Weighted-traffic-network–based geographic profiling for serial crime location prediction

2011 EPL 93 68006

(http://iopscience.iop.org/0295-5075/93/6/68006)

A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

March 2011

EPL, **93** (2011) 68006
www.epljournal.org
doi: 10.1209/0295-5075/93/68006

Weighted-traffic-network–based geographic profiling for serial crime location prediction

CHENG QIAN[1], YUBO WANG[2], JINDE CAO[1(a)], JIANQUAN LU[1,3] and JÜRGEN KURTHS[3,4]

[1] *Department of Mathematics, Southeast University - Nanjing 210096, China*
[2] *School of Electronic Science and Engineering, Southeast University - Nanjing 210096, China*
[3] *Potsdam Institute for Climate Impact Research, Telegraphenberg - D-14415 Potsdam, Germany, EU*
[4] *Department of Physics, Humboldt University Berlin - Newtonstr. 15, 12489 Berlin, Germany, EU*

received 20 December 2010; accepted in final form 27 February 2011
published online 28 March 2011

PACS 89.65.Ef – Social organizations; anthropology
PACS 89.75.Hc – Networks and genealogical trees

Abstract – Geographic profiling plays a significant role in serial crime detection nowadays, in which Rossmo's formula is applied for future crime location prediction. However, the limited accuracy and demanding for vast data have largely impeded the efficiency of this technology. In this letter, a traffic network is introduced to geographic profiling. The problem is remodeled with weighted traffic network and the original Euclidean distance is replaced by the shortest path between nodes for better location prediction. A serial crime case is used to validate the correctness, efficiency and robustness of the proposed method. The main contributions of this letter can be concluded as follows: 1) the proposed model displays a higher accuracy and is less dependent on crime data; 2) strong robustness is testified by sensitive analysis, *i.e.* the developed model can produce an accurate prediction based on somewhat inaccurate former crime data; 3) further application in counter-terrorism is put forward with some adjustments.

Introduction. – Geographic profiling has been put forward in order to narrow down the searching area of suspects, and it has proven to be an efficient tool for the investigation of serial crime cases in the past decades [1]. Based on the previous crime sites, geographic profiling predicts an area in which the terrorists may live by following certain underlying patterns [2]. However, traditional geographic profiling aims to identify the terrorists' resident rather than the next possible crime area. Furthermore, in the traditional geographic profiling, the time of the implemented murders was not taken into consideration and this will make the prediction unreliable to some extent. Most importantly, Rossmo's model does not take into account topography, resulting in predicted centers of crime that fall in sparsely inhabited regions (*i.e.* lakes and mountains) where crimes are unlikely.

Motivated by the above discussions, we will use the weighted traffic network [3–6], which has been shown to have small-world network characteristics [7,8], to develop a more accurate next-crime-area prediction method. In this letter, "weight" represents the distance between nodes. We handle the geographic profiling from the perspective of next-crime-area prediction partly because of its potential application in counter-terrorism. Studying serial crime especially terrorism by network has become a hot research topic in recent years. Travers *et al.* introduced complex networks to investigate social systems [9]. Borgatti *et al.* have recently announced that terrorist groups can be widely seen as networks rather than organizations [10]. Ressler modeled the basic structure of terrorists' network in [11]. Likewise, Carley has done some research on the application of complex networks in counter-terrorism under financial support of the U.S. Military [12]. Siqueira presented a game-theoretic study of general terrorist organization [13]. However, these works viewed the terrorist attacks only from a social network perspective. Ou *et al.* discovered the important role traffic network play in the process of resource allocation [14], which provides us with a totally new perspective of the counter-terrorism issue. In addition, Lv *et al.* [15] devoted their research on weighted-network prediction. They pointed out that the weights of links will have a significant influence on the predicted result. Bianconi [16] presented a general framework for weighted networks, in which the evolution

(a) E-mail: jdcao@seu.edu.cn

of the weight of links is discussed. This research suggests us that the weight of a link may also play an important role in the crime location prediction, while former research in this area rarely take it into account. In this paper, we study traffic-network–based geographic profiling for future serial crime location prediction, and our proposed method can be used to successfully solve the topographic problem, making the crime location prediction more reliable and practicable. Moreover, a realistic serial crime case is given to demonstrate the correctness, efficiency and robustness of the traffic-network–based method with sensitive analysis on crime location and time. Furthermore, considering the similarity between serial crimes and terrorist attacks, the proposed model can also be applied to counter-terrorism with certain adjustments.

Construction of the model. –

Rossmo's-formula–based probability estimation. Rossmo's formula was originally put forward for anchor point estimation in serial crimes. It heavily depends on two hypotheses, which are, respectively, the distance decay and the circle hypothesis [1,2]:

i) Distance decay: it has been announced that offenders do not travel far from home to offend and the frequency of offending decreases with increased distance from an offender's home location.

ii) Circle hypothesis: it has been found out [1] that the majority of serial rapists did not typically commit crimes outside a circle with its diameter defined by the distance between the offender's two furthermost offences.

Analysis of different types of crime has given support for these hypotheses [17]. The above two assumptions have been used in psychological studies for a long time, and have made it possible for police officers to drawn a mental map of a suspect.

In this letter, the methodology of the psychological analysis is inherited but with several adjustment for future crime location estimation, we especially include:

i) A distance decay is basically adopted. Considering the fact that a suspect will not go too far for a future crime because of time, psychological and financial restrictions, the possibility of crime decreases as the distance from the offender's crime location increases.

ii) The adapted circle hypothesis is rectified. An offender will rarely commit a next crime just beside his latest one, which is surrounded by police and detectives for certain. Therefore, there exists a buffer area where the radius of this area is a function of time. The longer the elapsed time, the smaller the radius of the buffer zone will be, meaning the suspect might eventually commit a crime near the former crime location.

The traditional Rossmo's-formula–based prediction model does not take into account the influence of the roads and time, thereby it can be written as

$$p_{ijn} = \begin{cases} \dfrac{1}{S_{in}^{f}}, & S_{in} > B_0, \\ \dfrac{B_0^{f-g}}{2B_0 - S_{in}^{g}}, & S_{in} \leqslant B_0, \end{cases} \tag{1}$$

where f, g are empirically determined exponents that dictate the form of the curve; S_{in} denotes the Euclidean distance between the studied spot on the map and the n-th crime site (namely $S_{in} = \sqrt{(x_i - x_n)^2 + (y_i - y_n)^2}$); $(x_i,\, y_i)$ and $(x_n,\, y_n)$ are the coordinates of the studied spot on the map and the n-th crime site, respectively; B represents the average distance between the offenders' crime points. Therefore, we can calculate the hit score of every point $(x_i,\, y_j)$ by

$$P_{ij} = \sum_{n=1}^{c} p_{ijn}. \tag{2}$$

Here c is the number of past crimes and P_{ij} is the total hit score of the point $(x_i,\, y_i)$, *i.e.* the former crime data (1st to $(c-1)$-th) is used to predict the c-th crime location. Here we assume the hit score of the places on the path obeys linear variation between nodes approximately. The higher the hit score is in every point, the more likely the anchor point will be.

Traffic-network–based probability estimation. Now, we present our traffic-network–based probability estimation. It is mainly based on a weighted network which is obtained from satellite maps. The nodes represent either bends in roads or road junctions of 2 roads [18]. Nodes are connected if there are roads between them, and the distance between 2 nodes denotes the Euclidean distance in reality. With the above approach, a satellite map can be converted into an abstract map which is much easier to analyze. By sampling enough nodes, the abstracted node map can reach a satisfactory accuracy. The continuous surface is then discretized to make calculations practicable. In particular, in the new model, it should be pointed out that the buffer zone is not a constant value as the time passes by. With passing time, radius will become smaller according to the circle hypothesis, and here it is assumed to satisfy an exponential decay:

$$B(t) = B_0 e^{-a_0 t}, \tag{3}$$

where B_0 refers to the average distance between the offenders' crime points and a_0 represents the probability decrease rate.

Under the above discretization, the modified Rossmo's formula can be written as follows:

$$p_{ijn} = \begin{cases} \dfrac{1}{D(x_i, y_j, x_n, y_n)^f}, & D(x_i, y_j, x_n, y_n) > B(t), \\ \dfrac{B(t)^{f-g}}{2B(t) - D(x_i, y_j, x_n, y_n)^g}, & D(x_i, y_j, x_n, y_n) \leqslant B(t), \end{cases} \tag{4}$$

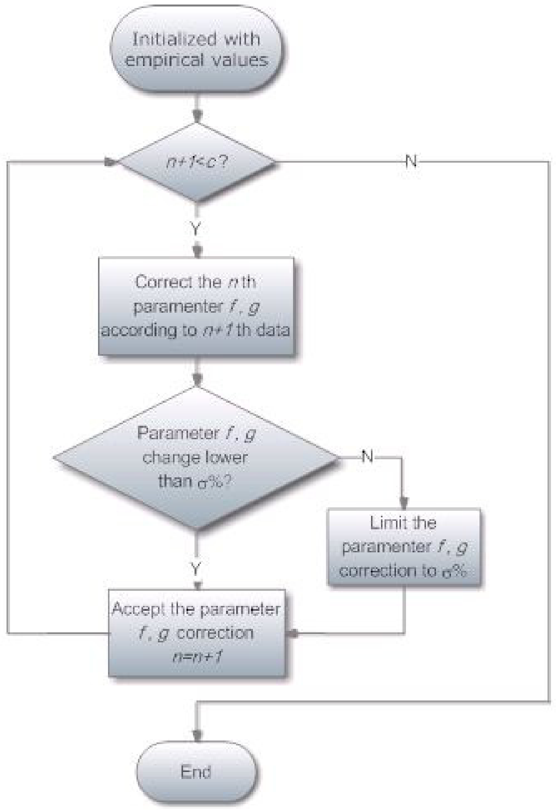


Fig. 1: Parameter tuning of c-th crime prediction.

where D donates the shortest distance between (x_i, y_i) and (x_n, y_n). In the equation above, the shortest path replaces the Euclidean distance between the studied spot in the original Rossmo's model. By adopting the distance decay hypothesis between nodes, the Floyd-Warshall algorithm is used to calculate the shortest path between $(x_i,\ y_i)$ and $(x_n,\ y_n)$ [19]. Through the above adjustment and following the remaining steps described above, a map of the jeopardized surface is obtained in which the hit score of every point can be figured out. The parameters B_0, a_0, f and g, which vary with different offenders and crime types, can be determined by former crime data with parameter tuning. Our procedure of the parameter tuning (fig. 1) is similar to the learning algorithm of neural network. Compared to the conventional Rossmo's model, the parameters f and g are not only determined empirically, in addition, they are revised by the crime information before. From the process of the parameter tuning, the offender's characteristic will be revealed by f and g more accurately.

Case study: serial killer. –

Background knowledge for a realistic serial crime case. This way we have provided a new traffic-network model to achieve next-crime-area prediction. Next, a real serial crime case, in which the offender committed 13 murders within a period of time, is applied to test the proposed model. The case studied is the serial murder case that lasted from 1975 to 1980. The offender, Peter William Sutcliffe, an English serial killer, was convicted in 1981 of murdering 13 women and dubbed The Yorkshire Ripper [20]. Figure 2(a) shows a satellite map of the body dumping places of the above 13 murders. A traffic network is generated as shown in fig. 2(b), which is abstracted from the satellite map. The roads of the studied area have not changed much, since the time serial murder was committed. It can be observed that these body dumping places are all very close to the path. Here, these crimes are calculated to be the nearest point on the path, because offenders are more likely to crime on the road and dump the body to an unobvious position near the path.

With the above traffic network, the possibility of the anchor point can be figured out with our model.

Table 1: Comparison of the two models.

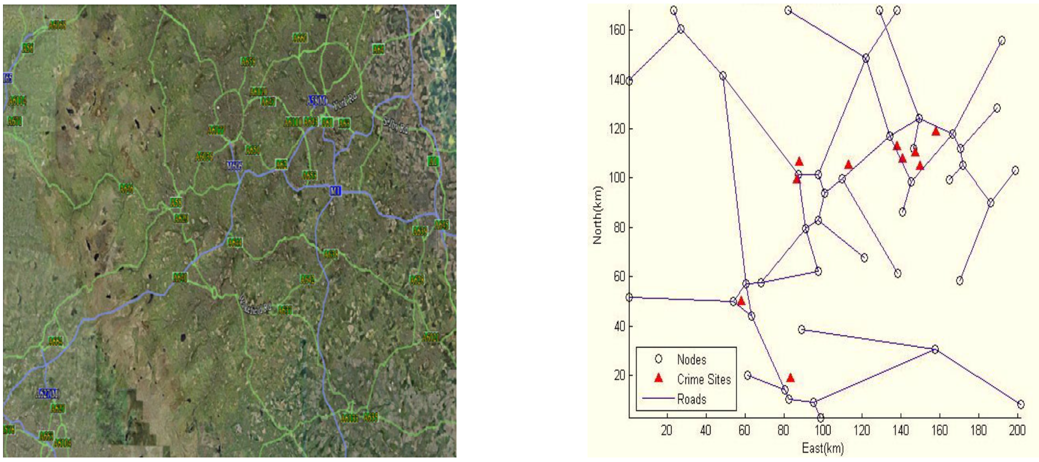
	Rossmo's-formula–based estimation	Traffic-network–based estimation
Cases with a	10	4, 5, 7
better prediction	12	8, 11, 13

Table 2: Distance between the center of the predicted location and the real crime location.

The number of murder	Rossmo's-formula–based estimation (km)	Traffic-network–based estimation (km)
4th	23.31	7.58
5th	19.02	10.87
7th	11.46	6.84
8th	20.28	15.96
10th	4.95	7.82
11th	11.28	8.97
12th	5.11	6.52
13th	7.72	7.32

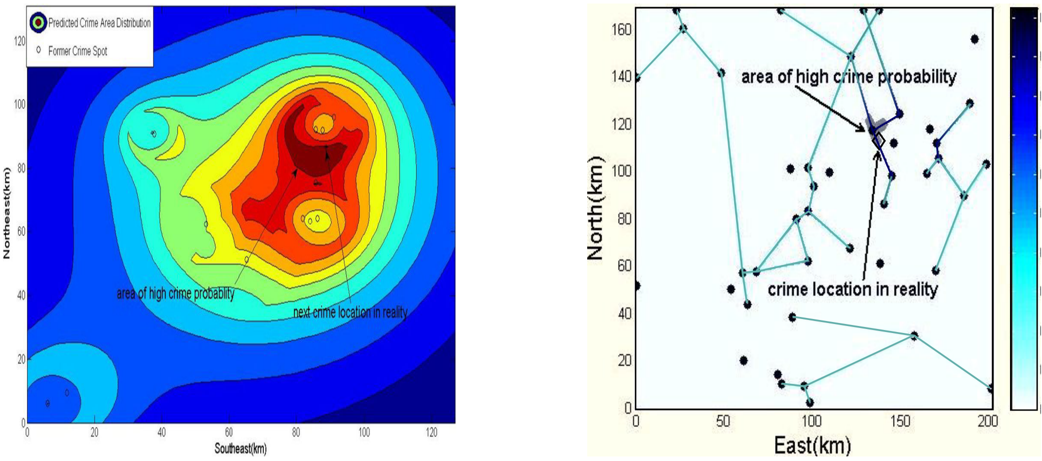
Predicted results. In order to validate the potentials and efficiency of the proposed model, a 13th crime location prediction is made with the former 12 crime data. Using our model, the results are presented in fig. 3. Furthermore, Rossmo's formula prediction without the utilization of traffic network is also presented in order to provide making a clear comparison with the proposed model. In addition, the 4th to 13th murders are correctly predicted by the above two models, and the corresponding results are presented in table 1[1]. In addition, table 2 presents a details of our results' comparison. However, it should be pointed out that if the prediction is to be made with less than 3 sets of data, the predicted area will be considerably large, which becomes less practical in reality. Therefore, the prediction is made in this letter from the 4th murder. It

[1]It should be pointed out that the 6th and 9th murder locations cannot be well predicted by both models.



(a) (b)

Fig. 2: (Colour on-line) Traffic-network generation: (a) satellite map with road and location of victims; (b) abstracted traffic network.



(a) (b)

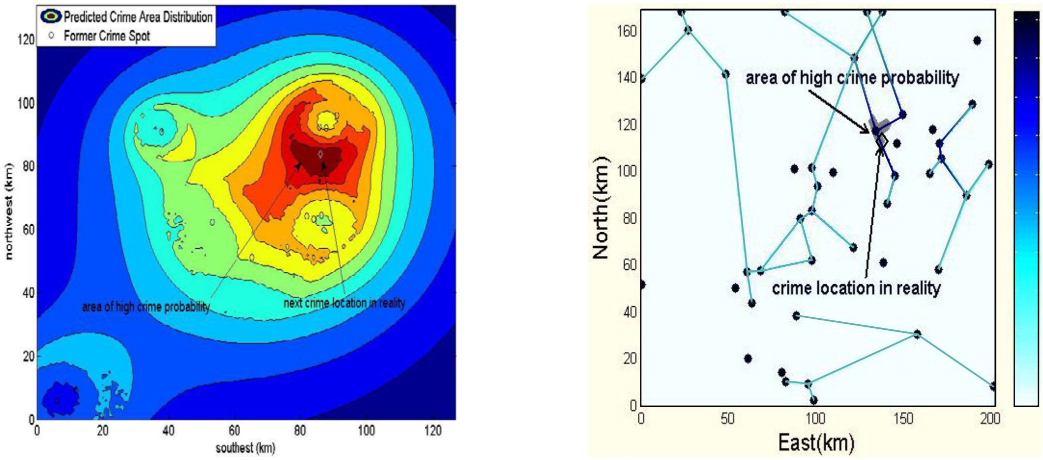
Fig. 3: (Colour on-line) Prediction of the 13th crime location: (a) estimation by Rossmo's formula; (b) estimation based on a complex network.

is clear that besides the 10th and the 12th murder location, the traffic-network–based probability estimation produces much better predictions than the Rossmo's-formula–based model.

From fig. 3 and table 1, it can be concluded that the traffic-network–based approach implies a smaller predicted region and a more accurate estimation than the model merely using Rossmo's formula. By introducing traffic information to the crime location prediction, the traffic-network–based model requires less data for an accurate prediction than Rossmo's formula. Furthermore, a smaller predicted region enables police agency consume a smaller human and financial resources in searching for the suspect.

The application of the proposed model on Peter William Sutcliffe relies on the fact that the offender's crime location are near the road. However, for some serial crime cases that do not occur near the road, networks should be resettled to precisely characterize the crime pattern. It should be pointed out that, since the traditional Rossmo's-formula–based model does not consider the topography, it might lead to a dilemma that the predicted area happens to be lakes or remote mountains which do not likely to commit a crime, while our traffic-network–based model can successfully avoid this problem.

Sensitive analysis. In the above discussion, we have made one basic assumption, which regards the body dump location as the murderer's crime location. However, in reality, more often than not, there exists a certain distance between the crime location and its body dumping place. Hence, a sensitive analysis should be performed to validate



(a) (b)

Fig. 4: (Colour on-line) Prediction of the 13th crime location with 8% white noise disturbance. (a) Estimation by Rossmo's formula; (b) estimation based on a complex network.

the robustness of the proposed model, where Peter's case is also used as data source. The uncertainty is added to the coordinate of the former crime locations in order to testify the robustness of the proposed models. Figure 4 shows the comparison of both models predicting the 13th murder under 8% white noise in their former crime sites coordinates. By comparing with fig. 3, the robustness of our model can be clearly observed: the contour line of Rossmo's formula prediction is changed, while our traffic-network–based prediction stays nearly the same. Therefore, the robustness of the traffic-network–based model is far better than Rossmo's model, *i.e.* the proposed model can still produce an accurate prediction even based on inaccurate former crime data.

Moreover, the time interval between crime committing and body discovery cannot be neglected in most of the cases. Therefore we have also taken time uncertainty into consideration, and found out that our method shows stronger robustness than Rossmo's-formula–based estimation method.

Application in counter-terrorism. – Another potential application of our approach is counter-terrorism. Terrorism, which threatens the safety of the citizens, becomes an important issue to be discussed nowadays. In fact, terror attacks have become highly organized and serial in the 21st century. These characteristics display a similarity to the serial crimes discussed above, which makes the application of the Rossmo's formula in counter-terrorism possible.

As terror attacks are world wide, the terrorist must travel to the destiny through transportation. Therefore, an airline-based counter-terrorism model can be developed with some adjustments to the serial crimes prediction model. Since airlines can be converted into a weighted network [21], the searcher for the terrorism or prevention of the terrorist attack can be implemented along the airlines. By predicting the future terror attacks on the basis of the former terror attacks, the authorities can enhance the level of security check in some airlines, which will prevent tragedy from happening.

If the authorities focus on countering terrorism in a smaller region, railway stations or terminals can also be added as the nodes. Through continuously increasing the nodes in the traffic network, the proposed model will reach a reliable accuracy as well as narrowing down the area with a high terror attack probability.

Conclusion and discussions. – In summary, we propose a traffic-network–based future crime location prediction method, which is derived from the Rossmo's formula by replacing the original Euclidean distance with the shortest path between nodes. The proposed model is tested with a realistic serial crime case which demonstrates better prediction performance over merely using Rossmo's formula. A sensitive analysis is also performed to show the robustness of the complex-network–based model. Its further application in counter-terrorism is also discussed and analyzed. However, such application is difficult to be tested without enough data source, and hence it demands further development.

* * *

JC was supported by the National Natural Science Foundation of China under Grant 60874088, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20070286003, and the Natural Science Foundation of Jiangsu Province of China under Grant BK2009271. JL was supported by the National Natural Science Foundation of China (NSFC) under Grant 11026182, the Natural Science Foundation of Jiangsu Province of China under Grant BK2010408, and the

Innovation Fund of Basic Scientific Research Operating Expenses of Southeast University 3207010501, and the Alexander von Humboldt Foundation of Germany. JK was supported by SUMO (EU) and ECONS (WGL).

REFERENCES

[1] Rossmo D., *Geographic Profiling* (CRC Press) 1999, pp. 67–80.
[2] Block R. and Bernasco W., *Offender Profiling*, **16** (2009) 187.
[3] Latora V. and Marchiori M., *Phys. Rev. Lett.*, **87** (2001) 198701.
[4] Arenas A., Díaz-Guilera A., Kurths J., Moreno Y. and Zhou C., *Phys. Rep.*, **469** (2008) 93.
[5] Rosvall M., Trusina A., Minnhagen P. and Sneppen K., *Phys. Rev. Lett.*, **94** (2005) 028701.
[6] Ashton D. J., Jarrett T. C. and Johnson N. F., *Phys. Rev. Lett.*, **94** (2005) 058701.
[7] Sen P. *et al.*, *Phys. Rev. E*, **67** (2003) 036106.
[8] Angeloudis P. and Fisk D., *Physica A*, **367** (2006) 553.
[9] Travers J. and Miligram S., *Sociometry*, **32** (1969) 425.
[10] Borgatti S., Mehra A., Brass D. and Labianca G., *Science*, **323** (2009) 892.
[11] Ressler S., *Homel. Secur. Aff.*, **2**, Issue No. 2 (2006).
[12] Bohannon J., *Science*, **325** (2009) 410.
[13] Siqueira K. and Sandler T., *Public Choice*, **142** (2010) 237.
[14] Ou Q., Jin Y., Zhou T., Wang B. and Yin B., *Phys. Rev. E*, **75** (2007) 021102.
[15] Lv L. and Zhou T., *EPL*, **89** (2010) 18001.
[16] Bianconi G., *Europhys. Lett.*, **71** (2005) 1029.
[17] Kocsis R., Irwin H., Hayes A. and Nunn R., *J. Interpers. Violence*, **15** (2000) 311.
[18] Montis A., Barthelemy M. and Chessa A., *Environ. Plan. B*, **34** (2007) 905.
[19] Floyd R., *Commun. ACM*, **5** (1962) 345.
[20] `http://en.wikipedia.org/wiki/Peter_Sutcliffe`.
[21] Bagler C., *Physica A*, **387** (2008) 2972.