

# Road layout in the KPZ class

Márton Balázs\*, Sudeshna Bhattacharjee<sup>†</sup>,  
Karambir Das<sup>‡</sup>, David Harper<sup>§</sup>

April 26, 2024

## Abstract

We propose a road layout and traffic model, based on last passage percolation (LPP). An easy naïve argument shows that coalescence of traffic trajectories is essential to be considered when observing traffic networks around us. This is a fundamental feature in first passage percolation (FPP) models where nearby geodesics naturally coalesce in search of the easiest passage through the landscape. Road designers seek the same in pursuing cost savings, hence FPP geodesics are straightforward candidates to model road layouts. Unfortunately no detailed knowledge is rigorously available on FPP geodesics. To address this, we use exponential LPP instead to build a stochastic model of road traffic and prove certain characteristics thereof. Cars start from every point of the lattice and follow half-infinite geodesics in random directions. Exponential LPP is known to be in the KPZ universality class and it is widely expected that FPP shares very similar properties, hence our findings should equally apply to FPP-based modelling. We address several traffic-related quantities of this model and compare our theorems to real life road networks.

## 1 Introduction

Road networks and traffic patterns are clearly important phenomena in everyday life. Cars start from many locations and generally take random destinations to travel to. One's first thought might be to model car movements as independent straight trajectories. We show in Section 1.1 that this naïve model results in divergent traffic densities in every region of the plane, which clearly does not align with observations.

Cars require roads to run on, and these cannot be built in arbitrary density and through arbitrary landscapes. Hence an interesting structure of road networks emerges, largely driven by geographic, historic, and social circumstances. The aim of this paper is to provide a mathematical model to describe some characteristics of such networks.

Characteristics of road networks have been considered in numerous works. As in this paper we are interested in a mathematical model to construct road networks, we only mention two papers, first Barthélemy and Flammini [3], who build urban networks by adding new centres and connecting them to the existing network in certain optimal way. Second, Molinero and Hernando [15] introduce an

---

\*University of Bristol [m.balazs@bristol.ac.uk](mailto:m.balazs@bristol.ac.uk)

<sup>†</sup>Indian Institute of Science [sudeshnab@iisc.ac.in](mailto:sudeshnab@iisc.ac.in)

<sup>‡</sup>Indian Institute of Science [karambirdas@iisc.ac.in](mailto:karambirdas@iisc.ac.in)

<sup>§</sup>Georgia Institute of Technology [dharp40@gatech.edu](mailto:dharp40@gatech.edu)

algorithm based on a succession of Delanuy triangulation of a dense network of the centres of interest plus auxiliary points, and only keeping paths that are shortest between the original nodes of interest in the resulting network.

We build a model where, rather than the locations to connect, the environment surrounding the roads has a crucial effect on the geometry. Points to connect are thought of as having a homogeneous configuration throughout the plane, whereas obstacles roads need to get around will be random which in turn causes nontrivial network characteristics.

The environment we live in presents geographical challenges to road building. If we model this in the simplest possible way with i.i.d. cost distribution on some lattice, and try to build roads that minimise overall cost, then we naturally arrive to first passage percolation (FPP) models.

FPP was originally proposed as a model for the flow of fluid through porous media by Hammersley and Welsh in 1965 [13]. Since then, FPP has evolved into a major area of interest in probability theory. See [1] for an extensive account on significant results in this field. Despite this active interest, many significant unanswered questions remain.

One of the fundamental objects in FPP are *geodesics*, paths that collect minimal costs connecting two points among a random penalty landscape on the plane. These will be our candidates to model roads. For quantitative analysis of our road model we need refined estimates on geodesics, which unfortunately are not yet available for FPP models. Instead, we switch to exponential last passage percolation (LPP), where extensive results on last passage times and geometry of geodesics are now available. FPP is widely believed to belong to the *KPZ universality class* – a property proved for LPP. This implies that the geometry of geodesics in the two models should share basic characteristics that we use throughout our arguments. Thus, switching to LPP from FPP seems a reasonable move. Of the vast literature on the KPZ universality class, we refer to the surveys by Corwin [5], Ferrari-Spohn [11], and Quastel [16].

LPP models have been subject of intensive research in the last decades. Instead of minimizing weights (which we called costs so far), geodesics are maximising them under the constraint that paths can only take up or right steps. We define this model precisely in Section 2.

The connection between roads and models in the KPZ universality class is not new. Solon, Bunin, Chu and Kardar [17] compared optimal paths in road networks with directed polymers in random medium (DPRM). LPP can be considered as extremal case of DPRM, and shares similar scaling exponents. Solon et al. find that shortest paths on the road network can be well approximated with DPRM – albeit with the environment exhibiting a power law rather than in exponential LPP. They also conclude that long-range correlations in the road network play an important role in the scaling properties of optimal paths.

To model cars, we augment the basic layer of LPP models with a network of half-infinite geodesics, the a.s. existence of which was established by Ferrari-Pimentel and Coupier [6, 12]. To be more precise, for any given direction and starting point, a.s. there exists a unique half-infinite geodesic from that point going into the given asymptotic direction. These form a perfect model of a car pursuing a distant destination in a given direction, while following the geodesic road network given to it. Each lattice point of  $\mathbb{Z}^2$  is thought of as the starting point of a car, and each car picks an independent random direction for its a.s. geodesic. As  $\mathbb{Z}^2$  is countable, the cars can jointly follow their own randomly oriented geodesics on a probability one event. In FPP models the direction could be  $\text{Uniform}(0, 2\pi)$ , in our LPP setup this must be restricted to  $\text{Uniform}(0, \frac{\pi}{2})$ . In fact we will assume some  $\epsilon$  separation from the trivial angles 0 and  $\frac{\pi}{2}$ .

We then discuss the following questions; notice that the model is translation-invariant, hence our inquiry for the origin is not restricting generality:

1. Probabilistic estimates for the furthest distance a car can come from to the origin (see Section 3).
2. Probabilistic bounds for the number of cars passing through a fixed point (see Section 4).
3. How far one needs to go from a point to see high-traffic roads i.e., geodesics used by several cars at the same time (see Section 5)?

The purpose of this paper is to introduce this theoretical traffic model based on LPP, and to rigorously derive some of its interesting characteristics. We do not claim a good fit with reality: in Section 6, we compare our findings to real-world traffic data and point to various reasons why our model stays in the theoretical domain rather than proving useful for practical applications about road networks.

### 1.1 A Poisson model for road network

Here we make the naïve assumption that cars go in a straight line i.e., the environment has no effect on their trajectories. The aim of this part is to demonstrate the need of a more elaborate model: this simple Poisson model cannot properly describe road networks.

Consider a homogeneous Poisson point process on  $\mathbb{R}^2$  with intensity 1. These Poisson points represent the starting points of cars. A car chooses a direction uniformly between 0 to  $2\pi$  independently of everything, and it travels along a straight line segment of length  $\ell$  in this direction. The length  $\ell$  of the trip follows the  $\text{Exponential}(\gamma)$  distribution with a positive fixed parameter  $\gamma$ , and is independent of all the other variables and cars.

This process can be thought of as a marked point process  $\xi$  where the starting point  $X_i$  of car  $i$  is a point in the homogeneous Poisson point process on  $\mathbb{R}^2$ , and this gets decorated with the mark  $(\theta_i, \ell_i)$  which represent the direction the car goes towards and the distance it travels respectively. The mark space is  $[0, 2\pi) \times [0, \infty)$ , equipped with the standard product Borel  $\sigma$ -algebra. By the marking theorem [14, Theorem 5.6],  $\xi$  is a Poisson point process on  $(\mathbb{R}^2 \times [0, 2\pi) \times [0, \infty))$ . We are after the number  $N_r$  of cars that come  $r$  close to the origin.

Without loss of generality we now assume that a car starts from coordinate  $(-z, 0)$  of  $\mathbb{R}^2$ , where  $z$  is a positive real. If  $z \leq r$  then the car does appear in the disc  $B(0, r)$  of radius  $r$  around the origin.

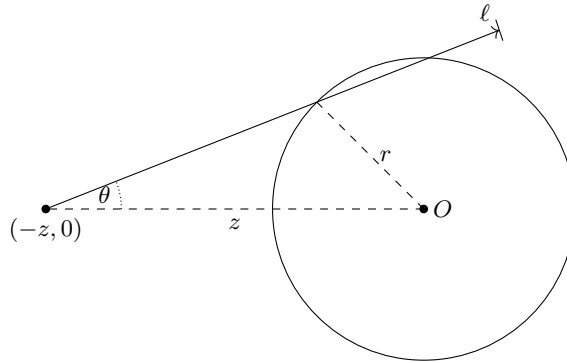


Figure 1: A car starting from  $(-z, 0)$  and intersecting the disc  $B(0, r)$  of radius  $r$  around the origin.

For  $z > r$ , as seen from Figure 1 and a bit of trigonometry, the car intersects  $B(0, r)$  if and only if  $|\sin \theta| \leq \frac{r}{z}$  and  $\ell \geq z \cos \theta - \sqrt{r^2 - z^2 \sin^2 \theta}$ . For a car in general position the quantity  $z$  is to be replaced by the distance of the car's starting point from the origin. This way, exactly cars of the marked point process  $\xi$  in the above region of  $(\mathbb{R}^2 \times [0, 2\pi) \times [0, \infty))$  will make it into  $B(0, r)$ . Hence the number  $N_r$  of such cars is Poisson distributed with parameter equal the mean of this number.

Assuming a car is starting at distance  $z > r$  from the origin, we estimate its probability to hit  $B(0, r)$  by

$$\mathbb{P}\{|\sin \theta| \leq \frac{r}{z} \cap \ell \geq z \cos \theta - \sqrt{r^2 - z^2 \sin^2 \theta}\} \geq \mathbb{P}\{|\theta| \leq \frac{r}{z} \cap \ell \geq z\} = \frac{r}{\pi z} \cdot e^{-\gamma z}.$$

Breaking up the homogeneous Poisson process on  $\mathbb{R}^2$  w.r.t. polar coordinates, we get from here

$$\mathbb{E}[N_r] \geq \int_0^{2\pi} \int_0^r 1 \cdot z \, dz \, d\varphi + \int_0^{2\pi} \int_r^\infty \frac{r}{\pi z} \cdot e^{-\gamma z} \cdot z \, dz \, d\varphi = \pi r^2 + \frac{2r}{\gamma} \cdot e^{-\gamma r}.$$

We conclude that the number of cars intersecting the disc is Poisson with mean at least the right-hand side of this display.

The mean distance travelled by cars is  $\frac{1}{\gamma}$ , and since we already fixed the Poisson intensity of cars at 1, we can think of this as a large number. Similarly, the radius  $r$  of interest e.g., for a homeowner wishing for a quiet house, can be considered  $\mathcal{O}(1)$ . Hence we find that even in a moderate sized garden, plenty of cars should pass. (For actual figures, the UK's 243,610 km<sup>2</sup> and 33.5m cars give 85 metres as unit length to get density 1 of cars, and the typical driving distance is around 10...15 km:  $\frac{1}{\gamma}$  is in the order of 100's. Of course spatial density fluctuates drastically between urban and rural areas.) Taking the driving distance to infinity ( $\gamma \searrow 0$ ) makes  $N_r$  divergent.

This is not what we find in real life: the coalescence of the cars' trajectories is a significant missing feature in this naïve model.

## 2 Model definitions, notations and results

We first define the exponential last passage percolation model on  $\mathbb{Z}^2$ . We assign i.i.d. random variables  $\{\tau_v\}_{v \in \mathbb{Z}^2}$  to each vertex of  $\mathbb{Z}^2$ , where  $\tau_v$ 's are distributed as  $\text{Exp}(1)$ . Let,  $u, v \in \mathbb{Z}^2$  be such that  $u \leq v$  (i.e., if  $u = (u_1, u_2), v = (v_1, v_2)$  then  $u_1 \leq v_1$  and  $u_2 \leq v_2$ ). For an up-right path  $\gamma$  between  $u$  and  $v$  we define  $\ell(\gamma)$  to be  $\sum_{v \in \gamma \setminus \{u, v\}} \tau_v$ . Let  $T(u, v) := \max\{\ell(\gamma) : \gamma \text{ is an up-right path from } u \text{ to } v\}$ .  $T(u, v)$  is the *last passage time* between  $u$  and  $v$ . Clearly, as the number of up-right paths between  $u$  and  $v$  is finite, the maximum is always attained. Between any two points  $u, v \in \mathbb{Z}^2$  maximum attaining paths are called *geodesics*. As  $\tau_v$  has a continuous distribution, between any two points  $u \leq v \in \mathbb{Z}^2$  almost surely there exists a unique geodesic. As a consequence of this uniqueness geodesics do not form loops. Hence, geodesics starting from ordered vertices (ordered in the  $x+y=0$  direction) always stay ordered. This property of geodesics is known as *planarity* and will be used in our proofs. The a.s. unique geodesic will be denoted by  $\Gamma_{u,v}$ . An infinite up-right path  $u_0, u_1, u_2, \dots$  in  $\mathbb{Z}^2$  is called a *semi-infinite geodesic starting from*  $u_0$ , if every finite segment of it is a geodesic. A semi-infinite geodesic starting from  $u \in \mathbb{Z}^2$  is said to have *direction*  $\theta$  if  $\lim_{n \rightarrow \infty} \frac{u_n}{|u_n|}$  exists and equals  $\theta$  (we will always identify an element on the unit circle with its corresponding angle) and is denoted by  $\Gamma_u^\theta$ . It is a fact that for a fixed direction  $\theta$ , almost surely starting from any point  $u \in \mathbb{Z}^2$  there exists a unique semi-infinite geodesic at the direction  $\theta$  [6, 12]. Now we will explain the kind of traffic problems we

are interested in. Let  $\epsilon > 0$  be fixed and imagine cars starting from each vertex of  $\mathbb{Z}^2$  and picking up uniform direction from the interval  $(\epsilon, \pi/2 - \epsilon)$  independent of each other, also independent of the vertex weights. The cars then travel via the a.s. unique semi-infinite geodesic in the chosen direction from the starting point. Notice that we simply assume that cars travel infinitely far. While in the naïve Poisson model they travelled much further than the radius  $r$  of the neighborhood we considered, infinite driving distances would have made that model obsolete. As we shall see, this is not the case for our LPP traffic model.

See Figure 2 for a simulation of some cars in the model. The main phenomenon we see is the empty regions towards the middle of the picture: most points around there will not see any cars passing from the distance to the bottom of the picture. When cars are started from all of  $\mathbb{Z}^2$ , that translates into a limited close neighbourhood from where cars can arrive to a point say, the origin. All other cars from further away will have coalesced into busy roads somewhere else that avoid passing through the origin.

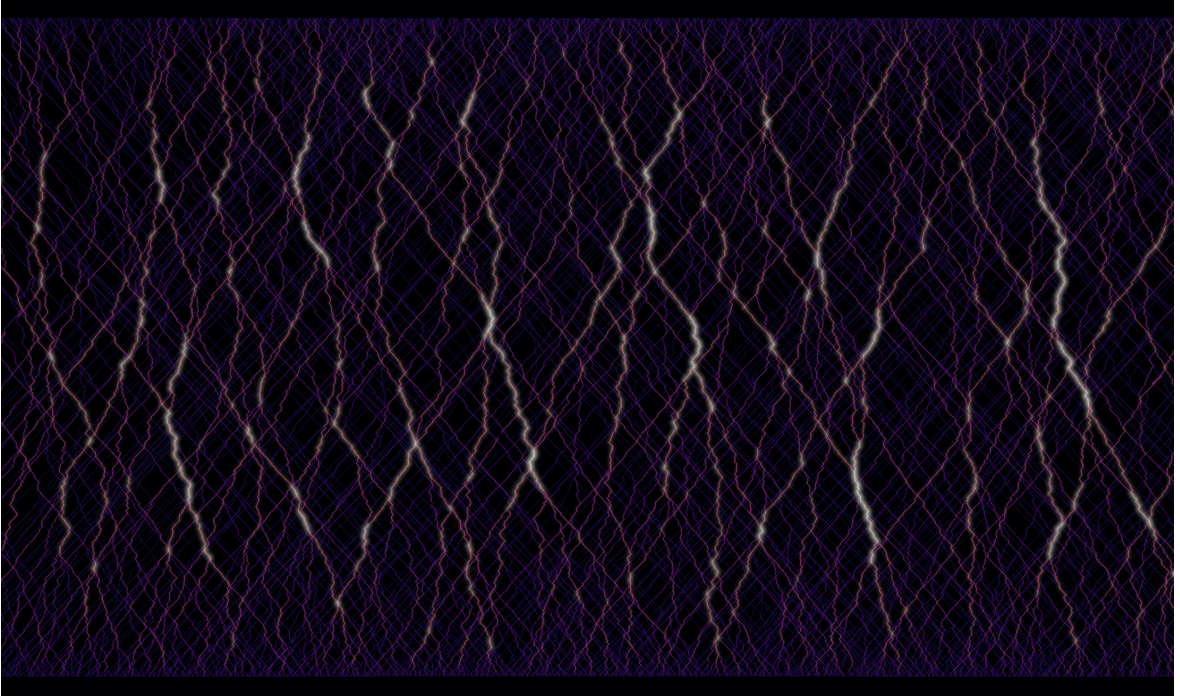


Figure 2: Simulation of geodesics in random directions, the picture is rotated left by  $45^\circ$ . Each point on the bottom line picks another one randomly on the top within the admissible angle range  $0^\circ - 90^\circ$  and sends a geodesic there (i.e., this simulation only launches cars from one line, as opposed to our model where this happens from each point of  $\mathbb{Z}^2$ ). Many geodesics coalesce for most of their travel, forming the heavy traffic paths which are marked in lighter colors. These leave gaps towards the middle of the picture where no geodesics pass. The dark mesh of low-traffic geodesics close to  $0^\circ$  and  $90^\circ$  demonstrate the need of separation from the edge of the angle domain: geodesics with such angles have little room to wiggle, hence do not coalesce much.

Of the several questions that can be asked, we will address the tail behaviour of the following quantities:

- How many cars pass through the origin (Theorems 2.1, 2.4)?
- What is the farthest distance a car arrives from to the origin (Theorem 2.5)?
- How far do we have to walk to find a busy road that carries certain volume of traffic (Theorems 2.6, 2.7)?

It is this last question that seemed tractable from the observations point of view. In Section 6 we compare our result to UK road layouts and traffic statistics, and we discuss the limitations of this experiment.

## 2.1 Further notations and results

Let  $\epsilon > 0$  be fixed; from now on all the constants we will obtain from the theorems will depend only on this  $\epsilon$ . From each vertex  $v \in \mathbb{Z}^2$  an angle is chosen according to i.i.d. uniform  $(\epsilon, \frac{\pi}{2} - \epsilon)$  distribution and we consider the a.s. unique semi-infinite geodesic in the chosen direction. We will denote the angle chosen by  $v$  by  $\theta_v$  and the semi-infinite geodesic in that direction by  $\Gamma_v^{\theta_v}$ . For  $r \in \mathbb{Z}$  let  $\mathbf{r}$  denote the vertex  $(r, r) \in \mathbb{Z}^2$ . For convenience sometimes we will work with the rotated axes  $x + y = 0$  and  $x - y = 0$ . They will be called space axis and time axis respectively. For a vertex  $v \in \mathbb{Z}^2$ ,  $\phi(v)$  will denote the time coordinate of  $v$  and  $\psi(v)$  will denote the space coordinate of  $v$ . Precisely, for  $u, v \in \mathbb{Z}^2$ , we have

$$\phi(u, v) = u + v, \quad \psi(u, v) = u - v.$$

For  $u, v \in \mathbb{Z}^2$ ,  $\Gamma_{u,v}$  denotes the geodesic between  $u$  and  $v$ . For a deterministic direction  $\alpha$ ,  $\Gamma_v^\alpha$  denotes the a.s. semi-infinite geodesic starting from  $v$  in the direction  $\alpha$ . The line  $x + y = T$  will be denoted by  $\mathcal{L}_T$ . For a geodesic  $\Gamma$ ,  $\Gamma(T)$  will denote the random intersection point of  $\Gamma$  and  $\mathcal{L}_T$ .

Let  $N$  denote the number of cars passing through origin. More precisely we define,

$$N := |\{u \in \mathbb{Z}^2 : \mathbf{0} \in \Gamma_u^{\theta_u}\}|.$$

We have the following theorem.

**Theorem 2.1.**  *$N$  is finite almost surely.*

*Proof.* This will follow from Theorem 2.4 which we will prove in Section 4. □

But  $N$  has infinite expectation. To show this we first need the following definition. For  $n \in \mathbb{N}$ , define  $N_n$  to be the number of cars starting from the line  $\mathcal{L}_{-n}$  and going through origin. We have the following theorem.

**Theorem 2.2.**  $\mathbb{E}(N_n) = 1$ .

*Proof.* For  $x \in \mathcal{L}_{-n}$ , let  $\mathcal{G}_x$  denote the event that the car starting from  $x$  goes through origin. Then

$$N_n = \sum_{x \in \mathcal{L}_{-n}} \mathbb{1}_{\mathcal{G}_x}.$$

So,

$$\mathbb{E}(N_n) = \sum_{x \in \mathcal{L}_{-n}} \mathbb{P}(\mathcal{G}_x).$$

For  $x \in \mathcal{L}_{-n}$ , we define  $\mathcal{G}_{-x}$  to be the event that the car starting from  $\mathbf{0}$  goes through  $-x$ . We have

$$\mathbb{P}(\mathcal{G}_x) = \mathbb{P}(\mathcal{G}_{-x}).$$

So,

$$\mathbb{E}(N_n) = \sum_{x \in \mathcal{L}_{-n}} \mathbb{E}(\mathbb{1}_{\mathcal{G}_{-x}}) = 1.$$

□

The following corollary is not surprising then, and is also in line with the divergence of the naïve Poisson model.

**Corollary 2.3.**  $\mathbb{E}(N) = \infty$ .

*Proof.* This is clear using Theorem 2.2. Indeed we have,

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} \mathbb{E}(N_n) = \infty.$$

□

We now dive a bit deeper into the scaling properties of the model.

**Theorem 2.4.** *There exist constants  $C_1, c_1 > 0$  (depending on  $\epsilon$ ) and  $n$  sufficiently large such that*

$$\frac{c_1}{n^{1/4}} \leq \mathbb{P}(N \geq n) \leq \frac{C_1 \log n}{n^{1/4}}.$$

To get estimates for the furthest distance a car can come from we define the following random variable.

$$D := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } \mathbf{0} \in \Gamma_u^{\theta_u}\}.$$

**Theorem 2.5.** *There exists  $C, c > 0$  (depending on  $\epsilon$ ) such that the following holds*

$$cn^{-1/3} \leq \mathbb{P}(D \geq n) \leq Cn^{-1/3}.$$

Finally, to investigate the distance to the nearest busy road, we define

$$T_n := \min\{|\psi(v)| : v \in \mathcal{L}_0, N_v \geq n^{4/3}\}.$$

We have the following theorem which is direct consequence of Theorem 2.4.

**Theorem 2.6.** *For  $\delta > 0$  sufficiently small (depending on  $\epsilon$ ) we have there exists constant  $C > 0$  (depending on  $\epsilon$ ) and for sufficiently large  $n$  (depending on  $\epsilon$ )*

$$\mathbb{P}(T_n \geq \frac{\delta n^{1/3}}{\log n}) \geq 1 - C\delta.$$

*Proof.* The proof follows from a union bound. let  $V$  denote the line segments with  $\lceil \frac{\delta n^{1/3}}{\log n} \rceil$  vertices and with midpoint  $\mathbf{0}$  on the line  $\mathcal{L}_0$ . Then

$$\{T_n \leq \frac{\delta n^{1/3}}{\log n}\} \subset \bigcup_{v \in V} \{N_v \geq n^{4/3}\}.$$

Hence, by Theorem 2.4 upper bound we have for some constant  $C > 0$  (depending on  $\epsilon$ )

$$\mathbb{P}(T_n \leq \frac{\delta n^{1/3}}{\log n}) \leq C\delta.$$

Choosing  $C\delta < 1$  we get Theorem 2.6. □

The above theorem says that, with high probability, to see a vertex having  $n^{4/3}$  many cars through it, we need to go at least sufficiently small multiple of  $\frac{n^{1/3}}{\log n}$  distance from origin. The following theorem shows that, with non-vanishing probability, there is at least one vertex within  $n^{1/3}$  distance around the origin that has  $n^{4/3}$  many cars through it. Precisely, for a fixed constant  $\ell_0$  we make a slight modification to the definition of  $T_n$  and by abuse of notation we will still call the new variable  $T_n$ .

$$T_n := \min\{|\psi(v)| : v \in \mathcal{L}_0, N_v \geq \frac{n^{4/3}}{\ell_0}\}.$$

We then have the following theorem.

**Theorem 2.7.** *There exists  $c > 0, \ell_0$  (depending on  $\epsilon$ ) such that for sufficiently large  $n$  (depending on  $\epsilon$ ) we have*

$$\mathbb{P}(T_n \leq n^{1/3}) \geq c.$$

The rest of the paper is devoted to proving these theorems and to compare the latter two to actual road traffic statistics.

### 3 Bounds for the furthest distance a car can come from

#### 3.1 Proof of Theorem 2.5 upper bound

First we divide the interval  $(\epsilon, \frac{\pi}{2} - \epsilon)$  into disjoint sub-intervals. Precisely, for  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  and for a uniformly bounded sequence  $\{c_n\}$  we partition the interval  $(\epsilon, \frac{\pi}{2} - \epsilon)$  into intervals  $A_i$  each of equal length  $\frac{c_n(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$ . For  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  we define the following random variables.

$$D_i := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } \theta_u \in A_i \text{ and } \mathbf{0} \in \Gamma_u^{\theta_u}\}.$$

Clearly, we have

$$\{D \geq n\} \subset \bigcup_{i=1}^{\lfloor n^{1/3} \rfloor} \{D_i \geq n\}.$$

So, we have

$$\mathbb{P}(D \geq n) \leq \sum_{i=1}^{\lfloor n^{1/3} \rfloor} \mathbb{P}(D_i \geq n). \tag{3.1}$$

We fix  $1 \leq i \leq \lfloor n^{1/3} \rfloor$ . We want to find an upper bound for  $\mathbb{P}(D_i \geq n)$ . We will apply an averaging argument. Let  $V$  denote the line segment with  $\lceil n^{2/3} \rceil$  vertices with  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$ . For  $v \in V$  we define the following random variables.

$$D_i^v := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } \theta_u \in A_i \text{ and } v \in \Gamma_u^{\theta_u}\}.$$

Clearly, for all  $v \in V$  we have

$$\mathbb{P}(D_i \geq n) = \mathbb{P}(D_i^v \geq n).$$

Hence,

$$\lceil n^{2/3} \rceil \mathbb{P}(D_i \geq n) = \sum_{v \in V} \mathbb{P}(D_i^v \geq n) = \mathbb{E}(\widehat{D}_i), \quad (3.2)$$

where  $\widehat{D}_i$  is defined as follows.

$$\widehat{D}_i := \sum_{v \in V} \mathbb{1}_{\{D_i^v \geq n\}}.$$

We will show that for sufficiently large  $\ell$ ,  $\mathbb{P}(\widehat{D}_i \geq \ell)$  has a stretched exponential in  $\ell$  upper bound. We prove it precisely now. Let  $\alpha_i, \beta_i, \theta_i$  denote the left end point, right end point and midpoint of  $A_i$  respectively and  $n'_i \in \mathcal{L}_{-n}$  be the intersection point of  $\mathcal{L}_{-n}$  and the line  $y = (\tan \theta_i)x$ . Consider the line segment  $\tilde{V}$  with  $\lfloor \ell^{1/96} n^{2/3} \rfloor$  vertices on  $\mathcal{L}_{-n}$  with midpoint  $n'_i$ . Let  $v_1, v_2$  be the end points of  $\tilde{V}$  with  $\psi(v_1) \leq \psi(v_2)$ . Consider  $\Gamma_{v_1}^{\alpha_i}$  and  $\Gamma_{v_2}^{\beta_i}$  (see Figure 3). We define the following event.

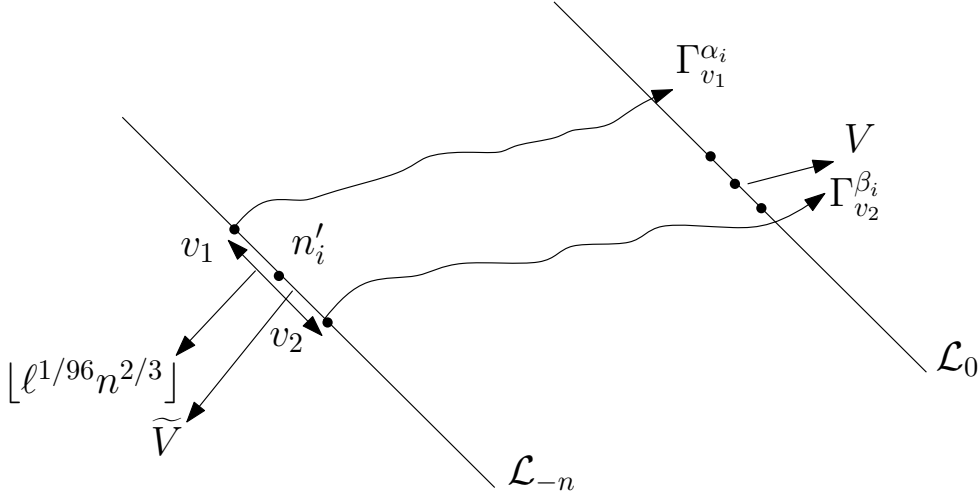


Figure 3: To prove the upper bound in Theorem 2.5 we fix  $1 \leq i \leq \lfloor n^{1/3} \rfloor$ . The event that there are more than  $\ell$  many  $u \in \mathcal{L}_{-n}$  with  $\theta_u \in A_i$  and  $\Gamma_u^{\theta_u}$  intersecting  $V$  can happen in two ways. Either the above  $u$  lies in  $\mathcal{L}_{-n} \setminus \tilde{V}$ . For sufficiently large  $\ell$  this will imply large transversal fluctuation for either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$ . Proposition 3.3 gives a stretched exponential upper bound for this event. The other possibility is there are more than  $\ell$  distinct vertices inside on  $V$  that carries a geodesic from  $\tilde{V}$ . Proposition 3.4 gives a stretched exponential upper bound for this event.

- $\mathcal{A}_i := \{\exists u \in \mathcal{L}_{-n} \setminus \tilde{V} \text{ such that } \theta_u \in A_i \text{ and } v \in \Gamma_u^{\theta_u} \text{ for some } v \in V\}.$

We have

$$\mathbb{P}(\widehat{D}_i \geq \ell) \leq \mathbb{P}(\mathcal{A}_i) + \mathbb{P}(\{\widehat{D}_i \geq \ell\} \cap \mathcal{A}_i^c).$$

Now, by planarity the event  $\mathcal{A}_i$  implies either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$  will intersect  $V$  (see Figure 3). Since,  $(\beta_i - \alpha_i)$  is of order  $\frac{1}{n^{1/3}}$ , for sufficiently large  $\ell$  (depending on  $\epsilon$ ), the event  $\mathcal{A}_i$  will imply that either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$  will have transversal fluctuation larger than  $\frac{\ell^{1/96} n^{2/3}}{4}$ . We have the following proposition. Let  $\Gamma^\alpha$  ( resp.  $\mathcal{J}^\alpha$ ) denote the semi-infinite geodesic (resp. the straight line) in the direction  $\alpha$  starting from  $\mathbf{0}$  and  $\Gamma^\alpha(T)$  (resp.  $\mathcal{J}^\alpha(T)$ ) denote the intersection points of  $\Gamma^\alpha$  (resp.  $\mathcal{J}^\alpha$ ) with  $\mathcal{L}_T$ . We have the following.

**Proposition 3.3.** [2, Proposition 1.5] *For  $\epsilon > 0$  there exist  $C_1, c_1 > 0$  such that for all  $T > 0, \ell > 0, n \geq 1$  and for all  $\alpha \in (\epsilon, \frac{\pi}{2} - \epsilon)$  we have*

- (i)  $\mathbb{P}(|\psi(\Gamma^\alpha(T)) - \psi(\mathcal{J}^\alpha(T))| \geq \ell T^{2/3}) \leq C_1 e^{-c_1 \ell^3},$
- (ii)  $\mathbb{P}(\sup\{|\psi(\Gamma^\alpha(t)) - \psi(\mathcal{J}^\alpha(t))| : 0 \leq t \leq T\} \geq \ell T^{2/3}) \leq C_1 e^{-c_1 \ell^3}.$

Hence, for sufficiently large  $\ell$  we have

$$\mathbb{P}(\mathcal{A}_i) \leq C e^{-c \ell^{3/96}}.$$

Now, we consider the event  $\{\widehat{D}_i \geq \ell\} \cap \mathcal{A}_i^c$ . On this event there are at least  $\ell$  distinct vertices on  $V$  that has a geodesic coming from  $\tilde{V}$ . We have the following proposition.

For  $k \in \mathbb{Z}$ , let  $L_n$  (resp.  $L_n^*$ ) denote the line segment on  $\mathcal{L}_0$  (resp.  $\mathcal{L}_{2n}$ ) of length  $2\ell^{1/32} n^{2/3}$  with midpoint  $\mathbf{0}$  (resp.  $\mathbf{n}_k = (n - kn^{2/3}, n + kn^{2/3})$ ). For  $u, u' \in L_n$  and  $v, v' \in L_n^*$  we say that  $(u, v) \sim (u', v')$  if the geodesics  $\Gamma_{u,v}$  and  $\Gamma_{u',v'}$  coincide between the lines  $\mathcal{L}_{n/3}$  and  $\mathcal{L}_{2n/3}$ . It is easy to see that  $\sim$  is an equivalence relation. Let  $M_n^k$  denote the number of equivalence classes.

**Proposition 3.4.** [2, Proposition 1.6] *For  $\psi < 1$  there exist  $C, c > 0$  such that for all  $k$  with  $|k| + \ell^{1/32} < \psi n^{1/3}$ , all  $\ell < n^{0.01}$  sufficiently large and all  $n \in \mathbb{N}$  sufficiently large we have*

$$\mathbb{P}(M_n^k \geq \ell) \leq C e^{-c \ell^{1/128}}. \quad (3.5)$$

Hence, using Proposition 3.4 we have for sufficiently large  $\ell$  and  $n$

$$\mathbb{P}(\{\widehat{D}_i \geq \ell\} \cap \mathcal{A}_i^c) \leq C e^{c \ell^{1/384}}.$$

Combining all the above we have for  $\ell < n^{0.01}$ ,  $n$  sufficiently large we have

$$\mathbb{P}(\widehat{D}_i \geq \ell) \leq C e^{-c \ell^{1/384}}.$$

So, finally we have there exists  $\ell_0$  (depending only on  $\epsilon$ ) and  $\tilde{C}$  (depending only on  $\epsilon$ ) such that for all large  $n$

$$\mathbb{E}(\widehat{D}_i) \leq \sum_{\ell \geq 1} \mathbb{P}(\widehat{D}_i \geq \ell) \leq \ell_0 + \sum_{\ell_0 \leq \ell < \frac{n^{0.01}}{2}} C e^{-c \ell^{1/384}} + \sum_{\frac{n^{0.01}}{2} \leq \ell < \lceil n^{2/3} \rceil} C e^{-c' n^{1/38400}} < \tilde{C}. \quad (3.6)$$

So, from (3.2) we have there exists  $C > 0$  (depending only on  $\epsilon$ ) such that for all  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  we have

$$\mathbb{P}(D_i \geq n) \leq \frac{C}{n^{2/3}}.$$

So, from (3.1) there exists  $C > 0$  such that for large  $n$

$$\mathbb{P}(D \geq n) \leq \frac{C}{n^{1/3}}.$$

This proves the upper bound in Theorem 2.5.  $\square$

### 3.2 Proof of Theorem 2.5 lower bound

We again consider disjoint sub-intervals of  $(\epsilon, \frac{\pi}{2} - \epsilon)$ . First we fix large constant  $M$  which will be chosen later. For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , we consider the intervals  $A_i \subset (\epsilon, \frac{\pi}{2} - \epsilon)$ , each of length  $\frac{(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$  and midpoint  $\frac{Mi(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$ . Note that, unlike the previous case we are not partitioning the interval  $(\epsilon, \frac{\pi}{2} - \epsilon)$  this time. Instead we are taking points which are  $\frac{M(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$  distance apart from each other and we are taking intervals of length  $\frac{(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$  around them. We do this so that the constants that we get at the end depend only on  $\epsilon$  but not on  $M$ . This will be elaborated later. Corresponding to each  $A_i$  we define the following random variables.

$$D_i := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } \theta_u \in A_i \text{ and } \mathbf{0} \in \Gamma_u^{\theta_u}\}.$$

Clearly,

$$\bigcup_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \{D_i \geq n\} \subset \{D \geq n\}.$$

For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$  let

$$\mathcal{B}_i := \{D_i \geq n\}.$$

So, applying inclusion-exclusion principle and considering the first two terms we have

$$\mathbb{P}(D \geq n) \geq \sum_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{B}_i) - \sum_{1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{B}_i \cap \mathcal{B}_j). \quad (3.7)$$

For the first sum we do the following. We fix  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ . We consider the line segment  $V_M$  with  $\lceil 4Mn^{2/3} \rceil$  many vertices on  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$  (see Figure 4). For  $v \in V_M$ , we define

$$D_i^v := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } \theta_u \in A_i \text{ and } v \in \Gamma_u^{\theta_u}\}.$$

So, same as before we have

$$\lceil 4Mn^{2/3} \rceil \mathbb{P}(D_i \geq n) = \sum_{v \in V_M} \mathbb{P}(D_i^v \geq n) = \mathbb{E}(\widehat{D}_i), \quad (3.8)$$

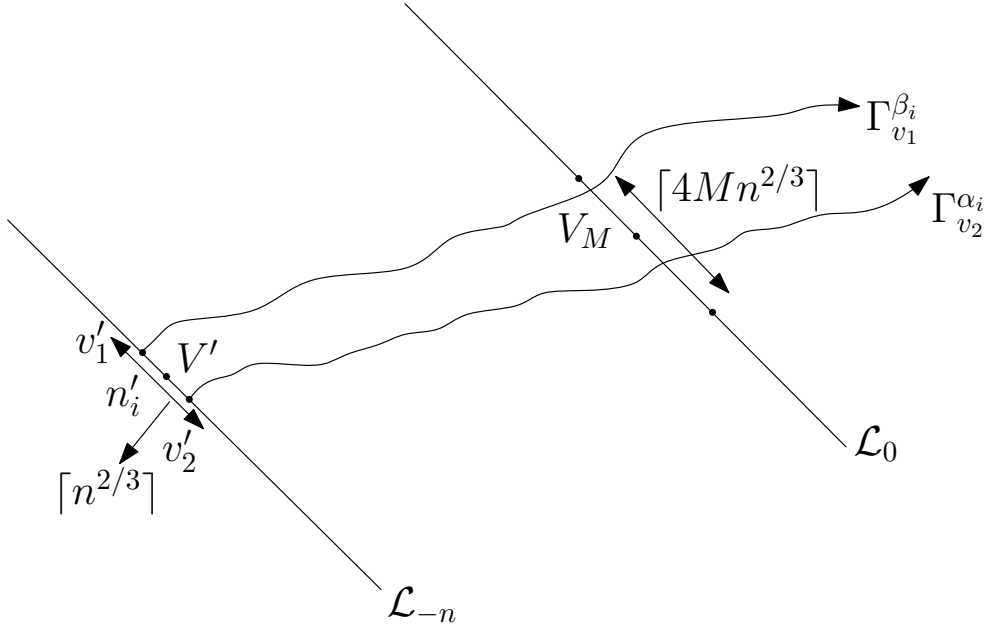


Figure 4: To prove the lower bound for the first sum in (3.7) we fix an  $i$ . Using Proposition 3.3 we can choose  $M$  large enough so that on a positive probability event  $\Gamma_{v'_1}^{\beta_i}$  and  $\Gamma_{v'_2}^{\alpha_i}$  intersect  $\mathcal{L}_0$  on  $V_M$ . Further, we consider an independent event on which there is at least one vertex on  $V'$  that chooses an angle in  $A_i$ . Due to central limit theorem the later event is a positive probability event. Hence, the intersection of these two events is a positive probability event and on this intersection  $\widehat{D}_i \geq 1$ .

where  $\widehat{D}_i$  is defined as

$$\widehat{D}_i := \sum_{v \in V_M} \mathbb{1}_{\{D_i^v \geq n\}}.$$

We will show on a positive probability event  $\widehat{D}_i \geq 1$ . We construct this event now. Consider  $V'$ , a line segment with  $\lceil n^{2/3} \rceil$  many vertices with midpoint  $n'_i$  on  $\mathcal{L}_{-n}$  and let  $v'_1, v'_2$  denote the end points of  $V'$  with  $\psi(v'_1) \leq \psi(v'_2)$ . Consider the geodesics  $\Gamma_{v'_1}^{\beta_i}$  and  $\Gamma_{v'_2}^{\alpha_i}$ . We define the following event.

$$\mathcal{C}_i := \{\Gamma_{v'_1}^{\beta_i}(0) \in V_M \text{ and } \Gamma_{v'_2}^{\alpha_i}(0) \in V_M\}.$$

Again as  $(\beta_i - \alpha_i)$  is of order  $\frac{1}{n^{1/3}}$ , the complement of the event  $\mathcal{C}_i$  will imply that either  $\Gamma_{v'_1}^{\beta_i}$  or  $\Gamma_{v'_2}^{\alpha_i}$  will have transversal fluctuation more than  $\frac{M}{2}n^{2/3}$ . So, using Proposition 3.3 we can choose  $M$  large enough so that

$$\mathbb{P}(\mathcal{C}_i) \geq 0.99.$$

Further, we consider the following event.

$$\mathcal{D}_i := \{\exists u \in V' \text{ such that } \theta_u \in A_i\} = \left\{ \sum_{u \in V'} \mathbb{1}_{\{\theta_u \in A_i\}} \geq 1 \right\}.$$

So, we consider the random variable  $X_i := \sum_{u \in V'} \mathbb{1}_{\{\theta_u \in A_i\}}$ . We have  $\mathbb{E}(X_n) = \frac{c_n \lceil n^{2/3} \rceil}{n^{1/3}}$  and  $\mathbb{V}\text{ar}(X_n) = (\frac{c_n}{n^{1/3}} - \frac{c_n^2}{n^{2/3}}) \lceil n^{2/3} \rceil$ . By the Berry-Esseen inequality we have for all  $n$  there exists a constant  $c > 0$  such that

$$|\mathbb{P}(X_n - \mathbb{E}(X_n) \geq 0) - \frac{1}{2}| \leq \frac{c}{(c_n n^{1/3} - c_n^2) n^{1/3}}.$$

Hence, for sufficiently large  $n$  we have

$$\mathbb{P}(X_n - \mathbb{E}(X_n) \geq 0) \geq \frac{1}{4}.$$

So, for sufficiently large  $n$  we have

$$\mathbb{P}(\mathcal{D}_i) \geq \frac{1}{4}.$$

Finally, note that the events  $\mathcal{C}_i$  and  $\mathcal{D}_i$  are independent. So,

$$\mathbb{P}(\mathcal{C}_i \cap \mathcal{D}_i) \geq \frac{0.99}{4}.$$

Clearly, by planarity of geodesics (see Section 2 for definition), on the event  $\mathcal{C}_i \cap \mathcal{D}_i, \widehat{D}_i \geq 1$ . So, from (3.8) we have there exists  $c_1 > 0$  such that

$$\mathbb{P}(D_i \geq n) \geq \frac{c_1}{M n^{2/3}}. \quad (3.9)$$

Now we consider the second sum in (3.7). We fix  $1 \leq i < j \leq \lceil \frac{n^{1/3}}{2M} \rceil$ , both large enough. For  $1 \leq i \leq \lceil \frac{n^{1/3}}{2M} \rceil$  we define the following events.

$$\mathcal{E}_i := \{\exists u \in \mathbb{Z}^2, \gamma \in A_i \text{ such that } \phi(u) = -n, \text{ and } \mathbf{0} \in \Gamma_u^\gamma\}.$$

Observe that for all  $1 \leq i \leq \lceil \frac{n^{1/3}}{2M} \rceil$

$$\{D_i \geq n\} \subset \mathcal{E}_i.$$

From now on we will work with  $\mathcal{E}_i$ 's. Consider the points  $n'_i, n'_j \in \mathcal{L}_{-n}$  defined as before. We have  $|n'_i - n'_j|$  is of order  $M|i - j|n^{2/3}$ . For simplicity let  $k_{i,j} = M|i - j|$ . We apply again an averaging argument. Consider the following parallelogram (see Figure 5).

$$W := \{w \in \mathbb{Z}^2 : |\phi(w)| \leq \lceil \frac{n}{100k_{i,j}} \rceil, |\psi(w)| < \lceil \frac{n^{2/3}}{2} \rceil\}.$$

Then for all  $w \in W$  define the following events.

$$\mathcal{E}_i^w := \{\exists u \in \mathbb{Z}^2, \gamma \in A_i \text{ such that } \phi(u) = -n + \phi(w), \text{ and } w \in \Gamma_u^\gamma\}.$$

We have for all  $w \in W$

$$\mathbb{P}(\mathcal{E}_i \cap \mathcal{E}_j) = \mathbb{P}(\mathcal{E}_i^w \cap \mathcal{E}_j^w).$$

Hence,

$$\frac{c_n n^{5/3}}{100k_{i,j}} \mathbb{P}(\mathcal{E}_i \cap \mathcal{E}_j) = \sum_{w \in W} \mathbb{P}(\mathcal{E}_i^w \cap \mathcal{E}_j^w) = \mathbb{E} \left( \sum_{w \in W} \mathbb{1}_{\mathcal{E}_i^w \cap \mathcal{E}_j^w} \right), \quad (3.10)$$

where  $\{c_n\}$  is a uniformly bounded sequence such that for each  $n$ , the number of vertices in  $W$  is  $\frac{c_n n^{5/3}}{100k_{i,j}}$ . Let us define the following random variable.

- $N_{i,j}$  is number of  $w \in W$  such that there exist  $u_i, u_j \in \mathbb{Z}^2$  with  $|\phi(u_i) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $|\phi(u_j) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $\theta_{u_i}$  (resp.  $\theta_{u_j}$ ) lies in  $A_i$  (resp.  $A_j$ ) and  $w \in \Gamma_{u_i}^{\theta_{u_i}} \cap \Gamma_{u_j}^{\theta_{u_j}}$ .

We have

$$\sum_{w \in W} \mathbb{1}_{\mathcal{E}_i^w \cap \mathcal{E}_j^w} \leq N_{i,j}.$$

We will show that there exists a constant  $C > 0$  (depending only on  $\epsilon$ ) such that for all  $n$  sufficiently large

$$\mathbb{E}(N_{i,j}) \leq \frac{Cn \log |k_{i,j}|}{k_{i,j}^3}. \quad (3.11)$$

We define two equivalence classes. For  $u, v \in \mathbb{Z}^2$  with  $|\phi(u) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $|\phi(v) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $\theta_u, \theta_v \in A_i$  we say  $u \sim v$  if  $\Gamma_u^{\theta_u}$  and  $\Gamma_v^{\theta_v}$  coincide inside  $W$ . Let  $\widetilde{E}_i$  denote the number of equivalence classes. Similarly, we define the equivalence classes corresponding to  $A_j$  and denote it by  $\widetilde{E}_j$ . Further for  $u, v \in \mathbb{Z}^2$  with  $|\phi(u) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil, \theta_u$  (resp.  $\theta_v$ ) contained in  $A_i$  (resp.  $A_j$ ) let  $I_{u,v}$  denote the intersection size of  $\Gamma_u^{\theta_u}$  and  $\Gamma_v^{\theta_v}$  inside  $W$ . Clearly,

$$N_{i,j} \leq \widetilde{E}_i \widetilde{E}_j \max I_{u,v},$$

where the maximum is taken over all  $u, v \in \mathbb{Z}^2$  with  $|\phi(u) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $|\phi(v) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $\theta_u$  (resp.  $\theta_v$ ) lies in  $A_i$  (resp.  $A_j$ ). So, for  $\ell \geq 1$

$$\mathbb{P} \left( N_{i,j} \geq \frac{\ell n \log |k_{i,j}|}{k_{i,j}^3} \right) \leq \mathbb{P}(\widetilde{E}_i \geq \ell^{1/3}) + \mathbb{P}(\widetilde{E}_j \geq \ell^{1/3}) + \mathbb{P} \left( \max I_{u,v} \geq \frac{\ell^{1/3} n \log |k_{i,j}|}{k_{i,j}^3} \right). \quad (3.12)$$

We start with an upper bound for the first term. Let us consider the following parallelogram (see Figure 5).

- $R_i^\ell$  is the parallelogram whose opposite sides lie on  $\mathcal{L}_{-n-\lceil \frac{n}{100k_{i,j}} \rceil}$  (resp.  $\mathcal{L}_{-n+\lceil \frac{n}{100k_{i,j}} \rceil}$ ), each with  $\lceil \ell^{1/96} n^{2/3} \rceil$  many vertices with midpoints are the intersection points of  $y = \tan(\theta_i)x$  with  $\mathcal{L}_{-n-\lceil \frac{n}{100k_{i,j}} \rceil}$  (resp.  $\mathcal{L}_{-n+\lceil \frac{n}{100k_{i,j}} \rceil}$ ).

Let us consider the following two line segments  $\mathcal{J}_i'$  (resp.  $\mathcal{J}_i''$ ) with  $\lceil 4\ell^{1/96} n^{2/3} \rceil$  many vertices on  $\mathcal{L}_{-n+\lceil \frac{n}{100k_{i,j}} \rceil}$  (resp.  $\mathcal{L}_n$ ) with midpoints the intersection point of  $y = \tan(\theta_i)x$  and  $\mathcal{L}_{-n+\lceil \frac{n}{100k_{i,j}} \rceil}$  (resp.  $\mathcal{L}_n$ ). Let us define following two events (see Figure 5).

- $\mathcal{F}_i := \{\exists u \notin R_i^\ell : |\phi(u) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil, \theta_u \in A_i, \Gamma_u^{\theta_u} \text{ intersects } W\}$ .
- $\mathcal{G}_i := \{\exists u \in R_i^\ell : \theta_u \in A_i, \Gamma_u^{\theta_u} \text{ does not intersect either } \mathcal{J}_i' \text{ or } \mathcal{J}_i''\}$ .

Similarly, we define  $\mathcal{F}_j, \mathcal{G}_j$ .

Using a planarity argument and due to large transversal fluctuation (see also the argument we used in the upper bound to bound the probability of the event  $\mathcal{A}_i$ ) we have that there exist  $C, c > 0$  (depending only on  $\epsilon$ ) such that for sufficiently large  $\ell$  and  $n$

$$\mathbb{P}(\mathcal{F}_i \cup \mathcal{F}_j) \leq Ce^{-c\ell^{3/96}}.$$

Further, we have the following lemma.

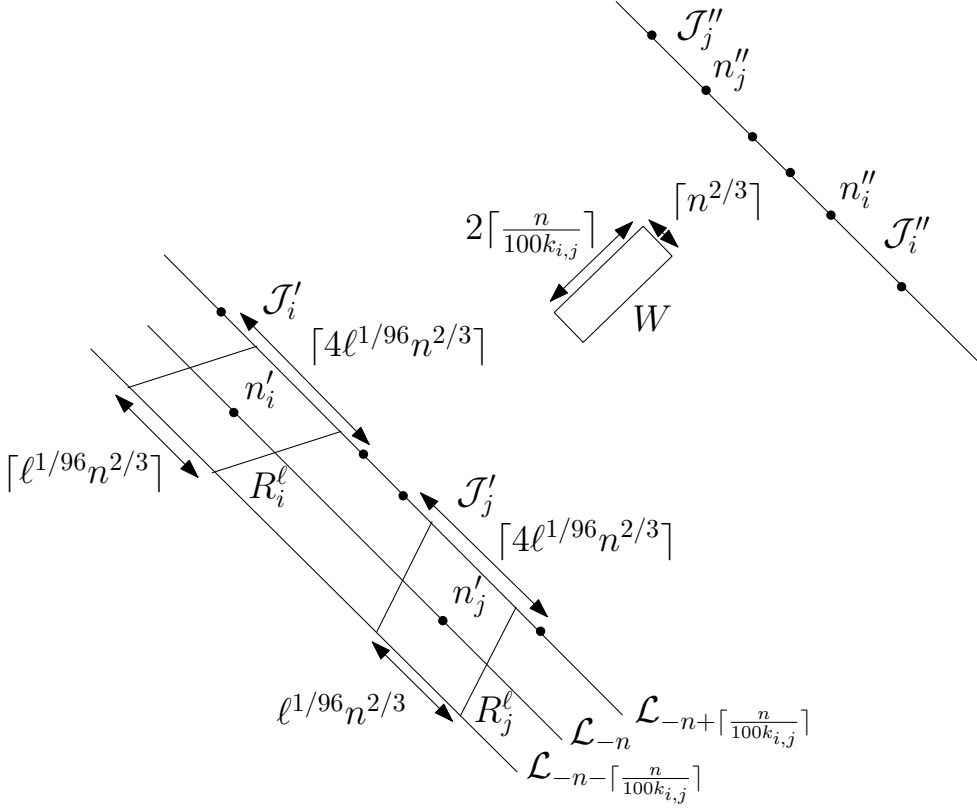


Figure 5: To find an upper bound for the second sum in (4.3) we fix  $1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ . Due to planarity and large transversal fluctuation there is a very small probability that there will be vertices  $u$  outside  $R_i^\ell$  (resp.  $R_j^\ell$ ) with  $|\phi(v) + n| \leq \lceil \frac{n}{100k_{i,j}} \rceil$  choosing angle in  $A_i$  (resp.  $A_j$ ) such that  $\Gamma_u^{\theta_u}$  will intersect  $W$ . These are the events  $\mathcal{F}_i$  and  $\mathcal{F}_j$ . Further, the same argument as in [2, Lemma 3.1] shows that, there is a very small probability that for any geodesic starting from  $R_i^\ell$  (resp.  $R_j^\ell$ ) and taking angle in  $A_i$  (resp.  $A_j$ ) will not intersect either  $\mathcal{J}_i'$  or  $\mathcal{J}_i''$  (resp. either  $\mathcal{J}_j'$  or  $\mathcal{J}_j''$ ). These are the events  $\mathcal{G}_i$  and  $\mathcal{G}_j$ . Further, due to Proposition 3.4, we have with very small probability that there are more than  $\ell^{1/3}$  distinct vertices in  $W$  that has a geodesic starting from  $\mathcal{J}_i'$  (resp.  $\mathcal{J}_j'$ ) and ending at  $\mathcal{J}_i''$  (resp.  $\mathcal{J}_j''$ ). Finally, we invoke [2, Lemma 3.2, Lemma 3.3] to conclude that with very small probability any geodesic starting from  $\mathcal{J}_i'$  (resp.  $\mathcal{J}_j'$ ) and ending at  $\mathcal{J}_i''$  (resp.  $\mathcal{J}_j''$ ) will have more than  $\frac{\ell n}{k_{i,j}^3}$  intersection size inside  $W$ . Combining all the above we get stretched exponential upper bound for each terms in (3.12).

**Lemma 3.13.** *For sufficiently large  $n, \ell$  there exist constants  $C, c > 0$  (depending only on  $\epsilon$ ) such that*

$$\mathbb{P}(\mathcal{G}_i \cup \mathcal{G}_j) \leq C e^{-c\ell^{3/96}}.$$

*Proof.* This lemma is already proved in [2, Lemma 3.1], the Reader can refer to this proof. We briefly describe the idea here. We consider two points  $a$  and  $b$  on  $\mathcal{L}_{-n-\lceil \frac{n}{100k_{i,j}} \rceil}$  which are  $\lceil 2\ell n^{1/96} n^{2/3} \rceil$

distance away (in the space direction) from the end points of  $R_i^\ell$ . Now, by transversal fluctuation estimate we can say that  $\Gamma_a^{\beta_i}$  and  $\Gamma_b^{\alpha_i}$  stays always  $\lceil \ell^{1/96} n^{2/3} \rceil$  distance away (in the space direction) from the rectangle  $R_i^\ell$  with probability at least  $1 - e^{-c\ell^{3/96}}$ . On this event, every geodesics starting in  $R_i^\ell$  will be sandwiched between  $\Gamma_a^{\beta_i}$  and  $\Gamma_b^{\alpha_i}$ . So, on this event by planarity argument the event  $\mathcal{G}_i$  will imply large transversal fluctuation of either  $\Gamma_a^{\beta_i}$  or  $\Gamma_b^{\alpha_i}$ . This gives us the desired upper bound.  $\square$

So, we have

$$\mathbb{P}(\widetilde{E}_i \geq \ell^{1/3}) \leq Ce^{-c\ell^{3/96}} + \mathbb{P}(\{\widetilde{E}_i \geq \ell^{1/3}\} \cap (\mathcal{F}_i \cup \mathcal{G}_i)^c).$$

Note that on the second event in the right hand side there are more than  $\ell^{1/3}$  distinct equivalence classes of geodesics starting from  $\mathcal{J}_i'$  and ending at  $\mathcal{J}_i''$ . Using Proposition 3.4, for  $\ell < n^{0.03}$  sufficiently large (depending on  $\epsilon$ ) we have

$$\mathbb{P}(\{\widetilde{E}_i \geq \ell^{1/3}\} \cap (\mathcal{F}_i \cup \mathcal{G}_i)^c) \leq Ce^{-c\ell^{1/384}}.$$

Same argument shows for  $\ell < n^{0.03}$  sufficiently large (depending on  $\epsilon$ ) we have

$$\mathbb{P}(\{\widetilde{E}_j \geq \ell^{1/3}\}) \leq Ce^{-c\ell^{1/384}}.$$

For the last term in (3.12) we consider the following.

$$\begin{aligned} & \mathbb{P}\left(\max I_{u,v} \geq \frac{\ell^{1/3} n \log |k_{i,j}|}{k_{i,j}^3}\right) \leq \mathbb{P}(\mathcal{F}_i \cup \mathcal{G}_i \cup \mathcal{F}_j \cup \mathcal{G}_j) + \\ & \mathbb{P}\left(\left\{\max I_{u,v} \geq \frac{\ell^{1/3} n \log |k_{i,j}|}{k_{i,j}^3}\right\} \cap (\mathcal{F}_i \cup \mathcal{G}_i \cup \mathcal{F}_j \cup \mathcal{G}_j)^c\right). \end{aligned}$$

Clearly, we have

$$\mathbb{P}(\mathcal{F}_i \cup \mathcal{G}_i \cup \mathcal{F}_j \cup \mathcal{G}_j) \leq 2Ce^{-c\ell^{3/96}}.$$

On the event  $\left\{\max I_{u,v} \geq \frac{\ell^{1/3} n \log |k_{i,j}|}{k_{i,j}^3}\right\} \cap (\mathcal{F}_i \cup \mathcal{G}_i \cup \mathcal{F}_j \cup \mathcal{G}_j)^c$ , we want to estimate the maximum size of the intersection of geodesics inside  $V$  that are starting from  $\mathcal{J}_i'$  (resp.  $\mathcal{J}_j'$ ) and ending at  $\mathcal{J}_i''$  (resp.  $\mathcal{J}_j''$ ). For this we have the following lemma.

**Lemma 3.14.** *[2, Lemma 3.2, Lemma 3.3] Consider any geodesic  $\Gamma_i$  (resp.  $\Gamma_j$ ) starting from  $\mathcal{J}_i'$  (resp.  $\mathcal{J}_j'$ ) and ending at  $\mathcal{J}_i''$  (resp.  $\mathcal{J}_j''$ ). If  $I_{\Gamma_i, \Gamma_j}$  denote the intersection size of  $\Gamma_i$  and  $\Gamma_j$  inside  $W$  then for  $\ell$  and  $n$  large enough there exist  $C, c > 0$  such that we have*

$$\mathbb{P}\left(\max I_{\Gamma_i, \Gamma_j} \geq \frac{\ell \log |k_{i,j}| n}{k_{i,j}^3}\right) \leq Ce^{-c\ell}.$$

Hence, we have that there exist constants  $C, c$  (depending on  $\epsilon$ ) such that for large enough  $\ell$  and  $n$

$$\mathbb{P}\left(\left\{\max I_{u,v} \geq \frac{\ell^{1/3} n \log |k_{i,j}|}{k_{i,j}^3}\right\} \cap (\mathcal{F}_i \cup \mathcal{G}_i \cup \mathcal{F}_j \cup \mathcal{G}_j)^c\right) \leq Ce^{-c\ell^{1/3}}.$$

Hence, combining the above we get that there exists a constant  $C > 0$  (depending on  $\epsilon$ ) such that for all large enough  $n$  we have

$$\mathbb{E}(N_{i,j}) \leq \frac{Cn \log |k_{i,j}|}{k_{i,j}^3}.$$

hence, from (3.10) we have there exists  $C_2 > 0$  such that for sufficiently large  $n$

$$\mathbb{P}(\mathcal{E}_i \cap \mathcal{E}_j) \leq \frac{C_2 \log |k_{i,j}|}{k_{i,j}^2 n^{2/3}}. \quad (3.15)$$

We now go back to (3.7). From (3.9) and (3.15) we have there exists constants  $c_1, C_2 > 0$ , large  $M$  such that for sufficiently large  $n$  (all depending only on  $\epsilon$ )

$$\mathbb{P}(D \geq n) \geq \frac{c_1}{M^2 n^{1/3}} - \sum_{1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor} \frac{C_2 \log(M|i-j|)}{M^2 |i-j|^2 n^{2/3}}.$$

Hence, we have for some  $c > 0, C > 0$

$$\mathbb{P}(D \geq n) \geq \frac{c}{M^2 n^{1/3}} - \frac{C}{M^3 n^{1/3}}.$$

As  $C, c$  depend only on  $\epsilon$  (as we have chosen all the interval length to be  $\frac{1}{n^{1/3}}$ ) we can choose  $M$  large enough so that there exists  $c > 0$  such that for sufficiently large  $n$

$$\mathbb{P}(D \geq n) \geq \frac{c}{n^{1/3}}.$$

This completes the proof of lower bound in Theorem 2.5.  $\square$

## 4 Bounds for number of cars

In this section we prove the bounds for the tail distribution of  $N$ . In particular, we will prove there exist  $c_1, C_1 > 0$  (depending on  $\epsilon$ ) such that for all  $n$  sufficiently large

$$\frac{c_1}{n^{1/3}} \leq \mathbb{P}(N \geq n^{4/3}) \leq \frac{C_1 \log n}{n^{1/3}}.$$

Note that, this will imply Theorem 2.4.

### 4.1 Proof of Theorem 2.4 upper bound

First we restrict ourselves to smaller events. We have

$$\mathbb{P}(N \geq n^{4/3}) \leq \mathbb{P}(D \geq 2n) + \mathbb{P}(\{N \geq n^{4/3}\} \cap \{D \leq 2n\}).$$

We have already proved

$$\mathbb{P}(D \geq 2n) \leq \frac{C}{n^{1/3}}.$$

So, we will only consider  $\mathbb{P}(\{N \geq n^{4/3}\} \cap \{D \leq 2n\})$ . We want to apply an averaging argument. Let  $V$  denote the line segment with  $\lceil n^{2/3} \rceil$  vertices on  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$ . For  $v \in V$  define the following random variables.

$$D_v := \max\{n \in \mathbb{N} : \exists u \in \mathcal{L}_{-n} \text{ such that } v \in \Gamma_u^{\theta_u}\}$$

$$N_v := |\{u \in \mathbb{Z}^2 : v \in \Gamma_u^{\theta_u}\}|$$

Then we have for all  $v \in V$

$$\mathbb{P}(N \geq n^{4/3}, D \leq 2n) = \mathbb{P}(N_v \geq n^{4/3}, D_v \leq 2n).$$

Hence,

$$\lceil n^{2/3} \rceil \mathbb{P}(N \geq n^{4/3}, D \leq 2n) = \sum_{v \in V} \mathbb{P}(N_v \geq n^{4/3}, D_v \leq 2n) = \mathbb{E}(\widehat{N}),$$

where  $\widehat{N}$  is defined as follows.

$$\widehat{N} := \sum_{v \in V} \mathbb{1}_{\{N_v \geq n^{4/3}, D_v \leq 2n\}}.$$

For  $\ell \geq 1$  we want to estimate  $\mathbb{P}(\widehat{N} \geq \ell n^{1/3} \log n)$ . We define some more random variables.

$$\widetilde{N} := |\{u \in \mathbb{Z}^2 : -2n \leq \phi(u) \leq 0 \text{ and } \exists v \in V \text{ such that } v \in \Gamma_u^{\theta_u}\}|.$$

Observe that

$$\{\widehat{N} \geq \ell n^{1/3} \log n\} \subset \{\widetilde{N} \geq \ell n^{5/3} \log n\}.$$

Same as before, for  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  and for a uniformly bounded sequence  $\{c_n\}$  we partition the interval  $(\epsilon, \frac{\pi}{2} - \epsilon)$  into intervals  $A_i$  each of equal length  $\frac{c_n(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$ . For  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  we define the following random variables.

$$\widetilde{N}_i := |\{u \in \mathbb{Z}^2 : -2n \leq \phi(u) \leq 0, \theta_u \in A_i \text{ and } \exists v \in V \text{ such that } v \in \Gamma_u^{\theta_u}\}|$$

Clearly, we have

$$\{\widetilde{N} \geq \ell n^{5/3} \log n\} \subset \bigcup_{i=1}^{\lfloor n^{1/3} \rfloor} \{\widetilde{N}_i \geq \ell n^{4/3} \log n\}.$$

We will show that for each  $1 \leq i \leq \lfloor n^{1/3} \rfloor$  there exist  $C, c > 0$  (depending only on  $\epsilon$ ) such that for sufficiently large  $n$  (depending only on  $\epsilon$ ) and  $\ell < n^{1/6}$

$$\mathbb{P}(\widetilde{N}_i \geq \ell n^{4/3}) \leq C e^{-c\ell}. \quad (4.1)$$

We fix  $1 \leq i \leq \lfloor n^{1/3} \rfloor$ . Let  $\alpha_i, \beta_i, \theta_i$  denote the left end point, right end point and midpoint of  $A_i$  respectively and  $n_i \in \mathcal{L}_{-2n}$  be the intersection point of  $\mathcal{L}_{-2n}$  and the line  $y = (\tan \theta_i)x$ . Consider the line segment  $\widetilde{V}$  with  $\lfloor \frac{\ell}{32} n^{2/3} \rfloor$  vertices on  $\mathcal{L}_{-2n}$  with midpoint  $n_i$ . Let  $v_1, v_2$  be the end points of  $\widetilde{V}$  with  $\psi(v_1) \leq \psi(v_2)$ . Consider  $\Gamma_{v_1}^{\alpha_i}$  and  $\Gamma_{v_2}^{\beta_i}$ . Let  $R_i$  denote the parallelogram whose one pair of opposite sides are line segments with  $\lfloor \frac{\ell}{16} n^{2/3} \rfloor$  many vertices and they lie on  $\mathcal{L}_{-2n}$  (resp.  $\mathcal{L}_0$ ) with midpoint  $n_i$  (resp.  $\mathbf{0}$ ) (see Figure 6). Let  $\mathcal{H}_i$  be the event defined as follows.

- $\mathcal{H}_i := \{\exists u \in \mathbb{Z}^2 \setminus R_i \text{ such that } \theta_u \in A_i \text{ and } v \in \Gamma_u^{\theta_u} \text{ for some } v \in V\}.$

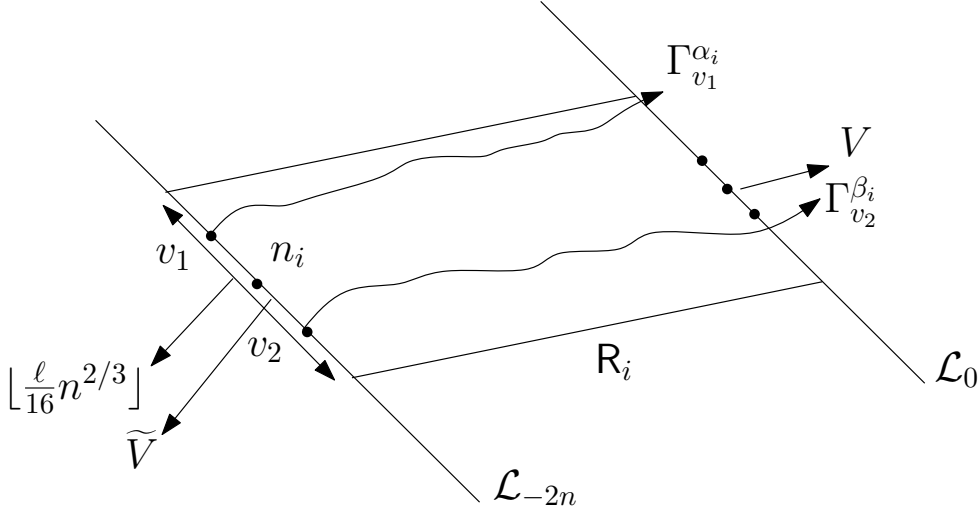


Figure 6: To prove the upper bound in Theorem 2.4 we fix  $1 \leq i \leq \lfloor n^{1/3} \rfloor$ . The event that there are more than  $\ell n^{4/3}$  many  $u$  with  $\theta_u \in A_i$  and  $\Gamma_u^{\theta_u}$  intersecting  $V$  can happen in two ways. Either the above  $u$  lies outside  $R_i$ . For sufficiently large  $\ell$  this will imply large transversal fluctuation for either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$ . Proposition 3.3 gives a stretched exponential upper bound for this event. The other possibility is there are more than  $\ell n^{4/3}$  many vertices inside  $R_i$  that choose an angle in  $A_i$ . Using Hoeffding's inequality we get a stretched exponential upper bound for this event. Combining the above we get (4.1).

Then

$$\mathbb{P}(\tilde{N}_i \geq \ell n^{4/3}) \leq \mathbb{P}(\mathcal{H}_i) + \mathbb{P}(\{\tilde{N}_i \geq \ell n^{4/3}\} \cap (\mathcal{H}_i)^c). \quad (4.2)$$

Note that by planarity the event  $\mathcal{H}_i$  can happen in two ways. Either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$  will leave  $R_i$  or they will intersect  $V$  (see Figure 6). Since,  $(\beta_i - \alpha_i)$  is of order  $\frac{1}{n^{1/3}}$ , for sufficiently large  $\ell$  (depending on  $\epsilon$ ), the event  $\mathcal{H}_i$  will imply that either  $\Gamma_{v_1}^{\alpha_i}$  or  $\Gamma_{v_2}^{\beta_i}$  will have transversal fluctuation larger than  $\frac{\ell}{64} n^{2/3}$ . We have the following proposition.

Hence, for sufficiently large  $\ell$

$$\mathbb{P}(\mathcal{H}_i) \leq C e^{-c\ell^3},$$

for some  $C, c > 0$  depending on  $\epsilon$ . We consider the second term in (4.2). On the event  $\{\tilde{N}_i \geq \ell n^{4/3}\} \cap (\mathcal{H}_i)^c$  there are at least  $\ell n^{4/3}$  many vertices inside  $R_i$  that chooses an angle from  $A_i$ . For large  $\ell$  and  $n$  there are at most  $\frac{\ell}{4} n^{5/3}$  many vertices inside  $R_i$ . Consider the collection of i.i.d. random variables  $\{\mathbb{1}_{\{\theta_u \in A_i\}}\}_{u \in R_i}$  with mean  $\frac{c_n}{n^{1/3}}$ . So,

$$\{\tilde{N}_i \geq \ell n^{4/3}\} \cap (\mathcal{H}_i)^c \subset \left\{ \sum_{u \in R_i} \mathbb{1}_{\{\theta_u \in A_i\}} \geq \ell n^{4/3} \right\}.$$

By Hoeffding's inequality we have, for some  $c > 0$

$$\mathbb{P} \left( \left\{ \sum_{u \in R_i} \mathbb{1}_{\{\theta_u \in A_i\}} \geq \ell n^{4/3} \right\} \right) \leq \mathbb{P} \left( \sum_{u \in R_i} \mathbb{1}_{\{\theta_u \in A_i\}} - \mathbb{E} \left( \sum_{u \in R_i} \mathbb{1}_{\{\theta_u \in A_i\}} \right) \geq \frac{\ell}{2} n^{4/3} \right) \leq e^{-c\ell}.$$

Hence, we proved (4.1). Now, for  $\ell < n^{1/6}$  sufficiently large

$$\mathbb{P}(\tilde{N}_i \geq \ell n^{4/3} \log n) \leq C e^{-c\ell \log n}.$$

So, we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\lfloor n^{1/3} \rfloor} \{\tilde{N}_i \geq \ell n^{4/3} \log n\}\right) \leq \sum_{i=1}^{\lfloor n^{1/3} \rfloor} \mathbb{P}(\tilde{N}_i \geq \ell n^{4/3} \log n) \leq C n^{1/3} e^{-c\ell \log n} \leq C' e^{-c'\ell}.$$

Thus,

$$\mathbb{P}(\hat{N} \geq \ell n^{1/3} \log n) \leq C' e^{-c'\ell}.$$

So, we have there exists  $\ell_0$  (depending only on  $\epsilon$ ) and  $\tilde{C}$  (depending only on  $\epsilon$ )

$$\mathbb{E}\left(\frac{\hat{N}}{n^{1/3} \log n}\right) \leq \sum_{\ell \geq 1} \mathbb{P}(\hat{N} \geq \ell n^{1/3} \log n) \leq \ell_0 + \sum_{\ell_0 \leq \ell < n^{1/6}} C' e^{-c'\ell} + \sum_{n^{1/6} < \ell < \frac{n^{1/3}}{\log n}} C' e^{-c' n^{1/6}} < \tilde{C}.$$

This proves the upper bound in Theorem 2.4 after an  $n \leftrightarrow n^{4/3}$  change of variable.  $\square$

## 4.2 Proof of Theorem 2.4 lower bound

Same as before, first we fix large constants  $M$  and  $\ell_0$  which will be chosen later. For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , we consider the intervals  $A_i \subset (\epsilon, \frac{\pi}{2} - \epsilon)$ , each of length  $\frac{(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$  and midpoint  $\frac{Mi(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$ . Corresponding to each  $A_i$  we define the following random variables.

$$N_i := |\{u \in \mathbb{Z}^2 : \theta_u \in A_i \text{ and } \mathbf{0} \in \Gamma_u^{\theta_u}\}|.$$

Then we have the following.

$$\bigcup_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \{N_i \geq \frac{n^{4/3}}{\ell_0}, n \leq D \leq 2n\} \subset \{N \geq \frac{n^{4/3}}{\ell_0}\}.$$

For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , let

$$\mathcal{I}_i := \{N_i \geq \frac{n^{4/3}}{\ell_0}, n \leq D \leq 2n\}.$$

Applying inclusion exclusion principle on the union of  $\mathcal{I}_i$  and considering the first two terms we have the following lower bound.

$$\mathbb{P}(N \geq \frac{n^{4/3}}{\ell_0}) \geq \sum_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{I}_i) - \sum_{1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{I}_i \cap \mathcal{I}_j). \quad (4.3)$$

First we consider the first sum. We fix  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ . Let us consider the line segment  $V_M$  with  $\lceil 4Mn^{2/3} \rceil$  many vertices on  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$ . For  $v \in V_M$ , define

$$N_i^v := |\{u \in \mathbb{Z}^2 : \theta_u \in A_i \text{ and } v \in \Gamma_u^{\theta_u}\}|.$$

$$\lceil 4Mn^{2/3} \rceil \mathbb{P}\left(\left\{N_i \geq \frac{n^{4/3}}{\ell_0}, n \leq D \leq 2n\right\}\right) = \sum_{v \in V_M} \mathbb{P}\left(\left\{N_i^v \geq \frac{n^{4/3}}{\ell_0}, n \leq D_v \leq 2n\right\}\right) = \mathbb{E}(\widehat{N}_i),$$

where

$$\widehat{N}_i := \sum_{v \in V_M} \mathbb{1}_{\{N_i^v \geq \frac{n^{4/3}}{\ell_0}, n \leq D_v \leq 2n\}}.$$

We will show on a positive probability event  $\widehat{N}_i \geq 1$ . As before let  $\alpha_i, \beta_i, \theta_i$  denote the left end point,

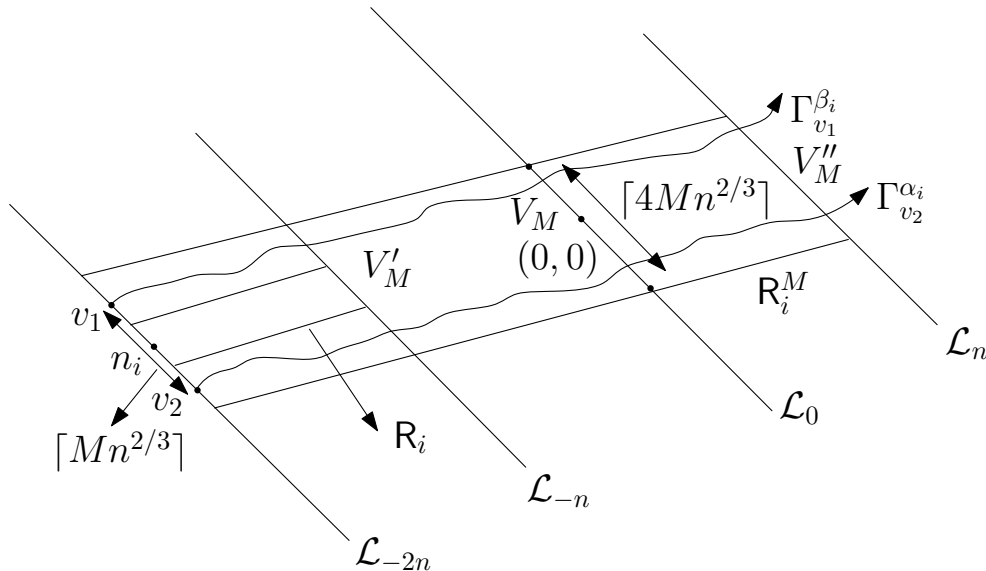


Figure 7: To find lower bound for the first sum in (4.3) we fix an  $i$ . Using Proposition 3.3 we can fix  $M$  large enough so that with high probability  $\Gamma_{v_1}^{\beta_i}$  and  $\Gamma_{v_2}^{\alpha_i}$  are contained in  $R_i^M$  and do not intersect  $R_i$ . On this event any car starting from  $R_i$  and taking angle in  $A_i$  will intersect  $\mathcal{L}_0$  on  $V_M$ . Further, using [4, Theorem 3.10] depending on this  $M$  we can chose  $\ell_0$  large enough so that with large probability there are at most  $\ell_0$  many distinct points on  $V_M$  carrying a geodesic starting from  $V'_M$  and ending at  $V''_M$ . Also, with positive probability there are at least  $n^{4/3}$  any vertices  $u \in R_i$  such that  $\theta_u \in A_i$ . The last event is independent of the first two events. Hence, combining all these we get on a positive probability event there exists  $v \in V_M$  such that there are at least  $\frac{n^{4/3}}{\ell_0}$  many  $u$  with  $\theta_u \in A_i$  and  $v \in \Gamma_u^{\theta_u}$ . Now, applying an averaging argument and summing over all  $i$  give a lower bound for the first sum in (4.3).

right end point, mid point of  $A_i$  respectively. Consider  $n_i$  which is the intersection point of  $\mathcal{L}_{-2n}$  and  $y = \tan(\theta_i)x$ . Consider the line segment  $\widetilde{V_M}$  with  $\lceil Mn^{2/3} \rceil$  many vertices on  $\mathcal{L}_{-2n}$  with midpoint  $n_i$ .

Let  $v_1$  (resp.  $v_2$ ) denote the end points of  $\widetilde{V}_M$  with  $\psi(v_1) \leq \psi(v_2)$ . Further consider the following two parallelograms (see Figure 7). Let  $n'_i$  (resp.  $n''_i$ ) denote the intersection point of  $y = \tan(\theta_i)x$  with  $\mathcal{L}_{-n}$  (resp.  $\mathcal{L}_n$ ).

- $R_i$  is the parallelogram whose two opposite pair of sides are line segments with  $\lceil \frac{n^{2/3}}{4} \rceil$  vertices and they lie on  $\mathcal{L}_{-2n}$  (resp.  $\mathcal{L}_{-n}$ ) with midpoint  $n_i$  (resp.  $n'_i$ ).
- $R_i^M$  is the parallelogram whose two opposite pair of sides are line segments with  $\lceil 4Mn^{2/3} \rceil$  many vertices, they lie on  $\mathcal{L}_{-2n}$  (resp.  $\mathcal{L}_n$ ) with midpoint  $n_i$  (resp.  $n''_i$ ).

We define the event  $\mathbf{TF}_i$  as follows.

- $\mathbf{TF}_i$  is the event that  $\Gamma_{v_1}^{\beta_i}$  and  $\Gamma_{v_2}^{\alpha_i}$  are contained within  $R_i^M$  and  $\Gamma_{v_1}^{\beta_i}$  and  $\Gamma_{v_2}^{\alpha_i}$  do not enter  $R_i$ .

As  $(\beta_i - \alpha_i)$  is of order  $\frac{1}{n^{1/3}}$ , the event  $\mathbf{TF}_i$  will imply for sufficiently large  $M$  either  $\Gamma_{v_1}^{\beta_i}$  or  $\Gamma_{v_2}^{\alpha_i}$  will have transversal fluctuation more than  $\frac{M}{4}n^{2/3}$ . Using Proposition 3.3 we can chose  $M$  (depending only on  $\epsilon$ ) large enough and  $n$  sufficiently large enough (depending on  $M$ ) so that the following happens.

$$\mathbb{P}(\mathbf{TF}_i) \geq 0.99.$$

On  $\mathbf{TF}_i$ , for any  $u \in R_i$  with  $\theta_u \in A_i$ ,  $v \in \Gamma_u^{\theta_u}$ , for some  $v \in V_M$ . Now we define an equivalence class. Let  $n''_i$  denote the intersection point of  $y = \tan(\theta_i)x$  with  $\mathcal{L}_n$ . Consider the two line segments  $V'_M$  (resp.  $V''_M$ ) with  $\lceil 4Mn^{2/3} \rceil$  vertices on  $\mathcal{L}_{-n}$  (resp.  $\mathcal{L}_n$ ) with midpoint  $n'_i$  (resp.  $n''_i$ ). For  $u, v \in V'_M$  and  $u', v' \in V''_M$ , we say  $(u, v) \sim (u', v')$  if  $\Gamma_{u,v}(0) = \Gamma_{u',v'}(0)$ . Let  $E_i$  denote the number of equivalence classes. Using [4, Theorem 3.10], we can chose  $\ell_0$  large enough (depending on  $M$ ) so that for all sufficiently large  $n$

$$\mathbb{P}(E_i \leq \ell_0) \geq 0.99.$$

Now, we consider the following random variable.

$$Y_n := \sum_{u \in R_i} \mathbb{1}_{\{\theta_u \in A_i\}}.$$

Note that for large  $n$ ,  $\mathbb{E}(Y_n) \geq \frac{n^{4/3}}{4}$  and  $\text{Var}(Y_n) = c_n(n^{4/3} - n)$  for some uniformly bounded sequence  $\{c_n\}$ . By Berry-Esseen inequality we have the following for all  $n$  and a constant  $c$

$$|\mathbb{P}(\{Y_n - \mathbb{E}(Y_n) \geq 0\}) - 1/2| \leq \frac{c}{(n^{4/3} - n)^{3/2}n^{5/6}}.$$

Hence, for all sufficiently large  $n$  we have

$$\mathbb{P}(\{Y_n - \frac{n^{4/3}}{4} \geq 0\}) \geq \frac{1}{4}.$$

Finally we observe,

$$\mathbf{TF}_i \cap \{E_i \leq \ell_0\} \cap \{Y_n - \frac{n^{4/3}}{4} \geq 0\} \subset \{\widehat{N}_i \geq 1\}. \quad (4.4)$$

Further, the event  $\{Y_n - \frac{n^{4/3}}{4} \geq 0\}$  is independent of  $\mathbf{TF}_i \cap \{E_i \leq \ell_0\}$ . So, combining all the above for sufficiently large  $n$ , there exists  $c_1 > 0$  ( $c_1$  is a fixed constant, does not depend on anything) such that

$$\mathbb{E}(\widehat{N}_i) \geq \mathbb{P}\left(\mathbf{TF}_i \cap \{E_i \leq \ell_0\} \cap \{Y_n - \frac{n^{4/3}}{4} \geq 0\}\right) \geq c_1.$$

Hence,

$$\mathbb{P}\left(\left\{N_i \geq \frac{n^{4/3}}{\ell_0}, n \leq D \leq 2n\right\}\right) \geq \frac{c_1}{4Mn^{2/3}}. \quad (4.5)$$

Now, we consider the second sum in (4.3). We fix  $1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , both large enough. We recall the events  $\mathcal{E}_i$  defined in the proof of lower bound of Theorem 2.5. Clearly, for all  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$

$$\mathcal{I}_i \subset \mathcal{E}_i.$$

So, as we did in the proof of lower bound of Theorem 2.5, from (4.5), (3.15) and (4.3) we have

$$\mathbb{P}(N \geq \frac{n^{4/3}}{\ell_0}) \geq \frac{c_1}{8M^2n^{1/3}} - \sum_{1 \leq i < j \leq \frac{n^{1/3}}{2M}} \frac{C_2 \log(M|i-j|)}{M^2|i-j|^2n^{2/3}}.$$

As argued before, choosing  $M$  sufficiently large completes the proof of Theorem 2.4 lower bound (with  $n \leftrightarrow n^{4/3}$ , as before).  $\square$

## 5 Distance to find a road with large number of cars

In this section we analyse how far one needs to go from origin to see a road with large number of cars.

### 5.1 Proof of Theorem 2.7

Let  $I$  denote the line segment with  $\lfloor n^{1/3} \rfloor$  many vertices on  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$ . We divide the interval  $(\epsilon, \frac{\pi}{2} - \epsilon)$  as we did in the proof of lower bound of Theorem 2.4. i.e., we fix large constants  $M$  and  $\ell_0$  which will be chosen later. For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , we consider the intervals  $A_i \subset (\epsilon, \frac{\pi}{2} - \epsilon)$ , each of length  $\frac{(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$  and midpoint  $\frac{Mi(\frac{\pi}{2} - 2\epsilon)}{n^{1/3}}$ . Let us define the following events. For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$  we define

- $\mathcal{K}_i := \{\exists w \in I \text{ such that there are } \frac{n^{4/3}}{\ell_0} \text{ many vertices } v \text{ with } -2n \leq \phi(v) \leq -n, \theta_v \in A_i \text{ and } w \in \Gamma_v^{\theta_v}\}.$

Clearly,

$$\bigcup_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \mathcal{K}_i \subset \{T_n \leq n^{1/3}\}.$$

Same as before, considering the first two terms after applying inclusion-exclusion principle we have

$$\mathbb{P}(T_n \leq n^{1/3}) \geq \sum_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{K}_i) - \sum_{1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor} \mathbb{P}(\mathcal{K}_i \cap \mathcal{K}_j). \quad (5.1)$$

We first consider the first sum. We fix  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ . We will again apply an averaging argument. But this time we will average over disjoint intervals containing  $\lfloor n^{1/3} \rfloor$  vertices each. Precisely, we do the following. We consider  $V_M$ , line segment with  $\lfloor 4Mn^{1/3} \rfloor \lfloor n^{1/3} \rfloor$  many vertices on  $\mathcal{L}_0$  with midpoint  $\mathbf{0}$ . We partition it into  $\lfloor 4Mn^{1/3} \rfloor$  many mutually disjoint sub-intervals each with  $\lfloor n^{1/3} \rfloor$  many vertices inside  $V_M$ . i.e., consider the mutually disjoint intervals  $I_j$  each with  $\lfloor n^{1/3} \rfloor$  many vertices. For each of these intervals we define the following events.

- $\mathcal{K}_i^j := \{\exists w \in I_j \text{ such that there are } \frac{n^{4/3}}{\ell_0} \text{ many vertices } v \text{ with } -2n \leq \phi(v) \leq -n, \theta_v \in A_i \text{ and } w \in \Gamma_v^{\theta_v}\}.$

We have

$$\mathbb{P}(\mathcal{K}_i) = \mathbb{P}(\mathcal{K}_i^j).$$

So,

$$\lfloor 4Mn^{1/3} \rfloor \mathbb{P}(\mathcal{K}_i) = \sum_{1 \leq j \leq \lfloor 4Mn^{1/3} \rfloor} \mathbb{P}(\mathcal{K}_i^j) = \mathbb{E} \left( \sum_{1 \leq j \leq \lfloor 4Mn^{1/3} \rfloor} \mathbb{1}_{\mathcal{K}_i^j} \right). \quad (5.2)$$

We will show, on a positive probability set, for large  $n$  the following holds.

$$\sum_{1 \leq j \leq \lfloor 4Mn^{1/3} \rfloor} \mathbb{1}_{\mathcal{K}_i^j} \geq 1.$$

But observe that, precisely this we have shown in the proof of Theorem 2.4 lower bound. Specifically, in (4.4) we have shown that on the event  $\mathbf{TF}_i \cap \{E_i \leq \ell_0\} \cap \{Y_n - \frac{n^{4/3}}{4} \geq 0\}$  there is at least one vertex  $w \in V_M$  such that there are  $\frac{n^{4/3}}{4\ell_0}$  many  $v \in \mathbb{Z}^2$  with  $-2n \leq \phi(v) \leq -n, \theta_v \in A_i$  and  $w \in \Gamma_v^{\theta_v}$ . So, we have

$$\mathbf{TF}_i \cap \{E_i \leq \ell_0\} \cap \{Y_n - \frac{n^{4/3}}{4} \geq 0\} \subset \left\{ \sum_{1 \leq j \leq \lfloor 4Mn^{1/3} \rfloor} \mathbb{1}_{\mathcal{K}_i^j} \geq 1 \right\}.$$

Hence, there exists  $c > 0$  such that for sufficiently large  $n$  we have

$$\mathbb{E} \left( \sum_{1 \leq j \leq \lfloor 4Mn^{1/3} \rfloor} \mathbb{1}_{\mathcal{K}_i^j} \right) \geq c.$$

So, from (5.2) we have for large  $n$ , and for some  $c > 0$ ,

$$\mathbb{P}(\mathcal{K}_i) \geq \frac{c}{Mn^{1/3}}.$$

Now, we consider the second sum in (5.1). We fix  $1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ . Same as before we consider the following events. For  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$ , we define

- $\mathcal{M}_i := \{\exists u \in \mathbb{Z}^2, \gamma \in A_i \text{ such that } \phi(u) = -n, \text{ and } \exists w \in I \text{ such that } w \in \Gamma_u^\gamma\}.$

Clearly,

$$\mathcal{K}_i \subset \mathcal{M}_i.$$

So, we will work with  $\mathcal{M}_i$  from now on. We will again apply an averaging argument in both space and time directions. But this time in the space direction we will average over disjoint intervals of length  $n^{1/3}$ . Precisely, we do the following. We fix  $m$  such that  $-\lceil \frac{n}{100k_{i,j}} \rceil \leq m \leq \lceil \frac{n}{100k_{i,j}} \rceil$ . We consider the line segments  $\{v : |\psi(v)| \leq \lfloor n^{1/3} \rfloor \lfloor n^{1/3} \rfloor\}$  on  $\mathcal{L}_m$  and partition them into line segments with  $\lfloor n^{1/3} \rfloor$  many vertices. So, we get  $\frac{c_n n^{4/3}}{100k_{i,j}}$  many mutually disjoint line segments each with  $\lfloor n^{1/3} \rfloor$  many vertices for some uniformly bounded sequence  $\{c_n\}$ . We will denote the line segments by  $I_{k,m}$ , if the line segment lies on  $\mathcal{L}_m$  and  $k$  is the index to specify the partition in the space direction on  $\mathcal{L}_m$ . For all  $-\lceil \frac{n}{100k_{i,j}} \rceil \leq m \leq \lceil \frac{n}{100k_{i,j}} \rceil$  and  $1 \leq k \leq 2\lfloor n^{1/3} \rfloor$  and for all  $1 \leq i \leq \lfloor \frac{n^{1/3}}{2M} \rfloor$  we define the following events.

- $\mathcal{M}_i^{k,m} := \{\exists u \in \mathbb{Z}^2, \gamma \in A_i \text{ such that } \phi(u) = -n + m, \text{ and } \exists w \in I_{k,m} \text{ such that } w \in \Gamma_u^\gamma\}.$

Clearly,

$$\mathbb{P}(\mathcal{M}_i \cap \mathcal{M}_j) = \mathbb{P}(\mathcal{M}_i^{k,m} \cap \mathcal{M}_j^{k,m}).$$

So, we have

$$\begin{aligned} \frac{c_n n^{4/3}}{100k_{i,j}} \mathbb{P}(\mathcal{M}_i \cap \mathcal{M}_j) &= \sum_{\substack{-\lceil \frac{n}{100k_{i,j}} \rceil \leq m \leq \lceil \frac{n}{100k_{i,j}} \rceil \\ 1 \leq k \leq 2 \lfloor n^{1/3} \rfloor}} \mathbb{P}(\mathcal{M}_i^{k,m} \cap \mathcal{M}_j^{k,m}) \\ &= \mathbb{E} \left( \sum_{\substack{-\lceil \frac{n}{100k_{i,j}} \rceil \leq m \leq \lceil \frac{n}{100k_{i,j}} \rceil \\ 1 \leq k \leq 2 \lfloor n^{1/3} \rfloor}} \mathbb{1}_{\mathcal{H}_i^{k,m} \cap \mathcal{H}_j^{k,m}} \right) \leq \mathbb{E}(N_{i,j}). \end{aligned}$$

Recall that we had define  $N_{i,j}$  in the proof of lower bound in Theorem 2.5 and from (3.11) we have there exists constant  $C > 0$  such that for large  $n$

$$\frac{c_n n^{4/3}}{100k_{i,j}} \mathbb{P}(\mathcal{M}_i \cap \mathcal{M}_j) \leq \frac{C n \log |k_{i,j}|}{k_{i,j}^3}.$$

Combining all the above and from (5.1) we have there are  $C, c > 0$  (depending only on  $\epsilon$ )

$$\mathbb{P}(T_n \leq n^{1/3}) \geq \sum_{i=1}^{\lfloor \frac{n^{1/3}}{2M} \rfloor} \frac{c}{M n^{1/3}} - \sum_{1 \leq i < j \leq \lfloor \frac{n^{1/3}}{2M} \rfloor} \frac{C \log(M|i-j|)}{M^2 |i-j|^2 n^{1/3}}.$$

Hence, again we have for some constant  $C, c > 0$  we have

$$\mathbb{P}(T_n \leq n^{1/3}) \geq \frac{c}{M^2} - \frac{C}{M^3}.$$

Finally, we can choose  $M$  large enough so that the right side is positive. This completes the proof.  $\square$

## 6 Empirical analysis of road networks

This section describes our approach for empirically validating theoretical predictions about the alignment of road networks with geodesic paths and the statistics that follows. A significant basis for our model is the hypothesis that in hilly terrain, real-world road network configurations substantially align with geodesic shortest paths when elevation variations are considered. Our road network model leverages the Last Passage Percolation (LPP) model for its theoretical tractability in predicting road network structures. However, we acknowledge that road networks more realistically conform to the dynamics of First Passage Percolation (FPP) models, despite their lesser tractability.

The interconnection between LPP and FPP models within the Kardar-Parisi-Zhang (KPZ) universality class offers a theoretical justification for this approach. This KPZ linkage underpins our

assumption that insights derived from the LPP framework can provide predictive insights for FPP models.

Our validation consists of two parts. First we verify that, on the input side of the model, FPP in a hilly environment indeed predicts where roads are built, this is done in Section 6.1. We then turn to compare the scaling behavior in Theorems 2.6 and 2.7 to actual road layout and traffic statistics in Section 6.2.

## 6.1 Road networks and elevation data

To ensure the relevance and accuracy of FPP models in capturing the essence of real-world road layouts, this section is dedicated to validating this assumption. By examining the alignment of road networks against geodesic paths influenced by elevation data, we aim to validate the predictive capability of FPP models and, in turn, our LPP variant.

### 6.1.1 Data acquisition and processing

Our study leverages high-resolution elevation data sourced from the Shuttle Radar Topography Mission (SRTM), accessed through the Earth Explorer platform [18]. The SRTM dataset provides comprehensive global elevation data. From this dataset, we extract elevation data encoded into image files, which serve as input in building an edge weighted graph.

### 6.1.2 Graph construction and path analysis

In this study, we leveraged a computational framework found within the GitHub repository [8] to facilitate our analysis of road networks within a designated study area. Utilizing the code from this repository enabled us to construct a weighted graph that models a square lattice segment of the terrain under investigation. Each vertex within this graph represents a precise point on the terrain, and the weights assigned to the edges are calculated based on both the elevation differences and the distances between the vertices at the endpoints.

Specifically, if  $\Delta$  denotes the grid spacing between elevation samples and  $h$  is a function expressing the elevation at a vertex then the edge weight on the edge  $\{x, y\}$  of the graph is taken to be

$$w_{\{x,y\}} = \sqrt{\Delta^2 + (h(x) - h(y))^2}.$$

Moreover, we have found that in some cases more consistent results can be achieved between the predicted paths and the observed roads if we treat  $\Delta$  as a tuneable parameter.

### 6.1.3 Figures

We provide Figures 8, 9 and 10 to illustrate the comparison between actual road segments and theoretical geodesic paths derived from our graph model. Each figure shows a map provided by ESRI World Topo Map [10] with a geodesic path, calculated based on elevation data and shortest path algorithms, overlaid on selected road segments. The intent behind these figures is to visually highlight where and how actual road networks align with or deviate from the calculated geodesic paths. This direct comparison serves as evidence for the hypothesis that road layouts tend to follow geodesic paths, considering elevation.

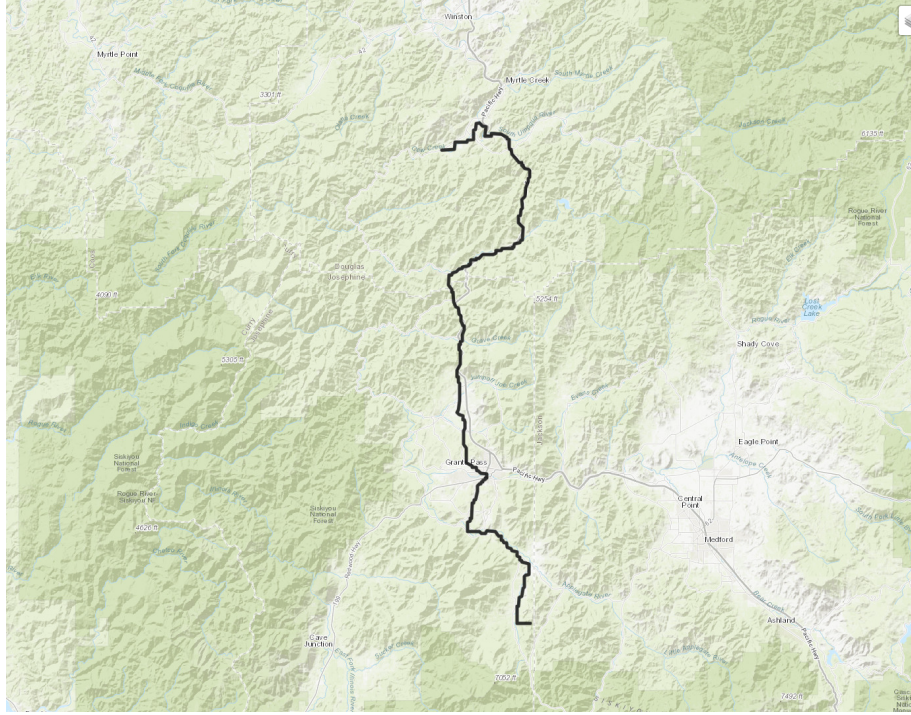


Figure 8: A path computed between coordinates  $(42.9251086, -123.4220748)$  and  $(42.2017416, -123.2366819)$ .

The figures largely corroborate our hypothesis that real-world road networks tend to follow the theoretical shortest paths deduced from our model, with particular emphasis on elevation data. Nonetheless, certain regions exhibit discrepancies, prompting a discussion of potential causes for these divergences.

**Finite Lattice Effects** One factor to consider is the graph model’s finite lattice structure, which simulates the continuous real-world terrain with a discrete grid. This approximation introduces potential errors in shortest path calculations, where the model’s sensitivity to edge weight value can manifest as notable deviations from expected paths. Such minor inaccuracies in terrain representation, therefore, might explain the observed misalignment between theoretical predictions and actual road configurations. This effect is made apparent between figures 9 and 10.

**Beyond Shortest Path Considerations** Additionally, the real-world intricacies of road design decision-making extend beyond the simplistic criteria of minimizing path length or elevation variations. Factors such as legal restrictions, environmental conservation, and socio-economic impacts heavily influence road planning, often leading to routes that diverge from those predicted by our shortest path model. Elements like land ownership, ecological preservation areas, and pre-existing structures

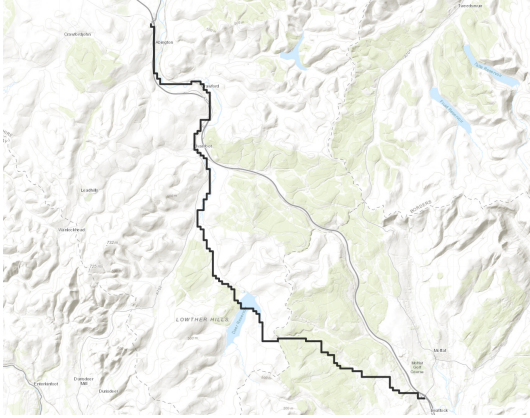


Figure 9: A path computed between  $(55.5028, -3.6964)$  and  $(55.3124, -3.4540)$ . We see an apparent disagreement between portions of the road and geodesic caused by low resolution lattice spacing.

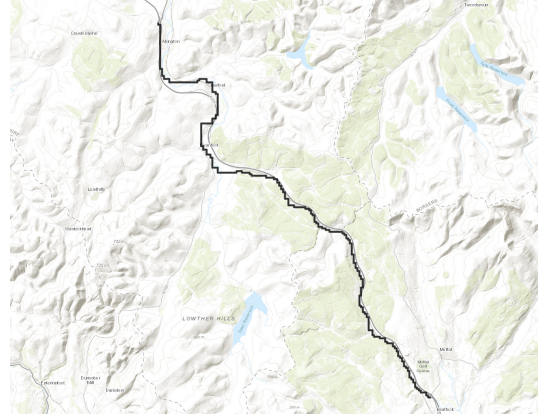


Figure 10: A path computed between  $(55.5028, -3.6964)$  and  $(55.3124, -3.4540)$ . This time we see better agreement when refining lattice scale.

necessitate alterations to the theoretically optimal paths. These considerations, absent from the shortest path algorithm, might account for any discrepancies between our model’s predictions and the existing road networks.

Nevertheless, the clear alignment observed between the model predictions and actual road layouts, despite some discrepancies, would seem to support the use of the FPP model for characterizing road network structures. Then by the KPZ correspondence mentioned earlier, we can feel more confident in the predictive capability of the LPP variant. We now turn to verifying some of these predictions on actual road networks.

## 6.2 Comparison of the model output and real-world traffic data

We use the government traffic dataset from the United Kingdom accessed at [9]. This consists of annual average daily traffic flow at count points scattered around the roads. We are interested in investigating the validity of Theorems 2.6 and 2.7 for this dataset. We will disregard the logarithmic correction in the former. Fix a startpoint and consider distances  $d$  along a spatial direction of LPP. These results say that, with probability separated away from both 0 and 1, we are expected to get at least one road with traffic count proportional to  $d^4$  within this distance  $d$ . As opposed to LPP, we consider real road networks to be isotropic, which means that the direction we choose should be irrelevant. We therefore work with East as our preferred direction. The task is to pick points on the map randomly, draw an Easterly line from each, record where this intersects actual roads as well as the traffic volumes carried by these roads.

It is quite difficult to determine an accurate traffic count on a specific intersection point with our Easterly line because the closest count point on that road might be far away from here. It might also happen that there are many junctions between the intersection point and the nearest count point on

the road, harming the accuracy of the count point to the traffic count at the intersection of interest. Figure 11 illustrates the difficulties.

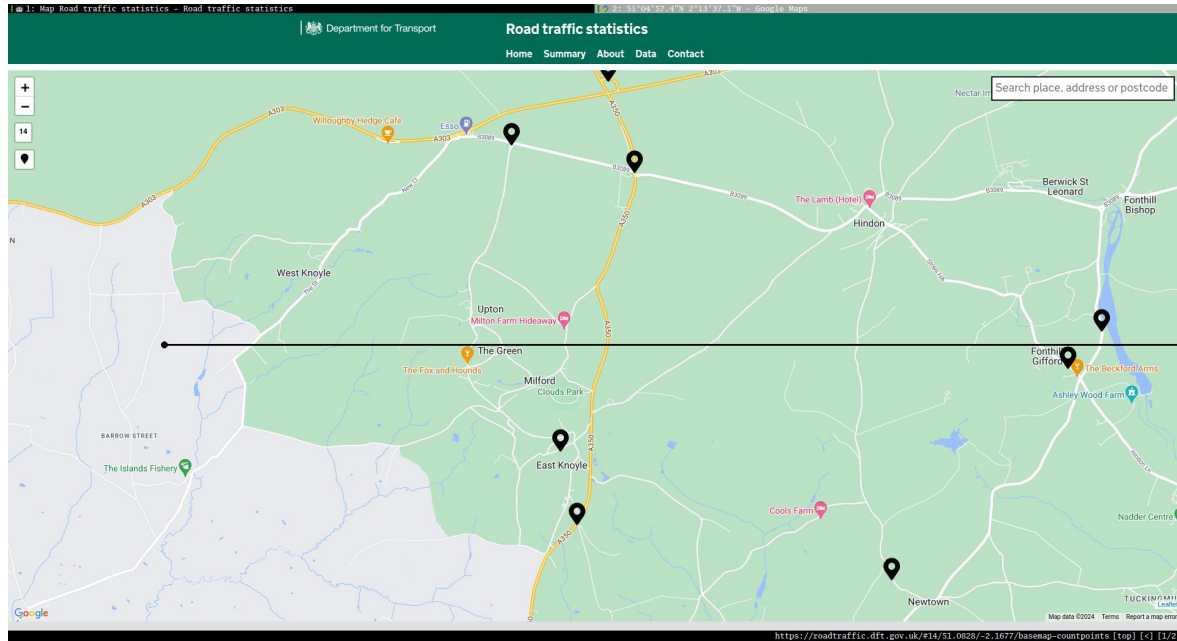


Figure 11: Manual measurement of distances and road traffic. This is the interface of the *Road traffic statistics* webpage, with a startpoint and its Easterly line overlaid by us in black. The pinpoints are the locations of the traffic count points. They can be clicked, and their data is also available to bulk-download. To be recorded are the distances between the startpoint at the left end of the black line and the intersection points of roads with the black line, as well as the traffic carried by those roads.

Careful inspection of the map and of the location of traffic count points allow for such data to be collected with reasonable accuracy, especially for busier roads. We will refer to this as the *manual method*. Smaller lanes are often not measured, we had no choice but to simply leave them out of our analysis. However, this process is really slow and would not give sufficient volumes of data for proper statistical analysis. Unfortunately, we failed to develop an algorithm that could perform this data collection for us.

Instead, we consider a strip around the Easterly line starting from the chosen random point  $x$  and take all the count points that lie inside this strip. We record their distances from  $x$  and the traffic count they possess, and pretend that these are actual locations of intersections between the Easterly line and the roads that they measure. We will refer to this process as the *automated method*.

In the rest of the paper we first calibrate the strip width to give a good match between outcomes of the manual method and the automated method. We do that on a few examples where the manual method was carried out. We then run the automated method on about 2 000 randomly chosen points on the map, and compare the statistics gained with the statements of our Theorems 2.6 and 2.7.

### 6.2.1 Calibrating the automated method

We performed the manual method at four locations of the UK, near Cirencester (51.7076,-1.90897), Gillingham (51.0826,-2.22697), Glasgow (55.606053,-4.0506979), and a point in Wales (52.1577699,-3.5958771). The code to support this analysis, along with additional simulation results is found in a dedicated GitHub repository created by the authors for this project [7].

While the startpoint should be completely random, large-scale inhomogeneity of the UK and its harmful effects soon became apparent. First, big cities are obviously much denser than rural areas, distorting the measurement. Hence we made sure to avoid startpoints from where Easterly lines would have traced into larger urban areas.

Second, the UK has a number of estuaries that needed to be avoided as they form long natural barriers to road building.

Finally, we found that a dense part in England, roughly in the rectangle Drochester – Hastings – Hull – Blackpool, has a significantly denser road network than the North and the West of the UK. Hence we performed our analysis separately for this South-East rectangle, and the North and West regions.

Of the four locations above for the manual method, Cirencester and Gillingham belong to the denser South-East block, while the Wales and the Glasgow startpoints are in the sparser West and North region.

We then ran the automated method on the same four startpoints, with different strip width. Strip width is very important. If the strip is very wide, then it will contain many countpoints which might wrongly include some high-traffic roads that don't actually intersect the Easterly line. Similarly, if the strip is too narrow, then it might miss some important countpoints of high-traffic roads because those countpoints can lie outside of the strip even if that road intersects the Easterly line.

To check the theoretical predictions, we need to answer whether or not there is at least one road above certain traffic threshold within certain distance of the startpoint we chose. This is equivalent to considering the running maximum of crossing traffic volumes as we move alongside the Easterly line from the startpoint. It is therefore this running maximum which we tried to align between the manual and the automated method. The figures below illustrate this graph on the left hand-side. For better intuition we also included violin plots on the right hand-side, which show the traffic distribution measured in the two methods.

Figures 12, 13 and 14 refer to the startpoint near Cirencester (51.7076,-1.90897) in the South-East region. It is noted that sometimes the violin plot takes a peak when there are multiple traffic counts around the same distance. For example, consider the orange line in Figure 14. From the left side, it shows that there is a new maximum that is observed around 60 km after a peak at 40 km. But on the right side, traffic distribution shows that the traffic around 60 km is lower than the traffic at a distance of 40 km. This happens because multiple traffic counts of nearly the same values are observed near 40 km. The plots indicate that the appropriate strip width is between 3 km and 5 km.

Similarly, also in the South-East region Figures 15, 16 and 17 are from the startpoint (51.0826,-2.22697) near Gillingham. Figures 18, 19 and 20 are the plots with the startpoint near Wales (52.1577699,-3.5958771) in the sparser, West region of UK. Finally, Figure 21 is the plot with the startpoint near Glasgow, which belongs to the North region of UK. When taking the manual measurement for this plot, we get a high-traffic road within a distance of 10km, but when considering the automated measurement of a strip width of 10 km, we encounter another, much higher-traffic road, and these remain the busiest crossings until the distance of 150 km. This makes the running maximum very different between the manual and the automated method. However, if we neglect the first few

mismatched measurements, we get a good comparison. Figure 21 illustrates it all. As the points are randomly chosen, rarely bad points like this one are expected to occur, and we do not have an efficient way to eliminate this issue.

After observing these plots, we have concluded that 3 km would be the ideal strip width for the South-East region and 10 km for the North and West regions of the UK.

### 6.2.2 Statistical analysis

We were now ready to deploy the automated method in large scale. As mentioned above, the plots (Figures 12 - 21) indicate that the statistics are different between the South-East and the North and West regions of the UK. We therefore divide the UK into these two regions for our statistical analysis. Around 2000 random startpoints were considered inside the UK, and those points are at least 80 km horizontally away from big cities (like London, Birmingham, etc.), big rivers or oceans to avoid distorting effects to the measurements. For a fixed strip width and distance  $d$ , we observe how many points  $x$  have traffic counts greater than  $k$ , and calculate the relative frequency by dividing by the total number of points, where  $k$  runs from 1000 to 100000. Figures 22 and 23 show the traffic threshold encountered with relative frequency between  $[0.49, 0.51]$  as a function of the distance travelled on the Easterly line.

Figure 22 is the plot when the South-East region is considered with strip width 3 km, whereas Figure 23 for the West and North region with the strip width 10 km. By Theorem 2.6 and Theorem 2.7, with positive probability, we should expect  $Cd^4$  graph from these plots, where  $C$  is constant, but  $C$  depends on the region where we are doing the analysis. We saw similar plots when relative frequency thresholds different from  $[0.49, 0.51]$  were tried. The plots do not resemble a quartic function, which we explain next.

### 6.2.3 Limitations of our analysis

There are obvious simplifications where our model and statistical procedure do not align with reality. The first of these is that geography is not an i.i.d. environment. Long valleys are carved by rivers, introducing obvious correlations in the landscape.

The second observation is the difficulties with extracting geographic information on intersection points between Easterly lines and actual roads. We discussed this in details earlier.

None of these two issues appear to fundamentally change the main ideas of this paper to the extent seen by our UK-wide analysis. Instead, we believe that the main problem is a combination of driving distances and the nature of the road statistics data that we worked with. As apparent from Figure 11, small roads are most often not measured, while the bigger, well-documented roads tend to be 20...30 km apart. Combine this with the typical driving distances of 10...15 km for a UK car trip, and add to the picture that two LPP geodesics towards the same direction, started from distance  $d$  apart, need to travel in the order of  $d^{3/2}$  distance before they have a reasonable chance to coalesce. The typical UK road trip is simply too short to have a good opportunity to join one of the well-measured, big roads. Hence, we are missing the upper tail of the traffic count distribution that would show the sharp growth of a  $d^4$  function. We suspect that more detailed measurements of smaller roads would match the predictions better but such data just do not seem to be available.

Further observations distancing our model from real road networks can be found in Solon et al. [17]. As already mentioned in the introduction, they found that for road networks the environment is best described as having a heavy-tailed distribution, which is not the case for our framework of

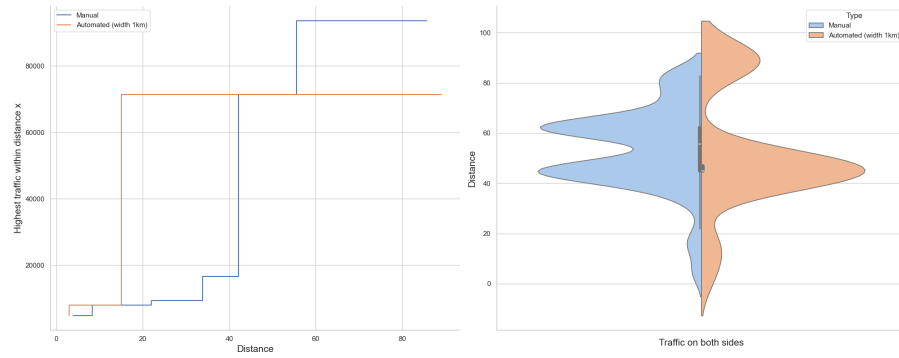


Figure 12: Busiest traffic from the startpoint (51.7076,-1.90897) (to the left) and violin plot (to the right) with strip width 1 km.

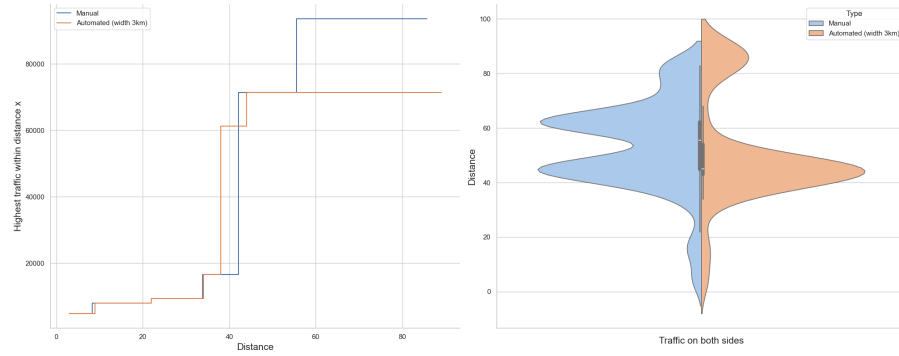


Figure 13: Busiest traffic from the startpoint (51.7076,-1.90897) (to the left) and violin plot (to the right) with strip width 3 km.

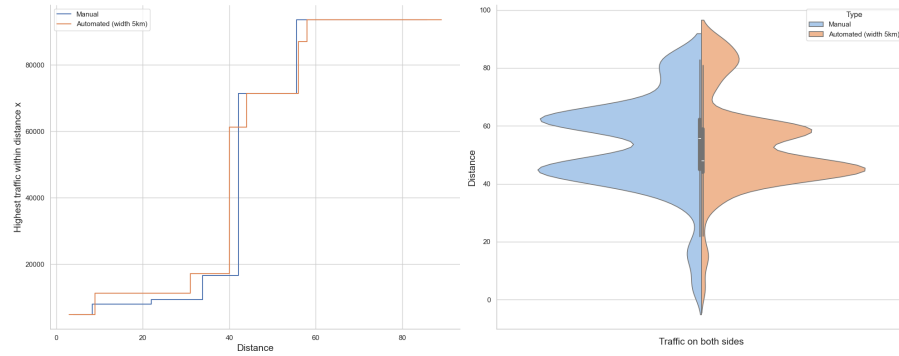


Figure 14: Busiest traffic from the startpoint (51.7076,-1.90897) (to the left) and violin plot (to the right) with strip width 5 km.

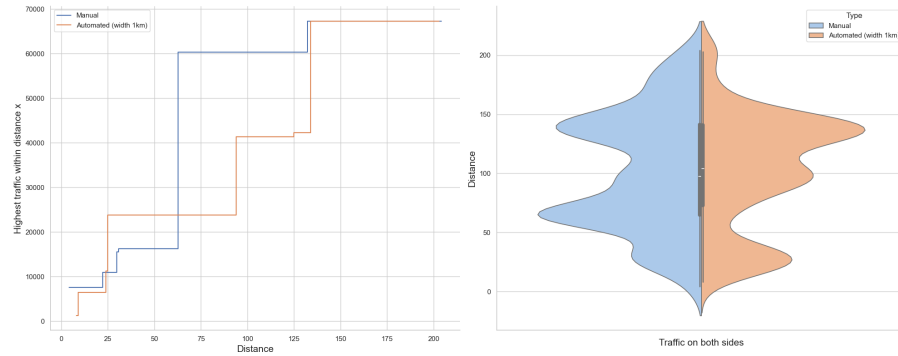


Figure 15: Busiest traffic from the startpoint (51.0826,-2.22697) (to the left) and violin plot (to the right) with strip width 1 km.

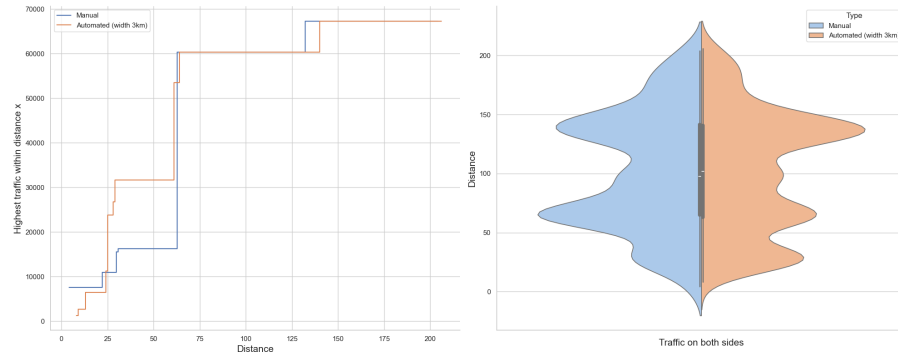


Figure 16: Busiest traffic from the startpoint (51.0826,-2.22697) (to the left) and violin plot (to the right) with strip width 3 km.

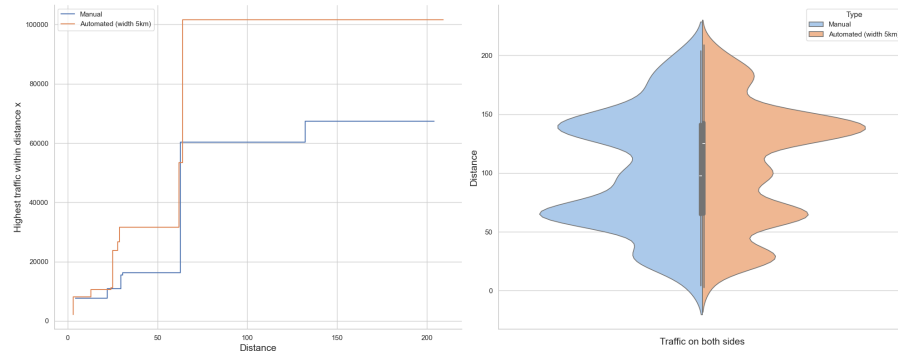


Figure 17: Busiest traffic from the startpoint (51.0826,-2.22697) (to the left) and violin plot (to the right) with strip width 5 km.

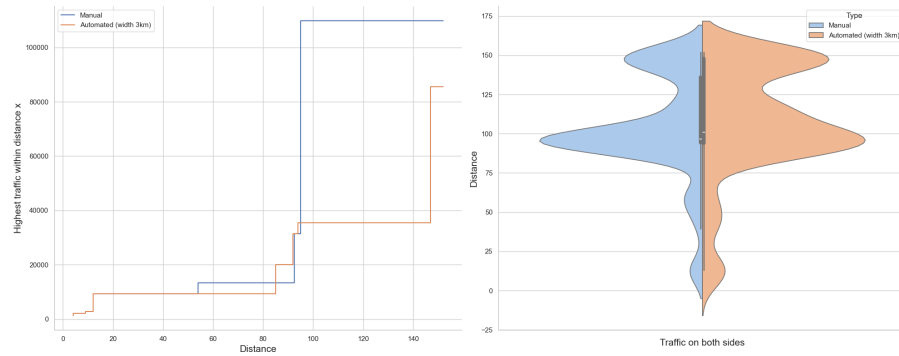


Figure 18: Busiest traffic from the startpoint (52.1577699,-3.5958771) (to the left) and violin plot (to the right) with strip width 3 km.

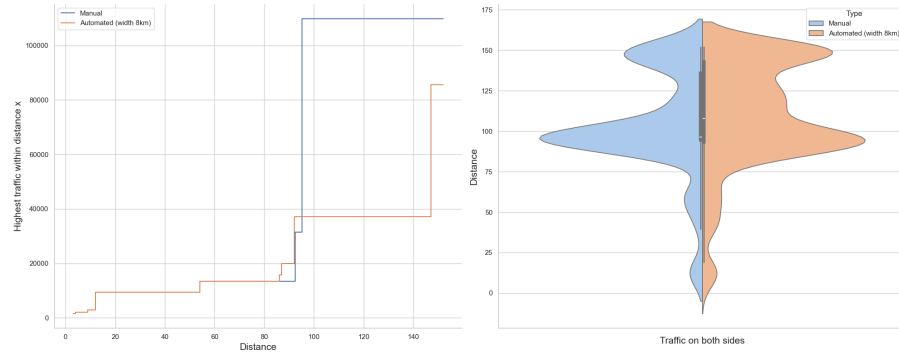


Figure 19: Busiest traffic from the startpoint (52.1577699,-3.5958771) (to the left) and violin plot (to the right) with strip width 8 km.

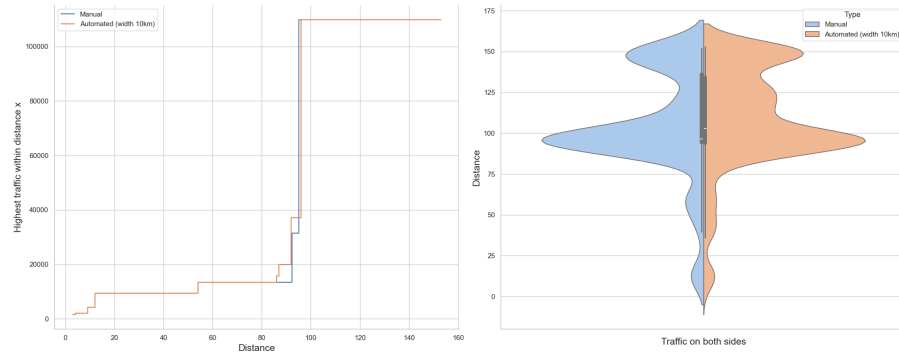


Figure 20: Busiest traffic from the startpoint (52.1577699,-3.5958771) (to the left) and violin plot (to the right) with strip width 10 km.

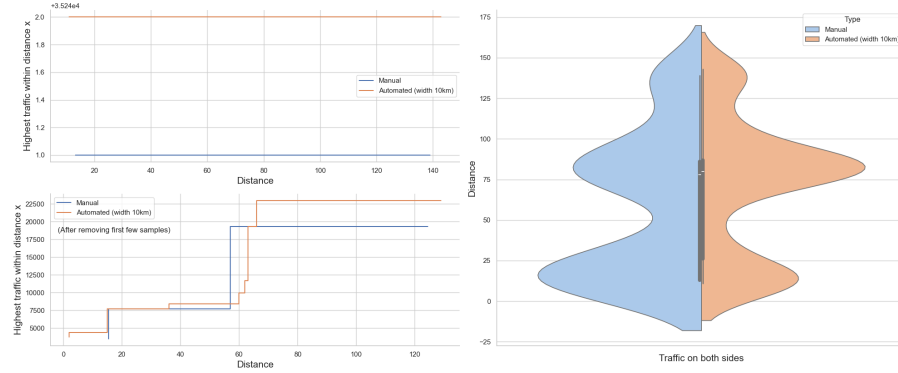


Figure 21: To the left, the busiest traffic from the startpoint (55.606053,-4.0506979) and to the right, the violin plot with a strip width 10 km. On the left side, the top figure indicates that the busiest traffic was captured within the first 10 km distance by the automated method, but this was a false count point that shouldn't have been included. Inaccuracies like this are expected from time to time. However, the bottom figure shows a much better fit after removing the closest few traffic points.

Exponential LPP. They also found long-range positive correlations in road density, spanning across hundreds of kilometers. Without traffic counts we do not have a meaningful way to define road density in our model as we assume that cars start their journeys from each point of  $\mathbb{Z}^2$ . As for busier roads in our model, we would guess negative correlations rather than positive ones. This might be due to a busy road representing a stretch of preferable environment, which might attract other traffic hence make busy roads less likely nearby; this could be a question for future mathematical research.

## Acknowledgements

The authors thank Riddhipratim Basu for useful discussions throughout the project. We also thank for discussions with Yuri Bakhtin, Atal Bhargava, Sanchari Goswami and Aquib Molla in the early phases of this work, and Janosch Ortmann later. MB was partially supported by the EPSRC EP/W032112/1 Standard Grant of the UK. SB was supported by scholarship from National Board for Higher Mathematics (NBHM) (ref no: 0203/13(32)/2021-R&D-II/13158). DH was partially supported by NSF grant DMS-2054559. KD was supported by the Prime Minister's Research Fellowship PM/PMRF-22-18686.03 (PMRF ID: 0202965) from the Ministry of Education of India. This project was initiated at the International Centre for Theoretical Sciences (ICTS), Bengaluru, India during the program "First-passage percolation and related models" in July 2022 (code: ICTS/fpp-2022/7), the authors thank ICTS for the hospitality.

All data used in this study are publicly available at [9] and [18].

## References

- [1] Antonio Auffinger, Michael Damron, and Jack Hanson. *50 years of first-passage percolation*. Vol. 68. University Lecture Series. American Mathematical Society, Providence, RI, 2017, pp. v+161.

- ISBN: 978-1-4704-4183-8. DOI: [10.1090/ulect/068](https://doi.org/10.1090/ulect/068). URL: <https://doi.org/10.1090/ulect/068>.
- [2] Márton Balázs, Riddhipratim Basu, and Sudeshna Bhattacharjee. “Geodesic trees in last passage percolation and some related problems”. Preprint, arXiv:2308.07312. 2023.
  - [3] Marc Barthélemy and Alessandro Flammini. “Modeling Urban Street Patterns”. In: *Phys. Rev. Lett.* 100 (13 Apr. 2008), p. 138702. DOI: [10.1103/PhysRevLett.100.138702](https://doi.org/10.1103/PhysRevLett.100.138702). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.100.138702>.
  - [4] Riddhipratim Basu, Christopher Hoffman, and Allan Sly. “Nonexistence of Bigeodesics in Planar Exponential Last Passage Percolation”. In: *Communications in Mathematical Physics* 389.1 (Jan. 2022), pp. 1–30. ISSN: 1432-0916. DOI: [10.1007/s00220-021-04246-0](https://doi.org/10.1007/s00220-021-04246-0). URL: <https://doi.org/10.1007/s00220-021-04246-0>.
  - [5] I. Corwin. “Kardar-Parisi-Zhang universality”. In: *Notices Amer. Math. Soc.* 63.3 (2016), pp. 230–239. ISSN: 0002-9920,1088-9477. DOI: [10.1090/noti1334](https://doi.org/10.1090/noti1334). URL: <https://doi.org/10.1090/noti1334>.
  - [6] David Coupier. “Multiple geodesics with the same direction”. In: *Electronic Communications in Probability* 16.none (2011), pp. 517–527. DOI: [10.1214/ECP.v16-1656](https://doi.org/10.1214/ECP.v16-1656). URL: <https://doi.org/10.1214/ECP.v16-1656>.
  - [7] David Michael H. *LPP Traffic Model*. GitHub repository. Year the repository was last updated, e.g., 2024. URL: <https://github.com/DavidMichaelH/lpp-traffic-model>.
  - [8] David Michael H. *Statistical Mechanics Models*. GitHub repository. Year the repository was last updated, e.g., 2024. URL: <https://github.com/DavidMichaelH/statistical-mechanics-models>.
  - [9] UK Department for Transport. *Road traffic statistics*. URL: <https://roadtraffic.dft.gov.uk/>.
  - [10] ESRI. *World Topographic Map*. [Tile layer]. Sources: Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community. Retrieved from <https://www.arcgis.com/home/item.html?id=6e850093c837475e8c23d905ac43b7d0>. 2024.
  - [11] P. L. Ferrari and H. Spohn. “Random growth models”. In: *The Oxford handbook of random matrix theory*. Oxford Univ. Press, Oxford, 2011, pp. 782–801. ISBN: 978-0-19-957400-1.
  - [12] Pablo A. Ferrari and Leandro P. R. Pimentel. “Competition Interfaces and Second Class Particles”. In: *The Annals of Probability* 33.4 (2005), pp. 1235–1254. ISSN: 00911798. URL: <http://www.jstor.org/stable/3481728> (visited on 03/04/2024).
  - [13] J. M. Hammersley and D. J. A. Welsh. “First-Passage Percolation, Subadditive Processes, Stochastic Networks, and Generalized Renewal Theory”. In: *Bernoulli 1713 Bayes 1763 Laplace 1813: Anniversary Volume*. Ed. by Jerzy Neyman and Lucien M. Le Cam. Berlin, Heidelberg: Springer Berlin Heidelberg, 1965, pp. 61–110. ISBN: 978-3-642-99884-3. DOI: [10.1007/978-3-642-99884-3\\_7](https://doi.org/10.1007/978-3-642-99884-3_7). URL: [https://doi.org/10.1007/978-3-642-99884-3\\_7](https://doi.org/10.1007/978-3-642-99884-3_7).
  - [14] Günter Last and Mathew Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2017.

- [15] Carlos Molinero and Alberto Hernando. *A model for the generation of road networks*. 2020. arXiv: [2001.08180](#) [[physics.soc-ph](#)].
- [16] Jeremy Quastel. “Introduction to KPZ”. In: *Current developments in mathematics, 2011*. Int. Press, Somerville, MA, 2012, pp. 125–194. ISBN: 978-1-57146-239-8.
- [17] A. P. Solon et al. “Optimal paths on the road network as directed polymers”. In: *Phys. Rev. E* 96 (5 Nov. 2017), p. 050301. DOI: [10.1103/PhysRevE.96.050301](#). URL: [https://link.aps.org/doi/10.1103/PhysRevE.96.050301](#).
- [18] U.S. Geological Survey. *EarthExplorer*. URL: [https://earthexplorer.usgs.gov/](#).

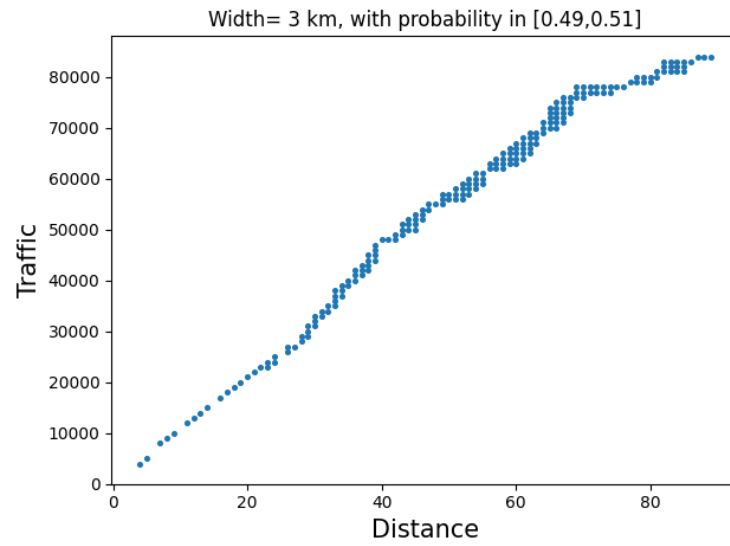


Figure 22: Statistical plot for the South-East region of the UK

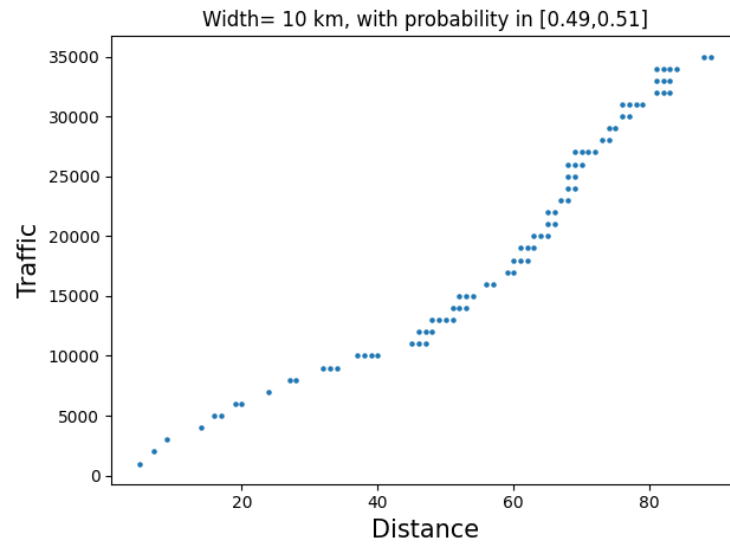


Figure 23: Statistical plot for the West region of the UK