CCTV as Smart Sensor Networks

Giles Oatley, Tom Crick, Dee Bolt
Department of Computing, Cardiff Metropolitan University, UK
{goatley,tcrick,dbolt}@cardiffmet.ac.uk

***Abstract*—With the emergence of so-called "smart CCTV" being able to recognise the precursors for disorder and civil disobedience, we present a study into using available CCTV networks augmented with social media datasets.**

We examine the existing CCTV infrastructure in the UK, and use an agent-based simulation to model interactions between people based on friendship networks and features derived from their social media usage, proposing a novel algorithm for detection of psychopathy. Finally, we explore the frequency of crimes occurring within CCTV viewsheds using available UK police crime datasets to illustrate the current limitations of the CCTV infrastructure, as well as the potential ramifications of the stealthy emergence of CCTV networks as the fifth utility in smart cities.

Keywords*—*CCTV, Smart Cities, Sensors, Networks, Crowd Behaviour, Traits, Agent-Based Modelling, Social Media

I. Introduction

In this paper we explore ways to model areas in a landscape, for example a city, in terms of cyber-physical networks and communities who frequently communicate through social networks with location-aware information. We present a survey of data sources that can be of potential value in this exploration, and conduct some proof-of-principle experiments. We are interested in exploring how closed-circuit television (CCTV) can be combined with analysis of large-scale social media datasets to determine the general "mood" of a crowd, and to explore the potential and limitations of this hybrid approach to behaviour modelling. To this end we present a background to the CCTV infrastructure in the UK, including the numbers and quality of the cameras networks involved. We also provide a review of the latest research deriving behaviours and traits from social media datasets.

Smart cities [1], [2] are an emerging research, policy and planning challenge, with the potential to generate huge amounts of data [3]; of particular interest to this study, big social data [4]. We want to see what is happening in a city by exploring cyber-physical crowd behaviour in an agent-based system [5], using data from location-based social networks.

We are interested in the relationship between digital footprint and behaviour and personality [6], [7]. A wide range of pervasive and often publicly available datasets encompassing digital footprints, such as social media activity, can be used to infer personality [8], [9]. Big social data offers the potential for new insights into human behaviour and development of robust models capable of describing individuals and societies [10]. Academic research in image or video analysis includes promising studies on YouTube videos for classification of specific behaviours and indicators of personality traits [11]. This work uses crowdsourced impressions, social attention, and audiovisual behavioural analysis on slices of conversational 'vlogs' (video blogs) extracted from YouTube.

We are also interested in profiling insider threats, a situation where legitimate access is used criminally, often in situations where it is known that actions are scrutinised closely (by amongst others, machine learning algorithms). How best then to learn a profile of an individual, so that criminal behaviour, which is assumed to be different in some way to normal operating behaviour, can be detected. This has been of particular interest in the UK, with high-profile incidents such as the riots in London that spread across the UK in 2011 [12] and the Woolwich terrorist attack in London in 2013 [13]. The data footprint will change significantly based on either time, location or role, as the individual legitimately passes through their range of activities, for instance an operator accessing a computer terminal at one location in the morning and another in the afternoon. Likewise the data footprint will change according to shifting emotional states, for instance the same operator working on a single terminal differently on different days, perhaps one day performing the 'harder' tasks first, and another day, the 'easier' tasks first. More generally then we have the problem of profiling complex behaviours [6].

The proliferation of CCTV camera networks across urban communities in the UK has received a mixed reception. There has been significant criticism as local authorities have spent £515m on CCTV and associated infrastructure in four years [14], with mixed results. Furthermore, they are often reluctant to reveal how effective they have been for monitoring of crime and anti-social behaviour[1]. Nevertheless, with the increased focus on the development (and associated e-infrastructure) of smart cities, there is often significant funding available for "smart" CCTV systems, claiming to prevent crime by detecting crowd characteristics indicative of criminal behaviour [15], [16].

CCTV is currently used as a deterrent in the UK and many other countries; in terms of crime prevention, an examination of crime in the viewshed of publicly funded CCTV cameras in Philadelphia, USA, found that the introduction of cameras was associated with a 13% reduction in crime [17]. Research found that while there appears to be a general benefit to the cameras, there were as many sites that showed no benefit of camera presence as there were locations with a positive outcome on crime [18]. A further study in Newark, USA, found that strategically-placed cameras were not any different from randomly-placed cameras at deterring crime within their viewsheds although there were significant improvements to location quotient values for gun shootings and automobile thefts after camera installations [19].

However, we are interested in CCTV as a sensor, for validation of models derived from sources such as social media

[1]For example: http://getthedata.org/questions/158/locations-of-council-operated-cctv-cameras-in-the-uk/

analysis. Over-reliance on sophisticated software products such 'intelligent network products' and geographical techniques such as 'hotspot analysis' can lead to weak critical thinking. Thus, the next-generation (social network) analysis must focus much more intensely on the content of the contacts, on the social context, and on the interpretation of such information [20].

Analysis of Competing Hypotheses (ACH[2]) is an important software tool used for intelligence analysis, and can ameliorate human operators misinterpreting results and preventing cognitive biases in inferences which may contaminate the process and ultimately the decision reached. Numerous data sources and techniques such as statistical profiling are available to the analyst, however we can often exhibit a "confirmatory bias" [21], [22] in that we focus upon data which conforms to the initial ideas formed (e.g. about who is the suspect) and so fail to test them by seeking evidence which contradicts our notions. Other errors of judgment include counterfactual thinking, illusory correlation, false consensus bias, ignoring base rate information, culture/gender biases, group effects and so on. Evaluating the validity, reliability and generalisability of any analyses and output from these data sources requires a knowledge of sampling issues. Base rates can be confidently calculated from large datasets. For an appropriate evaluation of evidence, a comparison of probabilities of the evidence under two different propositions is required, the system will provide this.

II. Social Media and Personality

The work of Schwartz et al. [23] analysed what people say in social media to find distinctive words, phrases, and topics as functions of known attributes of people such as gender, age, location, or psychological characteristics. This can thus be transposed, inferring gender, age and so on, from social media data. The negative implications of these developments are that they can easily be applied to large numbers of people without obtaining their individual consent or even being aware. Commercial companies, governmental institutions, or even ones Facebook friends could use software to infer personality (and other attributes, such as intelligence or sexual orientation) that an individual may not have intended to share [8].

There are now vast numbers of social media sites[3], with a number of attempted categorisations; for example: social networking sites (e.g. Facebook), professional networking (e.g. LinkedIn), microblogging (e.g. Twitter, Tumblr), wiki-based knowledge sharing sites (e.g. Wikipedia), social news/websites of news media (e.g. Huffington Post), forums, mailing lists, newsgroups, community media sites (e.g. YouTube, Flickr, Instagram), social Q&A sites (e.g. Quora, Yahoo Answers), user reviews (e.g. Yelp, Amazon, TripAdvisor), social curation sites (e.g., Reddit, Slashdot, Pinterest), location-based social networks (e.g., Foursquare), etc.

Caverlee et al. [24] discuss the knowledge discovery and data mining stages applied to geospatial data found in social media, developing geo-social intelligence. They describe the geo-social overlay of the physical environment of the planet; consider for example, the overlay of the six billion geotagged tweets[4]. Both they and Stefanidis et al. [25] discuss the need for new systems and techniques to leverage these footprints, as well as the wide range of challenges and opportunities to the geospatial intelligence community in particular.

Cheng, Caverlee & Lee [26] studied microblogging and location, with Twitter also able to be used as a sensor to detect frequent and diverse social and physical events in real-time [27]. While Twitter has been demonstrated to provide insight (and sociologically relevant demographics [28]) into major social and physical events such as riots, celebrity deaths and presidential elections [12], [13], Liang et al. [29] make the point that all is often not what it may seem, for instance many tweets may not a crowd make.

Kamath et al. [30] have studied a geo-tagged hashtag dataset of two billion tweets; although Twitter's hashtags can be used to indicate the topics of tweets, only a small fraction (c.11%) of tweets contain hashtags [31]. There are multiple challenges toward recognition of a sporting event using Twitter, not least that because a multitude of tweets are being generated every second, it is necessary to correctly filter out the important tweets [31]. *TwitterStand* [32] identifies current news topics and clusters the corresponding tweets into news stories, but it is unable to do it in near real-time.

Numerous studies have suggested certain key words and phrases can signal underlying tendencies and that this can form the basis of identifying certain aspects of personality [33]–[37]. Scherer [38] introduced a valuable classification with the following distinctions between emotions, moods, interpersonal stances, attitudes and personality traits:

- *Emotion:* short-lived, for instance being angry, sad, or joyful;
- *Mood:* longer-lived, low-intensity for instance being cheerful or gloomy;
- *Interpersonal stances:* duration linked to specific interaction, for instance friendly or supportive;
- *Attitudes:* long-lived linked to specific people or events, for instance loving and hating;
- *Personality traits:* stable personality dispositions and typical behaviour tendencies, for instance nervous, anxious, or hostile.

A wide corpus of research has been carried out establishing the link between personality and social media [34], [39], [40]; for example, the findings of a 2014 study reveal that by mining a person's Facebook Likes, an algorithm was able to predict a person's personality more accurately than most of their friends and family [41]. Only an individual's spouse came close to matching the algorithm's results. The predictions were based on which articles, videos, artists and other items the person had liked (or interacted with) on Facebook.

By observing the occurrences of words that related to these categories, we can conclude to certain degrees about the holder's psychological state. For instance, we have opinion mining or sentiment analysis at one end of the spectrum, by utilising open-source software such as *SentiWordNet*. At the

[2] http://www2.parc.com/istl/projects/ach/ach.html

[3] This list is by no means exhaustive: http://en.wikipedia.org/wiki/List_of_social_networking_websites

[4] https://www.mapbox.com/blog/twitter-map-every-tweet

other end of the spectrum we have [42] highlighting the use of features from psycholinguistic databases *LIWC* [43] and *MRC* [44] to create a range of statistical models for each of the Five-Factor personality traits [45], [46].

A. *Derived measure of psychopathy*

The Five-Factor model has known limits [47]–[49]: it has been criticised for its limited scope, methodology and the absence of an underlying theory, and attempts to replicate the Five-Factor model in other countries with local dictionaries have succeeded in some countries but not in others [50], [51]. Related to a study which analysed Twitter datasets for signs of psychotic behaviour in respondents, we wanted to investigate aspects of psychopathology, an interest related to insider threat and crime informatics [6], [52]–[54]. Mairesse et al. [42] demonstrated it is possible within various contraints to use psycholinguistic databases such as MRC and LIWC to create models for the Five-Factor model of personality. The text that someone writes, preferably containing as little 'unnatural' text as possible, will potentially reveal aspects of their personality.

We take this further by using Miller and Lynam's correlation [55], [56] between psychopathy and the Five-Factor's to determine the level of psychopathy of someone from their text (see Figure 1). There are of course many constraints and limitations to this. Also, this derived measure of psychopathy has consequences for civil liberties, privacy and possibilities of abuse from profiling in a smart city environment.

Psychopathy is a maladaptive personality disorder characterised by such traits as a lack of remorse, manipulativeness, egocentricity, superficial charm, and shallow affect. Correlates of psychopathy include high rates of both violent and nonviolent offending, violent and nonviolent recidivism, and substance use problems. Miller and Lynam's [55], [56] studied the relation between psychopathy and the Five-Factor dimensions of personality in adolescents from Pittsburgh, USA, confirming the hypothesis that the aspect of psychopathy representing selfishness, callousness, and interpersonal manipulation (Factor 1) is most strongly associated with low *Agreeableness*, whereas the aspect of psychopathy representing impulsivity, instability, and social deviance (Factor 2) is associated with low *Agreeableness*, low *Conscientiousness*, and high *Neuroticism*. This is represented in the following equation, and used in our simulation.

```
Psychopath if (Agreeableness is 'Very low') and
(Conscientiousness is 'Very low') and
(Extraversion is 'High') and
(Neuroticism.ANXIETY is 'Low') and
(Neuroticism.ANGER is 'High') and
(Neuroticism.DEPRESSION is 'Low') and
(Neuroticism.SELF-CONSCIOUSNESS is 'Low') and
(Neuroticism.IMMODERATION is 'High') and
(Neuroticism.VULNERABILITY is 'Low')
```

Figure 1. Derivation of "Psychopath" from Five-Factors

III. Data and Methodology

A. *CCTV data*

Figure 2 shows CCTV placements and viewsheds for the UK city of Leicester. The data was available in longitude/latitude, available through a Freedom of Information (FoI) request to Leicester City Council.

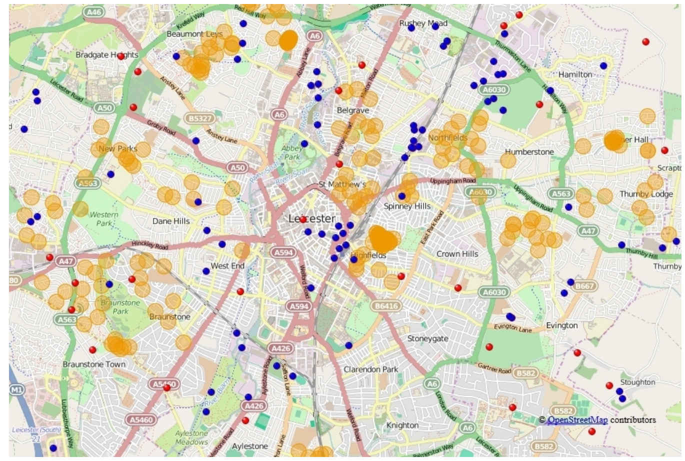

Figure 2. CCTV viewsheds in the city of Leicester, UK

For our purposes, there are several ways that network data, that is, vertices and edges, can be combined with geographical data. We applied two methods, firstly with connectivity (edges) between vertices being derived from viewsheds from individual vertices [57], and secondly, a density calculation built directly into the social network analysis measure of *betweenness* [58]. Our method thus combined line of sight visibility (viewshed analysis) with techniques from social network analysis to investigate this data. At increasing distances different nodes are connected creating a set of networks, which are subsequently described using centrality measures and clustering coefficients. This technique has significant relevance to CCTV cameras and their line of sight visibility, permitting construction of different levels of possible network structures.

B. *Crime data*

The crime data was obtained from the UK's open data repository[5] for crime and policing in England, Wales and Northern Ireland, which offers street-level crime and outcome data for individual police forces and neighbourhood teams. The UKs Home Office published crime data provided by the 43 geographic police forces in England and Wales, the British Transport Police, the Police Service of Northern Ireland and the Ministry of Justice. There are approximately four years of data available, and the classification of crimes includes the following: "*Anti-social behaviour*", "*Bicycle theft*", "*Burglary*", "*Criminal damage and arson*", "*Drugs*", "*Other crime*", "*Other theft*", "*Possession of weapons*", "*Public disorder and weapons*", "*Public order*", "*Robbery*", "*Shoplifting*", "*Theft from the person*", "*Vehicle crime*", "*Violence and sexual offences*", "*Violent crime*".

C. *Social media data*

There are an increasing number of applications, services and frameworks that will allow you to retrieve and analyse social data, some for a fee, some available under academic licenses. We considered the following sources of social media data for our experiments:

[5] http://data.police.uk

- *GNIP*[6] is a commercial company that serves customers in over 40 countries who serve over 95% of the companies in the Fortune 500 with data from numerous social media hosts, including Twitter, Tumblr, Foursquare, YouTube, Reddit, Google+, Facebook and Instagram;

- *Netlytic*[7] is a cloud-based text and social networks analyser that can automatically summarise and discover social networks from online conversations on social media sites, allowing free access for up to three datasets per month from Twitter (tweets matching a user specified query, including location, tweeter and media used), Facebook (posts and replies from public Facebook groups, pages, events, or profiles), Instagram and YouTube (video comments);

- *COSMOS*[8] (Collaborative Online Social Media Observatory) brings together social, computer, political, health, statistical and mathematical scientists to study the methodological, theoretical, empirical and technical dimensions of social media data in social and policy contexts [59]. The open source tool includes many pre-built packages for data mining and analysis, including the ability to plot and view tweet locations on a geo-spatial map.

After having various issues with securing access to the tools and datasets presented above, we opted to use the data from the *myPersonality* project[9]. *myPersonality* started out in June 2007 and collected Facebook data from the *myPersonality* 'app' questionnaires, ending in 2012. Nearly 7.5m people have completed a questionnaire, with more than 40 countries having had 1000 or more participants (but *myPersonality* is only available in English); users are able to retake *myPersonality* tests (for example, one test – the *Big 5* test – has been retaken over 900,000 times), providing longitudinal data. Users are able to rate the personalities of their FB friends, with over 300,000 friend-ratings. About 40% of users have granted *myPersonality* access to the data on their Facebook profiles (this includes their preferences, as expressed by Facebook Likes), providing more than 36,000,000 user-like pairs [60].

The list of variables is extensive[10]; of particular interest to our current study include: demographic and geo-location data (home and current location); religion, political views, and profile about section; psychological profiles; Facebook data; Facebook status updates; and, Facebook social networks [60]. We were primarily interested in the social media data, and the Big Five scores plus their constituent facet scores, for all the users in the dataset. For example: the number of posts on a user's wall (n= 2057455; mean = 122.03; SD = 318.48) – there are 4077428 records in the *Big 5* table, with 3386778 unique respondents, and there are 38330 people for whom both *Big 5* and self-monitoring scores are available. So for our purposes, we have personality traits, text from which we can determine sentiment, relationship data, and a 'triads' database numbering more than 3.5 million entries, which can be used to study dyads by ignoring the third actor. Unfortunately, the data or updates are not geocoded, although there are recorded coordinates for 'home' and occasionally 'current location'. When we attempted to use these values, it drastically reduced the database down to a few hundred compared to the original millions.

D. Agent-based methodology

Geospatial systems have long been paired with agent-based models [61], for instance Crooks [62] with the digital representation and physical environment of cities, along with Chorley, Whitaker & Allen [63] on investigating personality and location-based social networks based on check-in behaviour of volunteer Foursquare users. Crooks and Castle [64] present a list of software and guidelines for the integration of agent-based modelling and geographical information for geospatial simulation; for the agent simulation we used the *PyCX Project*[11] agent-based modelling software.

Our experiments utilised the social media data from the *myPersonality* project; also friends, families, social network analysis measures, personality traits, and our derived psychopathy value. Our aim was to simulate the movement of many users in a city by means of an agent-based model, and utilise the CCTV sensors in some way to gain a measure of 'ground-truth' or something to optimise against. The primary idea was to see how we might be able to represent these factors. Consider an agent, surrounded mainly by unknown people, perhaps some friends or family, a number of happy people, a number of sad people, the agent with their plans and beliefs, desires and intentions; where do they move to? Consider the difficulty of representing meaningful communication between agents in an 'unconstrained' environment, disambiguating (un)natural language, automatic detection of events.

IV. RESULTS

As we discussed previously, we derived a psychopathy score from the *myPersonality* dataset containing Five-Factors and subfactors, but with such strong criteria we found no psychopaths (a good result). We were therefore forced to successively 'weaken' the selection criteria, for instance *Conscientiousness.SELF-EFFICACY* being generalised from '*Very low*' to '*Low*'. Within the data there were approximately 7000 data points with all Five-Factors and subfactors, and there were approximately 3 million data points with just the Five-Factors. From c.7000 data points initially no users were identified. Removing the *Neuroticism* feature entirely, and changing '*Very High*' or '*Very Low*' to '*High*'/'*Low*', nine users were identified. Changing all values to '*Medium*' retrieved 66 users. From the 3 million dataset, which does not include the subfactors, then ignoring these 1640 users were identified.

We analysed all the data available for all crimes in the city of Leicester, with increasing sizes of viewsheds around the placement of CCTV cameras. Figure 4 shows viewsheds of sizes 10m, 100m and 200m, with the stacked bar-chart showing the proportion of crimes in that year which were within the respective viewsheds.

[6]http://gnip.com
[7]https://netlytic.org
[8]http://www.cosmosproject.net/
[9]http://myPersonality.org
[10]http://mypersonality.org/wiki/doku.php?id=list_of_variables_available
[11]http://pycx.sourceforge.net

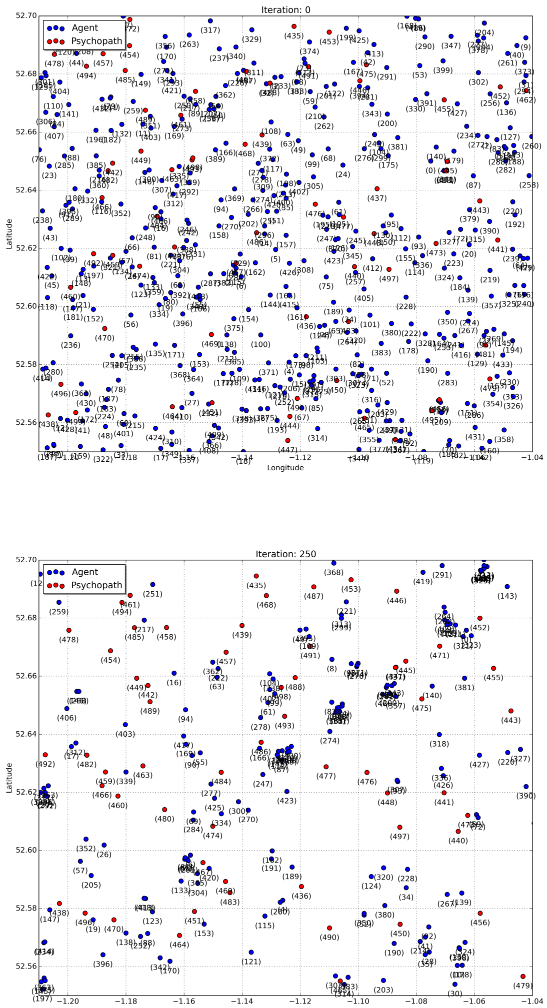

Figure 3. Agent start and finish positions. The data was comprised approximate equal numbers of psychopaths (using very weakened equations) and non-psychopaths. The reason for this was initially to build in other metrics of avoidance of CCTV and other criminal behaviour

V. Discussion and Conclusions

The technical issues surrounding smart CCTV, detecting crowd characteristics indicative of disorder and violence, are those related to image analysis and extracting crowd dynamics using motion estimation techniques. Learning crowd characteristics is a significant task without considering the problems caused by ambient weather conditions, improper viewing angle, installation position, image correction for lens distortion, and so on. A major problem with many CCTV installations is that the wrong lens is chosen, often resulting in people being too small to successfully identify. Unless a camera achieves 'Recognition of a Known Person' it is unlikely to be used by the police in the UK to identify a person for prosecution in a court of law, and the police have stated that over 80% of the video evidence that they collect fails to meet the required standard.

CCTV offers the ability to validate or verify theories about situations believed to be occurring derived from social

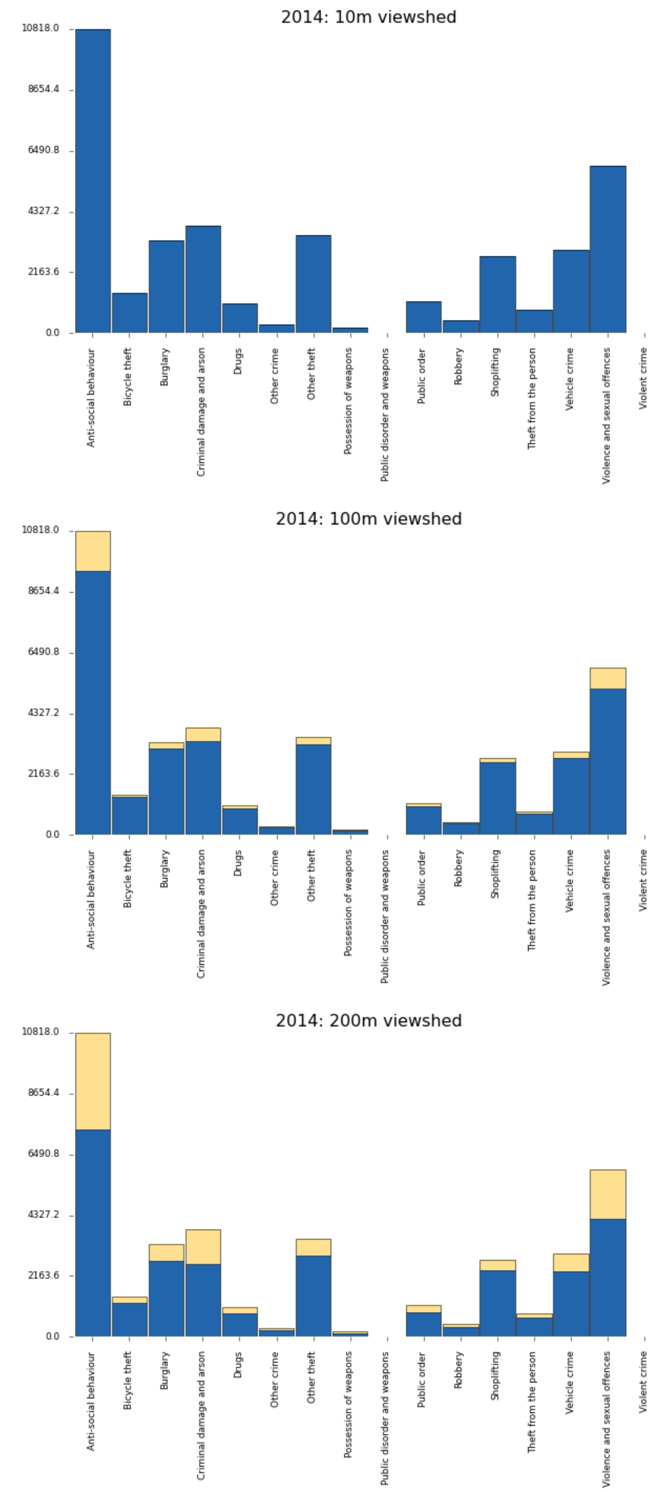

Figure 4. Proportion of crimes within CCTV viewsheds of 10m, 100m, 200m. The features from left to right are the crimes of: "*Anti-social behaviour*", "*Bicycle theft*", "*Burglary*", "*Criminal damage and arson*", "*Drugs*", "*Other crime*", "*Other theft*", "*Possession of weapons*", "*Public disorder and weapons*", "*Public order*", "*Robbery*", "*Shoplifting*", "*Theft from the person*", "*Vehicle crime*", "*Violence and sexual offences*", "*Violent crime*" (full data available from: http://data.police.uk)

media or other digital sources. Therefore our interest in CCTV is as a sensor, that can validate what is happening in the digital representation of a city. The UK Home Office *CCTV Operational Requirements Manual* [65] prescribes the relevant technical specifications. Camera and lens selections need to be chosen and based upon a 1.7 metre person. Detection requires the image of the person occupying at least 10% of picture screen height on the monitor; observation requiring 25% screen height, to follow a group of people such as in a town centre; recognition of a known person requires 50% screen height; and, identification of an unknown person requires 100% of picture screen height on the monitor. To observe a person at 50% ('Recognition of a Known Person') the CCTV can only view an area the width of two car park spaces (4.3m). The most popular lens sold to customers is a 1/3in (3.6mm) lens. This gives a wide angle view of 70%, but can only provide facial recognition of 3.2m so if it is located at a height of 5m it is no real use for recognition.

In general, there are too few cameras with too wide a field-of-view; cameras can view a wide area, or provide a high-level of detail, but not both. Many cameras are set to view an excessively large area (possibly for cost-savings), which makes it impossible to positively identify people at most points within the scene. Also, most cameras installed today have fields of view that are set too wide to allow facial recognition throughout most of their coverage area.

Utilising CCTV as a sensor to accurately model or give feedback on the reality of occurrences in digital space currently has several drawbacks, relating to both placement/position and quality of the equipment to also the complexities of image analysis. Limitations of our simulation include the lack of accurate/genuine geotagged data, and from lack of information about CCTV cameras – the UK city of Leicester is one of the few with accurately recorded positions for a study, but even of these we do not know the actual camera types (and therefore the correct viewshed to use).

With many CCTV systems having sophisticated recording and real-time logging/tracking capabilities, it raises wider questions about civil liberties, privacy and data retention [66] – particularly in light of anti-terrorist legislation in both the US and UK [67] – despite numerous Freedom of Information (FoI) requests to identify the specifics of systems in cities across the UK. While there is significant potential for governments and organisations to promote open and big data (for example, in supporting effective policy-making [68]), we raise the fact that it is now clearly possible, and will in the future become more accurate, that widespread data retention and aggregation could lead to abuses of civil liberties. The recent re-positioning of public-private partnerships in national cyber security strategies, particularly in the US and UK [69], also raises questions surrounding ambiguity of ownership, governance and responsibility. Consider the lingering stigma of false accusations in the press, or perhaps of being added in a gang-database (and never removed), etc. We note the work of key organisations in this space, such as the Electronic Frontier Foundation[12] and the Open Rights Group[13].

The results of the simulation are enough to demonstrate that many forms of evidence, rich enough to describe the range from transient moods to more permanent traits, can be derived from social media data; CCTV is a form of validation of the event. As of yet there does not exist a dataset that combines all of these features, for interest a civil unrest dataset containing 'live' geotagged tweets or status updates, with accompanying CCTV feed. However this should not be too far away, with CCTV set to become a commodity like other utilities [70], and smart CCTV touted as being able to recognise the precursors for civil disobedience and riots.

The work of Schwartz et al. [23] leverages what people say in social media to find distinctive words, phrases, and topics as functions of known attributes of people such as gender, age, location, or psychological characteristics. This can thus be transposed, inferring gender, age and so on, from social media data. The negative implications of these developments are that they can easily be applied to large numbers of people without obtaining their individual consent or even being aware. Commercial companies, governmental institutions, or even one's Facebook friends could use software to infer personality (and other attributes, such as intelligence or sexual orientation) that an individual may not have intended to share [8].

Smart CCTV networks will need to consider more than just postures and crowd positions, but also locations of unhappy or aggressive tweets or Facebook status updates. Indeed, from sufficient (un)natural text it is becoming possible to infer behaviours and even personality traits, and we have demonstrated the first attempts to extend this recognising psychopathy. Due to the explosive growth of cyber-real space and crowds over the burgeoning social networking domains, the real space in which we inhabit is also strongly tied with the virtual cyber-social space. We are thus interested in understanding the relationships and the interactions of crowds, in physical-real space and cyber-social space. By modelling areas in a landscape (e.g. cities/urban domains) in terms of cyber-physical crowds who are today commonly communicating through social networking, we expose geotagged interactions. We can explore urban characteristics which are reflected in the social networks through crowd behaviour.

We thus have a preliminary implementation of a model, with the intention of being able to successfully model complex dynamics among population, utilising features drawn from sources such as social media updates, knowledge about friendship, networks, family, location. The wider aims are broad; however initially we are interested merely to develop our research ideas around representation.

[12] https://www.eff.org/

[13] https://www.openrightsgroup.org/

References

[1] UK Government Office for Science, "What are future cities? Origins, meanings and uses," Tech. Rep., June 2014.

[2] E. Cosgrave, T. Tryfonas, and T. Crick, "The Smart City from a Public Value Perspective," in *Proceedings of 2nd International Conference on ICT for Sustainability (ICT4S 2014)*, 2014.

[3] Arup, The Climate Group, Accenture and Horizon, University of Nottingham, "Information Marketplaces: The New Economics of Cities," Tech. Rep., November 2011.

[4] UK Parliamentary Office of Science and Technology, "Social Media and Big Data," Tech. Rep. Report 460, March 2014.

[5] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 3, pp. 7280–7287, 2002.

[6] G. Oatley and T. Crick, "Changing Faces: Identifying Complex Behavioural Profiles," in *Human Aspects of Information Security, Privacy and Trust*, ser. Lecture Notes in Computer Science. Springer, 2014, vol. 8533, pp. 282–293.

[7] G. Oatley, B. Ewart, and J. Zeleznikow, "Decision Support Systems For Police: Lessons From The Application of Data Mining Techniques To "Soft" Forensic Evidence," *Artificial Intelligence and Law*, vol. 14, no. 1-2, pp. 35–100, 2006.

[8] R. Lambiotte and M. Kosinski, "Tracking the Digital Footprints of Personality," *Proceedings of the IEEE*, vol. 102, no. 12, pp. 1934–1939, 2014.

[9] G. Oatley, T. Crick, and M. Mostafa, "Digital Footprints: Envisaging and Analysing Online Behaviour," in *Proceedings of 2015 Symposium on Social Aspects of Cognition and Computing Symposium (SSAISB 2015)*, 2015.

[10] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational Social Science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[11] J. Biel and D. Gatica-Perez, "The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs," *IEEE Transactions on Multimedia*, vol. 15, no. 1, pp. 41–55, 2012.

[12] R. Procter, J. Crump, S. Karstedt, A. Voss, and M. Cantijoch, "Reading the riots: what were the police doing on Twitter?" *Policing and Society*, vol. 23, no. 4, pp. 413–436, 2013.

[13] P. Burnap, M. L. Williams, L. Sloan, O. F. Rana, W. Housley, A. Edwards, V. Knight, R. Procter, and A. Voss, "Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack," *Social Network Analysis and Mining*, vol. 4, no. 1, 2014.

[14] Big Brother Watch, "The Price of Privacy: How local authorities spent £515m on CCTV in four years," Tech. Rep., February 2012.

[15] B. C. Welsh and D. P. Farrington, "Public Area CCTV and Crime Prevention: An Updated Systematic Review and MetaAnalysis," *Justice Quarterly*, vol. 26, no. 4, pp. 716–745, 2009.

[16] E. di Bella, F. Odone, M. Corsi, A. Sillitti, and R. Breu, *Smart City: How to Create Public and Economic Value with High Technology in Urban Space*. Springer, 2014, ch. Smart Security: Integrated Systems for Security Policies in Urban Environments, pp. 193–220.

[17] J. Ratcliffe and T. Taniguchi, "The crime reduction effects of public CCTV cameras in the City of Philadelphia, PA installed during 2006," Temple University, Philadelphia, Tech. Rep., February 2008.

[18] J. H. Ratcliffe, T. A. Taniguchi, and R. Taylor, "The crime reduction effects of public CCTV cameras: A multi-method spatial approach," *Justice Quarterly*, vol. 26, no. 4, pp. 746–770, 2009.

[19] J. M. Caplan, L. W. Kennedy, and G. Petrossian, "Police-monitored CCTV cameras in Newark, NJ: A quasi-experimental test of crime deterrence," *Journal of Experimental Criminology*, vol. 7, no. 3, pp. 255–274, 2011.

[20] P. Klerks, "The Network Paradigm Applied to Criminal Organizations," *Connections*, vol. 24, no. 3, pp. 53–65, 2001.

[21] P. Lipton, *Inference to the Best Explanation*, 2nd ed. Routledge, 2002.

[22] ——, "Testing Hypotheses: Prediction and Prejudice," *Science*, vol. 307, no. 5707, pp. 219–221, 2005.

[23] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach," *PLoS ONE*, vol. 8, no. 9, 2013.

[24] J. Caverlee, Z. Cheng, D. Z. Sui, and K. Yeswanth Kamath, "Towards Geo-Social Intelligence: Mining, Analyzing, and Leveraging Geospatial Footprints in Social Media," *IEEE Data Engineering Bulletin*, vol. 36, no. 3, pp. 33–41, 2013.

[25] A. Stefanidis, A. Crooks, J. Radzikowski, A. Croitoru, and M. Rice, *Socio-Cultural Dynamics and Global Security: Interdisciplinary Perspectives on Human Geography in an Era of Persistent Conflict*, ser. USGIF, ch. Social Media and the Emergence of Open-Source Geospatial Intelligence, pp. 109–123.

[26] Z. Cheng, J. Caverlee, and K. Lee, "You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users," in *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM'10)*. ACM Press, 2010, pp. 759–768.

[27] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games," 2011, http://arxiv.org/abs/1106.4300.

[28] L. Sloan, J. Morgan, W. Housley, M. L. Williams, A. Edwards, P. Burnap, and O. F. Rana, "Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter," *Sociological Research Online*, vol. 18, no. 3, 2013.

[29] Y. Liang, J. Caverlee, Z. Cheng, and K. Y. Kamath, "How big is the crowd?: event and location based population modeling in social media," in *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HT'13)*. ACM Press, 2013, pp. 99–108.

[30] K. Y. Kamath, J. Caverlee, K. Lee, and Z. Cheng, "Spatio-Temporal Dynamics of Online Memes: A Study of Geo-Tagged Tweets," in *Proceeding of the 22nd Annual ACM World Wide Web Conference 2013 (WWW'13)*. ACM Press, 2013, pp. 667–678.

[31] L. Gannes, "The Short and Illustrious History of Twitter #Hashtags," https://gigaom.com/2010/04/30/the-short-and-illustrious-history-of-twitter-hashtags/, April 2010.

[32] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "TwitterStand: news in tweets," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS'09)*. ACM Press, 2009, pp. 42–51.

[33] M. Woodworth, J. Hancock, S. Porter, R. Hare, M. Logan, M. E. OToole, and S. Smith, "The Language of Psychopaths: New Findings and Implications for Law Enforcement," *FBI Law Enforcement Bulletin*, July 2012.

[34] F. Iacobelli, A. Gill, S. Nowson, and J. Oberlander, "Large Scale Personality Classification of Bloggers," in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII'11)*. Springer, 2011, pp. 568–577.

[35] J. Pennebaker and L. King, "Linguistic styles: language use as an individual difference," *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1296–1312, 1999.

[36] J. Oberlander and A. J. Gill, "Individual differences and implicit language: Personality, parts-of-speech and pervasiveness," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 2004, pp. 1035–1040.

[37] ——, "Language with character: A stratified corpus comparison of individual differences in e-mail communication," *Discourse Processes*, vol. 42, no. 3, pp. 239–270, 2006.

[38] K. R. Scherer, "Emotion as a Multicomponent Process: A model and some cross-cultural data," *Review of Personality and Social Psych*, vol. 5, pp. 37–63, 1984.

[39] S. Vazire and S. Gosling, "e-Perceptions: Personality Impressions Based on Personal Websites," *Journal of Personality and Social Psychology*, vol. 87, no. 1, pp. 123–132, 2004.

[40] B. Blamey, T. Crick, and G. Oatley, "R U :-) or :-(? Character- vs. Word-Gram Feature Selection for Sentiment Classification of OSN Corpora," in *Research and Development in Intelligent Systems XXIX, Proceedings of the 32nd SGAI International Conference on Artificial Intelligence (AI-2012)*. Springer, 2012, pp. 207–212.

[41] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 4, pp. 1036–1040, 2014.

[42] F. Mairesse, M. Walker, M. Mehi, and R. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research*, vol. 30, pp. 457–500, 2007.

[43] J. Pennebaker, M. Francis, and R. Booth, "Linguistic Inquiry and Word Count," Erlbaum Publishers, 2001.

[44] M. Wilson, "The MRC Psycholinguistic Database: Machine Readable

Dictionary, Version 2.00," *Behavior Research Methods, Instruments & Computers*, vol. 20, no. 1, pp. 6–10, 1988.

[45] W. T. Norman, "Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings," *Journal of Abnormal and Social Psychology*, vol. 66, no. 6, pp. 574–583, 1963.

[46] D. Peabody and L. Goldberg, "Some determinants of factor structures from personality-trait descriptor," *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 552–567, 1989.

[47] H. J. Eysenck, "Four ways five factors are not basic," *Personality and Individual Differences*, vol. 13, no. 6, pp. 667–673, 1992.

[48] S. V. Paunonen and D. N. Jackson, "What is beyond the Big Five? Plenty!" *Journal of Personality*, vol. 68, no. 5, pp. 821–836, 2000.

[49] J. Block, "The Five-Factor Framing of Personality and Beyond: Some Ruminations," *Psychological Inquiry*, vol. 21, no. 1, pp. 2–25, 2010.

[50] Z. Szirmák and B. De Raad, "Taxonomy and structure of Hungarian personality traits," *European Journal of Personality*, vol. 8, no. 2, pp. 95–117, 1994.

[51] F. De Fruyt, R. R. McCrae, Z. Szirmák, and J. Nagy, "The Five-Factor Personality Inventory as a Measure of the Five-Factor Model: Belgian, American, and Hungarian Comparisons with the NEO-PI-R," *Assessment*, vol. 11, no. 3, pp. 207–215, 2004.

[52] G. Oatley and T. Crick, "Measuring UK Crime Gangs," in *Proceedings of 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE Press, 2014.

[53] ——, "Exploring UK Crime Networks," in *Proceedings of 2014 International Symposium on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI 2014)*. IEEE Press, 2014.

[54] ——, "Measuring UK Crime Gangs: A Social Network Problem," *Social Network Analysis and Mining*, vol. 5, no. 1, 2015.

[55] J. Miller and D. Lynam, "Psychopathy and the Five-factor model of personality: a replication and extension," *Journal of Personality Assessment*, vol. 81, no. 2, pp. 168–178, 2003.

[56] D. Lynam, A. Caspi, T. Moffitt, A. Raine, R. Loeber, and M. Stouthamer-Loeber, "Adolescent psychopathy and the big five: results from two samples," *Journal of Abnormal Child Psychology*, vol. 33, no. 4, pp. 431–443, 2005.

[57] G. Oatley, T. Crick, and R. Howell, "Data Exploration with GIS Viewsheds and Social Network Analysis," in *Proceedings of 23rd GIS Research UK Conference (GISRUK 2015)*, 2015.

[58] G. Oatley, K. McGarry, and B. Ewart, "Offender Network Metrics," *WSEAS Transactions on Information Science & Applications*, vol. 12, no. 3, pp. 2440–2448, 2006.

[59] P. Burnap, O. Rana, M. Williams, W. Housley, A. Edwards, J. Morgan, L. Sloan, and J. Conejero, "COSMOS: Towards an integrated and scalable service for analysing social media on demand," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 30, no. 2, pp. 80–100, 2015.

[60] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 5802–5805, 2013.

[61] C. Gerritsen, "Agent-based modelling as a research tool for criminological research," *Crime Science*, vol. 4, no. 2, 2015.

[62] A. Crooks, *Complexity and Planning: Systems, Assemblages and Simulations*, ser. New Directions in Planning Theory. Ashgate, 2012, ch. The Use of Agent-Based Modelling for Studying the Social and Physical Environment of Cities.

[63] M. J. Chorley, R. M. Whitaker, and S. M. Allen, "Personality and location-based social networks," *Computers in Human Behavior*, vol. 46, pp. 45–56, 2015.

[64] A. T. Crooks and C. J. E. Castle, "The Integration of Agent-Based Modelling and Geographical Information for Geospatial Simulation," in *Agent-Based Models of Geographical Systems*. Springer, 2012, pp. 219–251.

[65] N. Cohen, J. Gattuso, and K. MacLennan-Brown, *CCTV Operational Requirements Manual 2009*, UK Home Office, Scientific Development Branch, 2009, publication No. 28/09.

[66] B. J. Goold, "Privacy rights and public spaces: CCTV and the problem of the "unobservable observer"," *Criminal Justice Ethics*, vol. 21, no. 1, pp. 21–27, 2002.

[67] C. Bowden, "Closed circuit television for inside your head: blanket traffic data retention and the emergency anti-terrorism legislation," *Computer and Telecommunications Law Review*, vol. 8, no. 2, pp. 21–24, 2002.

[68] T. Crick, J. Mateos-Garcia, H. Bakhshi, and S. Westlake, "Innovation Policy-Making in the Big Data Era," in *Proceedings of Data for Policy 2015*, 2015.

[69] M. Carr and T. Crick, "The Problem of the P3: Public-Private Partnerships in National Cyber Security Strategies," in *Proceedings of 1st International Conference on Cyber Security for Sustainable Society 2015*, 2015.

[70] S. Graham, "CCTV: The Stealthy Emergence of a Fifth Utility?" *Planning Theory & Practice*, vol. 3, no. 2, pp. 237–241, 2002.