Investigating the Repeat and Near-Repeat Patterns in Sub-categories of burglary crime

Monsuru Adepeju

University College London, Gower Street, WC1E 6BT
monsuru.adepeju.11@ucl.ac.uk

Abstract

The investigation of repeat and near-repeat (RNR) patterns of sub-categories of burglary crimes is of great importance to law enforcement since a distinct intervention strategy may be suitable in the application to a different sub-category. In this study, the Knox test is used to investigate the RNR patterns within the data set of three different sub-categories of burglary crime in relation to the city of San Francisco in the United States. The results of the investigation showed that each sub-category exhibits a distinct RNR patterns, based on spatio-temporal interactions revealed by the Knox test. This study suggests that similar characteristics may be obtainable in the sub-categories of other major crime types; the outcomes of which may enhance crime interventions.

Keywords: Spatio-temporal, crime hotspot, near-repeat, burglary, Knox test

1. Introduction

The repeat and near-repeat (RNR) patterns in crime data sets have been investigated widely in order to gain a better understanding of crime communicability within an urban environment (Johnson et al., 2007a). The RNR patterns result from the idea that if a location is the target of a particular crime, such as burglary, the houses within a relatively short distance have an increased risk of being burgled over a period of a limited number of weeks (Bowers and Johnson, 2004). The investigation of RNR usually involves detecting the spatial and temporal distances (signatures) at which crime events interact to form hotspots across an area. Many of these prior analyses have focussed on the major crime categories, for example burglary and assault crime types. In the case of the burglary crime type, 'burglary-in-residence' has been given some attention, while other sub-categories, such as 'burglary-of-hotels' and 'burglary-in-stores' are yet to be given any attention. The publicly available crime database of the city of San Francisco, which contains detailed sub-categorisation of all the major crime types, provides an opportunity to address this research gap. Therefore, the goal of this study is to investigate the RNR patterns in three sub-categories of burglary crime, namely: 'burglary-in-residence', 'burglary-of-hotels', and 'burglary-in-stores'.

There are a number of reasons to suggest that each sub-category of burglary crime may have distinct spatial and temporal interaction at which the RNR patterns are prominent. For example, 'burglary-in-residence' is known to take place mostly during the week days when people have gone to work, while 'burglary-in-stores' are most likely to occur during the night when stores (or warehouses) are closed. Also, the land use distribution, as well as the neighbourhood structural arrangement underlying each sub-category, differs. Townsley et al. (2003) found that homogeneous housing (i.e. identical properties) experience more RNR patterns of burglary. By revealing the spatial and temporal interactions in each sub-category, the intervention activities of the law enforcement may be further enhanced.

2. Knox test for spatio-temporal interaction analysis

Originally developed to detect space-time interaction in disease events (Knox 1964), the Knox test has been widely applied to crime event data to reveal the RNR patterns of different major crime types (Johnson and Bowers, 2007). Theoretically, the Knox test examines whether there are more observed pairs, n, of events within a defined spatio-temporal neighbourhood, than would be expected on the basis of chance. The neighbourhood is defined through measuring from every individual event, a critical spatial distance (δ) and a temporal distance (τ), along the spatial and temporal dimensions, respectively. For each pair of spatial and temporal distances, the closeness of all points j from a reference i can be examined. This is then repeated for every single points across the entire study area and finally added together in order to derive the Knox test statistic as follows:

$$n_{\delta,\tau} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-1} X_{ij}Y_{ij} \tag{1}$$

$$X_{ij} = \begin{cases} 1, & \text{if event } j \text{ is within } \delta \text{ of } i \\ 0, & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1, & \text{if event } j \text{ is within } \tau \text{ of } i \\ 0, & \text{otherwise} \end{cases}$$

The $n_{\delta,\tau}$ is the observed (also called the test statistics), with which the expected statistics $e_{\delta,\tau}$ are compared to estimate the critical value (p) through using the formula:

$$p = \frac{1 + \sum_{v=1}^{r} I(n_{\delta,\tau} \geq e_{\delta,\tau})}{r + 1} \tag{2}$$

Where r is the number of iterations generated, $e_{\delta,\tau}$ is the equivalent list of expected statistics, and I(.) is the indication function. The context is that pairs of cases will be near to one another when RNR patterns exist, and the test statistics, $n_{\delta,\tau}$, will be large. If $n_{\delta,\tau}$ is large enough, the null hypothesis of no space-time interaction may be rejected. In the previous studies, different approaches have been used to derive the event distribution under the null hypothesis, including the standardised residual analysis (Agresti and Finlay, 1997) and the Monte Carlo simulation method (MC) (Mantel, 1967). In this study, the latter will be used as it has been shown to minimise edge effects (Johnson et al., 2007b). The MC simulation method involves randomly shuffling the times among the fixed spatial locations. This process is repeated 999 times for a more statistically rigorous outcome.

In many applications of the Knox test to crime data, a contingency table is created which lists spatial distance (δ), bands and temporal distance (τ) bands (see examples in Townsley et al., 2003; Johnson and Bowers (2004, p. 297). The band sizes are usually arbitrary, although they should reflect some theoretical ideas from a criminological point of view. In this study, the focus is to enhance local neighbourhood short-term crime intervention. Therefore, the spatial bands will be set at 200m, while the temporal band is set at 7 days. The whole process is extremely computationally intensive considering the large combination of critical spatial distances (δ) and temporal distances (τ)

examined. Thus, the parallel computing system provided by the University College London will be used to speed up the process.

3. Application to the burglary crime of San Francisco

The area chosen for this investigation is the city of San Francisco in the US. The burglary data set of 2015 was downloaded from the City's official website, which contains the daily records of crimes across the city from 2001 to the present time. From the downloaded data sets, the top three sub-categories, based on land use, with the highest number of records were selected. They were: (i) 'burglary-in-residence' (2990 records), 'burglary-of-hotels' (138 records), and 'burglary-in-stores' (1166 records). The total number of records of all the remaining sub-categories, which include 'burglary-of-institution' and 'burglary-of-offices', were less than 30, and therefore, were not considered separately. The spatial distribution of the selected three burglary sub-categories is observed to reflect the population distribution, as well as the land use patterns of the area (Figure 1-2). The North-eastern (NE) districts of the area have the highest population, and also contain the highest concentration of burglary incidents (Figure 1). The NE districts (i.e. districts 3 and 6) comprise the major business areas of the city, as well as encompassing many tourist attractions. Thus, the numbers of 'burglary-of-hotels' and 'burglary-in-stores' are disproportionately skewed towards these two districts. While the spatial spread of both the 'burglary-in-residence' and 'burglary-in-stores' extend to every part of the city, 'burglary-of-hotels' are mostly concentrated around the NE parts, where most of the hotels are located.

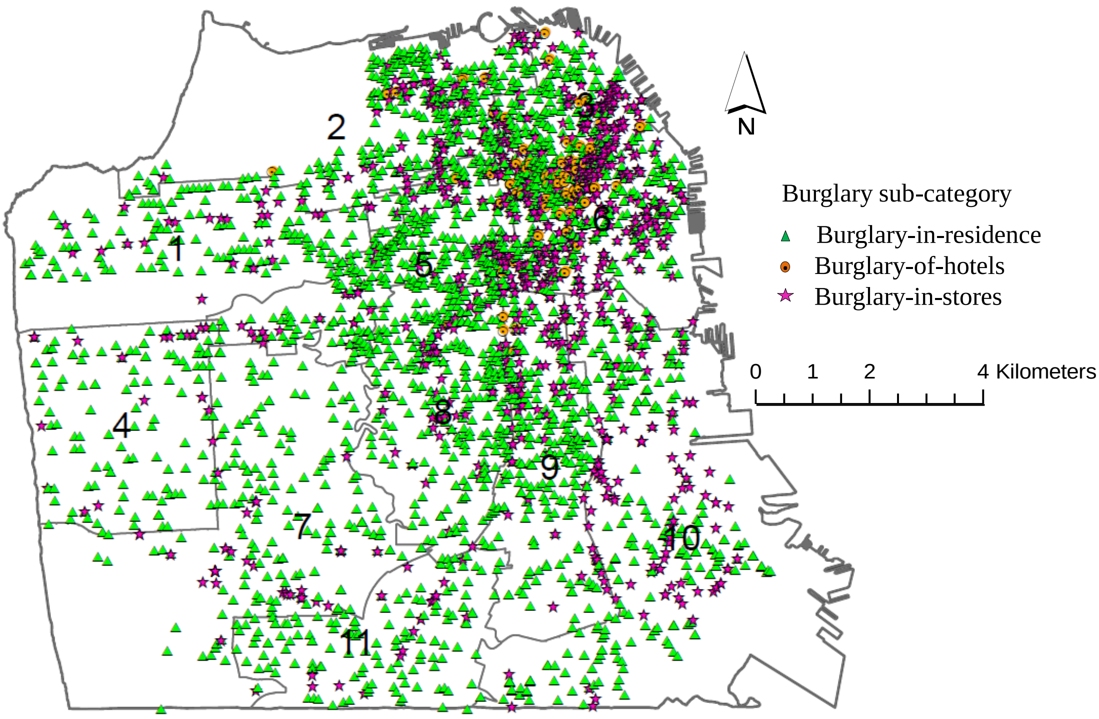

Figure 1: Map of San Francisco showing the spatial distribution of each sub-category of burglary crimes in 2015. The map also shows the boundaries of the eleven supervisorial districts of the city.

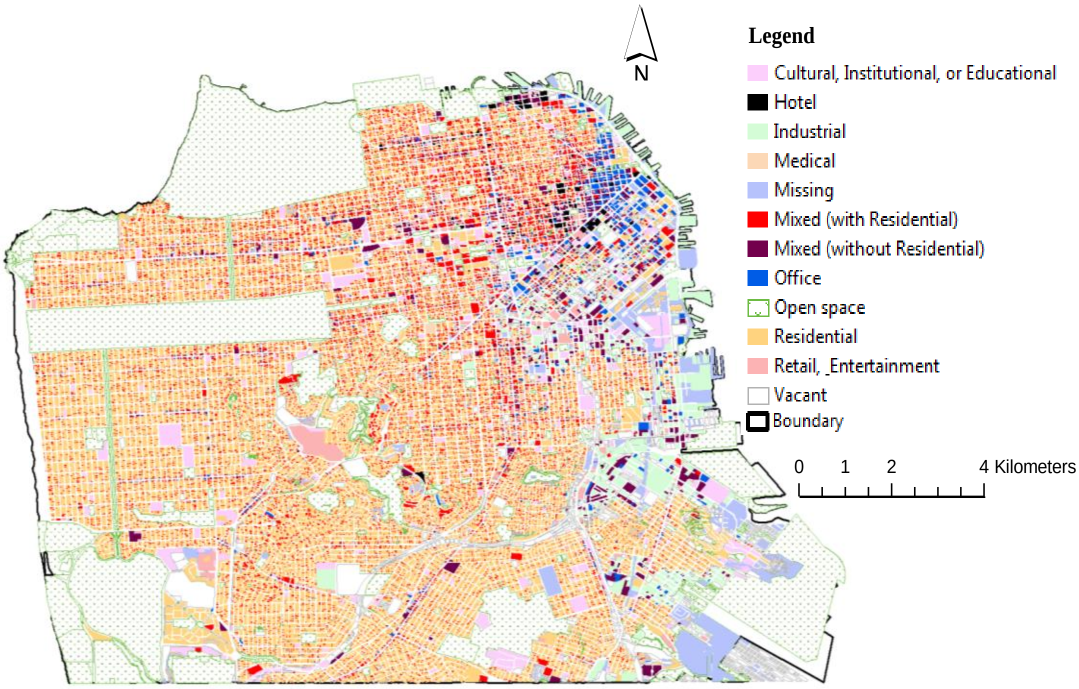

Figure 2: Land use map of San Francisco (2016). (Source: https://www.data.gov/). The map shows that the NE districts, particularly districts 3 and 6 are the most diverse in terms of land use, while the rest are largely residential.

4. Results and Discussion

Based on 999 Monte Carlo simulations, the smallest attainable $\hat{p}$ is given as $1/(999+1) = 0.001$. For this investigation, a $\hat{p}$ threshold of 0.05 was used as a cut-off for statistically significant space-time interaction. Hence, any cells with a $\hat{p} \leq 0.05$ will be interpreted as displaying strong evidence of repeat and near repeat (RNR) patterns. For a systematic analysis of the results, the result of the space-time interaction of the entire burglary crime (i.e. combining all sub-categories) is included. This is intended to provide a reference for the results of the sub-categories.

The results are, therefore, discussed in three parts: first, the space-time interaction of all the burglary data sets combined (Table 1); second, the space-time interaction of each sub-category (Tables 2 – 4); and third, sample 2D and 3D RNR maps for identifying hotspot regions (Figure 3). The shaded cells in the result tables highlight the critical distances at which the RNR patterns are statistically significant.

4.1 RNR patterns in the burglary data set (i.e. all sub-categories combined)

First, the investigation of RNR patterns in the entire burglary data set (i.e. all the sub-categories combined) is shown in Table 1. The results show multiple critical spatial and temporal bands (cells) at which the RNR patterns are prominent. For example, the $\hat{p} \leq 0.015$ at the cell intersection of spatial band 0-200m and temporal band 0-7 days indicates that the likelihood of burglary crime repeated by chance within these spatiotemporal bands is approximately one in a hundred. This indicates a very strong evidence of RNR patterns.

days / mts	0 - 7	8 - 14	15 - 21	22 - 28	29 - 35	>35
0-200	0.015	0.057	0.084	0.164	0.888	0.995
201-400	0.664	0.001	0.276	0.775	0.541	0.910
401-600	0.321	0.222	0.418	0.008	0.593	0.881
601-800	0.152	0.253	0.033	0.186	0.403	0.984
801-1000	0.258	0.907	0.151	0.380	0.012	0.889
1001-1200	0.693	0.514	0.007	0.236	0.831	0.875
1201-1400	0.979	0.279	0.952	0.253	0.940	0.023
1401-1600	0.013	0.711	0.428	0.710	0.003	0.961
1601-1800	0.239	0.160	0.823	0.461	0.742	0.617
1801-2000	0.516	0.911	0.460	0.032	0.522	0.594

Table 1: RNR patterns in the burglary data set (all sub-categories combined)

The results in Table 1 may be useful in terms of an intervention strategy for burglary crimes in general. However, for a more tactical intervention designed for a specific sub-category (type) of burglary crime, the results in Table 2-4 will be more useful.

4.2 RNR patterns in sub-categories of burglary data set

The results of RNR patterns of 'burglary-of-residence', 'burglary-of-hotels' and 'burglary-in-stores' are shown in Table 2, Table 3 and Table 4, respectively.

days / mts	0 - 7	8 - 14	15 - 21	22 - 28	29 - 35	>35
0-200	0.073	0.138	0.018	0.072	0.661	1.000
201-400	0.820	0.001	0.152	0.244	0.793	0.936
401-600	0.693	0.286	0.965	0.169	0.527	0.366
601-800	0.370	0.261	0.153	0.388	0.300	0.908
801-1000	0.006	0.984	0.203	0.350	0.019	0.960
1001-1200	0.976	0.212	0.063	0.045	0.928	0.673
1201-1400	0.954	0.671	0.934	0.136	0.653	0.075
1401-1600	0.021	0.512	0.250	0.123	0.269	0.986
1601-1800	0.378	0.254	0.845	0.698	0.924	0.223
1801-2000	0.055	0.539	0.245	0.372	0.836	0.312

Table 2: RNR patterns in the 'burglary-in-residence' data set

days / mts	0 - 7	8 - 14	15 - 21	22 - 28	29 - 35	>35
0-200	0.952	0.608	0.308	0.392	0.448	0.502
201-400	0.007	0.508	0.618	0.967	0.118	0.937
401-600	0.282	0.654	0.043	0.181	0.581	0.920
601-800	0.135	0.279	0.637	0.005	0.615	0.975
801-1000	0.665	0.650	0.012	0.223	0.789	0.848
1001-1200	0.692	0.436	0.232	0.762	0.285	0.738
1201-1400	0.794	0.686	0.811	0.792	0.321	0.161
1401-1600	0.435	0.874	0.916	0.994	0.102	0.072
1601-1800	0.890	0.094	0.731	0.159	0.611	0.642
1801-2000	0.501	0.106	0.317	0.327	0.991	0.587

Table 3: RNR patterns in the 'burglary-of-hotels' data set

days / mts	0 - 7	8 - 14	15 - 21	22 - 28	29 - 35	>35
0-200	0.313	0.273	0.331	0.286	0.928	0.748
201-400	0.300	0.333	0.932	0.665	0.936	0.089
401-600	0.046	0.943	0.346	0.240	0.653	0.668
601-800	0.122	0.974	0.045	0.243	0.077	0.934
801-1000	0.917	0.878	0.368	0.300	0.361	0.215
1001-1200	0.190	0.659	0.289	0.573	0.120	0.886
1201-1400	0.495	0.226	0.733	0.065	0.996	0.287
1401-1600	0.265	0.778	0.152	0.708	0.014	0.878
1601-1800	0.423	0.300	0.241	0.045	0.993	0.683
1801-2000	0.945	0.870	0.608	0.457	0.433	0.125

Table 4: RNR patterns in the ‘burglary-in-stores’ data set

Each sub-category of burglary displays different combinations of spatial and temporal bands at which RNR is statistically significant. Generally, the tables reflect the level of spatial spread of their respective data sets (Figure 1). For example, while ‘burglary-of-residence’ and ‘burglary-in-stores’ show some significant RNR patterns at large spatial bands (between 1,600m – 2000m), the ‘burglary-of-hotels’ only has a RNR pattern up to 1000m. The ‘burglary-of-residence’ shows that the RNR pattern is most serious ($\hat{p} \leq 0.001$) at between 201-400m and within 7-14 days (Table 2). This is the same result as generated by all the burglary crime in Table 1. This indicates that ‘burglary-in-residence’ dominates these bands over the entire burglary data set. Also, of significance is the RNR pattern shown by ‘burglary-of-hotels’, in which three statistically significant clusters of cells within the spatial range 401m to 1000m and temporal range of 15 days to 28 days. This provides not only information about the RNR, but also indicates the size of the space-time neighbourhoods at which clustering of this sub-category is at its highest.

4.3 Proposed RNR map for identifying hotspot regions

Each cell value in Tables 1-4 are generally described as global indicators. This means that each value is a representative of the general pattern across the entire study area, without pinpointing a specific region of interest. In order to enhance the use of these results in practice, RNR maps are proposed so that regions with highest contribution to the RNR pattern can be identified. Examples are shown in Figure 3a and 3b for 2D and 3D RNR maps, respectively. Basically, the map is generated by linking all crime points within specified and statistically significant spatial and temporal bands, based on the results of the Knox test. These linkages may help to identify regions with the highest concentration of RNR crimes, whereby urgent intervention activities may be focussed.

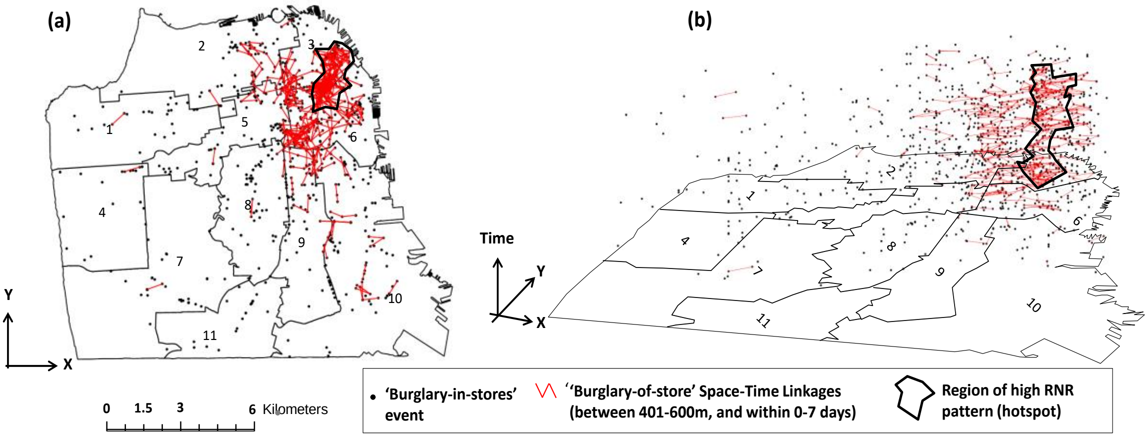

Figure 3: Generating a RNR map of 'burglary-in-stores' (a) 2D RNR map (b) 3D RNR map. The map is based on a statistically significant cell [row=3, col=1] with $\hat{p} = 0.046$ from Table 4 (i.e. linking pairs of 'burglary-in-stores' events that occur between 401 - 600m and within 0 - 7 days of one another). The delineated regions is visibly identifiable hotspot areas where a tailor-made intervention activity could be focussed.

Figure 3 is the RNR map showing space-time linkages of pairs of 'burglary-in-stores' events that occurred between 401-600m and 0-7 days of one another. The events in these linkages generated a statistically significant RNR pattern of $\hat{p} = 0.046$ (see Table 4) The region delineated can be observed as containing a visible concentration of linkages, which may be thought priority with regard to police intervention.

5. Conclusion and Current work

This study presents an investigation into the RNR patterns of three sub-categories of burglary crime, namely; 'burglary-in-residence', 'burglary-of-hotels', and 'burglary-in-stores'. The aim was to demonstrate that sub-categories of major crime types, such as burglary crime, also have different RNR patterns. Further, it is hoped that revealing these patterns may be a more informative way of tackling the RNR patterns for specific crimes. The Knox test was used to investigate these patterns at different combinations of critical spatial and temporal bands that were specifically selected with tactical application in mind. The results showed that sub-categories of a major crime types, for example burglary, may have a distinct RNR pattern and also suggests that in scenarios where specific intervention strategies exists for a sub-category, this mode of investigation has the potential to be of more use operationally.

From the operational view point, different intervention strategy may be applicable to different sub-categories of major crimes. By revealing the peculiar spatial and temporal signatures associated with different crime sub-categories, the law enforcement may be able to device different intervention strategy suitable for each sub-category. Thus, this study provides a potentially new insight into how crime data sets may be handled in the future.

Similar investigation is currently being carried out for other major crime types, including the assault sub-categories (aggravated assault, battery, gun violence etc.); and vehicle thefts sub-categories (theft-of-motor-vehicle, theft-from-motor vehicles, and vehicle damage etc.).

6. Acknowledgements

The author acknowledges the contribution of the government of the city of San Francisco for making the crime incident data publicly available. Also, the author thanks the University College London for providing the computing facilities (parallel computing) that have made this investigation possible.

7. Biography

Monsuru Adepeju is a completing PhD student in Geographic Information Science at the University College London, and a Research Fellow at the University of Leeds, UK. His PhD research focussed on the Spatio-temporal modelling of sparse point processes, with a special application in predictive crime policing. His general research interests include; space-time analytics, geo-computation, and big spatio-temporal data mining.

References

Agresti, A. and Finlay, B. 1997. Statistical Methods for the Social Sciences, 3rd edn. New Jersey: Prentice Hall.

Bowers, K.J. and Johnson, S.D., 2004. Who commits near repeats? A test of the boost explanation. Western Criminology Review, Vol. 5(3), pp. 12-24.

Johnson, S.D. and Bowers, K.J., 2004. The Burglary as Clue to the Future: The Beginnings of Prospective Hotspotting . European Journal of Criminology . Vol. 1 , pp. 237 – 255 .

Johnson, S.D., Birks, D.J., McLaughlin, L., Bowers, K.J. and Pease, K., 2007a. Prospective crime mapping in operational context: Final report. UCL, Jill Dando Institute of Crime Science.

Johnson, S.D., Bernasco, W., Bowers, K.J., Elffers, H., Ratcliffe, J., Rengert, G., Townsley, M. 2007b. Space-time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23:201–219

Knox, G., 1964. Epidemiology of Childhood Leukaemia in Northumberland and Durham. Brit. J. of Prev. and Social Med, Vol. 18, pp. 17-24.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res., Vol. 27, pp.209–220.

Townsley, M., Homel, R., and Chaseling, J., 2003. Infectious burglaries: a test of the near repeat hypothesis. Brit J Criminol, Vol. 43, pp. 615–633.