

# Random road networks: the quadtree model

David Eisenstat\*

## Abstract

What does a typical road network look like? Existing generative models tend to focus on one aspect to the exclusion of others. We introduce the general-purpose *quadtree model* and analyze its shortest paths and maximum flow.

## 1 Introduction

Many optimization problems naturally feature a road network. Computing driving directions can be viewed as a shortest-path problem, and at a higher level, a typical vehicle routing problem is to schedule a fleet of capacity-limited trucks making deliveries from a depot to several customers. Algorithms that solve these kinds of problems on general graphs are often impractical, leading algorithm designers to consider the special structure of road networks. How can we characterize this structure?

Viewing the US road network as a metric space, Bast, Funke, and Matijevic found a small set of *transit nodes* that covers almost all of the shortest paths [3]. The existence of such a set underlies their point-to-point shortest-paths algorithm and was later formalized by Abraham, Fiat, Goldberg, and Werneck as *highway dimension* in order to explain the performance of several competing algorithms [1]. Kalapala, Sanwalani, Clauset, and Moore examined several national road networks and discovered evidence of scale invariance [14]. Aldous has explored the axiomatic consequences of scale invariance in road networks [2]. Masucci, Smith, Crooks, and Batty studied the London road network, finding, among other things, that the length of a shortest path is usually competitive with the distance between its endpoints and that most pairs of points are connected by a path with few turns [16].

Topologically, road networks have few crossings involving roads at different elevations, leading many researchers, including Masucci et al., to view road networks as planar graphs. Eppstein and Goodrich, however, dispute both this characterization and a narrower one that excludes highways, citing evidence that in the US road network, crossings between local roads number on the order of  $\sqrt{n}$ , where  $n$  is the number of intersections [8]. They introduce as an alternative the notion of *multiscale-dispersed* graphs. Algorithmically, one key property of planar graphs [15] and multiscale-dispersed graphs [8] is that they have *separators* of size  $O(\sqrt{n})$ , whose removal leaves two subgraphs of approximately the same size disconnected from one another. Fred-

erickson was the first of many to design separator-exploiting graph algorithms [11]. Other properties exclusive to planar graphs are exploited by algorithms such as the  $O(n \log n)$  maximum *st*-flow algorithm due to Borradaile and Klein [5], which can be used for large-scale empirical experiments with the quadtree model.

Despite the role of road networks in practical optimization, there have been few attempts to synthesize these properties into a generative model. Ideally, such a road network model would be realistic enough to yield new insights, yet mathematically tractable so as to be useful in validating new specialized algorithms. Motivated by the success of generative models in the study of social networks [17], we introduce a new model of random road networks based on quadtrees.

### 1.1 Existing models

One natural proposal is the uniform random planar graph. It is well studied and has efficient sampling algorithms [12, 18], but it has more high-degree nodes than real road networks [16] and does not come with a geometric embedding. Gerke, Schlatter, Steger, and Taraz analyze an incremental random planar graph process where uniform random edges that do not complete a forbidden subgraph are added to an initially empty graph [13]. It has the same drawbacks, and a variant where straight edges are added between nodes chosen uniformly at random in the unit square is considered and rejected by Masucci et al. as being too difficult to navigate [16].

Abraham et al. describe an incremental road network construction algorithm that takes as input a sequence of points in a metric space [1]. Sets of input points  $C_0 \supseteq C_1 \supseteq \dots$  are maintained in such a way that  $C_i$  is a maximal set of points at distance greater than  $2^i$  from one another. For all  $i$ , roads are added between points  $p, q \in C_i$  with  $\text{dist}(p, q) \leq 6 \cdot 2^i$ . For all  $p \in C_{i-1} - C_i$ , a road is also added from  $p$  to the closest point in  $C_i$ . In essence, this algorithm creates local road networks at different scales, together with connections between the scales. It is introduced as motivation for the authors' hypothesis that real road networks have small highway dimension and is not intended to capture the topology of a road network.

Kalapala et al. introduce a simple model of fractal road networks in order to study scale invariance [14]. The quadtree model differs from this model in using a random tree whose distribution is controlled by a sprawl parameter, instead of a tree complete to a specified depth. Visually, the road networks produced by the model of Kalapala et al. are too uniform to resemble real road networks.

\*Department of Computer Science, Brown University. Supported in part by NSF grant CCF-0964037. [eisenstatdavid@gmail.com](mailto:eisenstatdavid@gmail.com)

Masucci et al. propose the Growing Random Planar Graph (GRPG) model, which is based on their statistical analysis of the London road network [16]. The GRPG model starts with a single vertex and constructs a planar-embedded tree incrementally. At each step, a random parent vertex is chosen along with a random offset, and a child is placed offset from the parent as long as the new edge does not cross an existing edge. Short edges are added to obtain the finished product. This model seems difficult to analyze formally in light of its brute-force approach to ensuring that the graph is planar, and its creators make no attempt to do so.

## 1.2 The quadtree model

The quadtree model has two parameters:  $n$ , the number of intersections, and  $r$ , which controls the amount of sprawl. Imagine a square tract delimited by roads, to which  $n$  people immigrate in succession. Each new arrival chooses an existing square at random, builds two road segments bisecting it horizontally and vertically, and settles at the newly-created intersection. This intersection is at the corner of four new delimited squares. If people want to be far from other people ( $r < 1$ ), then they prefer to settle larger squares. Conversely, if they want to be close ( $r > 1$ ), then they prefer smaller. Figure 1 shows the effects of different preferences. Figure 2 presents instances of the quadtree model and the GRPG model side by side with a real road network.

Road networks generated by the quadtree model have several desirable features. They are embedded in the plane, self-similar, and for some choices of  $r$ , they display realistic variations in density. Between any two points there exists a hierarchically organized path whose length is within a constant factor of their Euclidean distance and that with high probability when  $r \leq 1$  makes  $O(\log n)$  turns (Theorem 1; Section 4). This conforms with the observation of Masucci et al. that London is efficiently navigable in both senses [16]. Unlike the GRPG model, the quadtree model is simple and avoids an unrealistic abundance of acute-angled intersections. Although the quadtree model does not specify speeds or capacities for roads, the technique of Kalapala et al. can be used to assign both based on the depth of each road in the tree [14]. Different subdivision patterns can be used to avoid the monotony of square lots; we anticipate that our techniques will still apply with straightforward modifications.

## 1.3 Maximum flow on road networks

The capacity of road networks has received far less attention than shortest paths, yet for urban planners who must manage rush hour traffic and the possibility of natural disasters, it is a very important quantity. Imagine that a large number of drivers wish to travel from the west edge of the road network to the east: how does sprawl affect the rate at which they succeed?

For simplicity, we assume that each road has unit capacity. We anticipate that our techniques extend straightforwardly to more realistic settings where low-level roads have less capacity. We measure the capacity of a road network by the

maximum value of an east–west flow, which by our assumption equals the maximum number of edge-disjoint east–west paths. By the max-flow min-cut theorem [7, 10], the capacity of a road network is also equal to the minimum total capacity of roads that when removed, disconnect the east from the west.

We obtain a partial characterization of maximum flow in the quadtree model by analyzing the capacity of a random vertical cut for upper bounds and by embedding a partial grid for lower bounds. When  $r < 1$ , the road network asymptotically resembles a grid, with maximum flow on the order of  $\sqrt{n}$  in expectation (Theorem 2; Section 6). When  $r > 1$ , the squares nest very deeply and the maximum flow is constant in expectation (Theorem 4; Section 8). These results are proved by analyzing a continuous-time variant of the quadtree process with exponential random variables. When  $r = 1$ , we have been able to prove only that a lower bound is  $\Omega(n^{0.086\dots})$  and an upper bound is  $O(n^{1/3})$ , confirming the existence of a phase transition (Theorem 3; Section 7). The proofs of both bounds involve a generating function. Experimentally, the true exponent lies somewhere between the two, but its exact identity is unknown. (Note that the quadtree in this case is uniform random.)

Previously, Eppstein, Goodrich, and Trott studied the expected number of roads intersected by a random line, both theoretically (for multiscale-dispersed graphs) and experimentally [9]. They proved for arbitrary multiscale-dispersed graphs that this number is  $O(\sqrt{n})$  but found that for real road networks it is significantly smaller. The same number for the quadtree model with  $r = 1$  is  $\Theta(n^{1/3})$ , which together with visual evidence, suggests that  $r \triangleq 1 - \epsilon$  may be a realistic choice.

## 2 Formal model description and summary of results

Let  $\Sigma \triangleq \{N_W, N_E, S_W, S_E\}$  be an alphabet on four symbols. We view strings as nodes in an infinite quadtree rooted at the empty string  $\lambda$ , where the children of each node  $w \in \Sigma^*$  are  $wa$  for all  $a \in \Sigma$ . In turn, we view these nodes as squares. Starting with the unit square  $\lambda$ , we cut each square horizontally and vertically to obtain its four children, as shown in Figure 3. Each square  $w \in \Sigma^*$  so obtained is axis-aligned and has side length  $2^{-|w|}$ .

Let  $r \in (0, \infty)$  be a parameter. We specify a random growing road network as a sequence of sets of squares  $\text{Leaves}_n$ . Initially, there is only the unit square.

$$\text{Leaves}_0 \triangleq \{\lambda\}.$$

For all integers  $n \geq 0$ , we choose a random square  $\text{Split}_n \in \text{Leaves}_n$  to subdivide. When  $r < 1$ , larger squares are more likely to be chosen. When  $r > 1$ , smaller squares are more likely. When  $r = 1$ , the choice is uniform. The distribution

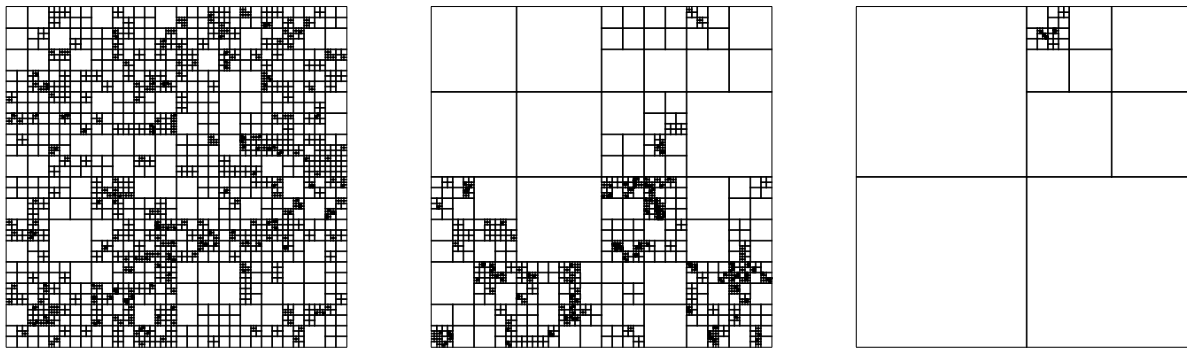


Figure 1: the quadtree model. Left:  $r = 3/4$  (large-square preference). Middle:  $r = 1$  (neutral). Right:  $r = 5/4$  (small-square preference).

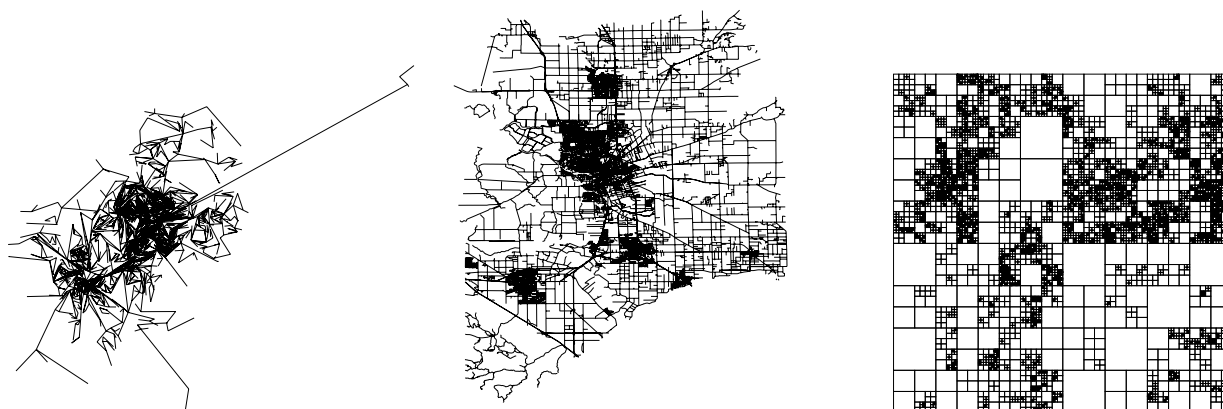


Figure 2: Left: the Growing Random Planar Graph model [16]. Middle: the roads of San Joaquin County, California, USA [19]. Right: the quadtree model with  $r = 15/16$ .

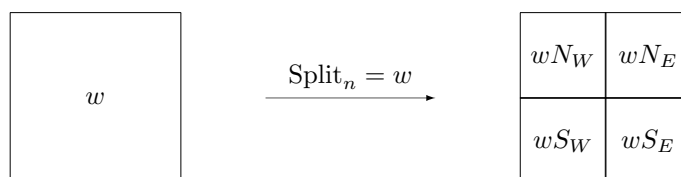


Figure 3: Splitting square  $w$ .

of  $\text{Split}_n$  is given by

$$\forall v \in \text{Leaves}_n,$$

$$\Pr(\text{Split}_n = v \mid \text{Leaves}_n) \triangleq \frac{r^{|v|}}{\sum_{w \in \text{Leaves}_n} r^{|w|}}.$$

To obtain the next set of squares, we replace  $\text{Split}_n$  by its children.

$$\text{Leaves}_{n+1} \triangleq (\text{Leaves}_n - \{\text{Split}_n\}) \cup \{\text{Split}_n a : a \in \Sigma\}.$$

We call this model the *quadtrees model* for random road networks. In the terminology of Bertoin [4, 1.1], it is a conservative fragmentation chain with index of self-similarity  $\alpha \triangleq \log_4(1/r)$ .

For all integers  $n \geq 0$ , let  $G_n$  be the graph of the square subdivision corresponding to  $\text{Leaves}_n$ . Let  $\text{Left}_n$  be the set of vertices with  $x = 0$  and  $\text{Right}_n$  be the set of vertices with  $x = 1$ . Define  $\text{MaxFlow}_n$  to be the maximum size of a set of pairwise edge-disjoint paths in  $G_n$  from vertices in  $\text{Left}_n$  to vertices in  $\text{Right}_n$ , that is, the maximum flow in  $G_n$  from  $\text{Left}_n$  to  $\text{Right}_n$  assuming unit-capacity edges.

**Theorem 1.** *Let  $w, x \in \text{Leaves}_n$  be squares, let  $p$  be a point on the boundary of  $w$ , and let  $q$  be a point on the boundary of  $x$ . Define  $v$  to be the least common ancestor of  $w$  and  $x$ . There exists a path in  $G_n$  from  $p$  to  $q$  that changes direction at most  $3(d_w + d_x) + 6$  times and has length at most  $2 \text{dist}_1(p, q)$ , where  $d_w, d_x$  are the depths of  $w, x$  in the subtree rooted at  $v$ .*

**Theorem 2.** *If  $r < 1$ , then*

$$\frac{n^{1/2}}{\exp O(\sqrt{\ln n})} \leq \mathbf{E}[\text{MaxFlow}_n] \leq O(n^{1/2}).$$

The lower bound for  $r < 1$  is very close to the upper bound; in this case, for all  $\epsilon > 0$ , we have  $\mathbf{E}[\text{MaxFlow}_n] = \Omega(n^{1/2-\epsilon})$ .

**Theorem 3.** *If  $r = 1$ , then*

$$\Omega(n^{0.086\dots}) \leq \mathbf{E}[\text{MaxFlow}_n] \leq O(n^{1/3}).$$

**Theorem 4.** *If  $r > 1$ , then  $\mathbf{E}[\text{MaxFlow}_n] = \Theta(1)$ .*

### 3 Continuous-time reformulation

When  $r \neq 1$ , it is convenient to work with a continuous-time variant of the quadtree process. In switching to continuous time, we lose fine control over the number of intersections but gain independence between disjoint subtrees. The former is not an issue for us.

For all squares  $w \in \Sigma^*$ , let  $\text{Lifespan}(w)$  be an independent exponential random variable with rate  $r^{|w|}$ . Define

$$\begin{aligned} \text{Birth}(w) &\triangleq \sum_{v \text{ is a proper ancestor of } w} \text{Lifespan}(v) \\ \text{Death}(w) &\triangleq \sum_{v \text{ is an ancestor of } w} \text{Lifespan}(v) \\ &= \text{Birth}(w) + \text{Lifespan}(w). \end{aligned}$$

For all times  $t \in [0, \infty)$ , let

$$\text{Alive}(t) \triangleq \{w : \text{Birth}(w) \leq t < \text{Death}(w)\}.$$

For all integers  $n \geq 0$ , let

$$T(n) \triangleq \inf\{t : \#\text{Alive}(t) = 3n + 1\}.$$

We assume, as is the case with probability 1, that  $T$  is everywhere defined. The following proposition can be proved inductively by coupling the discrete process with the continuous process.

**Proposition 5.** *For all integers  $n \geq 0$ , the random sets  $\text{Leaves}_n$  and  $\text{Alive}(T(n))$  are identically distributed.*

In the proofs that follow, we assume that  $\text{Leaves}_n = \text{Alive}(T(n))$ .

### 4 Navigation (Proof of Theorem 1)

The key to short paths with few turns is using the hierarchical organization of the road network. In essence, the proofs in this section work by replacing squares on a path in the quadtree with pieces of their boundaries (and possibly their siblings'). Note that when  $r \leq 1$ , with high probability the quadtree has height  $O(\log n)$ .

**Lemma 6.** *Let  $w \in \text{Leaves}_n$  be a square. Let  $p$  be a point on the boundary of  $w$  and let  $v$  be an ancestor of  $w$ . Define  $q$  to be the unique point on the boundary of  $v$  that is due north (respectively, west, east, south) of  $p$ . There exists a path in  $G_n$  from  $p$  to  $q$  that changes direction at most  $3d + 2$  times and has length at most  $2 \text{dist}_1(p, q)$ , where  $d$  is the depth of  $w$  in the subtree rooted at  $v$ .*

*Proof.* (See Figure 4.) We induct on  $d$ . If  $d = 0$ , then  $w = v$ , and  $p$  is on the boundary of  $v$ . Take the shortest path on the boundary of  $v$ .

If  $d > 0$ , then let  $a \in \Sigma$  be the unique symbol such that  $v' \triangleq va$  is an ancestor of  $w$ . Define  $q'$  to be the unique point on the boundary of  $v'$  that is due north of  $p$ . By the inductive hypothesis, there exists a path in  $G_n$  from  $p$  to  $q'$  that changes direction at most  $3d - 1$  times and has length at most  $2 \text{dist}_1(p, q')$ . If  $a \in \{N_W, N_E\}$ , then  $q' = q$ , and no more travel is required. Otherwise, from  $q'$ , we change directions at most 3 more times in taking the shortest path to  $q$  on the boundary of  $vN_W$  (if  $a = S_W$ ) or  $vN_E$  (if  $a = S_E$ ). The total path length is at most  $2 \text{dist}_1(p, q') + 2 \text{dist}_1(q', q) = 2 \text{dist}_1(p, q)$ , since  $q'$  is on the line segment from  $p$  to  $q$ .  $\square$

*Proof of Theorem 1.* Recall that  $w, x \in \text{Leaves}_n$  are squares whose least common ancestor is  $v$ . The depths of  $w, x$  in the subtree rooted at  $v$  are  $d_w, d_x$ . The point of departure  $p$  lies on the boundary of  $w$ , and the point of arrival  $q$  lies on the boundary of  $x$ . Our goal is to find a path from  $p$  to  $q$  that changes direction at most  $3(d_w + d_x) + 6$  times and has length at most  $2 \text{dist}_1(p, q)$ .

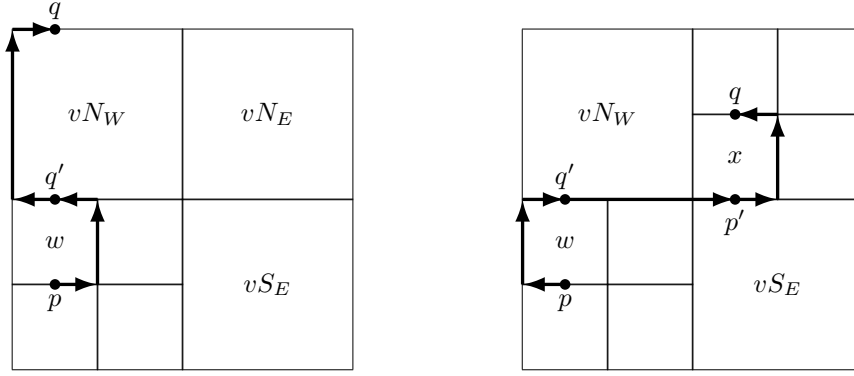


Figure 4: Left: Lemma 6 with  $w \triangleq vS_WN_W$  and  $p$  as shown. Right: Theorem 1 with  $w \triangleq vS_WN_W$  and  $x \triangleq vN_ES_W$  and  $p, q$  as shown.

If  $v = w = x$ , then take the shortest path on the boundary of  $v$ . Otherwise,  $w, x \neq v$ , since no square in  $\text{Leaves}_n$  has a proper ancestor in  $\text{Leaves}_n$ . Assume by rotational symmetry that  $p$  is on or below the horizontal bisector of  $v$  and that  $q$  is on or above the bisector. Let  $q'$  be the unique point on the bisector due north of  $p$  and let  $p'$  be the unique point on the bisector due south of  $q$ . Apply Lemma 6 to obtain paths from  $p$  to  $q'$  and from  $p'$  to  $q$ . Join these paths with a segment from  $q'$  to  $p'$  to obtain a path that changes directions at most  $(3d_w + 2) + 2 + (3d_x + 2)$  times and has length at most  $2 \text{dist}_1(p, q') + \text{dist}_1(q', p') + 2 \text{dist}_1(p', q) \leq 2 \text{dist}_1(p, q)$ .  $\square$

## 5 Maximum flow

To bound  $\text{MaxFlow}_n$  below, we look to embed a regular grid in  $G_n$ . When  $r = 1$ , it is advantageous to consider an incomplete grid, with *holes*. To bound  $\text{MaxFlow}_n$  above, we invoke the max-flow min-cut theorem and look for a vertical cut with low capacity.

For all integers  $d \geq 0$ , let

$$\text{Holes}_n(d) \triangleq \{w : w \in \text{Leaves}_n, |w| < d\}$$

be the set of *holes* at depth  $d$ , that is, the set of squares that obstruct the embedding of a  $2^d \times 2^d$  regular grid by having height larger than  $2^{-d}$ .

**Lemma 7.** For all integers  $n, d \geq 0$ ,

$$\text{MaxFlow}_n \geq 2^d + 1 - \sum_{w \in \text{Holes}_n(d)} (2^{d-|w|} - 1),$$

*Proof.* (See Figure 5.) We attempt to send  $2^d + 1$  units of flow horizontally. Each square  $w \in \text{Holes}_n(d)$  may block up to  $2^{d-|w|} - 1$  units.  $\square$

**Lemma 8.** For all integers  $n \geq 0$ ,

$$\text{MaxFlow}_n \leq 1 + \sum_{w \in \text{Leaves}_n} 2^{-|w|}.$$

*Proof.* By the max-flow min-cut theorem [7, 10], every  $\text{Left}_n\text{-Right}_n$  cut in the graph  $G_n$  has capacity at least  $\text{MaxFlow}_n$ . Choose  $U \in (0, 1) - \mathbf{Q}$  uniformly at random

and form an  $\text{Left}_n\text{-Right}_n$  cut by partitioning the vertices into those with  $x < U$  and those with  $x > U$  (all vertices have rational coordinates). The capacity of this cut is equal to the number of squares intersected by the line  $x = U$ , plus one. Each square  $w \in \text{Leaves}_n$  is intersected with probability equal to its width  $2^{-|w|}$ , and the conclusion follows by an averaging argument.  $\square$

Lemma 8 leads to an upper bound on  $\text{MaxFlow}_n$  that is tight when  $G_n$  is a regular grid.

**Lemma 9.** For all integers  $n \geq 0$ ,

$$\text{MaxFlow}_n \leq (3n + 1)^{1/2} + 1.$$

*Proof.* By Lemma 8, it suffices to obtain an upper bound on the total square width plus one, which can be accomplished via the Cauchy–Schwarz inequality. Let  $\{w_1, \dots, w_{3n+1}\} \triangleq \text{Leaves}_n$  and consider the  $(3n + 1)$ -dimensional vector  $\mathbf{x}$  where  $\mathbf{x}_k \triangleq 2^{-|w_k|}$ . Denoting by  $\mathbf{1}$  the all-ones vector of the same dimension,

$$\sum_{w \in \text{Leaves}_n} 2^{-|w|} = \mathbf{1} \cdot \mathbf{x} \leq \|\mathbf{1}\|_2 \|\mathbf{x}\|_2$$

by Cauchy–Schwarz

$$\|\mathbf{1}\|_2 = (3n + 1)^{1/2}$$

$$\|\mathbf{x}\|_2 = \left( \sum_{w \in \text{Leaves}_n} 4^{-|w|} \right)^{1/2} = 1$$

$\square$

since  $4^{-|w|}$  is the area of  $w$ .

## 6 Maximum flow when $r < 1$ (Proof of Theorem 2)

Given that large squares are much more likely to be split than small squares, it is probable that all square sizes lie in a narrow interval, in which case we can embed a relatively fine regular grid in  $G_n$ . In fact, for any time  $t$ , we can say with some confidence what the sizes of the squares in  $\text{Alive}(t)$  will be. Lemma 10 bounds the probability that a small square is born early.

Case	Lower bound	Upper bound	Theorem
$r < 1$	$\frac{n^{1/2}}{\exp O(\sqrt{\ln n})}$	$O(n^{1/2})$	2
$r = 1$	$\Omega(n^{0.086\dots})$	$O(n^{1/3})$	3
$r > 1$	$\Theta(1)$		4

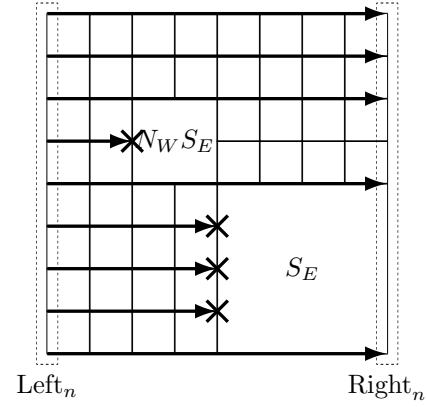


Figure 5: Left: the asymptotic behavior of  $\mathbf{E}[\text{MaxFlow}_n]$  in  $n$  for fixed  $r$ . Right: Lemma 7 with  $d \triangleq 3$ . Square  $N_W S_E$  blocks  $2^{3-2} - 1 = 1$  units of flow, and square  $S_E$  blocks  $2^{3-1} - 1 = 3$ .

**Lemma 10.** Assume that  $r < 1$ . For all squares  $w \in \Sigma^*$  and all integers  $b \in [0, |w|]$ ,

$$\Pr(\text{Birth}(w) < r^{-b}) \leq r^{\binom{|w|}{2} - b}.$$

*Proof.* If  $\text{Birth}(w) < r^{-b}$ , then for all proper ancestors  $v$  of  $w$ , it holds that  $\text{Lifespan}(v) < r^{-b}$ . As  $\text{Lifespan}(v)$  is exponential with rate  $r^{|v|}$ , we have  $\Pr(\text{Lifespan}(v) < r^{-b}) = 1 - \exp(-r^{|v|-b}) \leq \min\{r^{|v|-b}, 1\}$ . These events are independent, so we multiply their probabilities and obtain the stated bound.  $\square$

Lemma 11 bounds the probability that a large square dies late.

**Lemma 11.** Assume that  $r < 1$ . For all squares  $w \in \Sigma^*$  and all  $p \in (0, 1]$ ,

$$\Pr(\text{Death}(w) > \mathbf{E}[\text{Death}(w)] \ln(1/p)) \leq (|w| + 1)p.$$

*Proof.* If  $\text{Death}(w) > \mathbf{E}[\text{Death}(w)] \ln(1/p)$ , then there exists an ancestor  $v$  of  $w$  such that  $\text{Lifespan}(v) > \mathbf{E}[\text{Lifespan}(v)] \ln(1/p)$ . For all  $v$ , the probability of the latter is  $p$ , and  $w$  has  $|w| + 1$  ancestors.  $\square$

Lemma 12 establishes that with probability at least  $1/2$ , the last square at depth  $c - O(\sqrt{c})$  dies before the first square at depth  $c$  is born.

**Lemma 12.** For all  $r \in (0, 1)$ , there exist constants  $c_0, \alpha$  such that for all integers  $c \geq c_0$ , with probability at least  $1/2$ ,

$$\forall v \in \Sigma^a, w \in \Sigma^c, \quad \text{Death}(v) \leq \text{Birth}(w).$$

where  $a \triangleq c - \lfloor \alpha\sqrt{c} \rfloor$ .

*Proof.* We exhibit a relatively late time ( $r^{-b}$ ) at which with probability at least  $3/4$ , no square at depth  $c$  has been born. We then choose  $a$  so that with probability at least  $3/4$ , every square at depth  $a$  has been born. More details are in the appendix.  $\square$

We now relate these results to  $\text{MaxFlow}_n$ .

*Proof of Theorem 2.* Lemma 9 gives the upper bound  $\mathbf{E}[\text{MaxFlow}_n] = O(n^{1/2})$ . By an averaging argument, there exists a square  $w \in \text{Leaves}_n$  with area at most  $1/(3n + 1)$ . It thus holds that  $|w| \geq c \triangleq \lceil \log_4(3n + 1) \rceil$ . Defining  $d \triangleq c - \lfloor \alpha\sqrt{c} \rfloor + 1$ , where the constant  $\alpha$  comes from Lemma 12, it holds with probability at least  $1/2$  that  $\text{Holes}_n(d) = \{v : v \in \text{Leaves}_n, |v| < d\} = \emptyset$ . We apply Lemma 7 and obtain the lower bound

$$\square \quad \mathbf{E}[\text{MaxFlow}_n] \geq \frac{2^d + 1}{2} = \frac{n^{1/2}}{\exp O(\sqrt{\ln n})}.$$

## 7 Maximum flow when $r = 1$ (Proof of Theorem 3)

For all integers  $n \geq 0$ , define an ordinary generating function

$$F_n(t) \triangleq \sum_{w \in \text{Leaves}_n} t^{|w|}.$$

When  $r = 1$ , there is a nice closed form for  $\mathbf{E}[F_n(t)]$ . This is helpful because using  $F_n$ , we can approximate from below the lower bound of Lemma 7 and express exactly the upper bound of Lemma 8.

**Lemma 13.** Assume that  $r = 1$ . For all integers  $n \geq 0$  and all  $t$ ,

$$\mathbf{E}[F_n(t)] = \prod_{k=0}^{n-1} \left(1 + \frac{4t-1}{3k+1}\right) \leq \exp\left((4t-1) \sum_{k=0}^{n-1} \frac{1}{3k+1}\right).$$

*Proof.* For all integers  $k \geq 0$ ,

$$\mathbf{E}[F_{k+1}(t) \mid \text{Leaves}_k] = \left(1 + \frac{4t-1}{3k+1}\right) F_k(t).$$

Since the set  $\text{Leaves}_k$  determines  $F_k(t)$ , we can condition on  $F_k(t)$  alone. The conclusion follows by induction and the inequality  $1 + u \leq \exp u$ .  $\square$

*Proof of Theorem 3.* The upper bound is obtained by writing the total square width in terms of  $F_n$ . By Lemma 8,

$$\text{MaxFlow}_n \leq 1 + \sum_{w \in \text{Leaves}_n} 2^{-|w|} = F_n(1/2) + 1.$$

By Lemma 13 and integrating,

$$\begin{aligned} \mathbf{E}[\text{MaxFlow}_n] &\leq 1 + \exp \sum_{k=0}^{n-1} \frac{1}{3k+1} \\ &\leq \exp\left(1 + \frac{\ln(3n-2)}{3}\right) + 1 = O(n^{1/3}). \end{aligned}$$

The lower bound is more involved. To some extent, we emulate the proof of Devroye of the upper bound on the height of a random binary tree [6]. Let  $c, d, t$  be defined by

$$\begin{aligned} c \ln(3c/2) - c + 1/3 &\triangleq 1/\ln(3n+1) \\ &\quad (c = 0.124 \dots - O(1/\ln n)) \\ d &\triangleq \lfloor c \ln(3n+1) \rfloor \\ t &\triangleq 3c/4. \end{aligned}$$

We have

$$\begin{aligned} \text{MaxFlow}_n &\geq 2^d + 1 - \sum_{w \in \text{Holes}_n(d)} (2^{d-|w|} - 1) \\ &\quad \text{by Lemma 7} \\ &> 2^d - \sum_{w \in \text{Holes}_n(d)} (1/t)^{d-|w|} \\ &\quad \text{since } 1/t \geq 2 \text{ and } \forall w \in \text{Holes}_n(d), |w| \leq d-1 \\ &\geq 2^d - t^{-d} \sum_{w \in \text{Leaves}_n} t^{|w|} \\ &\quad \text{since } \text{Holes}_n(d) \subseteq \text{Leaves}_n \text{ and } t > 0 \\ &= 2^d - t^{-d} F_n(t). \end{aligned}$$

Take expectations, use Lemma 13, integrate, and substitute  $c, d, t$  (see the appendix).  $\square$

## 8 Maximum flow when $r > 1$ (Proof of Theorem 4)

Any square born after infinitely many other squares never enters the tree. When  $r > 1$ , there exists with probability 1 a least time  $U \in [0, \infty)$  for which there exist infinitely many squares  $v \in \Sigma^*$  with  $\text{Birth}(v) \leq U$ . This effect is known as the explosion of Markov chains in continuous time. At any given depth, most of the squares are born after time  $U$ , so the total width of all squares that ever enter the tree is small.

How can we bound  $U$  above? For all squares  $v \in \Sigma^*$ , define

$$\text{Countdown}(v) \triangleq \sum_{j=0}^{\infty} \text{Lifespan}(vN_W^j)$$

to be the amount of time elapsed between the birth of  $v$  and the time at which all of the infinitely many north-west descendants of  $v$  have been born. Given that  $r > 1$ , the sum converges with probability 1, and  $U \leq \text{Birth}(v) + \text{Countdown}(v)$ .

**Proposition 14.** *For all squares  $w \in \Sigma^*$ , if there exists a square  $v \in \Sigma^*$  such that  $\text{Birth}(v) + \text{Countdown}(v) < \text{Birth}(w)$ , then for all integers  $n \geq 0$ , we have  $w \notin \text{Leaves}_n$ .*

For Proposition 14 to be useful,  $\text{Countdown}(v)$  must be small some of the time. The significance of the constant  $\ln(15/14)$  is that for all  $w \in \Sigma^d$ ,

$$\Pr(\text{Lifespan}(w) \leq r^{-d} \ln(15/14)) = 1/15.$$

**Lemma 15.** *Assume that  $r > 1$ . There exists an integer  $k \geq 1$  such that for all integers  $d \geq 0$  and all sets of squares  $V \subseteq \Sigma^d$  with  $|V| = 4k$ , it holds with probability at least  $15/16$  that there exists a square  $v \in V$  such that*

$$\text{Countdown}(v) \leq r^{-d} \ln(15/14).$$

*Proof.* See the appendix.  $\square$

For all integers  $d \geq 0$ , define

$$\text{Born}_d \triangleq \Sigma^d \cap \bigcup_{n=0}^{\infty} \text{Leaves}_n$$

to be the set of all squares at depth  $d$  that ever enter the tree. In expectation, each set  $\text{Born}_d$  is small.

**Lemma 16.** *Assuming that  $r > 1$ ,*

$$\mathbf{E}[\#\text{Born}_d] = O(d).$$

*Proof.* Rather than analyze  $\#\text{Born}_d$  directly, we define sequences of sets  $\text{Candidates}_d \supseteq \text{Born}_d$  and  $\text{First}_d \subseteq \text{Candidates}_d$ . The set  $\text{First}_d$  is chosen so that we can apply Lemma 15 with  $V \triangleq \text{First}_d$ ; the set  $\text{Candidates}_{d+1}$  contains those squares that can still enter the tree given the quickest countdown in  $\text{First}_d$ . We prove that for all  $d$ ,

$$\mathbf{E}[\#\text{Candidates}_{d+1} \mid \mathcal{F}_d] \leq 16k(d+1) + \#\text{Candidates}_d/2,$$

where the  $\sigma$ -algebra  $\mathcal{F}_d$  belongs to a filtration that tracks the set of observed lifespans, and  $k$  is the constant from Lemma 15. It follows that  $\mathbf{E}[\#\text{Candidates}_d] \leq 32kd$ .

The sequences of sets  $\text{Candidates}_d$  and  $\text{First}_d$  are defined inductively. Let

$$\text{Candidates}_0 \triangleq \{\lambda\}$$

consist of the root square and let  $\mathcal{F}_0$  be the trivial  $\sigma$ -algebra (condition on nothing). The set  $\text{First}_d$ , which is deterministic given  $\mathcal{F}_d$ , consists of the  $4k$  squares  $v \in \text{Candidates}_d$  with the least values of  $\text{Birth}(v)$  for which  $\text{Countdown}(v)$  is

independent of  $\mathcal{F}_d$ . (If there are fewer than  $4k$  such squares, take all of them.) Let

$$\begin{aligned} \text{Parents}_d &\triangleq \\ &\{w : w \in \text{Candidates}_d, \\ &\quad \forall v \in \text{First}_d, \text{Death}(w) \leq \text{Birth}(v) + \text{Countdown}(v)\} \end{aligned}$$

be a set of squares whose children can still enter the tree. Finally, define

$$\text{Candidates}_{d+1} \triangleq \{wa : w \in \text{Parents}_d, a \in \Sigma\}$$

and let  $\mathcal{F}_{d+1}$  be the  $\sigma$ -algebra generated by  $\mathcal{F}_d$ , the random variables  $\text{Lifespan}(vN_W^j)$  for  $v \in \text{First}_d$  and  $j \geq 0$ , and the random variables  $\text{Lifespan}(w)$  for  $w \in \Sigma^d$ .

By construction and Proposition 14, it holds that  $\text{Born}_d \subseteq \text{Candidates}_d$ . Of the squares in  $\text{Candidates}_d$ , at most  $4kd$  have countdowns that are not independent of  $\mathcal{F}_d$ , one for each square in  $\text{First}_0 \cup \dots \cup \text{First}_{d-1}$ . At most  $4k$  belong to  $\text{First}_d$ . As for the remainder, with probability at least  $15/16$  there exists by Lemma 15 a square  $v \in \text{First}_d$  such that  $\text{Countdown}(v) \leq r^{-d} \ln(15/14)$ . For all remaining squares  $w$ ,

$$\begin{aligned} &\Pr(w \in \text{Parents}_d \mid \mathcal{F}_d) \\ &\leq \Pr(\text{Birth}(w) + \text{Lifespan}(w) \\ &\quad \leq \text{Birth}(v) + \text{Countdown}(v) \mid \mathcal{F}_d) \\ &\leq \Pr(\text{Lifespan}(w) \leq \text{Countdown}(v) \mid \mathcal{F}_d) \\ &\quad \text{since } \text{Birth}(v) \leq \text{Birth}(w) \\ &\leq 1/15, \end{aligned}$$

for a contribution of  $\#\text{Candidates}_d/4$  to  $\mathbf{E}[\#\text{Candidates}_{d+1} \mid \mathcal{F}_d]$ . The event where  $v$  does not exist contributes the same amount.  $\square$

*Proof of Theorem 4.* The lower bound  $\text{MaxFlow}_n = \Omega(1)$  is trivial. As for the upper bound,

$$\begin{aligned} \text{MaxFlow}_n &= 1 + \sum_{w \in \text{Leaves}_n} 2^{-|w|} \\ &\quad \text{by Lemma 8} \\ &\leq 1 + \sum_{d=0}^{\infty} \sum_{w \in \text{Born}_d} 2^{-|w|} \\ &\quad \text{since } \text{Leaves}_n \subseteq \bigcup_{d=0}^{\infty} \text{Born}_d \\ \mathbf{E}[\text{MaxFlow}_n] &= 1 + \sum_{d=0}^{\infty} 2^{-d} O(d) = O(1) \\ &\quad \square \quad \text{by Lemma 16.} \end{aligned}$$

## 9 Acknowledgment

The author would like to thank his advisor Claire Mathieu for much guidance and encouragement and the anonymous reviewers for their suggestions.

## References

- [1] Ittai Abraham, Amos Fiat, Andrew V. Goldberg, and Renato F. Werneck. Highway dimension, shortest paths, and provably efficient algorithms. In *Proceedings of the Twenty-First Annual ACM–SIAM Symposium on Discrete Algorithms*, pages 782–793, 2010.
- [2] David Aldous. Personal communication, 2010.
- [3] Holger Bast, Stefan Funke, and Domagoj Matijević. TRANSIT: ultrafast shortest-path queries with linear-time preprocessing. In *Proceedings of the Ninth DIMACS Implementation Challenge: The Shortest Path Problem*, 2006.
- [4] Jean Bertoin. *Random fragmentation and coagulation processes*. Cambridge University Press, 2006.
- [5] Glencora Borradaile and Philip N. Klein. An  $O(n \log n)$  algorithm for maximum  $st$ -flow in a directed planar graph. *Journal of the ACM*, 56(2), 2009.
- [6] Luc Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33(3):489–498, 1986.
- [7] P. Elias, A. Feinstein, and C.E. Shannon. A note on the maximum flow through a network. *IEEE Transactions on Information Theory*, 2(4):117–119, 1956.
- [8] David Eppstein and Michael T. Goodrich. Studying (non-planar) road networks through an algorithmic lens. In *Proceedings of the Sixteenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2008.
- [9] David Eppstein, Michael T. Goodrich, and Lowell Trotter. Going off-road: transversal complexity in road networks. In *Proceedings of the Seventeenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009.
- [10] L.R. Ford, Jr. and D.R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8, 1956.
- [11] Greg N. Frederickson. Fast algorithms for shortest paths in planar graphs, with applications. *SIAM Journal on Computing*, 16(6):1004–1022, 1987.
- [12] Éric Fusy. Uniform random sampling of planar graphs in linear time. *Random Structures & Algorithms*, 35(4):464–522, 2009.
- [13] Stefanie Gerke, Dirk Schlatter, Angelika Steger, and Anusch Taraz. The random planar graph process. *Random Structures & Algorithms*, 32(2):236–261, 2008.
- [14] Vamsi Kalapala, Vishal Sanwalani, Aaron Clauset, and Christopher Moore. Scale invariance in road networks. *Physical Review E*, 73(2):026130, 2006.

[15] Richard J. Lipton and Robert Endre Tarjan. A separator theorem for planar graphs. *SIAM Journal on Applied Mathematics*, 36(2), 1979.

[16] A.P. Masucci, D. Smith, A. Crooks, and M. Batty. Random planar graphs and the London street network. *European Physical Journal B*, 71(2):259–271, 2009.

[17] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[18] Gilles Schaeffer. Random sampling of large planar maps and convex polyhedra. In *Proceedings of the Thirty-First Annual ACM Symposium on the Theory of Computing*, pages 760–769, 1999.

[19] U.S. Census Bureau. TIGER/Line® Shapefiles [machine-readable data files], 2009.

## Appendix

*Proof of Lemma 12, continued.* Let  $b \in [0, c]$  be the greatest integer such that

$$4^c r^{\binom{c-b}{2}} \leq 1/4.$$

Let  $a \in [0, b]$  be the greatest integer such that

$$\frac{r^{-a}}{1-r} \ln \frac{1}{p(a)} \leq r^{-b},$$

where  $p(a) \triangleq 1/[4^{a+1}(a+1)]$ . Some analysis reveals that  $b = c - O(\sqrt{c})$  and  $a = b - O(\ln c)$ .

Since  $|\Sigma^c| = 4^c$ , by Lemma 10, it holds with probability at least  $3/4$  that

$$\forall w \in \Sigma^c, \quad \text{Birth}(w) \geq r^{-b},$$

that is, that no birth at depth  $c$  has occurred. Observe that for all  $v \in \Sigma^a$ ,

$$\mathbf{E}[\text{Death}(v)] = \sum_{k=0}^a r^{-k} < \frac{r^{-a}}{1-r}.$$

By Lemma 11 with  $p \triangleq p(a)$ , it holds with probability at least  $3/4$  that

$$\forall v \in \Sigma^a, \quad \text{Death}(v) \leq r^{-b},$$

that is, that every birth at depth  $a$  has occurred.  $\square$

*Proof of Theorem 3, continued.*

$$\begin{aligned} & \mathbf{E}[\text{MaxFlow}_n] \\ & > 2^d - t^{-d} \mathbf{E}[F_n(t)] \\ & \geq 2^d - t^{-d} \exp\left((4t-1) \sum_{k=0}^{n-1} \frac{1}{3k+1}\right) \\ & \quad \text{by Lemma 13} \\ & \geq 2^d - t^{-d} \exp\left((4t-1) \frac{\ln(3n+1)}{3}\right) \\ & \quad \text{by integrating, since } 4t-1 < 0 \\ & = 2^d \left(1 - (3c/2)^{-d} \exp((c-1/3) \ln(3n+1))\right) \\ & \geq 2^d \left(1 - \exp([-c \ln(3c/2) + c - 1/3] \ln(3n+1))\right) \\ & \quad \text{since } 3c/2 < 1 \text{ and } d \leq c \ln(3n+1) \\ & = 2^d (1 - \exp(-1)) \\ \square & = \Omega(n^{0.086\dots}). \end{aligned}$$

*Proof of Lemma 15.* Recalling that  $\lambda$  is the empty string, let

$$\begin{aligned} u & \triangleq \ln(15/14) \\ p & \triangleq \Pr(\text{Countdown}(\lambda) \leq u). \end{aligned}$$

For all squares  $w \in \Sigma^d$ , we have  $\Pr(\text{Countdown}(w) \leq r^{-d} \ln(15/14)) = p$ . After showing that  $p > 0$ , we finish by setting  $k \geq 1$  to the least integer such that  $(1-p)^{4k} \leq 1/16$ .

Define  $k \geq 0$  to be the least integer such that

$$\sum_{j=k}^{\infty} r^{-j} \leq u/4.$$

By Markov's inequality,

$$\Pr\left(\sum_{j=k}^{\infty} \text{Lifespan}(N_W^j) \leq u/2\right) \geq 1/2.$$

For all integers  $j \in [0, k-1]$ , it holds with positive probability that  $\text{Lifespan}(N_W^j) \leq u/(2k)$ , and all of the terms are independent.  $\square$