

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/157953/>

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

Citation for final published version:

Corcoran, Pdraig and Lewis, Rhyd ORCID: <https://orcid.org/0000-0003-1046-811X>
2023. A navigability entropy model for street networks. Environment and Planning B: Urban Analytics and City Science 10.1177/23998083231170191 file

Publishers page: <https://doi.org/10.1177/23998083231170191>

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 Navigability Entropy Model for Street Networks

Padraig Corcoran, PhD^{*1} and Rhyd Lewis, PhD²

¹School of Computer Science & Informatics, Cardiff University, Cardiff, Wales, CF24 4AG.

²School of Mathematics, Cardiff University, Cardiff, Wales, CF24 4AG.

Abstract

In this paper, we propose novel local and global models of street network entropy that measure levels of navigability given only limited local directional information. These models are defined for individual locations and entire street networks. Both models are derived using a generalised model of entropy from the field of game theory, which considers a decision-maker attempting to perform a task in the presence of incomplete information. We argue that the proposed models are more interpretable and useful than existing models of street network entropy since they measure the uncertainty of navigation, which is the task street networks are intended to facilitate. We demonstrate this utility by performing an empirical analysis of the entropy properties of UK city street networks.

Keywords. street network; spatial network; entropy; navigability.

1 Introduction

A street network is a type of transportation network corresponding to the set of streets contained in a given geographical region such as a city boundary. The properties of street networks directly influence many aspects of our society including where homes and services are located plus patterns of travel. Therefore, modelling properties of street networks represents an important research problem. There are several street network models that measure many types of properties including topological, geometrical and centrality properties. Each of these models has different applications and uses. For example, models of centrality may be used to help identify elements of a street network that a large proportion of routes pass through and therefore it may be useful to limit traffic congestion at these elements. In this work, we consider the problem of modelling the entropy of street networks. In the context of information theory, entropy is a measure of the uncertainty with respect to a given property (Gudmundsson and Mohajeri, 2013). The (Shannon) entropy of a given property, modelled as a random variable X on the set $\{x_1, \dots, x_n\}$ with probability mass function $p(x_1), \dots, p(x_n)$, is defined as follows:

$$- \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

A street network is most commonly represented as a graph embedded in the plane (Barthélemy, 2011). As such, there exists a large array of entropy models for general graphs that can be applied to street networks (Dehmer and Mowshowitz, 2011; Marin et al., 2022). However, these models do not consider the spatial properties of street networks and are therefore rarely considered. Instead, the two most commonly used models of street network entropy are defined with respect to the properties of street length and orientation (Mohajeri and Gudmundsson, 2014; Boeing, 2019b). These models have several useful applications because they have been demonstrated to be correlated with other properties of interest. Firstly, it has been demonstrated that these models of entropy

*Corresponding author - corcoranp@cardiff.ac.uk

can be used to identify the time period during which the street network in question was constructed (Gudmundsson and Mohajeri, 2013). Specifically, street networks constructed during older time periods tend to be less structured and have higher entropy than those constructed during more recent time periods. Secondly, these models of entropy can be considered an indicator of the directness of shortest routes (ratio of street network distance to great circle distance) and how easy it is to determine these routes given limited information. This is a consequence of the fact that street networks with a grid pattern, which exhibit high levels of both these properties, have relatively small entropy (Boeing, 2019b).

The principal purpose of a street network is to facilitate the task of navigating between locations. In this work, we propose novel local and global models of street network entropy that measure the navigability or uncertainty with respect to performing this task. Note that the model of Shannon entropy in Equation 1 is defined for a property or random variable and not a task. It cannot, therefore, be used to define a model of entropy defined for navigating. To overcome this challenge we consider a generalised model of Entropy from the field of game theory that measures the uncertainty with respect to a decision-maker attempting to perform a task in the presence of incomplete information (DeGroot, 1962; Grünwald and Dawid, 2004). We use this generalised model to derive the proposed models of entropy whereby we measure the uncertainty with respect to a decision-maker attempting to navigate given only local direction information. This information constraint is motivated by the fact that it is known to be an accurate model of how humans navigate street networks (Bongiorno et al., 2021). Since the proposed models of street network entropy measure the uncertainty with respect to navigation (as opposed to a more abstract feature) we argue that these models are more interpretable than existing models. This in turn increases the potential for useful analysis to be performed using these models. Furthermore, to the authors’ knowledge, the proposed local model of street network entropy is the first of its kind. The local nature of this model lends itself to many new useful applications. This includes identifying locations in a street network that exhibit poor navigability and therefore may require additional signage to prevent drivers, cyclists or pedestrians from taking incorrect turns.

The layout of this paper is as follows. In Section 2 we review related works on human navigation and motivate the proposed models of street network entropy. In Section 3 we review necessary background material on generalised entropy and, in turn, derive the proposed models. In Section 4 we demonstrate the usefulness of these models by performing an analysis of UK city street networks. In doing so, we examine the relationships existing between the proposed models of street network entropy, existing models of street network entropy, and other properties of street networks. Finally, in Section 5 we conclude this work and discuss some limitations of the proposed models.

2 Related Works on Human Navigation

Human navigation has been widely studied in many fields including network science, geographical information science and cognitive psychology. In this section, we review existing works that motivate the proposed models of street network entropy.

Individuals use many different criteria when selecting a route to navigate. This includes the length of, the number of turns in, and the amount of congestion along the route (Golledge, 1995; Golledge and Gärling, 2004). In many cases, the individual will have prior spatial knowledge of the geographical region in question that will help to complete the navigation task. This spatial knowledge is most commonly referred to as a cognitive map. A cognitive map will generally contain information about the locations of landmarks, which can be defined as distinctive features in the environment that help individuals to orient themselves (Siegel and White, 1975). Examples of landmarks include a church or a pub. Apart from a cognitive map, individuals will commonly use external sources of spatial knowledge to help navigation such as a GPS device and map (Corcoran et al., 2014). In many cases, these sources will contain references to landmarks.

With incomplete spatial knowledge, an individual needs to intelligently reason and make inferences to successfully navigate. It has been demonstrated that individuals have varying abilities for successfully navigating in this context, and this ability is correlated with factors including gender, age, economic wealth and gender inequality (Coutrot et al., 2018). Note that many experiments that evaluate an individual’s ability to navigate are performed in simulation, which is predictive

of an individual’s ability to navigate in the real-world (Coutrot et al., 2019). Coutrot et al. (2022) found that individuals that grew up in geographical regions where the corresponding street network had higher entropy had a superior ability to navigate more complex environments. In this work street network entropy was defined with respect to the property of street orientation.

It has been hypothesised that, with incomplete spatial knowledge, individuals will use heuristics to navigate street networks (Bongiorno et al., 2021). A heuristic is a useful method for making decisions when there are insufficient resources or knowledge for making an optimal decision. Hochmair (2005) and Lee and Holme (2012) proposed a heuristic that greedily navigates in the direction of the destination. Bongiorno et al. (2021) later empirically demonstrated, through an analysis of trajectory data, that this heuristic accurately models the navigation behaviour of humans. This finding motivates our proposed models of street network entropy, which will assume the use of this heuristic for navigation.

3 Navigability Entropy

This section is structured as follows. In Section 3.1 we describe a generalised model of entropy from the field of game theory. In Section 3.2 we use this model to derive local and global models of street network entropy.

3.1 Generalised Entropy

The generalised model of entropy described in this section uses notation and terminology drawn from (Grünwald and Dawid, 2004).

We consider the following decision problem, which involves a game between a *decision-maker* and an *adversary* (Grünwald and Dawid (2004) refer to the adversary as “nature”). Let \mathcal{X} denote a finite set of states, \mathcal{A} denote a finite set of actions and $L : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ denote a loss function. A decision-maker is required to select an action $a \in \mathcal{A}$. Following this, the adversary will reveal a state $x \in \mathcal{X}$. The decision-maker will, in turn, then suffer a loss $L(x, a)$. Note that the adversary does not consider the action a before revealing the state x . Grünwald and Dawid (2004) refer to the tuple $(\mathcal{X}, \mathcal{A}, L)$ as a *basic game*.

Let Γ denote the set of probability distributions defined on \mathcal{X} , and let X be a random variable defined on \mathcal{X} with probability distribution $P \in \Gamma$. If we assume that the adversary samples $x \in \mathcal{X}$ using the random variable X , the undesirability for the decision-maker to select any action $a \in \mathcal{A}$ is equal to the *expected loss*. This is defined as follows, where p denotes the probability mass function of P :

$$L(P, a) = \sum_{x \in \mathcal{X}} p(x)L(x, a). \quad (2)$$

An action $a \in \mathcal{A}$ that minimises Equation 2 is known as a *Bayes act* against P . The *Bayes loss* $H(P)$ of a distribution $P \in \Gamma$ is defined as follows:

$$H(P) = \inf_{a \in \mathcal{A}} L(P, a). \quad (3)$$

In words, the Bayes loss is the optimal expected loss when the decision-maker selects an action $a \in \mathcal{A}$. The Bayes loss generalises many important measures of uncertainty. This includes the Shannon entropy where $\mathcal{A} = P$ (that is, a equals the probability mass function p) and $L(x, a) = -\log a(x)$. It also includes the variance where $\mathcal{A} = \mathcal{X}$ and $L(x, a) = \|x - a\|_2^2$. For this reason, Equation 3 is also referred to as the *generalised entropy*. This is the term we use for the remainder of this article (Grünwald and Dawid, 2004).

3.2 Navigability Entropy

In this section, we describe how generalised entropy can be used to define local and global models of street network entropy. Let $G = (V, E)$ be the edge-weighted directed multigraph representation of a given street network where the set of vertices V correspond to intersections and dead-ends, the set of edges E correspond to street segments, and the weighting function $w : E \rightarrow \mathbb{R}$ maps

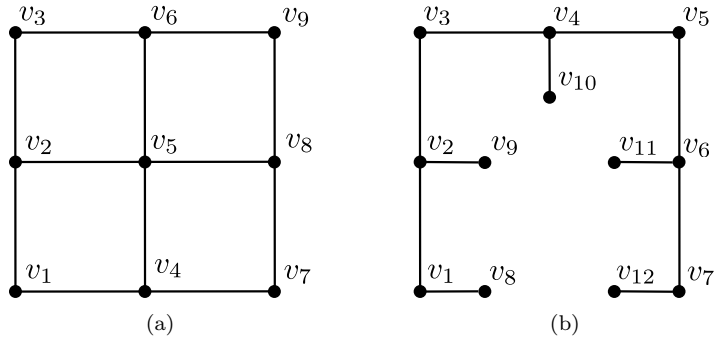


Figure 1: Two toy street network graphs are displayed in (a) and (b). The graph in (a) has a perfect grid pattern and therefore contains no cul-de-sacs. Each edge is bidirectional and has a length equal to 1. The graph in (b) does not have a perfect grid pattern and contains several cul-de-sacs. Each edge is bidirectional and has a length equal to 1 or 0.5.

each edge to its length, measured in metres. Let $N : V \rightarrow 2^V$ (the powerset of V) denote a map from a vertex to the set of its adjacent vertices.

We assume that a decision-maker wants to navigate from a source vertex $s \in V$ to a destination vertex $d \in V$ along a shortest route in G . Note that more than one shortest route may exist between a given pair of source and destination vertices. Let $S : V \rightarrow 2^V$ denote a mapping from a vertex appearing in a shortest route to the set of vertices that appear immediately after that vertex in a shortest route. To illustrate this mapping, consider the toy street network displayed in Figure 1(a). If we consider the case where the source vertex s equals v_1 and the destination vertex d equals v_8 , there are three shortest routes of length 3. These are the path v_1, v_4, v_7, v_8 , the path v_1, v_4, v_5, v_8 and the path v_1, v_2, v_5, v_8 . In this context, $S(v_1) = \{v_2, v_4\}$, $S(v_2) = \{v_5\}$ and $S(v_4) = \{v_5, v_7\}$. For the decision-maker to successfully navigate a shortest route, when at a vertex v_i in a shortest route, they must select an element from $N(v_i)$ that is also an element of $S(v_i)$. In layman's terms, they must select a vertex that keeps them on a shortest route to the destination d .

We model the above decision problem as a basic game $(\mathcal{X}, \mathcal{A}, L)$ as follows. First, we define \mathcal{X} to equal the set 2^V . When the decision-maker is at vertex v_i in a shortest route, the state $x \in \mathcal{X}$ is then equal to the set $S(v_i)$.

The definition of \mathcal{A} is based on the assumption that the decision-maker has access to only local bearing or direction information when navigating. Let $\text{lat} : V \rightarrow \mathbb{R}$ and $\text{lng} : V \rightarrow \mathbb{R}$ be mappings from vertices to corresponding latitude and longitude coordinates respectively. The bearing between two vertices equals the angle in degrees (clockwise) between north and the geodesic line from the first to the second vertex. It is defined by the function $\beta : V \times V \rightarrow [0, 360]$, where $\lambda = \text{lng}(v_2) - \text{lng}(v_1)$:

$$\beta(v_1, v_2) = \text{atan2}(\sin(\lambda) \cos(\text{lat}(v_2)), \cos(\text{lat}(v_1)) \sin(\text{lat}(v_2)) - \sin(\text{lat}(v_1)) \cos(\text{lat}(v_2)) \cos(\lambda)). \quad (4)$$

The bearing difference between two bearings is defined by the function $\Delta\beta : [0, 360] \times [0, 360] \rightarrow [0, 180]$:

$$\Delta\beta(a_1, a_2) = |(a_1 - a_2 + 180) \bmod 360 - 180|. \quad (5)$$

When at a vertex v_i in a shortest route, we assume the decision-maker can compute $\Delta\beta$ between the vertex v_i and each vertex in the set $N(v_i)$. We also assume that they can compute $\Delta\beta$ between the vertex v_i and the destination vertex d . These values could, for example, be computed using a GPS device given knowledge of the latitude and longitude coordinates for d . As discussed in Section 2, it has been demonstrated that a heuristic used by individuals when navigating is to select the vertex in $N(v_i)$ that minimises the angle difference between the bearing from v_i to this vertex and the bearing from v_i to d (Hochmair, 2005; Lee and Holme, 2012; Bongiorno et al., 2021). In layman's terms, this corresponds to navigating, as far as possible, in the direction of the

destination d . The heuristic in question is defined by the function $f : V \times V \rightarrow V$:

$$f(v_i, d) = \arg \min_{v' \in N(v_i)} \Delta\beta(\beta(v_i, v'), \beta(v_i, d)) \quad (6)$$

To illustrate this heuristic consider again the toy street network graph displayed in Figure 1(a), and the case where the source vertex s equals v_1 and the destination vertex d equals v_8 . When at vertex v_1 , the heuristic will select v_2 or v_4 as the next vertex. When at vertex v_2 , the heuristic will select v_5 as the next vertex. We assume that the decision-maker uses the above heuristic when attempting to navigate a shortest route. We model this assumption by defining the set of actions \mathcal{A} as the set of functions containing the single element f . That is, $\mathcal{A} = \{f\}$. This assumption allows us to simplify the definition of generalised entropy defined in Equation 3 as follows:

$$\begin{aligned} H(P) &= \inf_{a \in \mathcal{A}} L(P, a) \\ &= L(P, f). \end{aligned} \quad (7)$$

This equality results from the fact that \mathcal{A} only contains a single element f and, therefore, the infimum over this set is f . Finally, we define the loss function $L : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ as follows, where $\mathbb{1}$ is the indicator function:

$$L(x, a) = \mathbb{1}(a(v_i, d) \notin x) \quad (8)$$

The above loss function means that when the decision-maker is at a vertex v_i in a shortest route, they will incur a loss of 0 if the function $a \in \mathcal{A}$ returns an element in the set $S(v_i)$. Otherwise, the decision-maker will incur a loss of 1. In layman's terms, they incur zero loss if the function a keeps them on a shortest route to d .

As previously discussed, in this article we define both local and global models of street network entropy. Here, the local model is defined for individual vertices while the global model is defined with respect to an entire street network. The definition of these models is presented in the following two subsections and differs in how the probability mass function P is defined.

3.2.1 Global Model

We derive the global street network entropy model as follows. We define the probability mass function P to equal the probability mass function with which X occurs when the decision-maker is navigating the shortest routes. There are $2^{\binom{|V|}{2}}$ pairs of source and destination vertices and at least this number of shortest routes. This makes it infeasible to compute the expectation in Equation 2 exactly when the size of $|V|$ is large. Here, we therefore use plug-in estimation, which involves generating a random sample of shortest routes and computing the expectation with respect to this sample. To do this, we sample with replacement m pairs of distinct source and destination vertices $\{(s^1, d^1), \dots, (s^m, d^m)\}$ from V . For each pair (s^j, d^j) we compute using Dijkstra's algorithm a shortest route $v_1^j, \dots, v_{n(j)}^j$ from s^j to d^j where $n(j)$ equals the number of vertices in the path, $v_1^j = s^j$ and $v_{n(j)}^j = d^j$. Note that if more than one shortest path exists between s^j and d^j , one of these is selected arbitrarily. The overall complexity of this calculation is therefore $O(m(|E| + |V| \lg |V|))$, where $O(|E| + |V| \lg |V|)$ is the complexity of Dijkstra's algorithm using a Fibonacci heap. Given this, the plug-in estimation of the expected loss is defined as follows:

$$L(P, a) \approx \frac{1}{m} \sum_{j=1}^m \frac{1}{n(j)-1} \sum_{i=1}^{n(j)-1} L(S(v_i^j), a). \quad (9)$$

Using this, we can now derive a plug-in estimation of global street network entropy:

$$\begin{aligned}
H(P) &= L(P, f) \\
&\approx \frac{1}{m} \sum_{j=1}^m \frac{1}{n(j)-1} \sum_{i=1}^{n(j)-1} L(S(v_i^j), f) \\
&= \frac{1}{m} \sum_{j=1}^m \frac{1}{n(j)-1} \sum_{i=1}^{n(j)-1} \mathbb{1}(f(v_i^j, d) \notin S(v_i^j)) \\
&= \frac{1}{mn(j)-m} \sum_{j=1}^m \sum_{i=1}^{n(j)-1} \mathbb{1}(f(v_i^j, d) \notin S(v_i^j)).
\end{aligned} \tag{10}$$

Here, the first equality is the simplified definition of generalised entropy from Equation 7. The next approximate equality is formed by substituting the plug-in estimation defined in Equation 9. The next equality results from substituting the loss defined in Equation 8. From the final equality, we see that estimating global entropy reduces to the task of computing the proportion of times that the heuristic f **does not** select a vertex that keeps the decision-maker on a shortest route. This model returns a real value in the interval $[0, 1]$, where larger values indicate larger entropy. In this context, a larger entropy value indicates that shortest routes contained in the street network in question are more difficult to navigate using the heuristic f .

The toy street network displayed in Figure 1(a) has a global entropy value of 0. This is a consequence of the fact that the heuristic f will navigate shortest routes in a perfect grid pattern in all cases. In contrast, consider the toy street network displayed in Figure 1(b), which contains many cul-de-sacs. The global entropy value for this street network is greater than 0 and approximated by our model to equal 0.32. This is a consequence of the fact that the heuristic f does not navigate shortest routes in all cases. For example, consider the case where the source vertex s equals v_4 and the destination vertex d equals v_8 . When at vertex v_4 , the heuristic will select vertex v_{10} as the next vertex. This vertex corresponds to a cul-de-sac and is not on a shortest route to the destination v_8 .

3.2.2 Local Model

We now derive the local vertex entropy model as follows. For a given vertex s , we define the probability mass function P to equal the probability mass function with which X occurs when the decision-maker is located at s and is navigating to a vertex $d \neq s$. We again use plug-in estimation to estimate this value. To do this, we sample with replacement m vertices from the set $V - \{s\}$. Let us denote this set of vertices $\{d^1, \dots, d^m\}$. For each pair (s, d^j) we now compute a shortest route $v_1^j, \dots, v_{n(j)}^j$ from s to d^j . If more than one shortest route exists, we again select one of these paths uniformly at random. Given this, the plug-in estimation of the expected loss is defined as follows:

$$L(P, a) \approx \frac{1}{m} \sum_{j=1}^m L(S(v_1^j), a). \tag{11}$$

Using this, we can now derive a plug-in estimation of local street network entropy with respect to the vertex s :

$$\begin{aligned}
H(P) &= L(P, f) \\
&\approx \frac{1}{m} \sum_{j=1}^m L(S(v_1^j), a) \\
&= \frac{1}{m} \sum_{j=1}^m \mathbb{1}(f(v_1^j, d) \notin S(v_1^j)).
\end{aligned} \tag{12}$$

This calculation involves using Dijkstra's algorithm to compute a shortest path tree connecting s to each of the m selected vertices. Doing this for each $s \in V$ leads to an overall complexity of $O(|V|(|E| + |V| \lg |V|))$.

From the final approximate equality above, we see that estimating local vertex entropy reduces to computing the proportion of times that, when at the vertex s , the heuristic f **does not** select a vertex that keeps the decision-maker on a shortest route. Similar to the global street network entropy model defined above, this model returns a real value in the interval $[0, 1]$, where larger values indicate larger entropy.

Note that a vertex with degree one, which corresponds to a cul-de-sac, will always have a local vertex entropy value of 0.0. This is because its single neighbouring vertex must be on a shortest route if one exists. Also observe that computing the local entropy model can be computationally expensive for large street networks because it requires computing Equation 12 for each vertex in the street network.

Each vertex in the toy street network displayed in Figure 1(a) has a local vertex entropy value of 0. Again, this is a consequence of the fact that the heuristic f will navigate shortest routes in a perfect grid pattern in all cases. In contrast, each vertex in the toy street network displayed in Figure 1(b) has a local vertex entropy value greater than 0.0 apart from the vertices with degree one.

The local and global models of street network entropy presented above are related in the following way. The global model equals the expected loss L with respect to all decisions made when navigating from a random source vertex to a random destination vertex. On the other hand, the local model equals the expected loss L with respect to the first decision when navigating from a fixed source vertex to a random destination vertex. Note that, since both the local and global models approximate expected values, the entropy values in question for cities of different shapes and sizes can be directly compared without any form of normalisation.

4 Results and Analysis

In this section, we use the proposed models of local vertex and global street network entropy to perform an analysis of UK city street networks. Unfortunately, there exists no consensus on how to define the boundary of a city and, in turn, the boundary of a city street network (Gudmundsson and Mohajeri, 2013). In this work, we choose to define the boundaries of UK city street networks using the Urban Centre Database (UCD), which is a project supported by the European Commission’s Joint Research Centre and Directorate-General for Regional and Urban Policy (Florczyk et al., 2019). This database contains boundaries for the majority of the world’s major urban areas. These boundaries are determined by fusing census population and remotely sensed image data. At the date of writing, the UK contains 70 cities. However, some of these are very small in terms of geographical area and population and therefore are not represented in the UCD. The city of Armagh, which has a population of less than 15 thousand, is one such city. Furthermore, in several cases, a set of spatially close or touching cities are represented as a single urban area in the database. For example, the cities of Wolverhampton and Birmingham are represented as a single urban area entitled Birmingham. In this work, we only considered the 48 UK cities represented in the UCD. Given the UCD city boundaries, for each city we obtained the corresponding street network of drivable public streets from OpenStreetMap (OSM) which is a crowdsourcing project for geographical data (Boeing, 2017). Table 1 displays the name of each city considered, plus the number of vertices ($|V|$), the number of edges ($|E|$) and the geographical area (measured in km^2) of the corresponding street networks.

As previously noted, both the local and global models of street network entropy use a plug-in estimate parameterised by a value m . In our analysis, we computed local and global street network entropy using a value of m equal to 200 and 5,000 respectively. To better understand the properties of global street network entropy, for each city street network we also computed the street segment orientation entropy, the intersection density and the route directness. Street segment orientation entropy is the most commonly used model of street network entropy. It is therefore instructive to understand the relationship between this and the proposed models. We computed the street segment orientation entropy of each city street network using the algorithm described by Boeing (2019b) and implemented by the OSMnx software library (Boeing, 2017). This algorithm first computes the discrete empirical distribution of street network orientations, where this distribution contains 36 bins each representing a 10-degree interval of orientation. It

City	$ V $	$ E $	Area	City	$ V $	$ E $	Area
Aberdeen	6060	14228	57	Lancaster	3303	7599	37
Bangor	2083	4555	20	Leeds	49165	110377	472
Bath	2275	5032	22	Leicester	13024	29152	122
Belfast	17378	39283	163	Lincoln	4287	9228	38
Birmingham	69322	154165	668	Liverpool	45381	103130	418
Brighton	8342	19886	108	London	157835	364888	1865
Bristol	19731	43677	169	Manchester	74909	171593	674
Cambridge	3382	7307	37	Newcastle	27877	63378	241
Canterbury	1677	3495	15	Newport	4147	9089	57
Cardiff	10735	23810	109	Norwich	6801	14752	55
Carlisle	2046	4639	17	Nottingham	19848	43862	170
Chelmsford	3298	6929	31	Oxford	3315	7376	46
Chester	3036	6733	27	Peterborough	6897	14284	58
Coventry	13412	29212	126	Plymouth	8168	18048	88
Derby	8716	18931	83	Portsmouth	13325	29544	166
Derry	3172	6719	28	Preston	9532	21021	79
Dundee	5500	12761	62	Sheffield	22561	51089	247
Durham	1791	4782	99	Southampton	10310	22592	144
Edinburgh	11323	26537	123	Southend-on-Sea	5681	13102	75
Exeter	4248	9034	31	St Albans	2265	5062	27
Glasgow	29313	68232	306	Stoke-on-Trent	12448	27751	118
Gloucester	6287	13264	45	Sunderland	6599	15314	59
Hereford	1761	3841	17	Swansea	4049	9114	60
Hull	11142	24012	102	Worcester	3224	6984	30

Table 1: Details of the 48 UK city street networks considered in this study. For each street network, the number of vertices ($|V|$), the number of edges ($|E|$) and the geographical area (measured in km^2) is provided.

then computes the Shannon entropy of this distribution. The intersection density of a given street network equals the number of vertices in the street network divided by the size of the geographical area in question. Intersection density is considered an indicator of the level of connectivity and development of a street network (Dill, 2004). This is a consequence of the fact that street networks with a high density of intersections contain a larger number of distant paths and, in turn, a greater probability for the existence of more direct paths. Finally, the route directness of a given street network equals the ratio of shortest route distances to great circle distances (Huang and Levinson, 2015; Giacomini and Levinson, 2015; Boeing, 2019a). Route directness is considered a measure of the level of efficiency of a street network because more direct routes contain fewer detours. In this work, we approximate the route directness of a street network by sampling with replacement 1,000 distinct source and destination vertex pairs and computing the mean directness of the corresponding shortest routes. Table 2 displays the street segment intersection density values, the orientation entropy values, the route directness values and the proposed global street network entropy values for each city.

An analysis of the global entropy values for each of our city street networks provides some interesting insights. The mean of these values is 0.144. This is a relatively low global entropy value and indicates that the vast majority of decisions when navigating shortest routes can be made correctly using only local direction information. On the other hand, the variance of these values is quite large. The city of London has the lowest global entropy value of 0.088 while the city of Carlisle has the highest global entropy value of 0.191. This indicates that it is more than twice as difficult to navigate shortest routes using local direction information in the latter compared to the former. This can be attributed to the fact that the city of Carlisle, which is displayed in Figure 2, contains a relatively large number of cul-de-sacs adjacent to important transportation corridors. Three of these are highlighted in the figure. As previously illustrated using the toy street network in Figure 1(b), cul-de-sacs frequently do not belong to shortest routes and, in turn, contribute to higher entropy. Furthermore, the western and eastern regions of the street network are separated by a river, and all routes between these regions involve crossing bridges in the north. This requirement to travel north when travelling east to west and vice versa is not captured by the direction-based heuristic: in such cases, the heuristic will recommend moving in an easterly or westerly direction instead of a northerly direction. Figure 3 displays the street network for the city of Dundee which is of a similar size to the street network for the city of Carlisle. However,

City	ID	OE	RD	NE	City	ID	OE	RD	NE
Aberdeen	106.06	3.54	1.38	0.132	Lancaster	89.08	3.55	1.53	0.178
Bangor	103.66	3.54	1.53	0.150	Leeds	104.01	3.58	1.29	0.112
Bath	103.27	3.57	1.56	0.174	Leicester	106.64	3.54	1.35	0.113
Belfast	106.08	3.58	1.37	0.113	Lincoln	112.69	3.57	1.46	0.149
Birmingham	103.63	3.58	1.24	0.093	Liverpool	108.34	3.58	1.42	0.143
Brighton	77.21	3.47	1.31	0.100	London	84.61	3.57	1.29	0.088
Bristol	116.58	3.58	1.38	0.124	Manchester	111.12	3.58	1.27	0.108
Cambridge	91.33	3.55	1.57	0.151	Newcastle	115.67	3.55	1.48	0.156
Canterbury	111.71	3.52	1.63	0.159	Newport	72.62	3.58	1.63	0.178
Cardiff	98.3	3.58	1.48	0.157	Norwich	123.51	3.57	1.4	0.123
Carlisle	120.04	3.56	1.59	0.191	Nottingham	116.61	3.58	1.39	0.127
Chelmsford	106.3	3.57	1.58	0.162	Oxford	72.0	3.55	1.5	0.153
Chester	112.22	3.57	1.6	0.179	Peterborough	118.8	3.57	1.63	0.161
Coventry	106.32	3.57	1.36	0.131	Plymouth	92.58	3.56	1.48	0.152
Derby	104.88	3.58	1.51	0.154	Portsmouth	80.21	3.5	1.7	0.171
Derry	112.51	3.55	1.75	0.187	Preston	120.41	3.54	1.48	0.138
Dundee	88.46	3.51	1.33	0.114	Sheffield	91.22	3.58	1.35	0.125
Durham	18.06	3.19	1.32	0.127	Southampton	90.36	3.58	1.54	0.143
Edinburgh	91.81	3.54	1.34	0.133	Southend-on-Sea	75.69	3.45	1.38	0.130
Exeter	136.74	3.58	1.58	0.158	St Albans	83.82	3.58	1.49	0.150
Glasgow	95.78	3.56	1.37	0.131	Stoke-on-Trent	105.33	3.57	1.38	0.122
Gloucester	139.52	3.55	1.54	0.174	Sunderland	111.66	3.52	1.44	0.132
Hereford	103.41	3.57	1.62	0.189	Swansea	67.31	3.57	1.49	0.155
Hull	109.1	3.52	1.4	0.139	Worcester	107.31	3.56	1.63	0.186

Table 2: The intersection density (ID) values, the street segment orientation entropy (OE) values, the route directness (RD) values, and the global street network entropy (NE) values for each city street network are displayed.

this street network has a significantly lower global entropy value of 0.114. This can be attributed to the fact that this street network contains a relatively small number of cul-de-sacs adjacent to important transportation corridors. Furthermore, the environment in which the street network exists does not contain any barriers, such as a river, so shortest routes are more easily determined using only directional information.

The analysis of global entropy values could be used by urban planners to identify those cities that are more difficult to navigate and therefore in need of additional signage. It could also be used by urban planners to help suggest changes or additions to a given street network structure that improves navigability. It is likely that in some situations a trade-off exists between improving the navigability of a city and improving other aspects of navigation such as the length of shortest routes.

Table 3 displays the Pearson and Spearman’s rank correlation between the global entropy values and the street segment orientation entropy values, the intersection density values and the route directness values for our sample of cities. Each cell displays the correlation statistic in question,

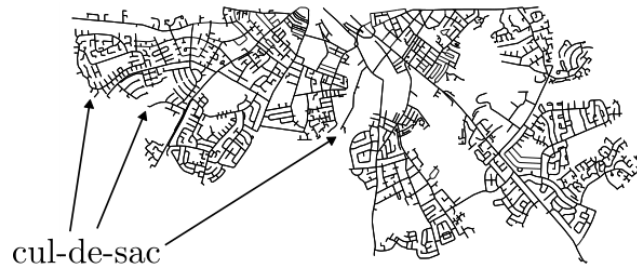


Figure 2: The street network for the city of Carlisle is displayed. This street network has the highest global entropy value of all city street networks considered. This can in part be attributed to a significant number of cul-de-sacs adjacent to important transportation corridors. A number of these are highlighted in the figure. Cul-de-sacs frequently do not belong to shortest routes and, in turn, contribute to higher entropy.

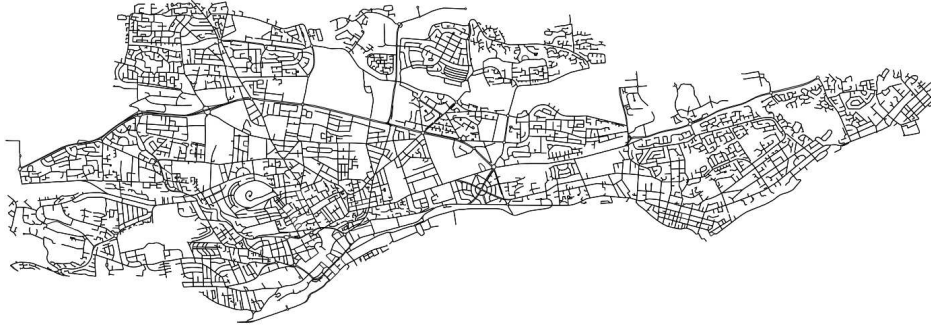


Figure 3: The street network for the city of Dundee is displayed. This street network has one of the smallest global entropy values of all city street networks considered. This can in part be attributed to a relatively small number of cul-de-sacs adjacent to important transportation corridors.

	Orientation entropy	Intersection density	Route directness
Pearson	0.14, 0.13, (-0.13, 0.31)	0.15, 0.28, (-0.10, 0.36)	0.91, < 0.001, (0.86, 0.93)
Spearman	-0.09, 0.36, (-0.12, 0.32)	0.13, 0.35, (-0.15, 0.41)	0.92, < 0.001, (0.88, 0.95)

Table 3: The Pearson and Spearman’s rank correlation between the global entropy values and the street segment orientation entropy values, the intersection density values and the route directness values. Each cell displays the correlation statistic in question, followed by the p-value and the 95% confidence interval.

followed by the p-value and the 95% confidence interval. The correlation between the global street network entropy values and the street segment orientation entropy values is weak and not statistically significant. This indicates that the former model measures a novel aspect of street network entropy even though both models are a function of direction information. The correlation between the global street network entropy values and the intersection density values is also weak and not statistically significant. This indicates that a greater degree of connectivity does not imply greater navigability. Finally, the correlation between the global street network entropy values and the route directness values is strong and statistically significant. This indicates that street networks that have more direct routes and, in turn, are more efficient, have lower entropy with respect to the task of navigating.

Figure 4(a) displays a histogram of the local vertex entropy values corresponding to the Cardiff city street network. Recall that these values equal the proportion of times that, when at the vertex in question, the direction-based heuristic does not select a vertex that keeps the decision-maker on a shortest route. The histogram is left-skewed, with a long tail to the right. Of the 10,735 vertices in the street network, 3,806 vertices (35%) have an entropy value equal to 0.0 while 765 vertices (7%) have an entropy value greater than or equal to 0.9. This demonstrates that, while a large percentage of vertices can be successfully navigated using only direction information, a significant percentage cannot.

Figure 5(a) displays the 2% of vertices that have the greatest local vertex entropy values. The two regions highlighted in this figure using blue dashed boxes are displayed in Figures 6(a) and 6(c). The regions in question correspond to a suburban residential region in the northwest, and a city-centre region respectively. As discussed previously, a vertex corresponding to a cul-de-sac will always have a local vertex entropy value of 0.0. Although some vertices identified in the above figures may appear to correspond to cul-de-sacs, they are not. Instead, this is a consequence of the red dots covering corresponding neighbouring streets. We can see that the suburban residential region has a significantly higher density of vertices with large entropy values than the city-centre region. This is a consequence of the fact that, by design, residential regions contain a relatively large number of streets leading to cul-de-sacs.

The local entropy model has many potential applications. It could potentially be used to help recommend routes that are easier to navigate and, in turn, easier to describe. It could also be

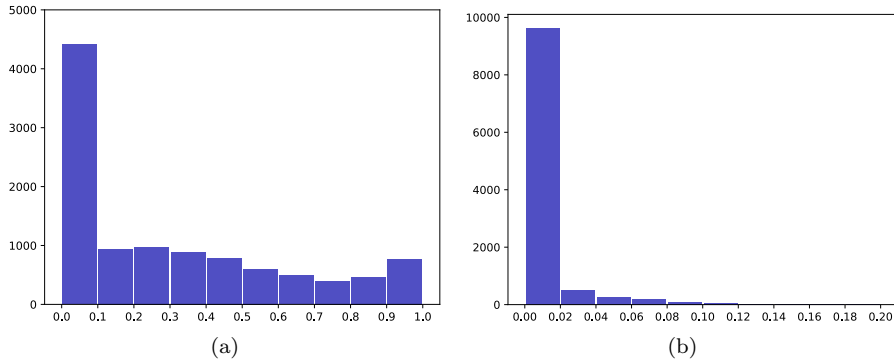


Figure 4: Histograms of the local vertex entropy and vertex betweenness centrality values for the Cardiff city street network are displayed in (a) and (b) respectively.

used to identify vertices where drivers are likely to frequently take an incorrect turn and therefore indicate a need for different road layouts or new signage. To help identify these vertices, we computed the vertex-wise product of local vertex entropy values and vertex betweenness centrality values. The betweenness centrality for a vertex equals the percentage of shortest routes that pass through that vertex (Brandes, 2001). Vertices with a large corresponding product value are those that are both more likely to be travelled frequently and more likely for an incorrect turn to be made at. Figure 4(b) displays a histogram of the betweenness centrality values corresponding to the Cardiff city street network. The histogram is skewed to the left indicating that most vertices have small centrality values. Figure 5(b) displays the 2% of street network vertices that have the greatest product of entropy and betweenness centrality values. Figures 6(b) and 6(d) display the same two highlighted regions considered above. By examining these figures we can identify the vertices that exist along important navigation corridors and have high entropy. For example, the most northerly vertex identified in Figure 6(d) corresponds to an intersection along a major navigation corridor that leads to a cul-de-sac. If not already in place, the street network intersections in question may benefit from clear signage to prevent drivers from frequently taking incorrect turns.

5 Conclusions

This article has proposed new models of street network entropy that are defined with respect to both individual network vertices and entire street networks. These models measure the uncertainty of navigating given only local direction information. This is distinct from existing models of street network entropy which, instead, measure the uncertainty of more abstract properties including street segment direction. Since navigating is the very task that street networks are intended to facilitate, we argue that the proposed models are more interpretable and this, in turn, increases the potential for useful analyses to be performed.

Despite the many potential applications, the proposed models have some limitations that should be considered if attempting to implement these applications. The proposed models assume that individuals navigate using only directional information. However, in many cases, individuals use other forms of information including landmarks, street signage, prior knowledge of the geographical region and satellite navigation systems. Further research is necessary to determine and model how humans integrate and reason using these different forms of information when navigating.

Code Availability Statement - The Python code used to compute the global and local model values described in Sections 3.2.1 and 3.2.2 is openly available on figshare at https://figshare.com/articles/software/Navigability_Entropy_Model_of_Street_Networks/21593688.

Conflicts of Interest - The authors declare no conflict of interest.

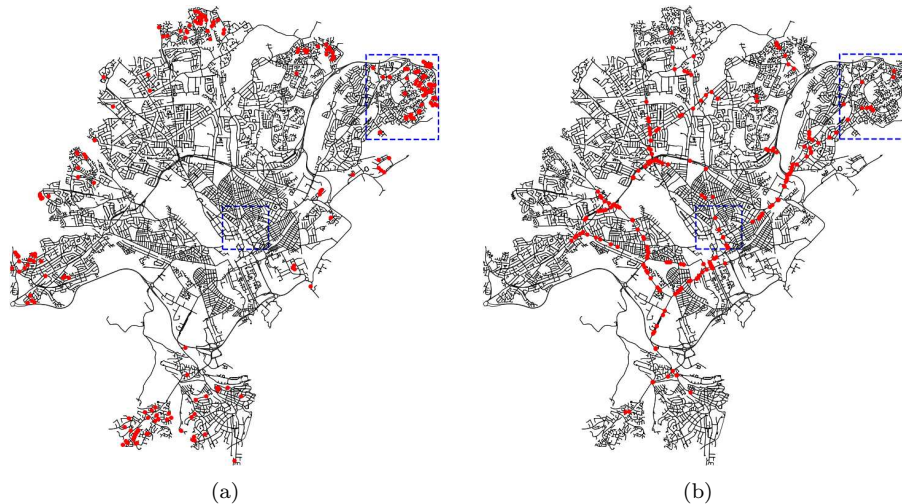


Figure 5: The Cardiff street network is displayed in both figures (a) and (b). In the former figure, the 2% of vertices that have the greatest local entropy value are represented by red dots. The local entropy of a vertex equals the likelihood of one taking an incorrect turn at that vertex when navigating. In the latter figure, the 2% of vertices that have the greatest product of local entropy value and betweenness centrality value are represented by red dots. The betweenness centrality for a vertex equals the percentage of shortest routes that pass through that vertex. The two regions highlighted in these figures using blue dashed boxes are examined in Figure 6.

Ethical statement - This article does not contain any studies with human or animal participants.

References

- Marc Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.
- Geoff Boeing. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 2017.
- Geoff Boeing. The morphology and circuitry of walkable and drivable street networks. In *The mathematics of urban morphology*, pages 271–287. Springer, 2019a.
- Geoff Boeing. Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, 4(1):1–19, 2019b.
- Christian Bongiorno, Yulun Zhou, Marta Kryven, David Theurel, Alessandro Rizzo, Paolo Santi, Joshua Tenenbaum, and Carlo Ratti. Vector-based pedestrian navigation in cities. *Nature Computational Science*, 1(10):678–685, 2021.
- Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.
- Padraig Corcoran, Peter Mooney, and Michela Bertolotto. Interactive cartographic route descriptions. *GeoInformatica*, 18(1):1–26, 2014.
- Antoine Coutrot, Ricardo Silva, Ed Manley, Will de Cothi, Saber Sami, Véronique D Bohbot, Jan M Wiener, Christoph Hölscher, Ruth C Dalton, Michael Hornberger, et al. Global determinants of navigation ability. *Current Biology*, 28(17):2861–2866, 2018.
- Antoine Coutrot, Sophie Schmidt, Lena Coutrot, Jessica Pittman, Lynn Hong, Jan M Wiener, Christoph Hölscher, Ruth C Dalton, Michael Hornberger, and Hugo J Spiers. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *PloS one*, 14(3):e0213272, 2019.

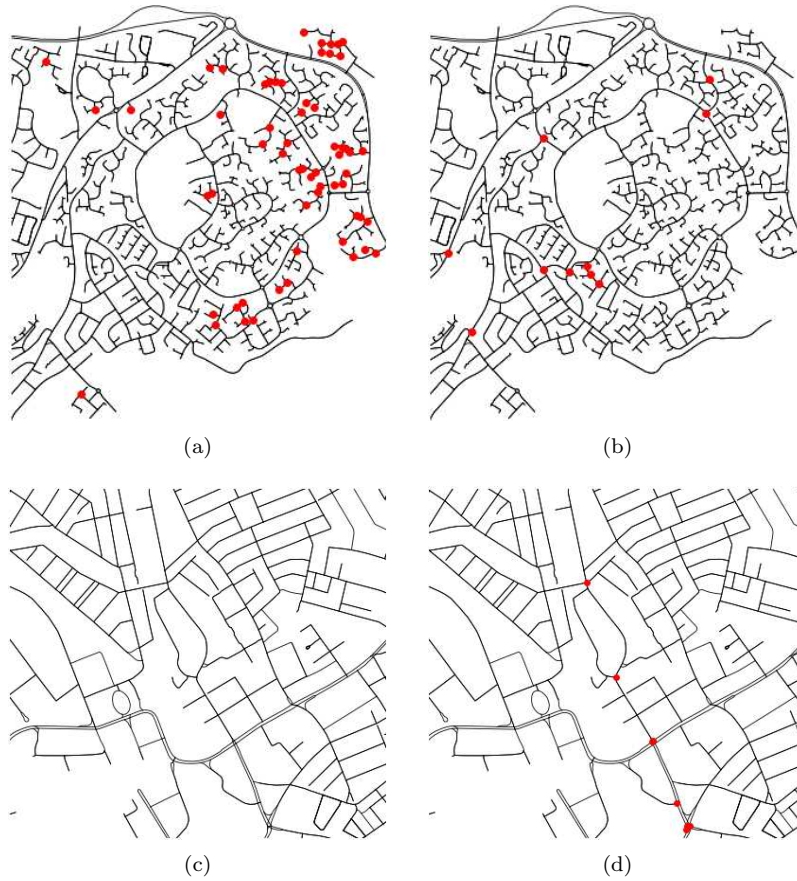


Figure 6: The two regions highlighted in Figure 5(a), corresponding to a suburban residential region in the northwest and a city centre region, are displayed in (a) and (c) respectively. The same two regions in Figure 5(b) are displayed in (b) and (d) respectively. The red dots in (a) and (c) indicate those vertices where the likelihood of one taking an incorrect turn when navigating is greatest. The red dots in (b) and (d) indicate those vertices where one is most likely to navigate through and take an incorrect turn when doing so.

- Antoine Coutrot, Ed Manley, Sarah Goodroe, Christoffer Gahnstrom, Gabriele Filomena, Demet Yesiltepe, RC Dalton, Jan M Wiener, Christian Hölscher, Michael Hornberger, et al. Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904):104–110, 2022.
- Morris H DeGroot. Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419, 1962.
- Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181(1):57–78, 2011.
- Jennifer Dill. Measuring network connectivity for bicycling and walking. In *83rd annual meeting of the Transportation Research Board, Washington, DC*, pages 11–15, 2004.
- A. Florczyk, M. Melchiorri, C. Corban, M. Schiavina, L. Maffenini, M. Pesaresi, P. Politis, F. Sabo, S.M. Carneiro Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. Description of the ghs urban centre database 2015. (KJ-02-19-103-EN-N (online)), 2019. doi: 10.2760/037310(online).
- David J Giacomini and David M Levinson. Road network circuitry in metropolitan areas. *Environment and Planning B: Planning and Design*, 42(6):1040–1053, 2015.
- Reginald G Golledge. Path selection and route preference in human navigation: A progress report. In *International conference on spatial information theory*, pages 207–222. Springer, 1995.
- Reginald G Golledge and Tommy Gärling. Cognitive maps and urban travel. In *Handbook of transport geography and spatial systems*. Emerald Group Publishing Limited, 2004.
- Peter D Grünwald and A Philip Dawid. Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory. *the Annals of Statistics*, 32(4):1367–1433, 2004.
- Agust Gudmundsson and Nahid Mohajeri. Entropy and order in urban street networks. *Scientific reports*, 3(1):1–8, 2013.
- Hartwig H Hochmair. Investigating the effectiveness of the least-angle strategy for wayfinding in unknown street networks. *Environment and Planning B: Planning and Design*, 32(5):673–691, 2005.
- Jie Huang and David M Levinson. Circuitry in urban transit networks. *Journal of Transport Geography*, 48:145–153, 2015.
- Sang Hoon Lee and Petter Holme. Exploring maps with greedy navigators. *Physical Review Letters*, 108(12):128701, 2012.
- Valentina Marin, Carlos Molinero, and Elsa Arcaute. Uncovering structural diversity in commuting networks: global and local entropy. *Scientific Reports*, 12(1):1–13, 2022.
- Nahid Mohajeri and Agust Gudmundsson. The evolution and complexity of urban street networks. *Geographical Analysis*, 46(4):345–367, 2014.
- Alexander W Siegel and Sheldon H White. The development of spatial representations of large-scale environments. *Advances in child development and behavior*, 10:9–55, 1975.