disorganization, Drug Market Activity, and Neighborhood Violent Crime. *Urban Affairs Review*, *43*(6), 846–874.

Marwick, A., boyd, danah. (2011). To See and Be Seen: Celebrity Practice on Twitter. *Convergence*, *17*, 139–158.

Marwick, A. E., boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media Society*, *13*, 114–133.

McCown, W., Johnson, J. (1991). Personality and chronic procrastination by university students during an academic examination period. *Personality and Individual Differences*, *12*(5), 413–415.

McCrae, R. R., Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81.

McEnery, T. (2005). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present (Routledge Advances in Corpus Linguistics)*. Routledge.

McNeeley, S., Wilcox, P. (2015). Street Codes, Routine Activities, Neighbourhood Context and Victimization. *The British Journal of Criminology*, *55*(5), 921–943. https://doi.org/10.1093/bjc/azu116

Mehl, M. R., Gosling, S. D., Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862.

Melde, C., Berg, M. T., Esbensen, F.-A. (2016). Fear, social interactions, and violence mitigation. *Justice Quarterly*, *33*, 481–509.

Melde, C., Taylor, T. J., Esbensen, F.-A. (2009). "I got your back": An examination of the protective function of gang membership in adolescence. *Criminology*, *47*, 565–594.

Merton, R. K. (1938). Social structure and anomie. *American Sociological Review*, *3*(5), 672–682.

Merton, R. K. (1983). Florian Znaniecki: A short reminiscence. *Journal of the History of the Behavioral Sciences*, *19*, 123–126.

Messner, S. F., Lu, Z., Zhang, L., Liu, J. (2007). Risks of Criminal Victimization in Contemporary Urban China: An Application of Lifestyle/Routine Activities Theory. *Justice Quarterly*, *24*, 496–522.

Messner, S. F., Mchugh, S., Felson, R. B. (2004). Distinctive Characteristics of Assaults Motivated by Bias*. *Criminology*, *42*, 585–618.

Miethe, T. D., McDowall, D. (1993). Contextual Effects in Models of Criminal Victimization. *Social Forces*, *71*, 741–759.

MIETHE, T. D., MEIER, R. F. (1990). Opportunity, Choice, and Criminal Victimization: A Test of a Theoretical Model. *Journal of Research in Crime and Delinquency*, *27*, 243–266. https://doi.org/10.1177/0022427890027003003

Miethe, T. D., Meier, R. F. (1994). *Crime and its social context: Toward an integrated theory of offenders, victims, and situations*. Suny Press.

Miethe, T. D., Stafford, M. C., Long, J. S. (1987). Social Differentiation in Criminal Victimization: A Test of Routine Activities/Lifestyle Theories. *American Sociological Review*, *52*, 184–194.

Moffatt, M. (2019). What Is Panel Data" The Definition and Relevance of Panel Data in Economic Research. *Thoughtco*.

Miller, J. M., Wright, R. A. (Eds.). (2005). *Encyclopedia of Criminology*. New York: Routledge.

Moore, K., McElroy, J. C. (2012). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior*, *28*, 267–274.

Morenoff, J. D., Sampson, R. J., Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, *39*, 517–558.

Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., " Rillig, M. C. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, *4*(**??**), 3514–3524. https://doi.org/10.1002/ece3.1155

Morstatter, F., Pfeffer, J., Liu, H., Carley, K. M. (2013). Is the Sample Good Enough" Comparing Data from Twitter's Streaming API with Twitter's Firehose. *ArXiv:1306.5204 [Physics]*.

Mustaine, E. E., Tewksbury, R. (1998). Predicting risks of larceny theft victimization: A routine activity analysis using refined lifestyle measures. *Criminology*, *36*, 829–858.

Nagler, J., Tucker, J. A. (2015). Drawing Inferences and Testing Theories with Big Data. *PS, Political Science Politics; Washington*, *48*, 84–88. http://dx.doi.org/10.1017/S1049096514001796

Nasar, J. L., Fisher, B., Grannis, M. (1993). Proximate physical cues to fear of crime. *Landscape and Urban Planning*, *26*(1–4), 161–178.

Naslund, J. A., Grande, S. W., Aschbrenner, K. A., Elwyn, G. (2014). Naturally occurring peer support through social media: The experiences of individuals with severe mental illness using YouTube. *PLOS One*, *9*(**??**), e110171.

Naveed, N., Gottron, T., Kunegis, J., Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on twitter. *Proceedings of the 3rd International Web Science Conference*, 8. ACM.

Newman, G. R., Freilich, J. D. (2013). Extending the reach of situational crime prevention. In *The Reasoning Criminologist* (pp. 234–247). Routledge.

Newman, O. (1972). *Defensible space; crime prevention through urban design*. Macmillan.

Newman, O. (1976). *Design guidelines for creating defensible space.*

Newton, A., Felson, M. (2015). Editorial: Crime patterns in time and space: the dynamics of crime opportunities in urban areas. *Crime Science*, *4*, 11. https://doi.org/10.1186/s40163-015-0025-6

NYC Annual Report 2018-19. (n.d.). Retrieved 20 August 2019, from https://indd.adobe.com/view/fcc4cd 7386-4b52-a39b-c401266a137f

NYC Visitation Statistics. (n.d.). Retrieved 24 March 2017, from Nycgo.com website: http://www.nycgo.com/research/nyc-statistics-page

Oberman, J., Johnson, K. (2015). Broken Windows: Restoring Social Order or Damaging and Depleting New York's Poor Communities of Color Symposium: The Underbelly of the Beast: Misdemeanor Practice in the Era of Broken Windows and Saturation Policing: Introduction. *Cardozo Law Review*, , 931–954.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Fourth International AAAI Conference on Weblogs and Social Media*.

Osgood, D. W., McMorris, B. J., Potenza, M. T. (2002a). Analyzing multiple-item measures of crime and deviance I: Item response theory scaling. *Journal of Quantitative Criminology*, *18*, 267–296.

Osgood, D. W., McMorris, B. J., Potenza, M. T. (2002b). Analyzing Multiple-Item Measures of Crime and Deviance I: Item Response Theory Scaling. *Journal of Quantitative Criminology*, *18*, 267–296.

Outlaw, M., Ruback, B., Britt, C. (2002a). Repeat and multiple victimizations: The role of individual and contextual factors. *Violence and Victims*, *17*, 187.

Outlaw, M., Ruback, B., Britt, C. (2002b). Repeat and Multiple Victimizations: The Role of Individual and Contextual Factors. *Violence and Victims*, *17*, 187–204. https://doi.org/10.1891/vivi.17.2.187.33648

Paoline III, E. A. (2003). Taking stock: Toward a richer understanding of police culture. *Journal of Criminal Justice*, *31*, 199–214.

Park, R. E., Burgess, E. W., McKenzie, R. D. (1984). *The City*. University of Chicago Press.

Parmelee, J. H. (2014). The agenda-building function of political tweets. *New Media Society*, *16*, 434–450.

Pattillo-McCoy, M. (1999). *Black picket fences*. Chicago: University of Chicago Press.

Cooper, P, (2014). Big data vs. crowdsourcing: The battle for the future. Retrieved 2 March 2017, from IT Pro Portal website.

Pekrun, R. (2009). Emotions at school *Handbook of motivation at school*, 575-604.

Pennebaker, J. W., King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296.

Perkins, D. D., Florin, P., Rich, R. C., Wandersman, A., Chavis, D. M. (1990). Participation and the social and physical environment of residential blocks: Crime and community context. *American Journal of Community Psychology*, *18*, 83–115. https://doi.org/10.1007/BF00922690

Perry, W. L., McInnis, B., Price, C. C., Smith, S., Hollywood, J. S. (2013). Predictive Policing [Product Page].

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*, 344–350.

Police powers and procedures England and Wales statistics. (n.d.). Retrieved 17 September 2019, from GOV.UK website.

Preoţiuc-Pietro, D., Liu, Y., Hopkins, D., Ungar, L. (2017). Beyond binary labels: Political ideology prediction of twitter users. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 729–740.

Preoţiuc-Pietro, D., Volkova, S., Lampos, V., Bachrach, Y., Aletras, N. (2015). Studying user income through language, behaviour and affect in social media. *PloS One*, *10*(**??**), e0138717.

Priks, M. (2010). Does frustration lead to violence" Evidence from the Swedish hooligan scene. *Kyklos*, *63*, 450–460.

Principles of Change: How Big Data is Shifting Paradigms One Industry at a Time | IBM Big Data Analytics Hub. (n.d.). Retrieved 27 September 2019, from https://www.ibmbigdatahub.com/blog/principles-change-how-big-data-shifting-paradigms-one-industry-time

Raamkumar, A. S., Erdt, M., Vijayakumar, H., Rasmussen, E., Theng, Y. L. (2018). Understanding the Twitter usage of humanities and social sciences academic journals. *Proceedings of the Association for Information Science and Technology*, *55*, 430–439. https://doi.org/10.1002/pra2.2018.14505501047

Raphael, S., Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, *44*, 259–283.

Research Analysis | NYC Company. (n.d.). Retrieved 20 August 2019, from https://business.nycgo.com/research/

Reynald, D. M. (2009). Guardianship in action: Developing a new tool for measurement. *Crime Prevention and Community Safety*, *11*, 1–20.

Reynald, D. M. (2010). Guardians on Guardianship: Factors Affecting the Willingness to Supervise, the Ability to Detect Potential Offenders, and the Willingness to Intervene. *Journal of Research in Crime and Delinquency*, *47*, 358–390. https://doi.org/10.1177/0022427810365904

Reynald, D. M. (2011). *Guarding Against Crime: Measuring Guardianship Within Routine Activity Theory*. Ashgate Publishing, Ltd.

Reynald, D. M., Elffers, H. (2009). The future of Newman's Defensible Space Theory: Linking Defensible Space and the routine activities of place. *European Journal of Criminology*, *6*, 25–46.

Ristea, A., Leitner, M., Martin, A. A. (2018). Opinion mining from Twitter and spatial crime distribution for hockey events in Vancouver. *Proceedings of the 21th AGILE International Conference on Geographic Information Science, Lund, Sweden*, 12–15.

Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., Harabagiu, S. M. (2012). EmpaTweet: Annotating and Detecting Emotions on Twitter. *Lrec*, *12*, 3806–3813. Citeseer.

Robinson, J. B., Lawton, B. A., Taylor, R. B., Perkins, D. D. (2003). Multilevel longitudinal impacts of incivilities: Fear of crime, expected safety, and block satisfaction. *Journal of Quantitative Criminology*, *19*, 237–274.

Roncek, D. W. (1981). Dangerous places: Crime and residential environment. *Social Forces*, *60*, 74–96.

Roncek, D. W., Maier, P. A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots". *Criminology*, *29*, 725–753.

Rountree, P. W., Land, K. C., Miethe, T. D. (1994). Macro-Micro Integration in the Study of Victimization: A Hierarchical Logistic Model Analysis Across Seattle Neighborhoods*. *Criminology*, *32*, 387–414.

Rountree, P. W., Warner, B. D. (1999). Social Ties and Crime: Is the Relationship Gendered"*. *Criminology*, *37*, 789–814. https://doi.org/10.1111/j.1745-9125.1999.tb00505.x

Routley, N. (2018). *Animation: The Population Pulse of a Manhattan Workday*. Visual Capitalist. https://www.visualcapitalist.com/population-pulse-manhattan/

Sacco, V. F. (2004). [Review of *Review of Criminal Circumstance: A Dynamic Multicontextual Criminal Opportunity Theory*, by P. Wilcox, K. C. Land, S. A. Hunt]. *The Canadian Journal of Sociology / Cahiers Canadiens de Sociologie*, *29*, 160–162. https://doi.org/10.2307/3341955

Sagiroglu, S., Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47.

Salzberg, S. (n.d.). Why Google Flu Is A Failure. Retrieved 17 September 2019, from Forbes website.

Sampson, R. J. (1991). Linking the micro-and macrolevel dimensions of community social organization. *Social Forces*, 43–64.

Sampson, R. J., Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology*, *94*, 774–802.

Sampson, R. J., Lauritsen, J. L. (1990). Deviant Lifestyles, Proximity to Crime, and the Offender-Victim Link in Personal Violence. *Journal of Research in Crime and Delinquency*, *27*, 110–139.

Sampson, R. J., Morenoff, J. D., Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review*, 633–660.

Sampson, R. J., Raudenbush, S. W. (1999). Systematic social observation of public spaces: A new look at disorder in urban Neighborhoods 1. *American Journal of Sociology*, *105*, 603–651.

Sampson, R. J., Raudenbush, S. W., Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*(5328), 918–924.

Sampson, R. J., Wooldredge, J. D. (1987). Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization. *Journal of Quantitative Criminology*, *3*, 371–393. https://doi.org/10.1007/BF01066837

Smolan, S. (2014). *The human fact of big data*.

Schiefele, U., Krapp, A., Winteler, A. (1992). *Interest as a predictor of academic achievement: A meta-analysis of research.*

Schneider, S. A. (2012). *Big data: Big challenge, big opportunity*.

Schreck, C. J., Stewart, E. A., Fisher, B. S. (2006). Self-control, Victimization, and their Influence on Risky Lifestyles: A Longitudinal Analysis Using Panel Data. *Journal of Quantitative Criminology*, *22*, 319–340.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., " Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Shalvi, S., Van Gelder, J.-L., Van der Schalk, J. (2013). Emotional justifi cations for unethical behaviour. In *Affect and cognition in criminal decision making* (pp. 197–210). Routledge.

Shannon, C. E., Weaver, W. (1949). A mathematical model of communication. *Urbana, IL: University of Illinois Press*, *11*.

Shaw, C. R., McKay, H. D. (1942). *Juvenile delinquency and urban areas.* Retrieved from http://psycnet.apa.org/psycinfo/1943-00271-000

Sheldon, A. L. (1969). Equitability indices: Dependence on the species count. *Ecology*, *50*, 466–467.

Sherman, L. W. (1995). Hot spots of crime and criminal careers of places. *Crime and Place*, *4*, 35–52.

Sherman, L. W., Gartin, P. R., Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, *27*, 27–56.

Silver, N. (2012). *The signal and the noise: The art and science of prediction*. Penguin UK.

Singer, S. I. (1981). Homogeneous Victim-Offender Populations: A Review and Some Research Implications. *The Journal of Criminal Law and Criminology (1973-)*, *72*, 779–788. https://doi.org/10.2307/1143015

Singh, J., Singla, V. (2015). Big data: Tools and technologies in big data. *International Journal of Computer Applications*, *112*(15).

Singh, S., Singh, N. (2012). Big Data analytics. *2012 International Conference on Communication, Information Computing Technology (ICCICT)*, 1–4. https://doi.org/10.1109/ICCICT.2012.6398180

Skogan, W. G. (1990). *Disorder and decline: Crime and the spiral of decay in American cities*.

Sloan, L., 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*, 10(11).

Small, H., Kasianovitz, K., Blanford, R., Celaya, I. (2012). What Your Tweets Tell Us About You: Identity, Ownership and Privacy of Twitter Data. *International Journal of Digital Curation*, *7*, 174–197.

Smith, C. A., Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813.

Solymosi, R., Bowers, K. (2018). The role of innovative data collection methods in advancing criminological understanding. *The Oxford Handbook of Environmental Criminology*, *210*, 210–237.

Solymosi, R., Bowers, K. J., Fujiyama, T. (2018). Crowdsourcing Subjective Perceptions of Neighbourhood Disorder: Interpreting Bias in Open Data. *The British Journal of Criminology*, *58*, 944–967. https://doi.org/10.1093/bjc/azx048

Spano, R., Nagy, S. (2005). Social Guardianship and Social Isolation: An Application and Extension of Lifestyle/Routine Activities Theory to Rural Adolescents*. *Rural Sociology*, *70*, 414–437.

Spellerberg, I. F., Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'Index. *Global Ecology and Biogeography*, *12*, 177–179.

St Jean, P. K. (2007). Pockets of Crime: Broken Windows. *Collective Efficacy, and The*.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research Evaluation*, *7*(**??**), 137–146.

Street Robbery | Center for Problem-Oriented Policing. (n.d.). Retrieved 17 September 2019, from https://popcenter.asu.edu/content/street-robbery-0

Stuntz, W. J. (2001). Local policing after the terror. *Yale LJ*, *111*, 2137.

Su, R., Rounds, J., Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*(6), 859.

Sul, H., Dennis, A. R., Yuan, L. I. (2014). Trading on twitter: The financial information content of emotion in social media. *2014 47th Hawaii International Conference on System Sciences*, 806–815. IEEE.

Sul, H. K., Dennis, A. R., Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*, 454–488.

Sullivan, C. J., McGloin, J. M. (2014). Looking Back to Move Forward: Some Thoughts on Measuring Crime and Delinquency over the Past 50 Years. *Journal of Research in Crime and Delinquency*, *51*, 445–466.

Sutherland, A., Brunton-Smith, I., Jackson, J. (2013). Collective efficacy, deprivation and violence in London. *British Journal of Criminology*, *53*(6), 1050–1074.

Swaray, R. (2006). Economic methodology and evaluations: The costs and benefits of criminal justice interventions. *Reducing Crime: The Effectiveness of Criminal Justice Interventions*, 143–162.

Tang, D., Wei, F., Qin, B., Liu, T., Zhou, M. (2014). Coooolll: A deep learning system for twitter sentiment classification. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 208–212.

Tausczik, Y. R., Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54.

Taylor, R. B. (1997). Social Order and Disorder of Street Blocks and Neighborhoods: Ecology, Microecology, and the Systemic Model of Social Disorganization. *Journal of Research in Crime and Delinquency*, *34*, 113–155. https://doi.org/10.1177/0022427897034001006

Taylor, R. B. (2001). Breaking away from broken windows: Evidence from Baltimore neighborhoods and the nationwide fight against crime, grime, fear and decline. *New York, NY: Westvie*.

Taylor, R. B., Gottfredson, S. D., Brower, S. (1984). Block crime and fear: Defensible space, local social ties, and territorial functioning. *Journal of Research in Crime and Delinquency*, *21*, 303–331.

Tewksbury, R., Mustaine, E. E., Covington, M. (2010). Offender Presence, Available Victims, Social Disorganization and Sex Offense Rates. *American Journal of Criminal Justice*, *35*(1–2), 1–14. https://doi.org/10.1007/s12103-010-9070-6

The Demographics of Social Media Users in 2016. (2016). Retrieved 22 March 2017, from Think Digital First website.

The Oxford Handbook of Criminological Theory. (2012).

Thelwall, M. (2008). Fk yea I swear: Cursing and gender in MySpace. *Corpora*, *3*, 83–107.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558.

Thelwall, M., Wilkinson, D., Uppal, S. (2010). Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology*, *61*, 190–199. https://doi.org/10.1002/asi.21180

Tillyer, M. S. (2010). Wilcox, Pamela, Kenneth C. Land, and Scott A. Hunt: Multicontextual opportunity Theory. In *Encyclopedia of Criminological Theory* (Vols 1–2, pp. 1009–1011). https://doi.org/10.4135/9781412959193

Titscher, S., Meyer, M., Wodak, R., Vetter, E. (2000). *Methods of text and discourse analysis: In search of meaning*. Sage.

Tompson, Lisa, Bowers, Kate. (2013). A Stab in the Dark": A Research Note on Temporal Patterns of Street Robbery.

Tone check your communication with Watson. (2016). Retrieved 26 August 2019, from Watson website.

Topalli, V., Wright, R. (2014). Affect and the dynamic foreground of predatory street crime. *Affect and Cognition in Criminal Decision Making*, 42.

Trudgill, P., Anderson, L. (1990). *Bad Language*. Cambridge, USA: Penguin books.

Tseloni, A., Wittebrood, K., Farrell, G., Pease, K. (2004). Burglary Victimization in England and Wales, the United States and the NetherlandsA Cross-National Comparative Test of Routine Activities and Lifestyle Theories. *The British Journal of Criminology*, *44*, 66–91. https://doi.org/10.1093/bjc/44.1.66

Tumasjan, A., Sprenger, T. O., Sandner, P. G., Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Fourth International AAAI Conference on Weblogs and Social Media*.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., Welpe, I. M. (2011). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, *29*, 402–418.

Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.

Tyler, T. R. (2011). Trust and legitimacy: Policing in the USA and Europe. *European Journal of Criminology*, *8*(4), 254–266.

Uittenbogaard, A. C. (2014). Assessing guardianship opportunities at underground stations. *Security Journal*, *27*, 147–163.

Using Twitter as a data source: An overview of current social media research tools | Impact of Social Sciences. (n.d.).

VCloudNews. (n.d.). Every Day Big Data Statistics – 2.5 Quintillion Bytes of Data Created Daily.

Wang, X., Yu, C., Wei, Y. (2012). Social media peer communication and impacts on purchase intentions: A consumer socialization framework. *Journal of Interactive Marketing*, *26*, 198–208.

Warner, B. D., Rountree, P. W. (1997). Local social ties in a community and crime model: Questioning the systemic nature of informal social control. *Social Problems*, *44*, 520–536.

Weber, R. P. (1990). *Basic content analysis*. Sage.

Weisburd, D. (2015). Small Worlds of Crime and Criminal Justice Interventions: Discovering Crime Hot Spots. In *Envisioning Criminology* (pp. 261–267). Springer.

Weisburd, D., Bushway, S., Lum, C., Yang, S.-M. (2004). Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle*. *Criminology*, *42*(2), 283–322.

Weisburd, D., Piquero, A. R. (2008). How well do criminologists explain crime" Statistical modeling in published studies. *Crime and Justice*, *37*(1), 453–502.

Weisburd, D., Telep, C. W., Braga, A. A., Cave, B., Bowers, K., Eck, J. E., others. (2016). *Place matters*. Cambridge University Press.

Weisburd, D., White, C. (2019). Hot spots of crime are not just hot spots of crime: Examining health outcomes at street segments. *Journal of Contemporary Criminal Justice*, *35*(2), 142–160.

Wells, S., Graham, K., Purcell, J. (2009). Policy implications of the widespread practice of 'pre-drinking'or 'pre-gaming'before going to public drinking establishments—Are current prevention strategies backfiring" *Addiction*, *104*(1), 4–9.

Weng, J., Lim, E.-P., Jiang, J., He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 261–270. ACM.

Wikstrom, P.-O. H. (2009). Questions of perception and reality. *The British Journal of Sociology*, *60*(1), 59–63.

Wilcox, P., Madensen, T. D., Tillyer, M. S. (2007). Guardianship in Context: Implications for Burglary Victimization Risk and Prevention*. *Criminology*, *45*(4), 771–803. https://doi.org/10.1111/j.1745-9125.2007.00094.x

Wild, B., Erb, M., Bartels, M. (2001). Are emotions contagious" Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research*, *102*(2), 109–124.

Wilson, J. Q., Kelling, G. L. (1982). Broken windows. *Critical Issues in Policing: Contemporary Readings*, 395–407.

Wilson, W. J. (1987). The truly disadvantaged Chicago. *IL: University of Chicago*.

Wooldredge, J. D., Cullen, F. T., Latessa, E. J. (1992). Research note victimization in the workplace: A test of routine activities theory. *Justice Quarterly*, *9*(2), 325–335. https://doi.org/10.1080/07418829200091391

Wortley, R. (2001). A classification of techniques for controlling situational precipitators of crime. *Security Journal*, *14*, 63–82.

Wortley, R. (2006). *Sexual Abuse of Children: Can Situational Prevention Work"*

Wortley, R. (2010). *Critiques of situational crime prevention*. Retrieved from http://discovery.ucl.ac.uk/1301877/1/Wortley_2010_SCP_criticisms.pdf

Wortley, R. (2013). Situational precipitators of crime. In *Environmental criminology and crime analysis* (pp. 70–91). Willan.

Wortley, R. K. (2015). Science and child sexual abuse: Navigating the pathway between emotion and objectivity. *Crime Science*, *4*, 18.

Wortley, R., Mazerolle, L. (Eds.). (2008). *Environmental Criminology and Crime Analysis* (1st edition). Cullompton, UK; Portland, Or: Willan.

Wortley, R., Townsley, M. (2016). *Environmental criminology and crime analysis*. Taylor Francis.

Wright, R. T., Decker, S. (1994). *Burglars on the Job*. Northeastern University Press Boston, MA.

Witten, K., McCreanor, T., Kearns, R., Ramasubramanian, L. (2001). The impacts of a school closure on neighbourhood social cohesion: narratives from Invercargill, New Zealand. *Health Place*, 7(4), 307-317.

Yau, N. (2013). Pizza Place Geography. FlowingData. https://flowingdata.com/2013/10/14/pizza-place-geography/

Yee, N., Ducheneaut, N., Nelson, L., Likarish, P. (2011). Introverted Elves Conscientious Gnomes: The Expression of Personality in World of Warcraft. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 753–762. https://doi.org/10.1145/1978942.1979052

Young, S. D., Rivers, C., Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive Medicine*, *63*, 112–115. https://doi.org/10.1016/j.ypmed.2014.01.024

Zafirovski, M. (2013). Beneath rational choice: Elements of 'irrational choice theory'. *Current Sociology*, *61*(1), 3–21.

Zhang, L., Messner, S. F., Liu, J. (2007). A multilevel analysis of the risk of household burglary in the city of Tianjin, China *The British Journal of Criminology*, 47(6), 918–937.