# Urban Analytics in the Context of Public Safety: The Case of Avoidance Patterns

Emre Eftelioglu, Cargill Inc., USA

**Abstract**

*Given a collection of geolocated activities, the goal of urban analytics in the context of public safety is to discover the underlying motives of people that affect their movement/activity patterns in space and time. Understanding the spatial patterns from urban mobility/activity datasets is an important task in public safety, city planning and sociology since these may reveal the underlying causes of crimes and safety issues, as well as behavior changes of individuals. Avoidance patterns are a type of behavioral change characterized by a lack of movement contrary to expectation. Avoidance pattern detection is a challenging task due to the lack of observations (e.g. lack of movement), defining the expected "normal" movement and large datasets (i.e. high number of GPS trajectories which are spread across the study area and large road network graphs). In addition, these challenges are exacerbated by the complicated and often hidden drivers of human activities and the complex relationships and dependencies between the spatially associated features.*

*In this paper, we will provide a brief overview of the state-of-the-art spatial data science approaches in the context of avoidance patterns. First, we introduce the background from the domain (i.e. public safety) perspective, followed by an overview of the current state-of-the-art work. Then we will discuss possible future directions that may help shape future research on the topic.*

## 1 Introduction and Motivation

With the increasing availability of geolocated data collected from a variety of sources, there is a tremendous opportunity to understand the movement and activity patterns of people [3]. These patterns are influenced by many motives which include the underlying demographics, goals (e.g. sightseeing, shopping, work-home commute, etc.), road conditions, etc. One of such patterns is avoidance. Avoidance patterns are the locations where drivers/pedestrians try to bypass when commuting. These are the result of a variety of driver concerns including rush hour traffic when there is congestion, road imperfections (e.g. potholes, etc.), safety of a neighborhood as well as hiding from detection (e.g. criminals' avoidance behavior).

Avoidance patterns are overlooked compared to other work on urban mobility analytics despite their importance to understand human behavior. This is due to the ease of focus on the observable phenomena rather than the lack of phenomena in urban analytics.

One way to define the avoidance pattern is as the area between the shortest path and the taken path by the driver [7]. Another definition may be the segment of the shortest path that is different than the taken path. These different definitions of avoidance patterns can be generalized as discovering a region (e.g. polygon, road segment, etc.) which lacks movement when it is expected.

Figure 1 shows an illustration of an avoidance. Suppose the blue line represents a shortest path between green-tagged start and end locations; and the red line represents the actual path taken by a driver. Depending on

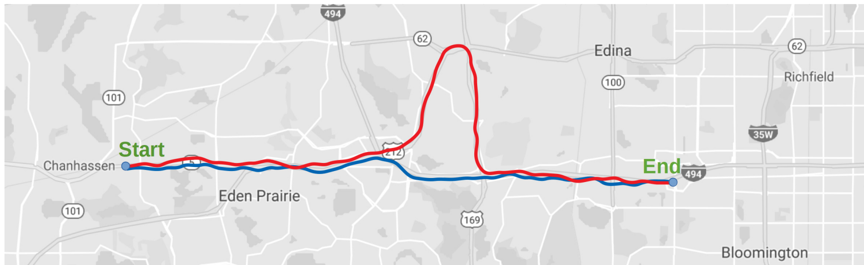



Figure 1: An illustration of an avoidance on a road network.

the definition of an avoidance, the output can be the region(s) enclosed by these two paths or a segment of the shortest (or expected) path that is different than the taken path.

In the following sections, we will provide some example application domains of avoidance pattern discovery as well as how these may help improve domain officials' work.

## 1.1 Avoidance Patterns in City Planning and Sociology

"Safety" or "Feeling Safe" is an often overlooked aspect that affects movement and activity patterns.

Feeling unsafe may affect the socialisation and activity patterns of individuals [8]. Thus, sociologists and city officials often investigate the neighborhoods by their demographic structures and take the necessary measures to mitigate the severity of feeling unsafe and preventing economic loss caused by the stigmatization. For example, some larger cities have distressed neighborhoods that are known to be riskier. These neighborhoods are not always spoken publicly but locals often know and avoid them. Such neighborhoods may not have been dangerous or risky before, but these characteristics may emerge gradually. Since, it is harder to do surveys/sociological analysis frequently, they may be undetected until it is too late. However, certain occupations are more sensitive to these changes and this fact can be leveraged instead of relying on the relatively sparse collection of surveys. One such occupation is taxi driving, where taxi drivers learn these neighborhoods through the experiences of each other and avoid entering them. For example, in the Chinese city of Kunming, taxi drivers try to not take customers from the regions where marginalized people are thought to be living [17]. Similarly, as shown in Figure 2, crowdsourced mobile applications such as Waze [1], allow users to flag some regions as dangerous, and let their users plan their routes accordingly.

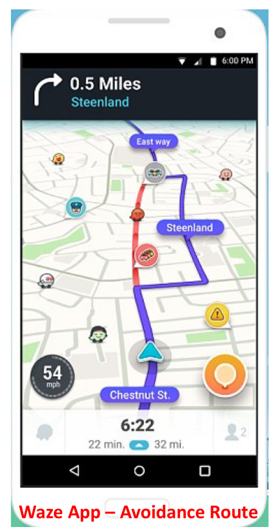



Figure 2: Screenshot of Waze app [1]

Nowadays, taxis are equipped with GPS devices due to their cheap availability and legal reasons to prevent conflicts with customers. One may leverage the GPS trajectory data collected from those devices [9] to identify the regions where taxi drivers avoid. In addition, since these datasets are collected in real time, the emergence of avoidance regions will be noticed much quicker than traditional surveys. Thus, the risk of a neighborhood being stigmatized can be prevented before the word of "unsafe" is widespread to all residents. City officials may mitigate the negative public opinion by updating their policies and planning more investments.

## 1.2 Avoidance Patterns in Law Enforcement

In the previous section, we provided a wide-lens perspective on the use cases of avoidance patterns for law-abiding citizens. However, sometimes it is not enough to solve the neighborhoods' sociological problems without getting into the criminal mind. Thus, in this section, we are going to introduce the avoidance patterns from a criminal mind's perspective.

A fundamental task in criminology is the analysis of crime locations and locating the criminal to prevent more crimes [4]. In the past, these analysis were done manually using paper maps and pins, but nowadays modern law enforcement agencies around the globe use specialized tools and spatial analysis techniques to automate this process as well as improve their accuracy and efficiency (e.g. CrimeStat [15]).

There is one particular problem with these tools: To successfully use them, multiple crimes should occur and this will cause a delay in the detection of a criminal behavior as well as cause more harm to society. In addition, even if a criminals' location is estimated by these tools, these will neither pinpoint the exact location of the criminal nor find their mobility behavior causing an extensive search by the security officials in field.

To overcome these issues, the mobility behavior of vehicle trajectories can be analyzed. Criminals often avoid some locations where there are security cameras or law enforcement checkpoints since these may lead to their arrest or identification [14]. Thus, given a set of trajectories and security checkpoints/cameras, suspicious behavior of a trajectory can be identified by its difference from the expected path.

Another preventive measure may be to identify loops in trajectories. These will also cause specific regions to be flagged as avoidance regions. However, the pattern may not have an intention of avoidance, but a result of surveillance. For example, criminals may do a surveillance around a target crime site (e.g. a bank for robbery). Such looping/circling behavior by an individual may be for surveillance. Thus, identifying such trajectories and the individuals who created them may help public security officials prevent crimes before they occur.

### 1.3 Avoidance Patterns in Transportation Planning

Transportation planners often deal with multiple data sources including cameras, road sensors, loop sensors, accident data, etc. to understand the flow of traffic throughout the day. These datasets are then used to improve design and synchronization of traffic lights, fix the flawed road segments, and plan new roads or increase the capacity of existing ones. One particular need of transportation planners is to understand the driver behavior under different road conditions [2].

For example, long term residents of cities know where and when the traffic congestion happens and avoid these locations even though this may result with a longer route to destination. Similarly, when there are structural problems (e.g. potholes, cracks, etc.) or there are construction zones on the road, locals know and avoid these areas. For example, the magnitude of potholes in Figure 3 may cause drivers' to use another road instead of a shorter one. Using the GPS trajectory data collected from location based applications (e.g. Google Maps, Apple Maps, Waze, etc.), transportation planners may better understand the driver behavior and the underlying causes.

Figure 3: Potholes that may cause drivers' avoidance behavior [19]

## 2 Related Work

There have been several attempts to use mobility datasets (i.e. trajectory datasets) to identify interesting patterns that can be further analyzed by domain scientists. Mobility datasets (i.e. GPS trajectories) are large sets of points with ordered timestamps (Figure 5(a)). Due to the imperfection of GPS devices as well as minor variability caused by driver behavior (e.g. lane changes, speed differences), it is hard to do analysis on these datasets without any pre-processing.

To overcome such difficulties as well as reduce the computational cost, some studies discretize the space into grids, and use grid cells to represent trajectories. For example, several works analyze anomalous trajectories to identify the taxi drivers who were using longer paths for their customers to increase the bill [24, 5]. This is done by representing the trajectories as grid cells and comparing these sets of cells with the set of cells that were representing the appropriate route for a source and destination pair. Grid based mobility analysis is used for other patterns as well. Some example work includes identifying the outlying trajectories by using their grid cell representations, and clustering these to understand the movement behaviors as well as the transportation modes [11, 27]. However, the output of these studies are sensitive to the selected grid cell sizes. Selecting a large grid cell size causes large areas to be flagged as outliers or miss them entirely. Also, selecting a too small cell size may make the comparison impossible. In addition, since the graph notion of the road network is not taken into account, the outputs can be unrealistic especially for vehicle trajectories.

Hence, there are other works that use the GPS points of trajectories, instead of their grid cell representations, to classify trajectories by their transportation modes (e.g. walking, cycling, driving, etc.) [28, 26, 21, 25], to infer the Points of Interests (POI), and to understand the public transportation behavior, i.e. finding preferred paths instead of shortest paths [6, 12, 23, 18, 22, 10].

However, the aforementioned approaches lack three important considerations. First, these works consider GPS trajectories as either a set of points or cells instead of a single trajectory entity. Second, they do not account for the underlying road network structure. However, most human mobility patterns are dependent on the roads and those roads affect mobility behavior. Third, they focus on the presence of mobility but sometimes the lack of movement, when it is expected, may be more interesting.

## 3 Avoidance Pattern Discovery

Avoidance patterns may be observed in different applications domains. Some avoidance patterns such as aircrafts avoiding extreme weather events, or a predator species avoiding another's territory [20] may occur in Euclidean space but most human activities on land occur on road networks. In addition, the minor perturbations as well as the driving behaviors (e.g. frequently changing to left, middle or right lanes) can be compensated by the help of the road network. Given a GPS trajectory with $tr = [p_1 \rightarrow p_2 ... \rightarrow p_n]$ where each point $p_i = (x_i, y_i, t_i) \in tr$ and a spatial network graph $G = (V, E)$ where each road intersection is represented by vertices ($v \in V$) and each street segment is represented by edges ($e \in E$), it is possible to match the trajectories on the road network to represent them by a collection of nodes and edges [16, 13] as shown in Figure 5(b).

Once trajectories are map-matched, the idea behind the Avoidance Region Discovery [7] is to compare them with a path which should be used by the normal drivers. Normal behavior can be defined as a shortest path between the source and destination of the trajectory. Thus, once a trajectory $tr_i$ and a shortest path $sp_i$ are compared, the edges and the nodes are used to create a set of polygons which represent an avoidance polygon set for that trajectory. Figure 5(d) shows an example avoidance polygon in red. When the road network graph is bigger and the trajectory-shortest path differences are in multiple locations, these polygons will create an avoidance polygon set for that pair.

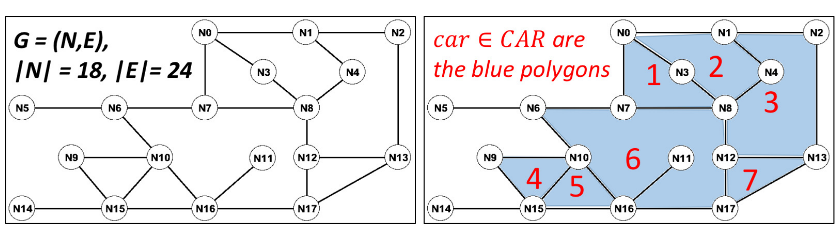
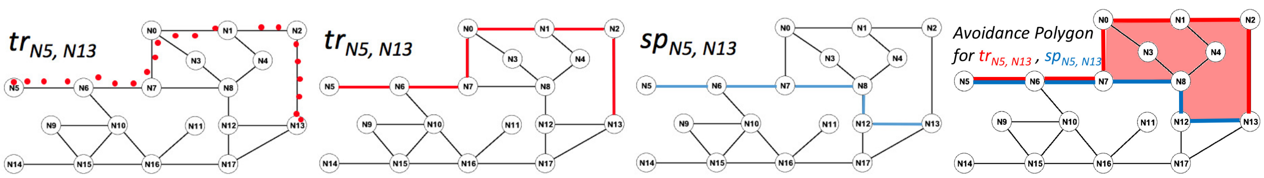



Figure 4: Illustration candidate avoidance regions (graph faces). Figure is excerpted from [7].

Avoidance polygons for each trajectory may be in different locations throughout the space (i.e. each trajectory may have different source and destination as well as shortest path). However, to evaluate the regions which are avoided by more than one trajectory, one will need the provide a consistent set of candidate avoidance regions. To overcome this issue, the space can be discretized to smaller polygons which are represented by the faces of a graph. Using Euler's theorem for planar graphs, the number of candidate avoidance patterns ($CAR$) on road network will be $|CAR| = |E| - |N| + 2$.

Using the count of the avoidance (denoted as $c$) for a region may be misleading. For example, close to a city center, there may be many candidate avoidance patterns that are avoided due to the higher number of trajectories intersecting them. However, in a rural area this number will plummet because of the fact that the number of trajectories in that location will be lower. Therefore, for each candidate avoidance region, the expectation of non-avoidance count should be known as well. To do this, the number of shortest paths that intersect avoidance polygons are counted (denoted as $nc$) and these counts are propagated to the candidate avoidance regions that are covered by those avoidance polygons.




(a) Trajectory Points (b) Map-matched Trajectory (c) Shortest Path (d) Avoidance Polygon

Figure 5: An example trajectory (5(a)), its map-matched edge representation(5(b)), corresponding shortest path(5(c)) and this pair's avoidance polygon(5(d)) (excerpted from [7]).

Finally, by defining a metric, i.e. interestingness ratio ($I = \left(\frac{c}{c+nc}\right) \times c$), which takes both avoidance and non-avoidance counts into account, is used to compute an interestingness score for each of the candidate avoidance polygons. The ones which exceed a specified threshold ($\lambda$) on this metric are flagged as interesting avoidance patterns.

This overall process is a challenging task due to the large number of candidate avoidance regions (CAR) which is related to the size of a road network graph (e.g. $10^6$ edges in a road graph) and the large number of trajectories that can be collected from GPS devices (e.g. $10^6$ trajectories per year for a large city's taxis). To overcome this challenge, [7] proposed an avoidance region miner algorithm that creates a road network subgraph that includes the road segments which were used by the trajectories instead of the whole road network graph. In addition, the authors proposed a pruning algorithm that eliminates the computation of the metric when it is proved to not exceed the specified threshold. Nevertheless, the proposed algorithm's pruning methods do no reduce the worst case complexity.

# 4 Discussion and Future Directions

The example work on Avoidance Region Discovery in the context of public safety is a starting point but there are still opportunities for improvement.

**Identifying and Comparing with the Non-Shortest Paths:** Although shortest paths are a logical choice for most of the drivers, sometimes non-shortest paths are more convenient due to the travel times, speed limits, road conditions (many turns vs. straight driving), rush hour, etc. Therefore, one may argue that the shortest path assumption for a comparison with trajectories may not be valid. In those cases, the preferred paths can be used.

**Minor vs. Major Deviations of Trajectories:** The current state of the art doesn't distinguish between minor and major deviation between shortest paths and trajectories. However, this may be particularly important depending on the use case. For example, the minor deviations may be used in the public safety domain since in case of avoiding security cameras/checkpoints the deviations may be minor but when there is a rush hour congestion the deviation may be greater (e.g. avoiding the city center at rush hour).

**Privacy Concerns:** One of the key things needed in the context of urban mobility analytics is datasets. Due to the privacy concerns, this is not usually possible unless Volunteered Geographic Information (VGI) sources are used. However, since the people with suspicious behavior may not be willing to share their locations, it is hard to collect these. In addition, tagging each trajectory with the driving individual's ID may violate privacy rights because these will point to the source and destination, consequently the locations where people live. Thus, some of the capabilities of the avoidance pattern discovery may be limited such as distinguishing between the population and individual avoidance behaviors.

**Statistical Significance:** In [7], the avoidance and non-avoidance counts were used but the interestingness ratio metric does not provide a statistical significance value for the output. Thus, spurious/chance patterns may exist in the output. One approach may be to understand the distribution of trajectories over the study area and using this distribution in to provide meaninful significance values for the output.

**Emerging Avoidance Regions:** One of the most interesting applications of avoidance region discovery is the detection of emerging such regions. For example, a structural damage to a road segment may occur over time by the traffic and/or weather conditions. Thus, the temporal information related to the trajectory datasets can be used to find some long term (e.g. structural damage) and short term (e.g. accidents causing congestion) emerging avoidance regions.


## Acknowledgments

This work is self supported and is not affiliated with any institution. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the affiliations of the authors. Author(s) would like to thank to the Data Science team at Cargill as well as Spatial Computing Research Group at the University of Minnesota for their constructive feedbacks.


## References


[1] W. Application. Waze map safety re-route, 2017. http://www.cbsnews.com/waze-map-app-suicide-straight-safety-reroute-high-crime-neighborhoods/.

[2] R. Arnott and K. Small. The economics of traffic congestion. *American scientist*, pages 446–455, 1994.

[3] S. Bouton, S. M. Knupfer, I. Mihov, and S. Swartz. Urban mobility at a tipping point. *McKinsey and Company*, 2015.

[4] P. J. Brantingham, P. L. Brantingham, et al. *Environmental criminology*. Sage Publications Beverly Hills, CA, 1981.

[5] C. Chen et al. Real-time detection of anomalous taxi trajectories from gps traces. In *Int. Conf. on Mobile and Ubiquitous Systems*, pages 63–74. Springer, 2011.

[6] C. Cheng et al. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*, volume 13, pages 2605–2611, 2013.

[7] E. Eftelioglu, X. Tang, and S. Shekhar. Avoidance region discovery: A summary of results. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 585–593. Society for Industrial and Applied Mathematics, 2018.

[8] K. F. Ferraro. *Fear of crime: Interpreting victimization risk*. SUNY press, 1995.

[9] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban

data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, 2013.

[10] Y. Fu et al. Real estate ranking via mixed land-use latent models. In *In Proc. of 21th SIGKDD*, pages 299–308. ACM, 2015.

[11] Y. Ge et al. Top-eye: Top-k evolving trajectory outlier detection. In *Proc. of 19th ACM Int. Conf. of CIKM*, pages 1733–1736. ACM, 2010.

[12] F. Giannotti et al. Trajectory pattern mining. In *In Proc. of 13th SIGKDD*, pages 330–339. ACM, 2007.

[13] P. Karich and S. Schröder. Graphhopper. *http://www. graphhopper. com*, 2014.

[14] S. Leman-Langlois. The local impact of police videosurveillance on the social construction of security. *Technocrime: Technology, Crime and Social Control*, pages 27–45, 2008.

[15] N. Levine et al. Crimestat iii: a spatial statistics program for the analysis of crime incident locations (version 3.0). *Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice*, 2004.

[16] P. Newson and J. Krumm. Hidden markov map matching through noise and sparseness. In *Proc. of the 17th SIGSPATIAL Conf.*, pages 336–343. ACM, 2009.

[17] B. E. Notar. 10 off limits and out of bounds taxi driver perceptions of dangerous people and places in kunming, china. *Rethinking Global Urbanism*, pages 190–207, 2012.

[18] P. Wang et al. Human mobility synchronization and trip purpose detection with mixture of hawkes processes. In *In Proc. of 23th SIGKDD*, pages 495–503. ACM, 2017.

[19] Wikimedia commons, Max Ronnersj. Pothole — Wikipedia, the free encyclopedia. [Online; accessed 10-23-2018].

[20] H. Wolf. Odometry and insect navigation. *Journal of Experimental Biology*, 214(10):1629–1641, 2011.

[21] C. Xu et al. Identifying travel mode from gps trajectories through fuzzy pattern recognition. In *Int. Conf. on Fuzzy Systems and Knowledge Discovery*, volume 2, pages 889–893. IEEE, 2010.

[22] J. Yuan et al. Discovering regions of different functions in a city using human mobility and pois. In *In Proc. of 18th SIGKDD*, pages 186–194. ACM, 2012.

[23] Q. Yuan et al. Time-aware point-of-interest recommendation. In *Proc. of 36th Int. SIGIR Conf. on Research and Development in Information Retrieval*, pages 363–372. ACM, 2013.

[24] D. Zhang et al. ibat: detecting anomalous taxi trajectories from gps traces. In *Proc. of 13th Int. Conf. on Ubiquitous Comp.*, pages 99–108. ACM, 2011.

[25] L. Zhang, S. Dalyot, D. Eggert, and M. Sester. Multi-stage approach to travel-mode segmentation and classification of gps traces. *Int. Arch. of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(W25):87–93, 2011.

[26] Y. Zheng et al. Understanding mobility based on gps data. In *Proc. of 10th Int. Conf. on Ubiquitous Comp.*, pages 312–321. ACM, 2008.

[27] Y. Zheng et al. Mining interesting locations and travel sequences from gps trajectories. In *Proc. of 18th Int. Conf. on WWW*, pages 791–800. ACM, 2009.

[28] Y. Zheng et al. Understanding transportation modes based on gps data for web applications. *Transactions on the Web (TWEB)*, 4(1):1, 2010.