ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/112872/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Hannah, Charlotte, Spasic, Irena and Corcoran, Padraig 2018. A computational model of pedestrian road safety: the long way round is the safe way home. Accident Analysis & Prevention 121 , pp. 347-357. 10.1016/j.aap.2018.06.004

Publishers page: http://dx.doi.org/10.1016/j.aap.2018.06.004

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

A Computational Model of Pedestrian Road Safety: the Long Way Round is the Safe Way Home

Charlotte Hannah

School of Mathematics, Cardiff University, Wales, UK.

Irena Spasić

School of Computer Science & Informatics
Cardiff University, Wales, UK.

Padraig Corcoran

School of Computer Science & Informatics
Cardiff University, Wales, UK.

Abstract

We propose a novel linear model of pedestrian safety in urban areas with respect to road traffic crashes that considers a single independent variable of *pedestrian path safety*. This variable is estimated for a given urban area by sampling pedestrian paths from the population of such paths in that area and in turn estimating the mean safety of these paths. We argue that this independent variable directly models the factors contributing to pedestrian safety. This contrasts previous approaches, which, by considering multiple independent variables describing the environment, traffic and pedestrians themselves, indirectly model these factors. Using data about 15 UK cities, we demonstrate that the proposed model accurately estimates numbers of pedestrian casualties.

Keywords: Pedestrian safety, Modelling, Road traffic crashes, Pedestrian paths, Pavement network.

1. Introduction

Each year about 1.24 million deaths worldwide result from road traffic crashes making this the eighth leading cause of death globally [1]. According to the UK Department for Transport, pedestrians accounted for 24% of all road deaths in Great Britain in 2015 [2]. Given this, modeling pedestrian safety with respect to road traffic crashes is an important research objective. In this work we only consider safety with respect to road traffic crashes and ignore other influences

Email address: corcoranp@cardiff.ac.uk (Padraig Corcoran)

such as crime [3, 4]. An accurate model of pedestrian safety has many potential uses. It may be used to identify a set of factors affecting pedestrian safety and in turn facilitate the improvement of safety. It may also be used in simulation experiments to evaluate the consequences to pedestrian safety under proposed changes to the environment such as the introduction of an increased speed limit [5]. Finally, it may be used as a recommender of safer pedestrian paths.

Typically, models of pedestrian safety consider the number of pedestrian casualties or crashes as a single dependent variable and a set of potential factors affecting pedestrian safety as independent variables. Existing models of pedestrian safety consider independent variables describing the environment (e.g. the number of pedestrian crossings), the traffic (e.g. the volume of heavy vehicles) and the pedestrians themselves (e.g. mean pedestrian age). In this article we propose a novel linear model that considers a single independent variable of *pedestrian path safety*. This variable is estimated by sampling pedestrian paths from the population of such paths and in turn estimating the mean safety of these paths. Relative to those independent variables considered by existing models, we argue that this independent variable more directly models the factors affecting pedestrian safety.

The remainder of this paper is organised as follows. In Section 2 we review related models of pedestrian safety. In Section 3 we describe the proposed model of pedestrian safety for a given urban area. In Section 4, we evaluate the accuracy of this model against data on pedestrians casualties in 15 UK cities. Finally in Section 5 we draw some conclusions from this study.

2. Related Work

To date, a number of models of pedestrian safety have been proposed; these are summarised in Table 1 and described below. Typically, these studies analyse historic data on pedestrians crashes to identify risk factors or assess likelihoods of crash related events such as injury severity or fatality. These analyses are supported by standard statistical methods, most commonly logistic regression. The independent variables or features considered can be classified as temporal, spatial, demographic and socio-economic. The types of crash locations considered are usually generic, but some studies focus on specific location types, e.g. unsignalised zebra crossings [3] or intersections [6], or the road network or road pattern [7, 8, 9] surrounding the crash location. Others aggregate the statistical findings at level of a borough [10] or a buffer of given size [11]. Even though the sidewalk network is found to be negatively associated with the pedestrian-vehicle crashes [12], some studies explicitly ignore crosswalks and sidewalks, e.g. [11]. Here we provide more details into individual studies.

Sze et al. [13] used a binary logistic regression model to identify factors contributing to pedestrian crashes in Hong Kong including pedestrian behaviour and traffic congestion. Similarly, Olszewski, et al. [5] used a binary logistic regression model to identify factors contributing to pedestrian crashes at unsignalised zebra crossings in Poland including darkness and non built-up area. Pulugurtha et al. [11] used a generalised linear model with a negative binomial

Study	Pedestrian safety type	Country	Number of cities	Entity	Independent variables	Model type	Output
[3, 4]	Crime	USA	2	Path	Crime activity spatial density	Density estimation	Safest path
[5]	Crash	Poland	N/A	Unsignalised zebra crossing	Spatial, temporal, demographic	Logistic regression	Fatality risk factors
[13]	Crash	Hong Kong	1	Crash location	Temporal, spatial, demographic	Logistic regression	Risk factors
[11]	Crash	USA	1	Buffer	Temporal, spatial, demographic socio-economic	Statistical tests	Risk factors
[6]	Crash	USA	N/A	Intersection	Temporal, spatial, demographic	Log-linear model	Injury severity likelihood
[14]	Crash	Australia	1	N/A	Socio-economic	Boosted regression tree	Risk factors
[15]	Crash	South Korea	N/A	Crash location	Temporal, spatial, demographic	Multinomial logit	Severity factors
[16]	Crash	Canada	1	Crash location	Temporal, spatial, demographic	Kernel density estimation	Temporal, spatial crash patterns
[7]	Crash	Hong Kong	1	Road network	Temporal, spatial, demographic	Bayesian (MCMC)	Pedestrian crash likelihood
[10]	Crash	USA	1	Borough	Temporal, spatial, demographic	Random parameter	Pedestrian-vehicle crash severity
[8]	Crash	Canada	1	Road network	Temporal, spatial, demographic	Multinomial logit model	Likelihood of injury and fatality
[9]	Crash	Canada	1	Road network	Temporal, spatial, demographic	Generalized ordered logit model	Likelihood of injury and fatality
[12]	Crash	Canada	1	Crash location	Road network characteristics	Generalized linear model	Macro-level crash prediction
[17]	Crime	USA	1	Path	Historical crime data	Path planning	Safest path
[18]	Crash	Canada	1	Crash location	Pedestrian action, crossing location	Logistic regression	Injury severity
[19]	Crash	USA	24	Road network	Spatial	Linear regression	Likelihood of injury and fatality
This study	Crash	UK	15	Path	Path safety	Linear regression	Number of casualties

Table 1: This table summarises related models of pedestrian safety. For each model we present the type of safety considered, the country for which the model is developed, the number of cities it is applied to, the entity for which safety is modelled, the independent variables used by the model, the type of model and finally the output of the model.

distribution to identify factors contributing to pedestrian crashes at signalised intersections in Charlotte North Carolina including the number of transit stops and approaches at an intersection. Lee et al. [6] used a log-linear model to identify factors contributing to pedestrian crashes at intersections in Florida including undivided roads with a greater number of lanes. Pour et al. [14] used a boosted decision tree model to identify socio-economic factors contributing to pedestrian crashes in Melbourne Australia. Tay et al. [15] used a multinomial logistic regression model to identify factors contributing to pedestrian crashes in South Korea including heavy vehicles, drivers who were drunk, male or under the age of 65. Pour et al. [16] performed a spatial and temporal analysis of pedestrian crashes in Melbourne Australia toward identifying hotspots and hot times respectively.

Recently a number of models of pedestrian safety have been proposed to identify those road network characteristics or features which contribute to pedestrian casualties or crashes. Marshall et al. [19] used a linear model to identify a number of road network characteristics contributing to pedestrian crashes in 24 California cities including low road density with low intersection counts. Guo et al. [7] used a Bayesian model to identify a number of topological road network characteristics contributing to pedestrian crashes in Hong Kong including connectivity. Aziz et al. [10] used a multinomial logistic regression model to identify a number of road network characteristics contributing to pedestrian crashes in New York City including the number of lanes and road surface. Rifaat et al. [8] and Rifaat et al. [9] used a multinomial logistic regression model and a partially constrained generalized ordered logistic regression model respectively to identify a number of high level road network patterns contributing to pedestrian crashes in Calgary Canada including loops and lollipops designs. Osama [12] used a generalised linear model to identify a number of road network characteristics contributing to pedestrian crashes in Vancouver Canada including lower continuity, linearity, coverage, and slope of road pavements.

The majority of the above outlined studies focus on a single city, thus making it virtually impossible to generalise the findings. In addition, in terms of traffic characteristics, the countries considered, i.e. USA, Canada, Poland, South Korea, Hong Kong and Australia, differ significantly from the UK. For example, the Hong Kong traffic system follows closely that of the UK, including driving on the left side of the road. However, its highly sophisticated infrastructure makes the traffic itself differ significantly from that of the UK. For example, at 90%, Hong Kong has got the highest use of public transport for daily journeys [20]. Furthermore, the high density population implies that the overcrowded roads pose greater risk to pedestrian safety. Therefore, the findings presented in these studies cannot be readily applied to the UK. We, therefore, focus specifically on the UK data. To make the findings generalisable within the UK, we focus on 15 cities in contrast to a great majority of the reviewed studies who focus on a single city.

Another aspect that makes our study different is the type of entities considered. The referenced studies focus on modelling the safety of a single point or a buffer surrounding it, whereas our study focuses on the safety of paths.

While there are other studies also focusing on the same problem [3, 17] our focus is on specifically on pedestrian road safety, i.e. pedestrian safety with respect to pedestrian-vehicle crashes. The other two studies focus on pedestrian safety with respect to crime. Our approaches are similar in terms of posing the problem as that of finding the least weight path in a graph, but differ in the type of graph considered and way in which weights are assigned to edges in the graph. Both Galbrun et al. [3] and Keler et al. [17] used crime statistics to estimate spatial density for crime activity and use its values to assign weights to edges in a graph representation of the road network. In our approach, we use classification of pedestrian crossings to stratify their risk and assign weights to edges in a graph representation of the pavement network.

3. Model of Pedestrian Safety

In this section, we propose a new model of pedestrian safety in urban areas which is based on the principle of utility maximisation from the theory of multi-criteria decision making [21]. In our decision making scenario, the pedestrian chooses a path that maximises its utility in terms of two criteria: overall length and risk of a road crash incident. We assume that the pedestrian will always attempt to minimise both length and risk. In other words, the pedestrian will chose a Pareto optimal solution in terms of length and risk. Taking into account the complexity of finding an optimal path, a pedestrian is assumed to construct a simplified model of decision making and behave rationally or optimally with respect to this model. We model this behaviour by creating a model of the pedestrian pavement network and selecting a path in this network which minimize a weighted sum of path length and risk. We in turn use this value as a measure of the safety of the path in question. The use of this as a measure is motivated by the fact that the probability of a pedestrian crash occurring along a path is proportional to both its length and risk. Note that, there is a positive correlation between distance walked and probability of being involved in a pedestrian crash.

Given the above model of pedestrian behaviour, the pedestrian safety of a given urban area is modelled in four steps. First, a pavement network is constructed. Next, an edge weighting function is defined which maps each edge in the pavement network to a weighted sum of corresponding risk and length. We assume that a pedestrian follows a path which minimizes the sum of edges with respect to this edge weighting function and that this value is a measure of the safety of the path in question. The pedestrian safety for a given urban area is then estimated as the mean safety of pedestrian paths in that area. Finally, a linear regression model of this variable is used to estimate the corresponding number of pedestrians casualties. These steps are described in turn in the following subsections.

3.1. Pavement Network Construction

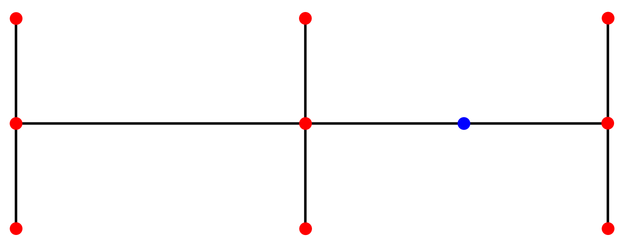
Given the fact that pedestrians mainly walk on pavements we consider a network representation of the pavement structure, which we refer to as a *pave-*

ment network. To facilitate the construction of this network we used data from OpenStreetMap (OSM) which is a crowdsourced project that provides free geographic data such as road maps [22]. For most regions in the UK, OSM provides an accurate road network but not necessarily an accurate pavement network. It is important to note that the Ordnance Survey, which is the national mapping agency for the UK, also does not provide an accurate pavement network. To compensate for this, we automatically construct a pavement network from the corresponding road network based on two assumptions regarding the relationship between the roads and pavements. This is known as a network buffering approach to the construction of a pavement network [23]. The pavement network construction comprises a set of steps which are now described.

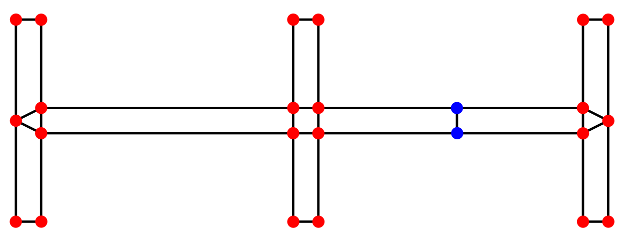
First, a road network $G^s = (V^s, E^s)$ with two vertex- and one edge-labelling function is constructed. Here, each vertex from V^s corresponds to a designated pedestrian crossing, road intersection, or dead-end. Whereas each edge from E^s is undirected and corresponds to a road segment connecting the corresponding vertices. The first vertex-labelling function $\xi : V^s \to \{0, 1\}$ maps each $v \in V^s$ to a value in the set $\{0, 1\}$ indicating if the vertex in question is a designated pedestrian crossing. The second vertex-labelling function $\lambda : V^s \to \mathbb{Z}^+$ maps each $v \in V^s$ to a unique positive integer; this value is used to uniquely identify the vertex in question when constructing the pavement network. Finally, the edge-labelling function $\mu : E^s \to (\mathbb{R}^+, \text{String})$ maps each $(v, v') \in E^s$ to a tuple containing a positive real number equalling the length of the corresponding road segment and a string equalling its type (e.g. 'primary', 'residential', etc.). Figure 1(a) illustrates a toy road network G^s which will be used as a running example throughout this paper.

Next, given a road network G^s, a corresponding pavement network $G^p = (V^p, E^p)$ with one edge-labelling function is constructed. This construction is based on two assumptions. First, each road segment is assumed to have a pavement on both sides. All road networks considered in this work correspond to urban areas in the UK and this assumption is a close approximation to reality in this case. Second, each road segment is assumed to have a designated or non-designated (jaywalking) pedestrian crossing at each end. A designated crossing is determined to exist at the end of a given road segment if such a crossing exists at that location in the corresponding road network. Otherwise a non-designated (jaywalking) crossing is determined to exist at the location in question. The placement of pedestrian crossings at the end of road segments is based on the fact that pedestrians generally do not perform crossings in the interior of road segments unless there is a designated crossing present at such locations.

Each vertex from V^p corresponds to a pavement segment endpoint and/or a designated pedestrian crossing endpoint. Each edge from E^p is undirected and corresponds to a pedestrian path connecting the corresponding vertices. The edge-labelling function $\delta : E^p \to (\mathbb{R}^+, \text{String})$ maps each $(v, v') \in E^p$ to a tuple containing a positive real number equalling the length of the corresponding pavement segment/pedestrian crossing and a string equalling its type (e.g. 'primary pavement', 'residential pavement', 'primary designated pedestrian cross-



(a)



(b)

Figure 1: A toy road network $G^s = (V^s, E^s)$ is illustrated in (a). Each vertex in V^s is represented using blue and red dots indicating if the vertex in question is respectively a designated pedestrian crossing or not. Each edge in E^s is a road segment connecting two vertices. The corresponding pavement network $G^p = (V^p, E^p)$ is illustrated in (b). Each vertex in V^p is represented using blue and red dots indicating if the vertex in question is respectively a designated pedestrian crossing endpoint or a pavement segment endpoint. Note that, this pavement network contains a single designated pedestrian crossing; all other pedestrian crossings are non-designated (jaywalk).

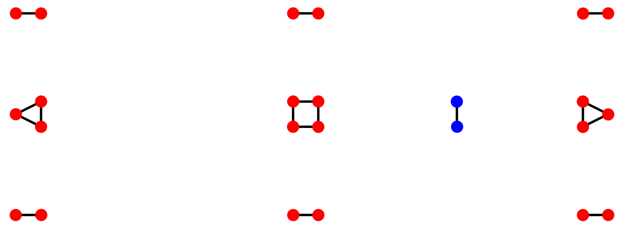


Figure 2: The result of iterating over and performing the actions specified in Algorithm 1 with respect to the vertices of the road network G^s in Figure 1(a) is illustrated.

ing', 'residential non-designated (jaywalk) pedestrian crossing'). Figure 1(b) illustrates the pavement network constructed from the toy road network shown in Figure 1(a).

The pavement network is constructed using Algorithm 1. Internally, this algorithm uses three functions which we now define. The function $\rho : V^s \rightarrow \mathbb{Z}^+$ maps each $v \in V^s$ to a positive integer equalling its degree. The function $\text{CyclicOrder} : E^s \rightarrow \mathbb{Z}^+$ maps each $(v, v') \in E^s$ to a positive integer in the range $[1, \rho(v)]$ equalling the position of (v, v') in the circular ordering of edges incident to vertex v. Such a circular ordering exists because the graph G^s is embedded in $\mathbb{R}^2$ [24]. The function $\tau : V^p \rightarrow (\mathbb{Z}^+, \mathbb{Z}^+)$ is an invertible function which maps each $v \in V^p$ to a unique tuple. Invertibility follows from the uniqueness property.

Algorithm 1 initially iterates over the vertices in V^s (lines 3 to 15). Recall that, each of these vertices corresponds to a pedestrian crossing, road intersection, or dead-end. For each, $v \in V^s$ the following actions are performed. If the degree of v is one (line 4), which happens when v corresponds to a dead-end, two copies of v are added to V^p (lines 5 and 6). An edge is subsequently inserted between these vertices (line 7). This edge corresponds to a pedestrian crossing at the dead-end of a road segment. If the degree of v is greater than one (line 8), $\rho(v)$ copies of v are added to V^p (line 10). Edges are subsequently inserted between these vertices to form a ring topology (line 13). Each of these edges corresponds to a pedestrian crossing. Figure 2 illustrates the result of iterating over, and performing the above actions with respect to, the vertices of the toy road network G^s in Figure 1(a).

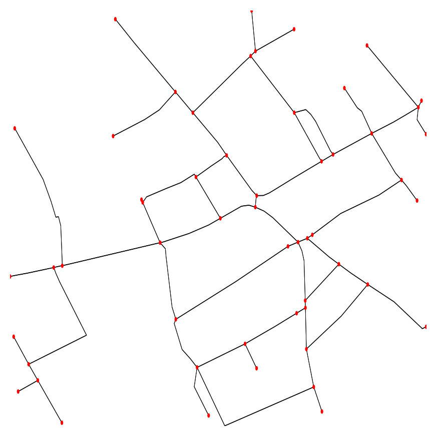
Next, the algorithm iterates over the edges in E^s. Recall that, each of these edges corresponds to a road segment. For each $e \in E^s$, the pavement segments on both sides of the road segment in question are added (lines 19 and 20). Here the circular ordering of edges is used to ensure that the correct pairs of pavement segment endpoints are connected (lines 17 and 18). Figure 1(b) illustrates the result of iterating over, and performing the above actions with respect to, the edges of the toy road network G^s in Figure 1(a). Finally the algorithm returns the network G^p (line 22). Figure 3 illustrates both the road network G^s and pavement network G^p corresponding to a subset of the city of York.

Algorithm 1: Pavement network construction.

Input: A road network graph $G^s = (V^s, E^s)$ and labelling functions $\lambda : V^s \to \mathbb{Z}^+$, $\xi : V^s \to \{0, 1\}$ and $\mu : E^s \to (\mathbb{R}^+, \text{String})$.

Output: A pavement network $G^p = (V^p, E^p)$.

```
begin
    Initialize sets V^p, E^p = ∅
    for v ∈ V^s do
        if ρ(v) == 1 then
            V^p = V^p ∪ {v : τ(v) = (λ(v), 1)}
            V^p = V^p ∪ {v : τ(v) = (λ(v), 2)}
            E^p = E^p ∪ {(τ^{-1}(λ(v), 1), τ^{-1}(λ(v), 2))}
        else if ρ(v) > 1 then
            for i ← 1 to ρ(v) do
                V^p = V^p ∪ {v : τ(v) = (λ(v), i)}
            end
            for i ← 1 to ρ(v) do
                E^p = E^p ∪ {(τ^{-1}(λ(v), i), τ^{-1}(λ(v), i + 1 mod ρ(v)))}
            end
    end
    for (v_1, v_2) ∈ E^s do
        i = CyclicOrder(v_1, v_2)
        j = CyclicOrder(v_2, v_1)
        E^p = E^p ∪ {(τ^{-1}(λ(v_1), i), τ^{-1}(λ(v_2), j + 1 mod ρ(v_2)))}
        E^p = E^p ∪ {(τ^{-1}(λ(v_1), i + 1 mod ρ(v_1)), τ^{-1}(λ(v_2), j))}
    end
    return G^p
end
```



(a) (b)

Figure 3: The road network G^s and pavement network G^p corresponding to a subset of the city of York are illustrated in (a) and (b) respectively. To simplify the visualisation process, the edges of G^p are drawn with straight line segments.

3.2. Computing Edge Safety

We assume that a pedestrian follows a path in the pavement network G^p which minimises the weighted sum of the length and risk of the corresponding edges. We assume this value is a measure of the safety of the path in question and therefore refer to this value as the *safety* of the path in question. To model this formally, for a given pavement network G^p, we construct an edge weighting function $S : E^p \to \mathbb{R}^+$. This function is defined in Equation 1 and maps each edge in E^p to the weighted sum of its length and risk as defined by the functions $L : E^p \to \mathbb{R}^+$ and $R : E^p \to \mathbb{R}^+$ respectively. The weighting in question is controlled by the parameter $\alpha \in \mathbb{R}^+$. We now describe each of these functions in turn.

$$S(e) = L(e) + \alpha R(e) \tag{1}$$

The function L assigns a length value to each edge in E^p. The set of edges in E^p can be partitioned into two subsets corresponding to pavement segments and pedestrian crossings. For example, the subset of edges in the pavement network of Figure 1(b) corresponding to pedestrian crossings are illustrated in Figure 2. If an edge in E^p corresponds to a pavement segment, then the function L returns a value equal to the length of the corresponding road segment. On the other hand, if an edge in E^p corresponds to a pedestrian crossings, then the function L returns a value equal to the width of the road crossed.

The function R assigns a risk value to each edge in E^p where smaller values indicate less risk. The set of edges E^p can be partitioned into four subsets corresponding to designated pedestrian crossings, dead-end pedestrian crossings, non-designated (jaywalk) pedestrian crossings and pavement segments.

Pedestrian crossing type	Risk value
Light controlled	4
Pelican	4
Zebra	3
Human controlled	2
Dead-end	1

Table 2: Safety values assigned to designated and dead-end pedestrian crossings.

Although edges corresponding to pedestrian crossings are assigned the highest risk values, non-zero risk values are also assigned to pavement segments. This is motivated by the fact that a pedestrian is subject to a number of significant risks while walking on a pavement. In the year 2015, *pedestrian impaired by alcohol* was the fifth most frequently reported contributory factor to pedestrian casualties in the UK [2]. The Queensland Department of Transport and Main Roads identified stumbling from the pavement into the adjacent road as being a major risk to intoxicated pedestrians [25]. A pedestrian walking on the pavement is also subject to the risk of a vehicle leaving the road and entering the adjacent pavement. In the year 2015, *vehicle travelling along pavement* was a reported contributory factor to 277 pedestrian casualties in the UK [2].

We assigned risk values to edges in each of the four subsets defined above using a combination of national statistics on pedestrian casualties and heuristics. The use of heuristics is a consequence of the fact that our model employs a finer grained categorisation of edges than that for which national statistics are available. To determine appropriate risk values for edges corresponding to designated pedestrian crossings, we obtained the number of annual pedestrian casualties on zebra, pelican, light controlled and human controlled crossings in Great Britain for the years 2011 to 2015 inclusive from the UK Department for Transport [2]. These are the four main types of designated pedestrian crossings in the UK defined by the UK Department for Transport as in Figure 4. A Repeated Measures ANOVA was performed to compare the differences between the mean number of casualties for each crossing type. These results confirmed that human controlled is the safest crossing type, followed by zebra and finally pelican and light controlled are jointly the least safe. No national statistics relating to pedestrian casualties at dead-end pedestrian crossings are available. However, due to the absence of traffic passing through the end of dead-end streets, we determined them to be safer than other crossing types. Using this ordering of pedestrian crossing types in terms of safety, we assigned risk values to each type such that safer types had lower risk values. These values are displayed in Table 2.

For edges corresponding to pedestrian jaywalk crossings and pavement segments, risk depends on the type of road which is crossed and adjacent to respectively. The UK Department for Transport do provide statistics on the number of pedestrian casualties for different road types. However, these types do not

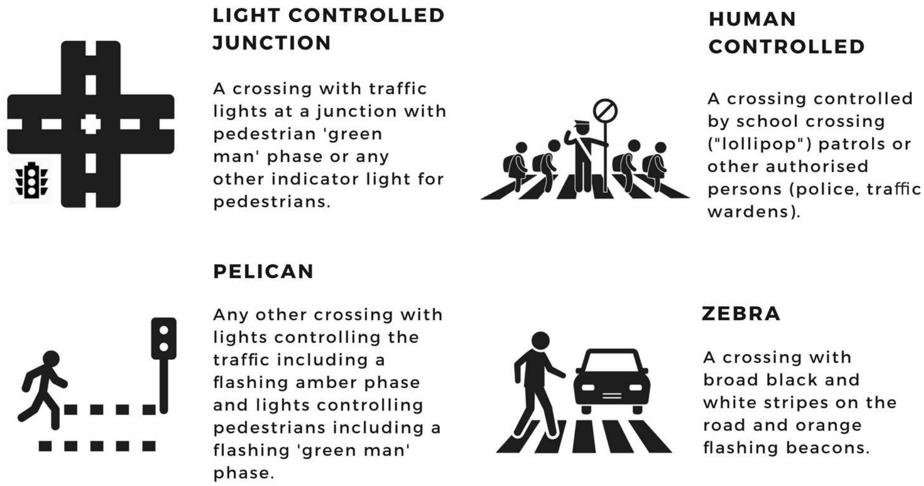


Figure 4: The four main types of designated pedestrian crossings in the UK defined by the UK Department for Transport [2].

naturally map to the road types used in OpenStreetMap and the statistics do not distinguish between casualties resulting from walking on the pavement and jaywalking. Therefore risk values for edges corresponding to jaywalk pedestrian crossings and pavement segments were determined by considering the level of the corresponding road type in the OpenStreetMap road type hierarchy and assigning higher risk values to road types higher in the hierarchy. For example, a pavement segment adjacent to a residential road is assigned a smaller risk value than a pavement segment adjacent to a primary road. There is significantly higher risk of a being involved in a pedestrian crash when jaywalking compared to walking on the pavement. Therefore, given a particular road type we assigned risk values to non-designated (jaywalk) crossings that were three times higher than those assigned to pavement segments. Table 3 summarises the risk values assigned to edges corresponding to pavement segments and pedestrian jaywalk crossings. Note that, a safety value of 0 was assigned to edges corresponding to pavement segments adjacent to, and jaywalk crossings of, all types of pedestrianised roads.

3.3. Pedestrian Path Safety

We model the safety of a given path in a pavement network G^p as the sum of the weights of its constituent edges as defined by the edge weighting function S in Equation 1. For a given origin-destination pair of locations, we assume that a pedestrian follows that path between these locations which minimises the corresponding sum of weights. In this work we compute such paths using Dijkstra's algorithm for finding the shortest path between two vertices in a given

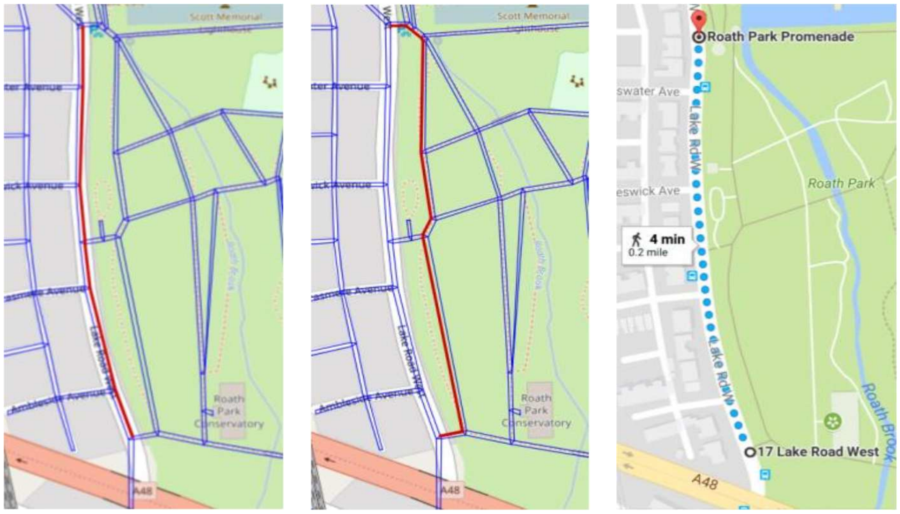
OpenStreetMap road type	Pavement segment risk value	Non-designated (jaywalk) crossing risk value
Trunk	7	21
Primary	7	21
Secondary	7	21
Trunk Link	6	18
Primary Link	6	18
Secondary Link	6	18
Tertiary	5	15
Tertiary Link	5	15
Residential	4	12
Service	4	12
Track	2	6
Pedestrian	0	0
Footway	0	0
Path	0	0
Steps	0	0

Table 3: Risk values assigned to edges corresponding to pavement segments and non-designated (jaywalk) crossings.

graph [26]. To illustrate this model we compare the *shortest path* with the *model pedestrian path* for a number of example origin-destination pairs in the city of Cardiff Wales. Here the shortest and model pedestrian paths are obtained using the edge weighting function S with an α value of 0 and 10 respectively.

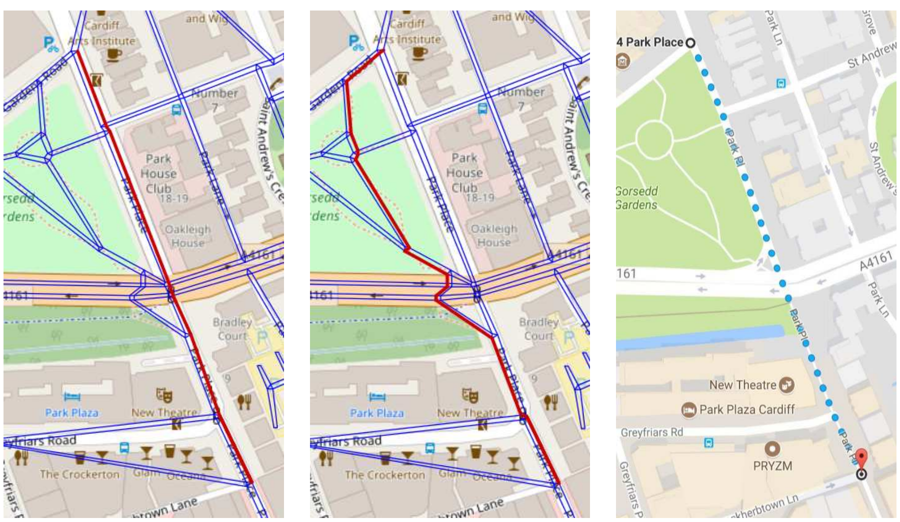
The first example considered is illustrated in Figure 5 and requires a pedestrian to navigate along a tertiary road. The shortest path is indicated in Figure 5(a) and takes the pedestrian along the tertiary pavement. The model pedestrian path is indicated in Figure 5(b) and takes the pedestrian via a footway through a park. With respect to geographical distance, the shortest and model pedestrian paths have lengths of 343 and 382 meters respectively. The pedestrian path recommended by Google Maps is illustrated in Figure 5(c). It is interesting to note that paths recommended by Google Maps do not state which pavement pedestrians should use. This is a consequence of the fact that Google Maps path planning algorithm is applied to a road network similar to G^s defined in this work and not to a pavement network.

The second example considered is illustrated in Figure 6 and requires a pedestrian to navigate across a primary road. The shortest path is indicated in Figure 6(a). This path takes the pedestrian along a tertiary pavement and across a relatively dangerous primary road jaywalk crossing. The model pedestrian path is indicated in Figure 6(b). This path takes the pedestrian via a footway instead of the tertiary pavement and across three sets of traffic light pedestrian crossings to avoid jaywalking across the primary road. With respect to geographical distance, the shortest and model pedestrian paths have lengths of 305 and 382 meters respectively.



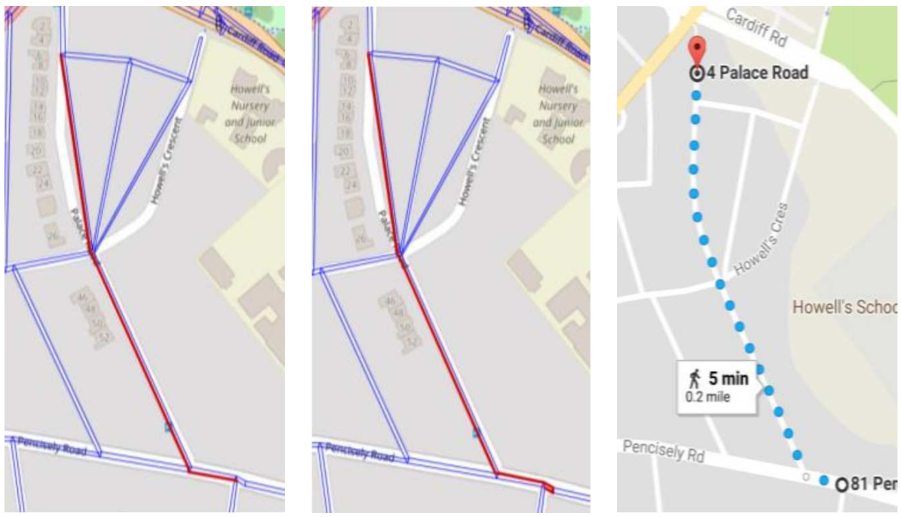
(a) (b) (c)

Figure 5: The shortest and model pedestrian paths from a location in the top to a location in the bottom are illustrated in (a) and (b) respectively. In both figures the pavement network is represented by a set of blue lines and the path in question is represented by a sequence of red lines. The pedestrian path recommended by Google Maps is illustrated in (c).



(a) (b) (c)

Figure 6: The shortest and model pedestrian paths from a location in the top to a location in the bottom are illustrated in (a) and (b) respectively. In both figures the pavement network is represented by a set of blue lines and the path in question is represented by a sequence of red lines. The pedestrian path recommended by Google Maps is illustrated in (c).

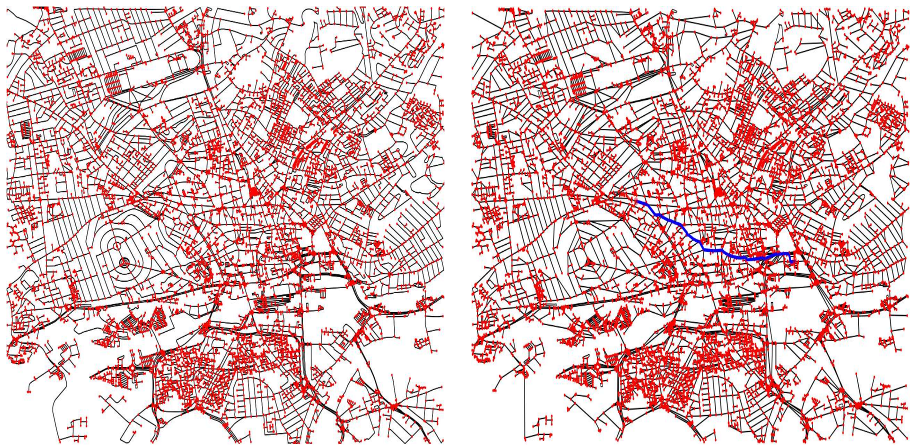


(a) (b) (c)

Figure 7: The shortest and model pedestrian paths from a location in the bottom to a location in the top are illustrated in (a) and (b) respectively. In both figures the pavement network is represented by a set of blue lines and the path in question is represented by a sequence of red lines. The pedestrian path recommended by Google Maps is illustrated in (c).

The final example considered is illustrated in Figure 7 and requires a pedestrian to navigate across a secondary road. The shortest path is indicated in Figure 7(a) and requires the pedestrian to jaywalk across the secondary road. The model pedestrian path is indicated in Figure 7(b) and avoids jaywalk by using a zebra crossing beside the origin location. With respect to geographical distance, both the shortest and model pedestrian paths have a length of 201 meters.

To estimate pedestrian path safety for a given urban area, as opposed to a single origin-destination pair, the following approach was employed. Rejection sampling was used to generate a sample of 500 origin-destination pairs corresponding to locations pedestrians would realistically consider walking between as opposed to using alternative means of transportation. Specifically, two locations in the given urban area were selected randomly. If the length of the shortest path between these locations in the pavement network was greater than fifty meters but less than three kilometres, the pair was accepted and added to the sample. Otherwise the pair was rejected and not added to the sample. This was repeated until the sample was sufficiently large. Next, for each origin-destination pair in the sample, the safety value of the model pedestrian path between the locations in question was calculated. Finally, pedestrian path safety for the urban area in question is estimated as the mean of these values. For example, Figure 8(a) illustrates the road network G^s for the centre of Nottingham city. Figure 8(b) illustrates the corresponding pavement network



(a) (b)

Figure 8: The road network G^s for the centre of Nottingham city is illustrated in (a). The corresponding pavement network G^p plus the safest path between a single origin-destination pair is illustrated in (b). The path in question is represented using a blue line.

G^p plus the safest path between a single origin-destination pair.

Given the estimated pedestrian path safety for a given urban area, we estimate the number of pedestrian casualties in that area using a linear regression model. Specifically, a linear least squares regression model is used. Many of the models previously reviewed in the related work section consider a non-linear as opposed to a linear model. In the context of the proposed model, we found a linear model to be of sufficient complexity to accurately fit our data and therefore did not consider a non-linear model.

4. Results

In this section we evaluate if, given the estimated *pedestrian path safety* of an urban area, the number of pedestrian casualties in that area can be accurately estimated using a linear regression model. We considered 15 urban areas corresponding to the centres of 15 UK cities. The cities in question are listed in the first column of Table 4. In this work we define the centre of a given city to be the area within a 2 kilometre radius of a manually defined point in the very centre of that city. Figure 8(a) illustrates the road network G^s for the city centre of Nottingham city.

For each urban area we estimated the pedestrian path safety using the model described in Section 3.3 and obtained the number of pedestrian casualties per million population for the year 2015 from the UK Department for Transport [2] (pages 220-223). These values are given in Table 4 and displayed as a scatter plot in Figure 9. It is evident from the scatter plot that the relationship between the estimated pedestrian safety and the number of pedestrian casualties per

	Estimated Pedestrian Safety	Pedestrian Casualties per Million Population
Bath	2824.16	297
Bedford	3026.10	385
Blackpool	3205.25	724
Bristol	3147.62	454
Coventry	3021.24	385
Leeds	3191.08	497
Leicester	3210.84	654
Liverpool	3261.51	702
Manchester	3029.55	371
Nottingham	3176.16	706
Reading	3125.53	458
Salford	2898.63	200
Sheffield	3088.05	455
Swindon	2952.80	272
York	3043.59	387

Table 4: For 15 UK cities the corresponding estimated pedestrian path safety and number of pedestrian casualties per million population for the year 2015 are stated.

million population is strongly linear. Using Pearson's correlation test, the null hypothesis that these statistics are uncorrelated is rejected with a p-value of 0.001. Furthermore, the Pearson's correlation coefficient between the statistics is 0.893. This demonstrates that, for a given urban area, a linear regression model of the estimated pedestrian path safety may be used to accurately estimate the corresponding number of pedestrian casualties. In this work we consider the linear least squares regression model. This model fitted to the entire data is illustrated in Figure 9.

To benchmark the proposed model we compared it to an alternative linear least squares regression model which uses the set of independent variables in Table 5 which describe the environment, traffic and pedestrians. This set of independent variables is similar to sets in other studies that also attempt to estimate numbers of pedestrian casualties or crashes [19, 27]. Therefore, we argue that their consideration represents a reasonable baseline model.

To compare the ability of both linear models to estimate the number of pedestrian casualties in a given urban area we performed a leave-one-out cross-validation using a mean squared error. Specifically, for each individual urban area we fitted the linear model in question using all remaining urban areas and subsequently used this model to estimate the number of pedestrian casualties for the urban area in question. The error of this individual estimate was measured as the squared difference between the estimated and actual number of pedestrian casualties. The overall mean squared error of the model equals the mean of these values for all urban areas. The proposed linear model, which uses a single independent variable of pedestrian path safety achieved a mean squared error

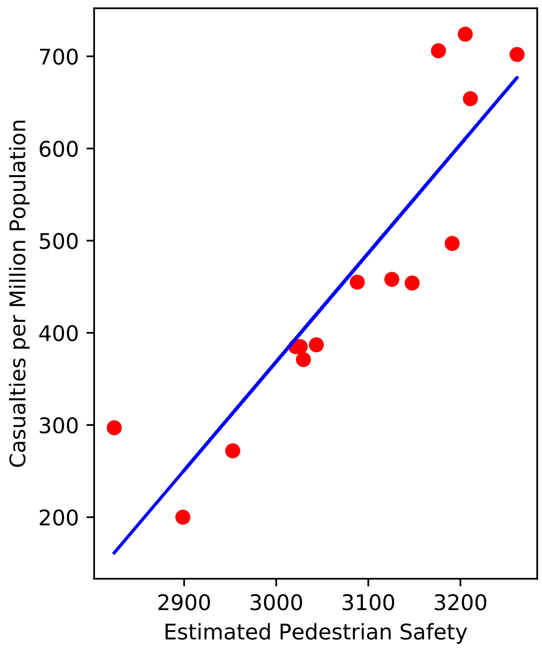


Figure 9: A scatter plot of the statistics contained in Table 4. Each city is represented by an individual red dot. The linear least squares regression model fitted to all data points is represented by a blue line.

Independent variable	Description
Road density	Number of edges in road network divided by the area.
Intersection density	Number of vertices in the road network with degree greater than one (dead ends are excluded) divided by the area.
Mean clustering coefficient	Mean clustering coefficient of vertices in the road network [28].
Mean betweenness centrality	Mean betweenness centrality of vertices in the road network [28].
Mean age	Mean age of residents. This data was obtained from the UK Census [29].
Population density	Number of persons per hectare. This data was obtained from the UK Census [29].
Traffic congestion index	Percentage of time that drivers spend in congestion. This data was obtained from [30].

Table 5: The set of independent variables used in the baseline linear least squares regression model.

of 8,150. The alternative linear model which uses a different set of independent variables achieved a mean squared error of 43,923. The error of the proposed linear model is significantly less than that of the alternative linear model. This demonstrates the usefulness of the proposed model with respect to estimating the number of pedestrian casualties in a given urban area.

5. Conclusions

In this article we proposed a linear model of pedestrian safety that can accurately estimate the number of pedestrian casualties in a given urban area. Models of pedestrian safety proposed to date have considered independent variables describing the environment, traffic and pedestrians themselves. The proposed model considers a single independent variable which is an estimate of pedestrian path safety. We argue that this independent variable more directly models the factors contributing to pedestrian safety.

As discussed in the introduction to this article, an accurate model of pedestrian safety has many potential uses. One such use would be as a recommender of pedestrian paths which are safer with respect to road traffic crashes. As discussed in the related work section of this paper, previous works have considered pedestrian path safety with respect to crime. To the authors knowledge, this article represents the first work which considers pedestrian path safety with respect to road traffic crashes. As such, we believe the proposed model has many potential novel applications.

Despite the good performance of the proposed model there are many ways in which it could potentially be improved. For example, the proposed model implicitly assumes that all locations are equally likely to be origin or destination locations which is an approximation. Furthermore, the proposed model does not consider pedestrian and vehicle volumes which may be good indicators of pedestrian casualties in a given area. The authors hope to investigate such potential improvements to the model in future work.

References

[1] World Health Organization, Make walking safe: a brief overview of pedestrian safety around the world, http://www.who.int/violence_injury_prevention/publications/road_traffic/make_walking_safe.pdf (Accessed 29 October 2017).

[2] Department for Transport, Reported Road Casualties Great Britain: 2015 Annual Report, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/568484/rrcgb-2015.pdf (Accessed 23 May 2018).

[3] E. Galbrun, K. Pelechrinis, E. Terzi, Safe Navigation in urban environments, in: International Workshop on Urban Computing (UrbComp 2014), New York, USA, 2014.

[4] E. Galbrun, K. Pelechrinis, E. Terzi, Urban navigation beyond shortest route: The case of safe paths, Information Systems 57 (2016) 160–171.

[5] P. Olszewski, P. Szagała, M. Wolański, A. Zielińska, Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland, Accident Analysis & Prevention 84 (2015) 83–91.

[6] C. Lee, M. Abdel-Aty, Comprehensive analysis of vehicle–pedestrian crashes at intersections in florida, Accident Analysis & Prevention 37 (4) (2005) 775–786.

[7] Q. Guo, P. Xu, X. Pei, S. Wong, D. Yao, The effect of road network patterns on pedestrian safety: a zone-based bayesian spatial modeling approach, Accident Analysis & Prevention 99 (2017) 114–124.

[8] S. M. Rifaat, R. Tay, A. De Barros, Effect of street pattern on the severity of crashes involving vulnerable road users, Accident Analysis & Prevention 43 (1) (2011) 276–283.

[9] S. M. Rifaat, R. Tay, A. de Barros, Urban street pattern and pedestrian traffic safety, Journal of urban design 17 (3) (2012) 337–352.

[10] H. A. Aziz, S. V. Ukkusuri, S. Hasan, Exploring the determinants of pedestrian–vehicle crash severity in new york city, Accident Analysis & Prevention 50 (2013) 1298–1309.

[11] S. S. Pulugurtha, V. R. Sambhara, Pedestrian crash estimation models for signalized intersections, Accident Analysis & Prevention 43 (1) (2011) 439–446.

[12] A. Osama, T. Sayed, Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety, Accident Analysis & Prevention 107 (2017) 117–125.

[13] N.-N. Sze, S. Wong, Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes, Accident Analysis & Prevention 39 (6) (2007) 1267–1278.

[14] A. T. Pour, S. Moridpour, R. Tay, A. Rajabifard, Neighborhood influences on vehicle-pedestrian crash severity, Journal of urban health (2017) 1–14.

[15] R. Tay, J. Choi, L. Kattan, A. Khan, A multinomial logit model of pedestrian-vehicle crash severity, International journal of sustainable transportation 5 (4) (2011) 233–249.

[16] A. T. Pour, S. Moridpour, R. Tay, A. Rajabifard, Influence of pedestrian age and gender on spatial and temporal distribution of pedestrian crashes, Traffic Injury Prevention 19 (1) (2018) 81–87.

[17] A. Keler, J. D. Mazimpaka, Safety-aware routing for motorised tourists based on open data and VGI, Journal of location Based services 10 (1) (2016) 64–77.

[18] L. Rothman, A. W. Howard, A. Camden, C. Macarthur, Pedestrian crossing location influences injury severity in urban areas, Injury prevention 18 (6) (2012) 365–370.

[19] W. E. Marshall, N. W. Garrick, Does street network design affect traffic safety?, Accident Analysis & Prevention 43 (3) (2011) 769–781.

[20] W. H. Lam, M. G. Bell, Advanced modeling for transit operations and service planning, Emerald Group Publishing Limited, 2002.

[21] A. Mas-Colell, M. D. Whinston, J. R. Green, et al., Microeconomic theory, Vol. 1, Oxford university press New York, 1995.

[22] P. Mooney, P. Corcoran, Using OSM for LBS - an analysis of changes to attributes of spatial objects, in: Advances in Location-Based Services, Springer, 2012, pp. 165–179.

[23] H. A. Karimi, P. Kasemsuppakorn, Pedestrian network map generation approaches and recommendation, International Journal of Geographical Information Science 27 (5) (2013) 947–962.

[24] A. Gagarin, W. Kocay, D. Neilson, Embeddings of small graphs on the torus, Cubo. v5 (2003) 351–371.

[25] Queensland Department of Transport and Main Roads, Pedestrian safety - get the facts, https://jointhedrive.qld.gov.au/pedestrians/factsheet (accessed 23 May 2018).

[26] E. W. Dijkstra, A note on two problems in connexion with graphs, Numerische mathematik 1 (1) (1959) 269–271.

[27] Y. Zhang, J. Bigham, D. Ragland, X. Chen, Investigating the associations between road network structure and non-motorist accidents, Journal of transport geography 42 (2015) 34–47.

[28] G. Boeing, OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks, Computers, Environment and Urban Systems 65 (2017) 126–139.

[29] UK Census, Nomis service provided by the Office for National Statistics, https://www.nomisweb.co.uk/ (Accessed 29 March 2018).

[30] G. Cookson, B. Pishue, INRIX global traffic scorecard, Tech. rep. (2017).