Exploring the homogeneity of theft offenders in spatio-temporal crime hotspots

Tongxin Chen[1], Kate Bowers[2], Tao Cheng[1*] , Yang Zhang[1] and Peng Chen[3]

Abstract

Offender homogeneity occurs when the same criminal group is composed of offenders with similar attributes (e.g., socio-economic-demographics). Exploring the homogeneity of offenders within spatio-temporal crime hotspots (STCHs) is useful for understanding not only the generational mechanisms of crime hotspots, but also has crime prevention implications. However, the homogeneity of offenders within STCHs has not been explored in criminological studies hitherto. Indeed, current techniques of STCH detection are limited to using statistical clustering methods in existing studies that lack the ability to identify the shape of STCHs or the distribution and variety of offences/offender activity with them. In this study, we utilise a spatio-temporal clustering algorithm called ST-DBSCAN to determine STCHs. We then propose novel entropy-based indices that measure the similarity of offenders (and offences) within STCHs. The method is demonstrated using theft crime records in the central area of Beijing, China. The results show that theft in the city is concentrated in a narrow space and time span (STCHs) and that within these associated offenders with similar social demographics, referred to as homogeneous offender groups are detectable.

Keywords: Spatio-temporal clustering, Homogenous offenders, ST-DBSCAN, Entropy, Near-repeat victimization, Unsupervised learning

Introduction

Crime is not distributed randomly but concentrated in places vis a vis the *law of crime concentration* (Weisburd 2015). *Spatio-temporal crime hotspots* (STCHs) refer to crime concentrations in space and time (Leong and Sung 2015). Empirically, STCHs exhibit that the infectious pattern of crime risk is elevated temporarily for prior burgled households and nearby households, referring to repeat and near-repeat victimisation, respectively (Bowers and Johnson 2005; Johnson and Bowers 2004b; Johnson et al. 2007; Townsley et al. 2003). It is contended that the offender's rational target-selecting strategies contribute to generating STCHs (Johnson 2014). Two alternative theoretical mechanisms–*flag* and *boost*–have

been proposed to explain how STCHs generate in various urban areas. If places are *flagged* in terms of their suitability as victims/targets, attracting multiple different offenders, then offences in such places are committed independently. A previous place suffering an offence will be *boosted* if the same offender or group of offenders will return to chase the benefits they have knowledge of at the original or nearby locations (Bowers and Johnson 2004; Johnson 2008; Pease et al. 1998). Based on the *boosted* mechanism, some analyses suggest that same offenders are more likely to be responsible for crimes occurring close in space and time (STCHs) (Bernasco 2008; Johnson et al. 2009).

These studies, however, seldom consider in depth that not merely the same offender but associated offenders from the same criminal group can be hypothesized as being likely to be involved in STCHs. Co-offending, or sharing of information across a group or network (for property crime such as theft or burglary) is an under-researched topic in the literature (Felson 2009). There is

*Correspondence: tao.cheng@ucl.ac.uk
[1] SpaceTimeLab for Big Data Analytics, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, UK
Full list of author information is available at the end of the article

Chen *et al. Crime Sci*   (2020) 9:9

Page 2 of 13

some evidence that group offenders are likely endowed with similar criminal behaviours, e.g., MOs (Modus Operandis) and location choices (Adderley and Musgrove 2003; Bernasco 2006; Lammers 2017). Further, there is evidence that near repeat burglaries are more likely to have common MO features – such as method and point of entry than those further apart in space and time (Bowers and Johnson 2004). Explaining this evidence through the *boosted* account it can be reasoned that subsequent offenders may receive opportunity information from previous offenders in a criminal group and select the originally victimized location or one nearby to forager for suitable targets. This foraging strategy depicts that the subsequent offender's hunting decision (and potentially method of crime) would be influenced by the initial offender's experience through information-sharing pathways existing within the criminal group.

Homogeneity (or similarity) refers to the presence of individuals or occurrences with similar attributes. For individuals, similar attributes can help form the ties linking people into a social group (or network) (Borgatti et al. 2009; McPherson et al. 2001). In criminology, studies show that similarity of social demographics of individuals are strongly correlated with group offending, such as age, gender, race, kinship, hometown (ethnic origin) and neighbourhoods (Carrington 2009; Ozgul and Erdem 2012; Reiss 1986; Van Mastrigt and Farrington 2009; Weerman 2003). So, similarity of offenders in terms of their attributes might indicate groups or networks in which information is shared and should therefore have an association with space-time clusters or STCHs. Regarding the *flag* and *boost* mechanisms, we can expect that high-level homogeneity of offender characteristics (homogenous offenders) in STCHs crimes could reflect a group consisting of associated offenders, one offender committing a series of offences or a combination of these under the *boost* explanation. In contrast, low-level homogeneity (heterogeneous offenders) might indicate that victimized places in STCHs are mainly *flagged* by different types of offenders. Although both types of account are observable in STCHs, *boost* is always evident in some form (Johnson et al. 2009).

Our intention is to determine whether there is evidence of the *offenders with similar characteristics* acting in STCHs, as opposed to focusing exclusively on the similarity of the features of the offences themselves (such as MO or theft type) which have been the subject of other studies. If such homogeneity of offenders is detectable in STCHs this might contribute to explanation of the generation mechanism of STCHs. There is a parallel here with the literature on collective efficacy and crime in the community. Theories based on the idea of social cohesion as a guardian against crime state that communities

where populations are similar and share common goals and values are those that are more likely to help guard against crime through informal social control (Sampson and Groves 1989; Sampson et al. 1997). Here it could be proposed that commonality among offenders might lead to stronger networks and sharing of information which can promote boost-related offending.

Practically speaking, homogeneity is significant as it helps us understand what kind of offender groups recommitting according to space-time crime patterns, which could in turn inform crime reduction operations. To be clear, we are not purely interested in identifying networks of offenders here – we are interested in whether concurrent offenders operating in similar places at similar times are more likely to share common characteristics. A second innovation in this paper is the development of practical methods of delineating space-time patterns into meaningful space-time hotspots and exploring the homogeneity of the resulting clusters. We will demonstrate that this method allows comparison of similarity within different types or scales of STCHs. In other words, we can compare levels of homogeneity in STCHs with the shortest time periods to those that exist for longer. To enable this, a density-based spatio-temporal clustering method is used to identify the shape of STCHs and the distribution of offences and innovative entropy-based indices are proposed to measure the homogeneity of offenders (and other crime characteristics). Both of these approaches are detailed and compared with previously used approaches in the methods section below.

In this study, we utilise police-detected theft datasets in the central urban area of Beijing for 2014 to illustrate the methods developed. The remainder of this paper file is organised as follows. In the methodology section, we employ a spatio-temporal clustering method called ST-DBSCAN to extract STCHs, then propose several novel indices to examine the homogeneity of offenders in a given STCH. Next, the case study section introduces the data and the study area used in this study. Results and discussion section depicts the results from several analyses and discussion, including selecting the optimal unit of space and time for STCHs and assessing the implications of detecting the pattern at different levels regarding homogeneity in STCHs. Lastly, the conclusions section summarizes the results and indicates further research work in this area.

Methodology

Methods developed in this section are introduced as follows. Fig. 1 illustrates the main procedures in this study. In terms of the workflow, we first leverage an unsupervised learning approach by a spatio-temporal clustering algorithm – ST-DBSCAN – introduced in the

Chen *et al. Crime Sci*   (2020) 9:9

Page 3 of 13

Fig. 1 The workflow in this study
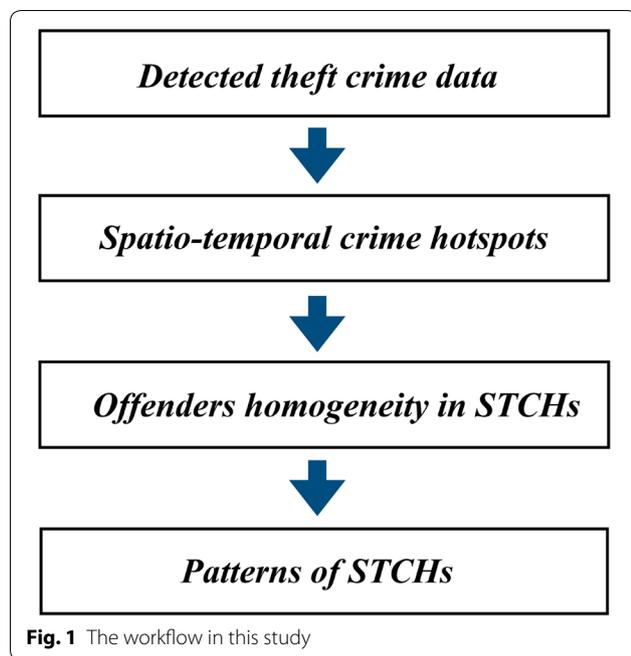
Spatio-temporal clustering section below – to extract STCHs from the detected theft crime data. Thereafter, we calculate the homogeneity of offenders with the indices detailed in the Entropy-based metric for homogeneity section. Following this we select the optimal spatio-temporal unit in our study area to extract STCHs and explore the patterns of offenders. Last, we portray the distinctive patterns of STCHs based on different levels of offender homogeneity and confirm that offender homogeneity can help to discriminate likely boost and flag account places in STCHs.

Spatio-temporal clustering

Spatio-temporal clustering is a process of grouping objects based on their spatial and temporal similarity that can discern several significant events or phenomena (Atluri et al. 2018; Cheng et al. 2014; Shi and Pun-Cheng 2019). Among the existing spatio-temporal clustering methods, some have been applied in detecting spatio-temporal patterns of crime events in previous criminological research. Most studies utilise the space-time interaction method known as *Knox's Index* or its modification, *Mantel's Index*, to test the phenomena of STCHs (Bowers and Johnson 2004, 2005; Haberman and Ratcliffe 2012; Johnson and Bowers 2004a, b; Johnson et al. 2007, 2009; Marchione and Johnson 2013; Ratcliffe and Rengert 2008; Townsley et al. 2003). Additionally, Grubesic and Mack (2008) employed the *Jacquez Test* (i.e., *k* nearest neighbours test) to detect urban

crime patterns. Malleson and Andresen (2015) used *SaTScan* software based on *space-time scan statistics* to extract STCHs whilst considering populations. However, the space-time scan method can only extract the spatio-temporal cylinder clusters and cannot detect the nuanced shape of data distributions in reality. Further, these aforementioned methods are all limited by statistical processes (Shi and Pun-Cheng 2019).

Herein, we utilise a density-based method called the *Spatial-Temporal Density Based Spatial Clustering of Applications with Noise* (ST-DBSCAN) algorithm proposed by Birant and Kut (2007). The major convenience of ST-DBSCAN is that it can find spatio-temporal clusters with arbitrary shape and noise points (Cheng et al. 2014). Effectively, the clusters, i.e., spatio-temporal crime hotspots, could be detected by ST-DBSCAN considering both space and time units. The ST-DBSCAN algorithm has three predefined parameters: spatial maximum reachable distance (SMRD), temporal maximum reachable distance (TMRD) and the minimum number of points (MinPts) should occur within SMRD and TMRD. For ready comprehension, if we set the SMRD at 300 m, TMRD at 5 days and MinPts at 3 for ST-DBSCAN, then STCHs extracted from the algorithm indicate that offences in such STCHs occur within 300 m and 5 days consecutively, and the minimum offence number in a STCH is 3. The SMRD and TMRD define the extent of STCHs distributed across space and time, so it is meaningful that selecting an appropriate space-time extent for STCHs by iterative calculations within ST-DBSCAN.

The entropy-based metric for homogeneity

Various indices have been developed to measure homogeneity/similarity or diversity of detected clusters or groups, such as the Simpson Index (Simpson 1949) in ecology and Shannon Entropy Index (Shannon 1948) in information science. In criminology, Bouhana et al. (2016) made use of the Simpson Index to evaluate the consistency of MOs in series burglars. Such indices are becoming more commonly used in assessing similarly and concentration in micro-level crime studies (Lee and Eck 2019).

Shannon Entropy is a metric for measuring the amount of information in information theory and has been widely applied as a popular similarity/diversity index in ecological studies. Herein, we chose Shannon Entropy as our representative metric of homogeneity, because variations of Entropy can be established from the additive calculations which can help to establish a global measurement for examining the homogeneity of STCHs. We define several indices based on Shannon Entropy to assess the homogeneity of offenders in a STCH:

Chen *et al. Crime Sci* (2020) 9:9

Page 4 of 13

Definition 1 Entropy of an Attribute (EOA)

Suppose attribute A is a random variable with possible values $\{a_0, a_1 \cdots a_i \cdots a_n\}$. Its entropy, H, is defined as:

$$H(A) = -\sum_{i=0}^{n-1} P(a_i) \log_2 P(a_i) \tag{1}$$

where $H(A)$ indicates the homogeneity of the feature, A, among offenders' demographic variables in the same STCH. A higher H indicates a lower homogeneity and vice versa. The homogeneity is the highest when H is 0 – referring to the case in which values are same in A. $P(a_i)$ is the probability of $A = (a_i)$. For example, suppose we categorise the age of offenders into levels, denoted as $A_{age} = \{level_0, level_1, level_2\}$. If a STCH committed by five offenders with age $\{level_0, level_1, level_1, level_2, level_2\}$, respectively, the homogeneity of the age of this STCH is then:

$$\begin{aligned} H(A_{age}) = &-\frac{1}{5} \times \log_2 \frac{1}{5} - \frac{2}{5} \\ &\times \log_2 \frac{2}{5} - \frac{2}{5} \\ &\times \log_2 \frac{2}{5} \approx 1.5219 \end{aligned}$$

Definition 2 Entropy of a Cluster (EOC)

If Cluster C consists of objects with m attributes $\{A_0, A_1 \cdots A_j \cdots A_m\}$, the entropy of a cluster considering all attributes is defined as:

$$H(C) = \sum_{j=0}^{m} H(A_j) \tag{2}$$

where $H(C)$ indicates the homogeneity of offenders in a cluster and A_j is the j-th attribute of offenders (e.g., age) in this cluster. m is the number of attributes' categories in a cluster. Therefore, higher H indicates a lower homogeneity of offenders in a STCH.

Definition 3 Entropy of Clusters (ECs)

To measure the quality of the clustering results, we propose to use the entropy of all detected clusters. Clusters, S, is the set of different clusters from a certain spatio-temporal pattern and could be a system that is also assessed by the overall entropy. ECs is defined as the weighted sum of EOC and denoted as:

$$H(S) = \sum_{k=0}^{K} H(C_k) * \omega_k, \omega_k = \frac{|c_k|}{|s|}, \quad S = \{C_0, C_1 \cdots C_K\} \tag{3}$$

where K is the number of obtained clusters; C_k is the k-th cluster in the clusters set, S; $|C_k|$ is the number of objects in cluster C_k; $|S|$ is the total number of objects for all clusters; and w_k is the weight of C_k. Globally, higher ECs indicates an overall lower homogeneity of the offenders for all STCHs detected.

Definition 4 Entropy of Noise (EON)

Of the outcomes from the ST-DBSCAN algorithm, we can obtain several clusters and noise, respectively. Noise is the set of data that does not belong to any clusters. Like a cluster, noise set, N, consists of objects with m attributes, $\{A_0, A_1 \cdots A_j \cdots A_m\}$. Hence, the calculation of EON is as same as the Definition 2 (EOC) and could be denoted as follows:

$$H(N) = \sum_{j=0}^{m} H(A_j), \quad |D| = |N| + |S| \tag{4}$$

where $|D|$ is the total number of objects across the whole dataset D; $|N|$ and $|S|$ is the total number of objects for noise and all clusters, respectively. It is obvious that EON follows Definition 4 with ECs also following Definition 3, and both are subjected to the units of space and time. So, we select optimal parameters (space and time units) in our study based on two considerations from iterative calculations within the algorithm. First, we select an optimal range of space and time units by examining the variation of STCHs numbers, ECs and EON from iterative calculations. Then, in such range, we discuss the EOC distribution and find the optimal set of space and time unit for STCHs in study area.

Case study–detected theft data from the city of Beijing

With rapid urbanisation, the capital city of China Beijing – attracts thousands of domestic migrants pursuing opportunities for jobs. Although this migration contributes to economic development, the growing number of migrants have also been linked to urban crime increases (Curran 1998; Liu 2006; Lo and Jiang 2006). Beijing is not an exception. Empirical studies indicate that most electric bicycle thefts are committed by migrants and offenders in a criminal groups who are mainly from the same hometown (Chen and Lu 2018). The study area in this paper is the central district of Beijing (latitude: $39.7574\,°$ $- 40.0287\,°$, longitudes: $116.1937\,° - 116.5519\,°$).

Chen *et al. Crime Sci*   (2020) 9:9

Page 5 of 13

The study dataset is the police-detected theft data in the study area ranging from 1st Jan 2014 to 30th Dec 2014. It comprises spatio-temporal information (i.e., date, time and coordinates) and offender's information, including ID, age and registered hometown address (RHA). There are many different classifications or interpretations of theft across various countries, in this dataset the crime of theft comprises pickpocketing, shoplifting, bicycles theft or theft from cars or bicycles. The detected crime data for theft crime consists of 7 802 theft offences committed by 6 754 offenders. The majority of theft crime (78 percent) involves a single offender, while the remainder, 22 percent, involves two or more offenders. There are some issues with the use of detection data for crime analysis. The most significant of these is that there could be systematic bias in the population of offenders and offences that are successfully detected by the police. Johnson et al. (2009) explored this possibility empirically and found no evidence of bias in that case, however in other cases this could be a possibility. We discuss the implications of this limitation where pertinent below.

In spatio-temporal clustering algorithm (ST-DBSCAN), the spatial unit (SMRD) (Euclidean distance used in this study) iterates over the range [100, 5 000] with steps of approximately 500 (metre) and temporal units (TMRD) iterates over the range [1, 60] with steps of approximately 5 (day) to determine the optimal parameter settings from clustering results. Further, MinPts is set to be 3 as the minimum number of cases in a STCH for this study. Additionally, We primarily studied the homogeneity of offenders age and RHA as the representation of social demographic homogeneity. The attribute age is categorised into four levels (under 18, 18 − 40, 40 − 60, above 60) and RHA (30 main provinces in China) represents where offenders come from. So, the EOC in this paper denotes the sum of age entropy and RHA entropy among offenders in one STCH (i.e, $m = 2$) following Definition 1 and Definition 2 in section.

Results and discussion

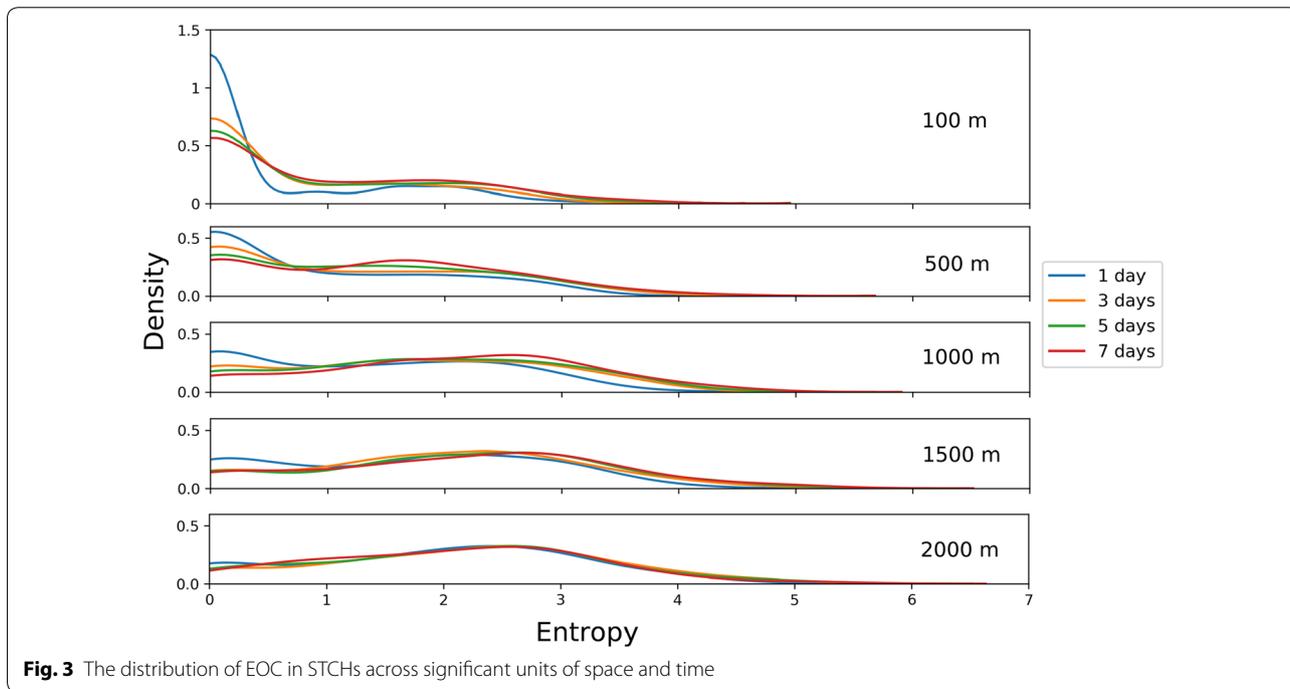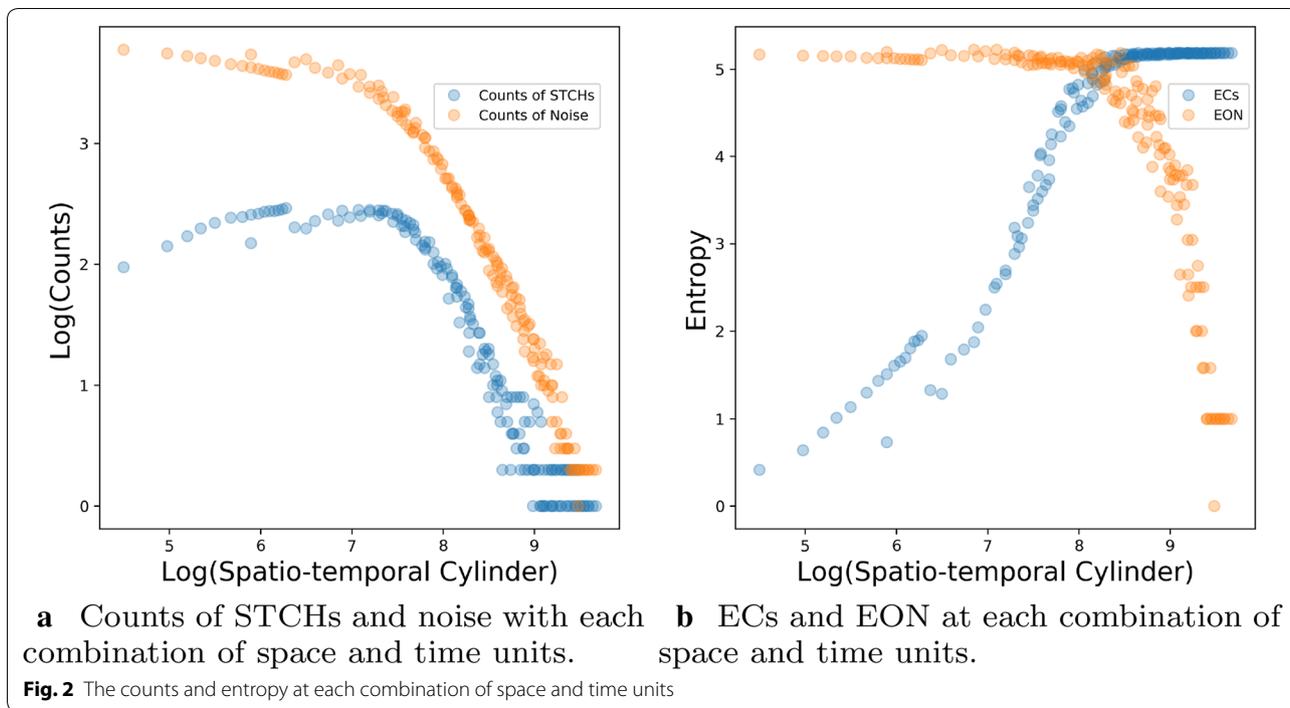### Space and time unit optimisation for STCHs

The phenomena of STCHs detected by previous empirical research shows that contagion tends to be distributed in small units of space and time, such as domestic burglary within 400 m and 2 months in Merseyside, UK (Bowers and Johnson 2005) or short dimensions across different nations (Johnson et al. 2007), like gun violence within 400 feet and 14 days or robbery within 400 feet and 7 days in Philadelphia, USA (Haberman and Ratcliffe 2012; Ratcliffe and Rengert 2008). Studies using data on detections demonstrate that the burglaries within 200 metres and 31 days in Hague, Netherlands (Bernasco

2008) or theft from motor vehicles within 100 m and 14 days in Bournemouth, UK (Johnson et al. 2009) are more likely to be cleared to the same offender.

In this study, homogeneity of offenders is considered helpful to determine the appropriate extent of STCHs spread over space and time. We first obtain the number of STCHs and noise from the outcome using different combinations of space and time units (SMRD and TMRD in ST-DBSCAN), then we check the ECs (i.e., total entropy of clusters) and EON (i.e., total entropy of noise) across different outcomes. For better visualisation of the outcome of iterative calculations, we collapse the space and time units from two dimensions to one dimension by calculating the spatio-temporal cylinder (referring to the combination of space and time units) that could be denoted as $STC = \pi(SMRD)^2 * TMRD$. The results for the Beijing theft data are shown in Fig. 2.

Fig. 2a denotes the changes of normalised counts of clusters and noise in each normalised spatio-temporal cylinder size. With a growth in the size of the defined spatio-temporal cylinder, the counts of identified STCHs initially increase at a stable level. After reaching a peak at approximately 7.5 of the normalised cylinder, there is then a rapid decrease in identified STCHs. In contrast, counts of noise see a continual reduction as the algorithm (ST-DBSCAN) compiles more theft crime within a larger cylinder (greater space and time units). In addition, Fig. 2b indicates the adverse tendencies between ECs(global homogeneity of clusters) and EON (global homogeneity of noise) with the increase of the defined cylinder size. Specifically, the STCHs are meaningless after a normalised cylinder reaches roughly 8.2 because the homogeneity of offenders in terms of noise is higher than those in STCHs (i.e, EON is less than ECs). At smaller cylinder sizes however (< 8.2) offender homogeneity is higher in the defined STCH than it is in the remaining noise.

The implication is that more homogenous offenders will be detected in smaller cylinder short-term STCHs than in long-term STCHs, which is consistent with previous studies. For this work, we expect to extract abundant STCHs with overall homogenous offenders (i.e., low values of ECs), so the significance of the range of the normalised spatio-temporal cylinder over the short term selected will be discussed further. Fig. 3 illustrates the distribution of EOC (homogeneity of the individual STCHs) across short-term cylinders (spatial units: 100m, 500m, 1000m, 1500m, 2000m; temporal units: 1 day, 3 days, 5 days, 7 days). For example, the blue line in the first subfigure represents the distribution of EOC in STCHs occurring within 100 m and 1 day. In order to clearly compare each space and time unit, the distributions are constructed using density estimation. With

Chen *et al. Crime Sci* (2020) 9:9

Page 6 of 13

a Counts of STCHs and noise with each combination of space and time units.

b ECs and EON at each combination of space and time units.

Fig. 2 The counts and entropy at each combination of space and time units

Fig. 3 The distribution of EOC in STCHs across significant units of space and time

regards to the blue line in the top subfigure, the density value is relatively high (approximately 1.25) when the EOC is 0, which denotes that most STCHs (within 100 metres and 1 day) involve an extremely high number of homogenous offenders population.

In detail, for a given spatial unit, the influence of the temporal units on the distribution of EOC is evident in every distribution in Fig. 3. It is remarkable that the space-time unit exhibits a narrow range (under 500 metres in spatial unit) that allows entropy to be

Chen *et al. Crime Sci*   (2020) 9:9

Page 7 of 13

meaningfully contrasted. Significantly, we can expect that homogenous offenders could be detected in STCHs occurring within 500 m and 7 days in our study area. In fact, the appropriate spatio-temporal units for STCHs are considered not only by a high level of homogenous offenders in STCHs, but the abundance of such

Table 1 The count of STCHs and theft offences across different levels of the EOC

| EOC level	Count of STCHs	Count of theft offences
0	69 (26.6%)	617 (21.7%)
$0 < EOC \leq 1$	40 (15.4%)	405 (14.2%)
$1 < EOC \leq 2$	80 (30.9%)	699 (24.5%)
$2 < EOC \leq 3$	52 (20.1%)	473 (16.6%)
$3 < EOC \leq 4$	15 (5.8%)	338 (11.9%)
$4 < EOC$	3 (1.2%)	317 (11.1%)
Sum	259 (100%)	2849 (100%)

phenomena in the study area. So, the STCHs with extremely short-term units are not recommended for detecting homogenous offenders because of the limited numbers of crime events involved. For example, the STCHs occurring within 100 m and 1 day include 9.3% of theft events, but the STCHs occurring with 500 m and 7 days account for 36.5%. To summarise, the optimal spatial and temporal units are determined as 500 m and 7 days, respectively.

Patterns of STCHs with offender homogeneity

Based on the spatio-temporal clustering results from the optimal unit of space and time in the study area (500 m, 7 days), we calculated the EOC of each of the 259 identified clusters to measure the homogeneity of theft offenders. The results show that all STCHs included 2849 theft crime offences committed by 698 offenders. The EOC of the STCHs are separated into six levels (0; $0 < EOC \leq 1$; $1 < EOC \leq 2$; $2 < EOC \leq 3$;
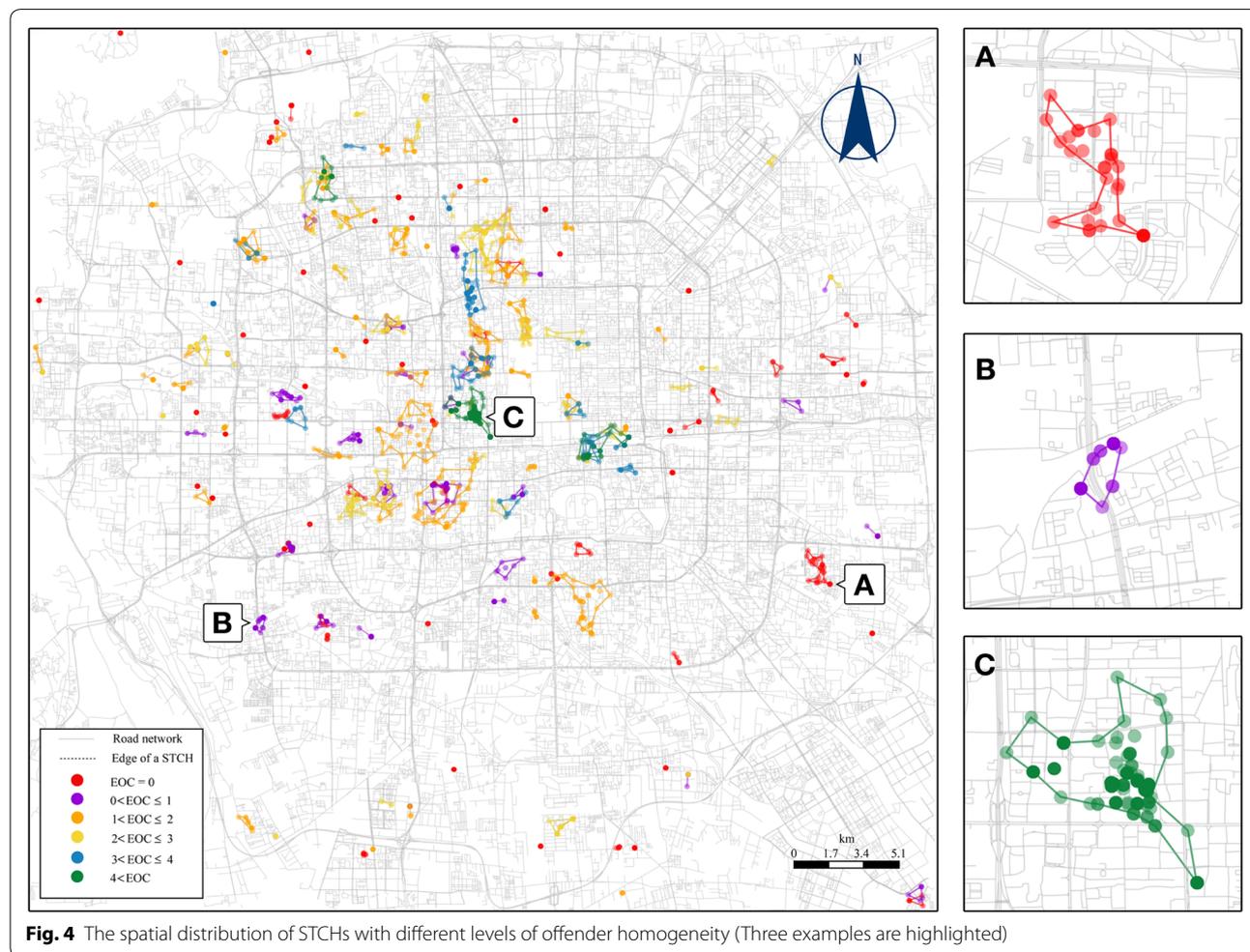
Fig. 4 The spatial distribution of STCHs with different levels of offender homogeneity (Three examples are highlighted)

Chen *et al. Crime Sci*   (2020) 9:9

Page 8 of 13

$3 < EOC \leq 4$; $4 < EOC$). Table 1 lists the counts for STCHs and theft offences at each level of the EOC.

The distribution of STCHs and their EOC levels in the study area are visualised in Fig. 4. The EOC levels (offenders' homogeneity in STCHs) are denoted by different colours. The transparency reflects the quantity of offences occurring in the same location (i.e., darker colour meaning repeat offences). Inspection of the distribution of homogeneity across STCHs is essential for understanding how the STCHs manifest in different place contexts.

Fig. 4 illustrates how STCHs are mainly distributed on the west side of the study area. Further, the STCHs with high-level homogeneity of offenders (i.e., red and purple, such as STCH A or B) are further from the city centre compared with the STCHs of the low-level homogeneity of offenders (i.e., blue and green, such as STCH C). Here we take STCH A, STCH B and STCH C as typical cases representing three levels of EOC and different characteristic districts in the study area.

In particular, STCH A covers some residential communities and consists of serial thefts from cars committed by a single offender over 2 weeks at cars park along the roadside lacking a guardian. Obviously, the offender comes back to this area for further targeting after prior benefit and boosts the generation of STCH A. In comparison, STCH B consists of several thefts from parked cars committed by two co-offenders with the same level of age and from the same hometown in a one week period. In STCH B, we also found another single shoplifting committed by an offender occasionally, whose RHA (hometown) and age level are different from the co-offenders. The area where STCH B is distributed is a residential community with limited guardianship, the same as STCH A. Lastly, STCH C consists of pickpocketing offences committed by a diversity of offenders from January to November within because the place of offences is located in the largest commercial district of Beijing–Xidan, which attracts not only many consumers but also potential theft offenders of diverse social demographics. Hence, this characteristic area is flagged by many theft offenders in Beijing given its distinct patterns of potential victims/targets–a dense crowd of people gathered after work or on weekends and holidays.

The risk duration of a STCH refers to the time period from the original offence to the last offence in one STCH, which also depicts the temporal length of crime risk propagation. Being able to distinguish between levels of risk duration in STCHs could be useful in differentiating the most appropriate crime prevention practices within them.

Table 2 shows descriptive statistics of STCH risk duration (in days) for the distinct levels of offender homogeneity. It demonstrates that the risk duration of STCHs in days are within a narrow range except for the last most heterogeneous group, i.e., $EOC > 4$ (e.g., STCH C). For the current result, 97.5% STCHs exhibit temporal periods of under 30 days suggesting that the most risk durations of STCHs are short in this study area. Further, Table 2 shows a clear trend – that the risk duration period increases with an increasing level of EOC grows (i.e., as the level of offender homogeneity declines).

The finding that homogenous offenders tend to generate a STCH with shorter temporal periods can be discussed through the lens of offenders' foraging strategies. If crime risk at places is boosted by homogenous offenders gathering for hunting or the same offender returning to the location, it makes sense that this will be limited in time. Foraging opportunities at the same or near locations may be seen as time limited benefits subject to improved guardianship and increased awareness after initial victimisation. In contrast, places flagged for victimisation are where heterogeneous offenders who exhibit relatively weak connections might co-exist in a STCH. The lack of dependency between these offenders is likely to result in a weaker temporal signature.

To explore this at a finer level of detail, Fig. 5 illustrates the three-dimensional (3D) projection of offences committed by the same offender or associated offenders in each STCH from 1st Apr 2014 to 21st May 2014.

In Fig. 5, the thefts in a STCH committed by a same offender and offender homogeneity are denoted by scatters with a black edge and colours, respectively. While each STCH features different levels of offender

Table 2 The descriptive statistics of STCH risk duration (in days) with different offender homogeneity levels

EOC level	Mean	SD	Median	Minimum	Maximum	N
0	6.9	6.5	5	1	28	69
$0 < EOC \leq 1$	9.4	5.6	8.5	2	24	40
$1 < EOC \leq 2$	12.0	6.7	11.0	2	40	80
$2 < EOC \leq 3$	15.1	8.2	13.0	3	42	52
$3 < EOC \leq 4$	22.3	22.0	17	7	98	15
$4 < EOC$	93.7	106.0	47	19	215	3
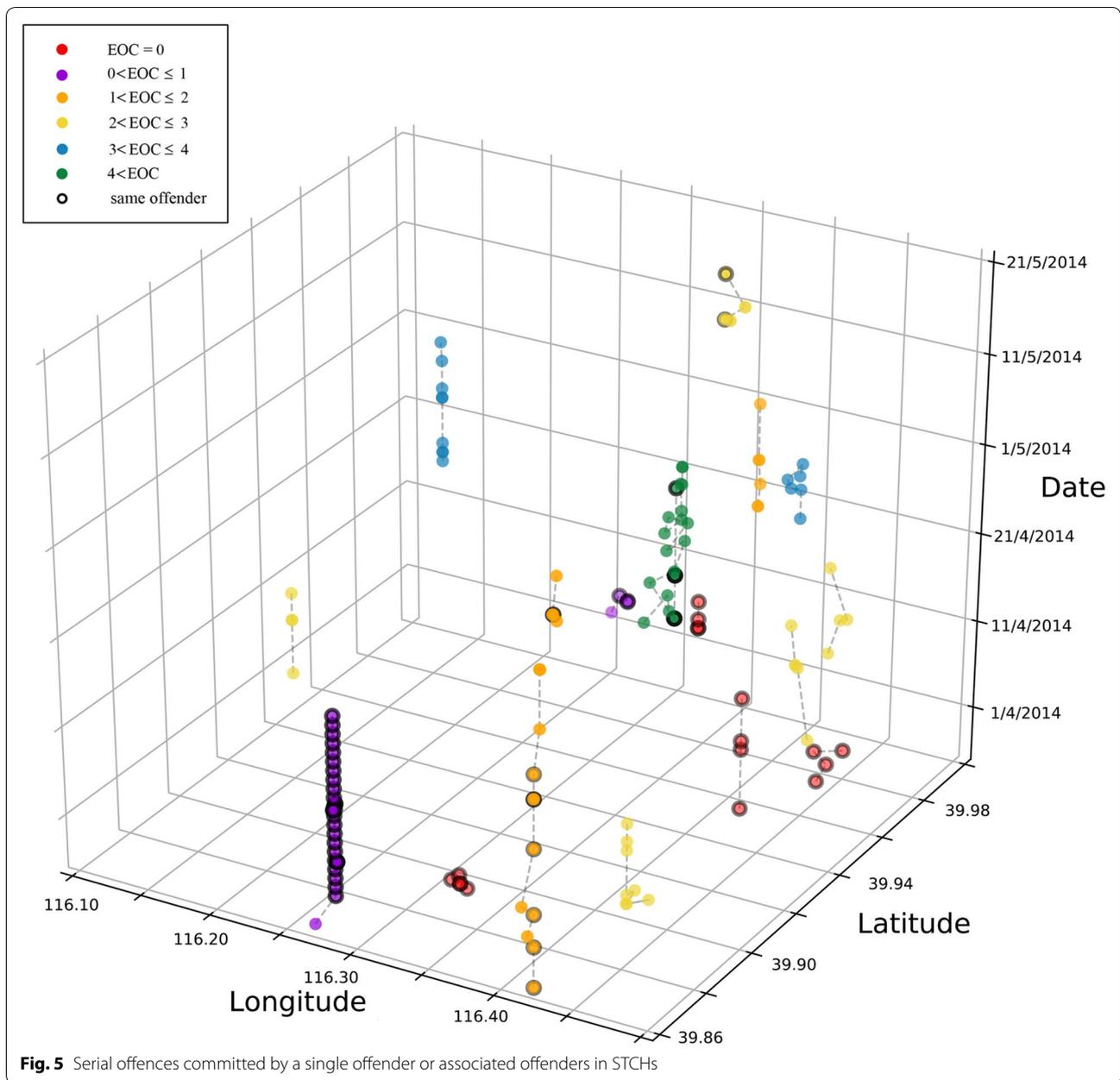
Fig. 5 Serial offences committed by a single offender or associated offenders in STCHs

homogeneity, a single offender committing a series of crimes may show up in a STCH. For example, in the selected red STCHs, each offender is responsible for each STCH so as homogeneity is (necessarily) extremely high. Series offenders show up in some STCHs with high homogeneity, e.g., the purple ones. Though certain STCHs have low offender homogeneity, such as orange or green STCHs, serial offences continue to emerge in some cases, thereby demonstrating the mixed mechanism of flag and boost. Additionally, even though no same-offenders offences are observed in some STCHs,

such as those that are blue and yellow, offenders from the same RHA or age level could contribute to the homogeneity in detected STCHs with particular spatio-temporal patterns.

Refocusing on offender homogeneity

In this section, we will further discuss whether offender homogeneity can generally represent the generation mechanisms of STCHs, i.e., how well it discriminates the boost or flag account places within STCHs. In our study, the offender homogeneity has been constructed
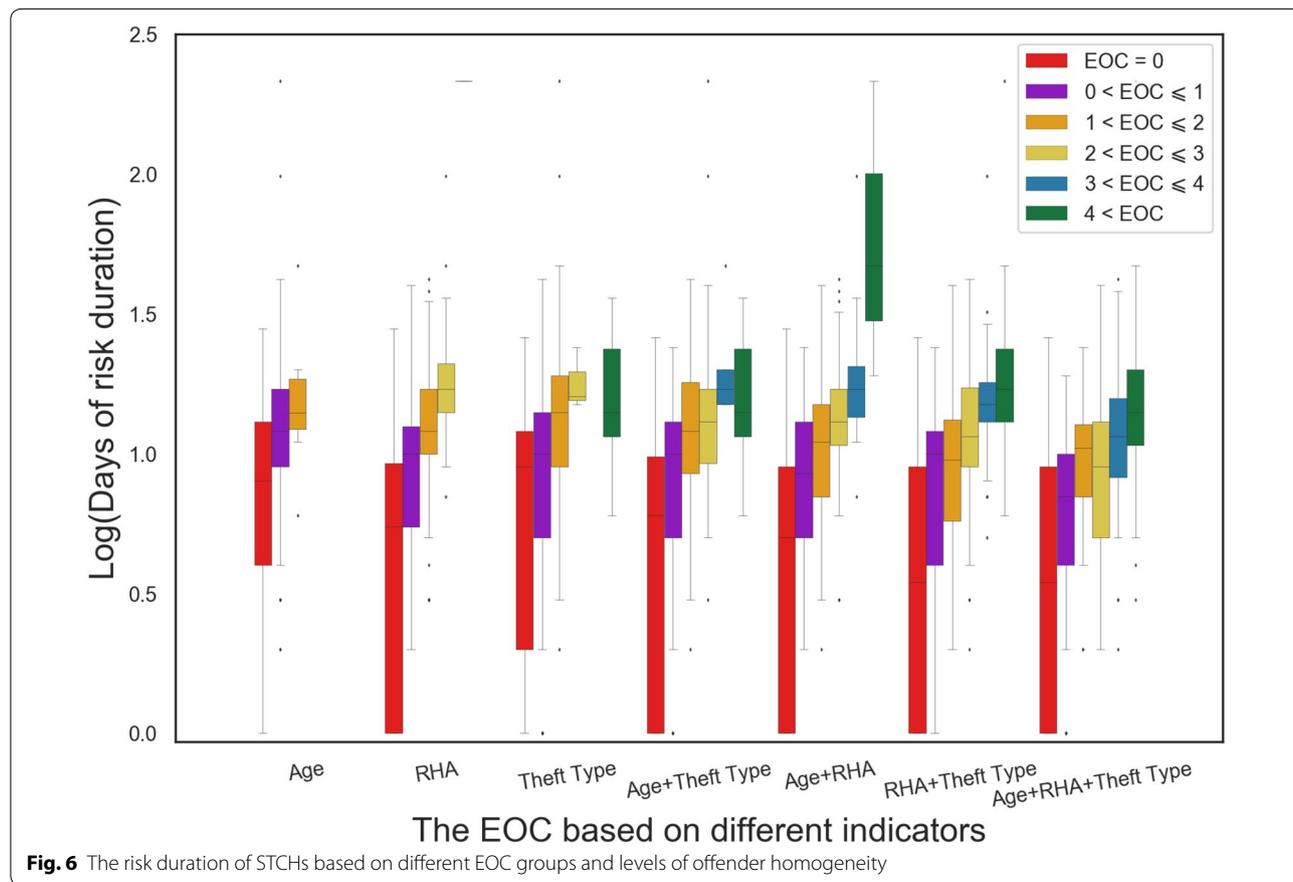
using two indicators (age and RHA (hometown)) from offender information. Results above have indicated that risk duration of STCHs rises as the level of offender homogeneity declines (i.e., level of EOC increases). However, it would be interesting to compare how this discriminatory power compares to that provided by other features of the thefts. The available data set enabled determination of similarity between crimes in STCHs on the basis of another feature – the type of theft undertaken.

We therefore added theft type (which had 10 categories) from our dataset to the existed indicators (Age and RHA). To explore discriminatory power, we calculated homogeneity levels for seven groups of indicator combinations: (i) Age; (ii) RHA; (iii) Theft Type; (iv) Age + Theft type; (v) Age +RHA; (vi) RHA + Theft Type; (vii) Age + RHA + Theft. For each defined group, we calculated EOC of STCH following Definitions 1, 2 in section ($m = 1, 2, 3$). For example, Age would be the indicator used in calculating EOC in group i, but RHA and theft type would be considered in group iv.

Fig 6 depicts the distribution of normalised STCH risk duration separated by levels in different EOC

groups for the seven different indicator combinations. Generally, the risk duration of STCH always rise as homogeneity level declines (the EOC level increases) in each group, regardless of what kind of indicators were used. In addition, the scale and range of homogeneity is subjected to the number of indicators, such as three levels of EOC in group i compared with six levels of EOC in group vii.

The calculated means of risk duration for the different indicator combinations, denote the same pattern as Table 2 with respect to offender homogeneity levels. Table 3 shows the means of risk duration (in days) separated by EOC levels for the different indicator groups. Of particular note is that in the EOC = 0 level (homogenous offender within STCHs), the means of STCHs risk duration declines as the number of indicators used increases. So, more matching features of offenders and offences appear to give better discriminatory power in isolating very short duration STCHs. Furthermore, the offender characteristics appear as useful in discriminating patterns as the type of theft. To analyse significance of the difference in the STCH risk duration (in days) according to offender homogeneity grouped by EOC

Fig. 6 The risk duration of STCHs based on different EOC groups and levels of offender homogeneity

Chen *et al. Crime Sci*   (2020) 9:9

Page 11 of 13

Table 3 The means of STCHs risk duration (in days) in EOC levels from different indicators and the results of Kruskal-Wallis H test

Indicators	EOC = 0 Mean	0–1 Mean	1–2 Mean	2–3 Mean	3–4 Mean	4 < EOC Mean	K-W H test Chi-square
Age	8.7	16.7	16.6				37.318***
RHA	6.9	10.3	13.7	22.5	215		67.504***
Theft type	8.2	10.7	18.6	18.3		18.1	37.593***
Age+Theft type	6.8	9.8	13.9	21.6	22.8	18.1	53.142***
Age+RHA	6.9	9.4	12.0	15.1	22.3	93.7	66.444***
Theft type+RHA	6.2	8.8	10.9	13.4	18.3	32.5	65.223***
Age+RHA+Theft type	6.2	7.3	10.4	10.6	13.0	19.5	64.792***

Notes: *** *p* < 0.001

levels, we conducted a Kruskal-Wallis H test in each indicator group. The results of the K-W H test in Table 3 show a p-value of less than 0.001 illustrating the statistically significant differences between EOC levels. Hence, the expected time period of a STCH can be classified by EOC levels with reasonable reliability. It therefore appears appropriate to utilise offender homogeneity to identify boosted STCHs for potential intervention within expected time periods for prospective policing purposes. Though significant differences in risk duration can be found for every type of indicator, it is evident that in general, the more indicators of homogeneity that are used, the better the discrimination in terms of STCH time periods.

Practically, the selection of homogeneity indicators will depend on available data and/or police interest. From a prevention point of view, it might be useful for police to know there are time restricted STCHs that emerge in certain areas that have certain crime-based features in common (e.g. MO, crime type, point of entry). It might also be useful to have an indication of whether a STCH is generated by a single offender, likely associated networks or a more heterogeneous offender population.

Conclusions

The aim of this study was to examine homogeneity of offenders involved in a specific pattern of spatial-temporal crime concentration, known as spatio-temporal crime hotspots (STCHs). Although previous studies suggested the same offender contributes to the generation of STCHs, without offender data this can only be assumed and limited discussion exists surrounding other types of offenders (such as potential networks or associates) involved in such patterns. That is, there are patterns whereby subsequent crimes occur at a location or nearby location within limited time periods as the consequence of homogenous offenders consisting of the same offender or similar social-demographic background offenders.

In our study, Shannon Entropy indices were employed to capture the homogeneity of the offenders of STCHs extracted from a spatio-temporal clustering algorithm–ST-DBSCAN. Such analysis can provide possible insight into the generation mechanism of certain STCHs.

In conclusion, the homogeneity of offenders provides support to the theoretical boost interpretation of the generation of STCHs. It also supports arguments for the-oretically-based generation of STCHs. It seems useful to consider the extent of homogeneity of STCHs as something to map in its own right, and the chosen indicators in this process will yield greater understanding of the offender's foraging strategies. The point here is that production of homogeneity-sensitive STCHs give a method of capturing and distinguishing more boost-based (or if desired flag-based) space-time clusters.

The significance of the findings in this work following our logical workflow are reported as follows. First, considering homogeneity indices in iterative calculations by ST-DBSCAN, the optimal extent of STCHs were selected to be 500 m and 7 days as significant spatio-temporal units for theft crimes in the city of Beijing. That is the selection of the spatial and temporal extent was *informed* by levels of similarity. Second, the level of offender homogeneity can depict the impact of an at risk location being boosted by the offender or flagged by victimisation opportunities. Many of the STCHs revealed a combination of these processes but there were also clusters with clearer delineation. Third, it is interesting that the STCHs with high-level offender homogeneity always undergo short-term risk duration compared with other low-level STCHs. Further, offender homogeneity can stably classify temporal patterns of STCH. We suggest that different selections of offender (and offence) features could contribute to development of more specific crime intervention strategies. The empirical analysis also suggests that shared characteristics between offenders are likely to help indicate associated offenders and suggest

Chen *et al. Crime Sci*   (2020) 9:9

Page 12 of 13

networks of similar offenders working together in more 'boost-based' contexts. There might also be implications for using homogeneity analysis of this type to determine undetected potential crime linkages

As discussed, the spatio-temporal algorithm leveraged in this study shows efficiency when extracting the clusters in our study area. Yet, ST-DBSCAN still has limitations such that it is time-consuming with a volumes dataset and is inefficient with datasets that have much noise. Some alternative algorithms, such as ST-OPTICS (Agrawal et al. 2016), could be tested as a useful comparison. Additionally, a quantitative method (e.g., evaluation index for STCHs) could be proposed to decide the optimal spatial and temporal unit for STCHs. In terms of other limitations, as mentioned above using detected data only can cause systematic bias. Offenders that are successfully detected might be atypical of the broader base of offenders–they might be less vigilant or more prolific for example. Given that much of the analysis above has used relevant comparison (looking at trends *within* the dataset) this might be less problematic than in other exercises. However, detected data is only one source of information on crime characteristics and the methods here could be usefully applied to other data sets (e.g., recorded rather than detected crimes).

In terms of the findings that homogeneous offenders are likely to be more evident in certain type of STCHs, whilst useful, caution in interpretation is advised. Whilst common characteristics in short duration STCHs might hint at networks and associates, the analysis here does not give direct evidence of these relationships. Moreover, while the pattern indicating there are similar offenders within STCHs is detectable in our study, the current analysis cannot clearly separate homogenous offenders or the same offenders in certain STCHs with longer-term duration (e.g., the green areas in Figs. 4, 5), such as those places afflicted by boost or flag simultaneously. Our future work will continue to explore the characteristics of both homogenous offenders and victims within STCHs in more detail with the goal of further interpreting the mechanisms and predicting such patterns for policing practices.

Authors' contributions
TC developed the data analysis and drafted the manuscript under the guidance of Prof. TC and Prof. KB. Specifically, Prof. TC contributed to the essential comments at the methodology section. Prof. KB contributed to the significant suggestions and advanced revisions at several sections based on the initial manuscript. YZ contributed to the suggestions in the analysis procedures and several helpful suggestions. PC contributed to several initial suggestions on the step of data processing. All authors read and approved the final manuscript.

Author details
[1] SpaceTimeLab for Big Data Analytics, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, UK. [2] Jill Dando Institute of Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK. [3] School of Information Engineering and Cyber Security, People's Public Security University of China, Muxidi Nanli 1st, Xicheng 100038, Beijing, China.

References
Adderley, R., & Musgrove, P. (2003). Modus operandi modelling of group offending: a data-mining case study. *International Journal of Police Science and Management*, *5*(4), 265–276.

Agrawal, K., Garg, S., Sharma, S., & Patel, P. (2016). Development and validation of optics based spatio-temporal clustering technique. *Information Sciences*, *369*, 388–401.

Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-temporal data mining: a survey of problems and methods. *ACM Computing Surveys (CSUR)*, *51*(4), 83.

Bernasco, W. (2006). Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling*, *3*(3), 139–155.

Bernasco, W. (2008). Them again? same-offender involvement in repeat and near repeat burglaries. *European Journal of Criminology*, *5*(4), 411–431.

Birant, D., & Kut, A. (2007). St-dbscan: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, *60*(1), 208–221.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895.

Bouhana, N., Johnson, S. D., & Porter, M. (2016). Consistency and specificity in burglars who commit prolific residential burglary: testing the core assumptions underpinning behavioural crime linkage. *Legal and Criminological Psychology*, *21*(1), 77–94.

Bowers, K. J., & Johnson, S. D. (2004). Who commits near repeats? a test of the boost explanation. *Western Criminology Review*, *5*(3).

Bowers, K. J., & Johnson, S. D. (2005). Domestic burglary repeats and space-time clusters: the dimensions of risk. *European Journal of Criminology*, *2*(1), 67–92.

Carrington, P. J. (2009). Co-offending and the development of the delinquent career. *Criminology*, *47*(4), 1295–1329.

Chen, P., & Lu, Y. (2018). Exploring co-offending networks by considering geographic background: an investigation of electric bicycle thefts in beijing. *The Professional Geographer*, *70*(1), 73–83.

Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G., & Wang, J. (2014). Spatiotemporal data mining. *Handbook of regional science*, *1173–1193*.

Curran, D. J. (1998). Economic reform, the floating population, and crime: the transformation of social control in china. *Journal of Contemporary Criminal Justice*, *14*(3), 262–280.

Felson, M. (2009). The natural history of extended co-offending. *Trends in Organized Crime*, *12*(2), 159–165.

Grubesic, T. H., & Mack, E. A. (2008). Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, *24*(3), 285–306.

Haberman, C. P., & Ratcliffe, J. H. (2012). The predictive policing challenges of near repeat armed street robberies. *Policing: A Journal of Policy and Practice*, *6*(2), 151–166.

Johnson, S. D. (2008). Repeat burglary victimisation: a tale of two theories. *Journal of Experimental Criminology*, *4*(3), 215–240.

Chen *et al. Crime Sci*   (2020) 9:9

Page 13 of 13

Johnson, S. D. (2014). How do offenders choose where to offend? Perspectives from animal foraging. *Legal and Criminological Psychology*, *19*(2), 193–210.

Johnson, S. D., & Bowers, K. J. (2004a). The burglary as clue to the future: the beginnings of prospective hot-spotting. *European Journal of Criminology*, *1*(2), 237–255.

Johnson, S. D., & Bowers, K. J. (2004b). The stability of space-time clusters of burglary. *British Journal of Criminology*, *44*(1), 55–65.

Johnson, S. D., Summers, L., & Pease, K. (2009). Offender as forager? A direct test of the boost account of victimization. *Journal of Quantitative Criminology*, *25*(2), 181–200.

Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., et al. (2007). Space-time patterns of risk: a cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology*, *23*(3), 201–219.

Lammers, M. (2017). Co-offenders' crime location choice: do co-offending groups commit crimes in their shared awareness space? *The British Journal of Criminology*, *58*(5), 1193–1211.

Lee, Y., & Eck, J. E. (2019). Comparing measures of the concentration of crime at places. *Crime Prevention and Community Safety*, *1–26*.

Leong, K., & Sung, A. (2015). A review of spatio-temporal pattern analysis approaches on crime analysis. *International E-journal of Criminal Sciences*, *9*, 1–33.

Liu, J. (2006). Modernization and crime patterns in china. *Journal of Criminal Justice*, *34*(2), 119–130.

Lo, T. W., & Jiang, G. (2006). Inequality, crime and the floating population in china. *Asian Journal of Criminology*, *1*(2), 103–118.

Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, *4*(1), 10.

Marchione, E., & Johnson, S. D. (2013). Spatial, temporal and spatio-temporal patterns of maritime piracy. *Journal of Research in Crime and Delinquency*, *50*(4), 504–524.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Ozgul, F., & Erdem, Z. (2012). Detecting criminal networks using social similarity. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, (pp. 581–585). IEEE Computer Society

Pease, K., et al. (1998). Repeat Victimisation: Taking Stock vol. 90. Home Office Police Research Group London.

Ratcliffe, J. H., & Rengert, G. F. (2008). Near-repeat patterns in philadelphia shootings. *Security Journal*, *21*(1–2), 58–76.

Reiss, A. J, Jr. (1986). Co-offender influences on criminal careers. *Criminal Careers and Career Criminals*, *2*, 121–160.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: testing social-disorganization theory. *American Journal of Sociology*, *94*(4), 774–802.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*, *277*(5328), 918–924.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical journal*, *27*(3), 379–423.

Shi, Z., & Pun-Cheng, L. S. (2019). Spatiotemporal data clustering: a survey of methods. *ISPRS International Journal of Geo-Information*, *8*(3), 112.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688.

Townsley, M., Homel, R., & Chaseling, J. (2003). Infectious burglaries. a test of the near repeat hypothesis. *British Journal of Criminology*, *43*(3), 615–633.

Van Mastrigt, S. B., & Farrington, D. P. (2009). Co-offending, age, gender and crime type: implications for criminal justice policy. *The British Journal of Criminology*, *49*(4), 552–573.

Weerman, F. M. (2003). Co-offending as social exchange. Explaining characteristics of co-offending. *British Journal of Criminology*, *43*(2), 398–416.

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, *53*(2), 133–157.

Publisher's Note