Route lengths in invariant spatial tree networks

David Aldous[*]

Abstract

Is there a constant r_0 such that, in any invariant tree network linking rate-1 Poisson points in the plane, the mean within-network distance between points at Euclidean distance r is infinite for $r > r_0$? We prove a slightly weaker result. This is a continuum analog of a result of Benjamini et al (2001) on invariant spanning trees of the integer lattice.

1 Introduction

Parts of classical stochastic geometry [12], for instance Delaunay triangulations on random points, implicitly concern random spatial networks but without direct motivation as real-world network models. Substantial recent literature, surveyed in the 2018 monograph [9], concerns toy models of more specific types of real-world spatial network, studied in statistical physics style rather than theorem-proof style. Intermediate between those styles, and envisioning examples such as inter-city road networks, one can model the city positions as a Poisson point process, and one can study the trade-off between a network's cost (taken as network length) and its effectiveness at providing short routes [4, 5, 6]. It is often remarked that tree networks are *obviously* very ineffective at providing short routes, and the purpose of this article is to give one formalization, as Theorem 1.2.

As background we mention two results for *lattice* models. Consider m^2 cities at the vertices of the $m \times m$ grid. Any connected network must have length $\Omega(m^2)$, and the mean route-length between two uniform random points must be $\Omega(m)$. Observe that these orders of magnitude can be attained by a tree-network; from each vertex create a unit edge to a neighbor vertex nearer to a central root. This type of construction extends readily to the Poisson model. But this apparent "linearity of mean route lengths" is in some ways misleading, in that it depends on a finite network having a central region. *Infinite* tree networks with a spatial stationarity property are different, as shown by the following elegant result of Benjamini et al. [10] in the infinite lattice setting. Here *invariant* means the distribution of the network is invariant under the automorphisms of the lattice.[1]

---

[*]U.C. Berkeley. E-mail: aldous@stat.berkeley.edu
[1]Theorem 1.1 was stated in [10] Theorem 14.3 for a particular model, but as noted in [15] Exercise 4.48 it holds in the general automorphism-invariant case.

Theorem 1.1 ([10]). *For any invariant random spanning tree in the infinite 2-dimensional square lattice, the (within-tree) route length D between lattice-adjacent vertices satisfies*

$$\mathbb{P}(D \geq i) \geq \tfrac{1}{8i}, \ i \geq 1.$$

In particular, $\mathbb{E}D = \infty$.

A relation between finite models and infinite invariant models is provided by local weak convergence, discussed briefly in section 3.2.

The proof of Theorem 1.1 exploits symmetries of the lattice which clearly are not directly applicable in the Poisson model. So what is the analog of Theorem 1.1 in the rate-1 Poisson model on the plane? Here *invariant* means the distribution of the network is invariant under the Euclidean group.[2] We would like to consider

$$\rho(r) := \text{ mean route length between two Poisson points at distance } r. \qquad (1.1)$$

As noted in section 3.1, the MST (minimum spanning tree) provides a model in which $\rho(r) < \infty$ for small r. It seems natural to conjecture that there exists a constant $r_0 < \infty$ such that, for all invariant tree networks over the rate-1 Poisson process, $\rho(r) = \infty$ for a.a. $r \geq r_0$. To avoid possible very artificial examples (see section 3.3) we actually prove a slightly weaker assertion, by considering instead the route-length D_r between Poisson points at distance *at most* r.

To be precise about the meaning of *tree-network*, we allow Steiner points (junctions, envisaging road networks) as vertices in addition to the given Poisson points. And we take edges to be line segments between vertices. The *tree* property is that there are no circuits.

Theorem 1.2. *There exist constants $r_0 < \infty$ and $\beta > 0$ such that, in every invariant tree-network connecting the points of a Poisson point process of rate 1 in the infinite plane, for $r \geq r_0$*

$$\mathbb{P}(D_r > d) \geq \beta r / d, \ r \leq d < \infty$$

and so $\mathbb{E}D_r = \infty$ for $r \geq r_0$.

So this is a continuum analog of Theorem 1.1. The proof in section 2 relies on the fact that a finite tree has a *centroid* from which each branch contains less than half the vertices; the route between two vertices in different branches must go via the centroid, so the route length is lower bounded by the sum of distances to the centroid. Consider the partition of a very large square into a large number of large subsquares. If there are a non-negligible number of subsquares in which points from more than one branch have non-negligible relative frequency, then the point-pairs within such subsquares provide the desired long routes. Otherwise almost all subsquares have almost all points from the same branch, but therefore (and this is the key intricate technical issue, Lemma 2.2) there must be some number of pairs of adjacent subsquares for which these are *different* branches, and so (by the easy Lemma 2.1) some overlapping square has a substantial proportion of its points from different branches, which as before provide the desired long routes.

Our proof is technically elementary, albeit rather intricate, using only very basic facts from percolation theory. It seems quite likely that some shorter proof could be found, using some more sophisticated percolation theory.

Remarks on analogous questions for general networks are given in section 3.5. Note also that, for Theorem 1.2 to be interesting in the sense of generality, one would like to know that there are many different ways to construct invariant tree-networks over Poisson points, and we discuss this in section 3.2.

---

[2]In fact only translation-invariance is needed in Theorem 1.2, because we can impose a uniform random rotation over a given translation-invariant model, and this does not change D_r.

2 Proofs

2.1 Technical lemmas

Here we give two lemmas. The first, which is elementary, will enable reduction to a lattice percolation setting, and the second is the key technical ingredient we need in that setting. To aid intuition we state these in terms of colorings, though with different interpretations in the two lemmas, and it is **not** the graph-theoretic *coloring* notion in which adjacent vertices must have different colors.

Fix a large integer m. Call a configuration of points in general position in the continuum $m \times m$ square $S_1 = [0, m]^2$ *balanced* if, in each of the ten sub-rectangles $[(i-1)m/5, im/5] \times [0, 1], 1 \leq i \leq 5$ and $[0, 1] \times [(i-1)m/5, im/5], 1 \leq i \leq 5$, the number of points is between $0.98m^2/5$ and $1.02m^2/5$. Make the analogous definition for the adjacent square $S_2 = [m, 2m] \times [0, m]$.

Lemma 2.1. *Suppose S_1 and S_2 each contain a balanced configuration of points. Consider a {blue, red} coloring of the points in $S_1 \cup S_2$, and suppose that neither*
(a) S_1 and S_2 both contain less than $0.1m^2$ blue points
nor (b) S_1 and S_2 both contain more than $0.88m^2$ blue points
is true. Then the number of blue-red point pairs at distance at most $2^{1/2}m$ apart is at least $0.088m^4$.

Note we are counting *all* such pairs, not asking for a matching where a point can be in only one pair.

Proof of Lemma 2.1. First, if either S_1 or S_2 contains between $0.1m^2$ and $0.88m^2$ blue points, say y blue points, then (from the definition of *balanced*) there are at least $0.98m^2 - y$ red points, and so at least $y(0.98m^2 - y) \geq 0.1m^2 \times 0.88m^2$ blue-red pairs within that square. Such a pair is at most $2^{1/2}m$ apart. The only remaining case is w.l.o.g where S_1 contains less than $0.1m^2$ blue points, and S_2 contains more than $0.88m^2$ blue points. In this case, consider the successive translated squares $[im/5, m + im/5] \times [0, m], i = 0, 1, 2, \ldots, 5$. At each step the number of blue points can increase by at most $1.02m^2/5$, so in at least one of the translated squares there are between $0.1m^2$ and $0.88m^2$ blue points, and the result follows as in the first case. $\square$

For our key technical lemma, fix a large integer k and consider the $k \times k$ grid graph with vertices $G_k = \{0, 1, , \ldots, k - 1\} \times \{0, 1, , \ldots, k - 1\}$.

Lemma 2.2. *Given an arbitrary subset ξ^k of G_k, let $c(\xi^k)$ be the minimum, over all {green-yellow} colorings of G_k with at least $k^2/4$ vertices of each color, of the number of green-yellow adjacent pairs where neither vertex is in ξ^k. Then there exists $q > 0$ such that, taking Ξ^k to be the random subset in which each vertex is present independently with probability q,*

$$\mathbb{P}(c(\Xi^k) < k/400) \to 0 \text{ as } k \to \infty.$$

As motivation, in the proof of Theorem 1.2 we will apply this where the vertices represent large squares and the two colors indicate a relatively large or relatively small number of points in a given tree-branch in the square. The proof of Lemma 2.2 is in essence just the classical *Peierls contour* method [16], in that it involves counting self-avoiding paths, but applied in two different ways at (2.1) and (2.7).

Proof of Lemma 2.2. To recall basic percolation theory, in any coloring a green-yellow adjacent pair specifies an edge in a dual graph, and these edges form the boundaries of colored components. More precisely, as illustrated in Figure 1 (left), the set of such edges is a disjoint union of
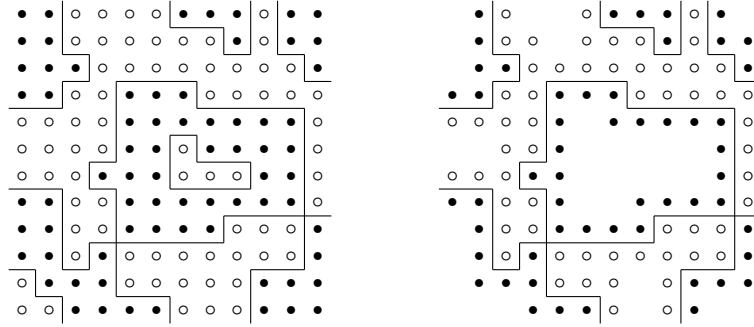(i) self-avoiding circuits within the $k \times k$ grid

Figure 1: Maximal circuits and paths in the finite grid.

(ii) self-avoiding paths starting and ending on the external dual boundary.

We write path* for "path or circuit". Fix $\ell > 4$ and consider a self-avoiding path* (in the dual graph) π of length ℓ in G_k. Each edge separates some pair of vertices in G_k. A pair of adjacent vertices overlaps with 6 other pairs, and so (by listing the ℓ pairs associated with π and greedily choosing pairs not overlapping with previously chosen pairs) we can find a set S_π of $\lfloor \ell/7 \rfloor$ *disjoint* adjacent vertex pairs separated by some edge within π.

Consider the event A_π that at most $\ell/20$ pairs within S_π have neither end-vertex in Ξ^k. This event has probability

$$\mathbb{P}(A_\pi) = \mathbb{P}(\mathrm{Bin}(\lfloor \ell/7 \rfloor, (1-q)^2) \le \ell/20).$$

The number of length-ℓ self-avoiding paths* π is at most $4k^2 3^{\ell-1}$. So the expected number of events A_π that occur is at most

$$4k^2 3^{\ell-1} \times \mathbb{P}(\mathrm{Bin}(\lfloor \ell/7 \rfloor, (1-q)^2) \le \ell/20).$$

We claim that for sufficiently small $q > 0$

$$\lim_{k \to \infty} \sum_{\ell > \log k} 4k^2 3^{\ell-1} \times \mathbb{P}(\mathrm{Bin}(\lfloor \ell/7 \rfloor, (1-q)^2) \le \ell/20) = 0. \tag{2.1}$$

To verify this, tidy by setting $p = 1 - (1-q)^2$ and $m = \lfloor \ell/7 \rfloor$; now we need to prove that for sufficiently small $p > 0$

$$\lim_{k \to \infty} k^2 \sum_{m > \frac{1}{7} \log k} 3^{7m} \times \mathbb{P}(\mathrm{Bin}(m, 1-p) \le 2m/5) = 0. \tag{2.2}$$

The Chernoff bound says that $\mathbb{P}(\mathrm{Bin}(m, 1-p) \le 2m/5) = \mathbb{P}(\mathrm{Bin}(m, p) > 3m/5) \le \exp(-m\psi(p))$ for a certain function $\psi(p) \uparrow \infty$ as $p \downarrow 0$. Then (2.2) follows easily.

The impact of (2.1) is that for sufficiently small q we may assume

(*) For every self-avoiding path* π of length $\ell > \log k$ in the dual graph of G_k, there exist at least $\ell/20$ disjoint adjacent vertex pairs separated by some edge within π and with neither vertex in Ξ^k.

Note this is a property of Ξ^k, not involving any coloring.

Now consider a green-yellow coloring of G_k with at least $k^2/4$ vertices of each color, By an elementary argument, the length of the boundary within G_k between colored regions, that is the sum of lengths of the paths* at (i,ii), is at least $k/2$. Split that sum

as $S_{long} + S_{short}$ according as the path* lengths are longer or shorter than $\log k$. If $S_{long} > k/10$ then property (*) implies there are at least $k/200$ green-yellow adjacent pairs where neither vertex is in Ξ^k, and because a vertex can be adjacent to at most 2 paths* these contain a subset of at least $k/400$ *disjoint* green-yellow adjacent pairs where neither vertex is in Ξ^k.

So it is enough to consider only colorings in which

$$S_{long} \le k/10 \text{ and } S_{short} \ge k/2 - k/10 = 2k/5. \qquad (2.3)$$

Fix such a coloring, and consider the associated paths and circuits, as described in (i,ii) above and illustrated in Figure 1. Note that a *circuit* splits G_k into an *exterior* and an *interior* region. Also a *path*, which by (2.3) has length $\le k/10$, splits G_k into a well-defined larger and a smaller region, where (somewhat confusingly) we designate the smaller region meeting the boundary of G_k as the *interior* of the path. Now a circuit of length c fits inside some square of side $c/2$ and so its interior has at most $c^2/4$ vertices. A path of length c fits inside some square of side c and so its interior has at most c^2 vertices. Because $\sum_i c_i^2 \le (\sum_i c_i)^2$, under (2.3) we have

$$\text{at most } (k/10)^2 = k^2/100 \text{ vertices are in the interior of long paths*.} \qquad (2.4)$$

A short *path* starts and ends on the boundary of the square and so the interior vertices are within a distance $\log k$ from the boundary. So by considering the number of vertices within distance $\log k$ from the boundary of the $k \times k$ square

$$\text{at most } 4k \log k \text{ vertices are in the interior of short paths.} \qquad (2.5)$$

We remark that the point of this argument is that, from the upper bounds above and the lower bound on S_{short} in (2.3), we will next be able to lower bound the number of vertices within short circuits, as (2.6) below.

A *maximal* circuit is one that is not contained inside another circuit or path, and a *maximal* path is one that is not inside[3] another path. Figure 1 (right) shows the 5 maximal paths and the 1 maximal circuit in that example. Note that, by definition, there is a single-color path[4] immediately inside and a single-opposite-color path immediately outside each maximal path or circuit. Moreover the colors of these immediately-inside paths are the same (say •) for each component, because a path in G between a vertex in each component must cross component boundaries an even number of times. Every vertex of color • at distance at least $\log k$ from the sides of G_k is either inside some path, or inside some circuit and therefore inside some maximal circuit. By hypothesis there are at least $k^2/4$ vertices of each color, so using (2.4, 2.5) and considering maximal circuits we have shown that property (2.3) implies that for large k

$$\text{there exist circuits, each of length less than } \log k, \text{ with disjoint interiors}$$

$$\text{and containing a total of at least } k^2/5 \text{ vertices.} \qquad (2.6)$$

So it is enough to consider only colorings with property (2.6). To analyze this case we need to set up some notation. Write Ξ^∞ for the random subset of the *infinite* square lattice $\mathbb{Z}^2$ in which each vertex is present independently with probability q. Define the *cost* of a dual circuit C in $\mathbb{Z}^2$ to be the number of edges for which neither adjacent vertex is in Ξ^∞, and similarly for dual circuits in G_k and Ξ^k. Consider the event

$$A_\infty^q := \text{some circuit in } \mathbb{Z}^2 \text{ around the origin has zero cost.}$$

---

[3] Here *inside* means "in the interior of".

[4] In this specific context the path may include diagonals.

By a simpler use of the Peierls contour method used for (2.1), $\mathbb{P}(A_\infty^q) \to 0$ as $q \downarrow 0$. So we can fix q sufficiently small that

$$\mathbb{P}(A_\infty^q) \le \tfrac{1}{20}. \tag{2.7}$$

Consider a coloring of G_k (depending on Ξ^k) satisfying (2.6): to complete the proof of Lemma 2.2 it will suffice to show that

$$N_k := \text{ number of green-yellow adjacent pairs in } G_k \text{ with neither vertex in } \Xi^k$$

satisfies

$$\mathbb{P}(N_k < k/400) \to 0 \text{ as } k \to \infty. \tag{2.8}$$

Write $\mathcal{C}_k$ for the set of circuits guaranteed by (2.6), and write $\mathcal{G}_k$ for the union of their interior vertices. For $v \in \mathcal{G}_k$ write $C_k(v)$ for the circuit in $\mathcal{C}_k$ containing v and $\mathbf{area}(C_k(v))$ for its area (= number of interior vertices). Now

$$
\begin{aligned}
N_k \;&\ge\; \sum_{C_k \in \mathcal{C}_k} \mathbf{cost}(C_k) \\
&=\; \sum_{v \in \mathcal{G}_k} \frac{1}{\mathbf{area}(C_k(v))} \mathbf{cost}(C_k(v)).
\end{aligned}
$$

Now, taking Ξ^k as the restriction of Ξ^∞, we have $\mathbf{cost}(C_k(v)) \ge 1_{A_k^c(v)}$ where $A_k(v)$ is the event that some circuit around v in Ξ^∞ with length $\le \log k$ has zero cost. Note by (2.7)

$$\mathbb{P}(A_k(v)) \le \mathbb{P}(A_\infty^q) \le \tfrac{1}{20} \tag{2.9}$$

and write

$$N_k \ge \sum_{v \in \mathcal{G}_k} \frac{1}{\mathbf{area}(C_k(v))} \, 1_{A_k^c(v)}.$$

By (2.6), the circuit $C_k(v)$ has length at most $\log k$, and so lies within a square of side $\log k$, and so each $\mathbf{area}(C_k(v))$ is at most $\log^2 k$, so

$$N_k \ge \frac{1}{\log^2 k} \left(|\mathcal{G}_k| - \sum_{v \in \mathcal{G}_k} 1_{A_k(v)} \right) \ge \frac{1}{\log^2 k} \left(|\mathcal{G}_k| - \sum_{v \in G_k} 1_{A_k(v)} \right). \tag{2.10}$$

By (2.9)

$$\mathbb{E}\left(\sum_{v \in G_k} 1_{A_k(v)} \right) \le \tfrac{k^2}{20}.$$

If v_1 and v_2 are more than $\log k$ apart, the events $A_k(v_1)$ and $A_k(v_2)$ are independent, so

$$\mathrm{var}\left(\sum_{v \in G_k} 1_{A_k(v)} \right) = O(k^2 \log^2 k)$$

and then Chebyshev's inequality gives

$$\mathbb{P}\left(\sum_{v \in G_k} 1_{A_k(v)} > k^2/10 \right) \to 0 \text{ as } k \to \infty. \tag{2.11}$$

By (2.6) we have $|\mathcal{G}_k| \ge k^2/5$, and combining with (2.10) we find

$$\mathbb{P}(N_k < k^2/(10 \log^2 k)) \to 0 \text{ as } k \to \infty$$

which is stronger than the desired bound (2.8).

To check the logic of this argument, note that the event in (2.11) involves only Ξ^∞. The other inequalities are deterministic, and show that, outside event (2.11), for every coloring satisfying (2.6) and for large k, we have $N_k \ge k^2/(10 \log^2 k)$. $\qquad\square$

We actually need the following modification of Lemma 2.2, to say that the same result holds if we insist that we count only pairs outside an arbitrary subsquare of side $0.001k$.

Corollary 2.3. *Let Ξ^k be the random subset of G_k in which each vertex is present independently with probability q. Let $\square_k$ be a subsquare of G_k of side asymptotic to $0.001k$, dependent on Ξ^k. Let $c'(\Xi^k)$ be the minimum, over all {green-yellow} colorings of G_k with at least $k^2/4$ vertices of each color, of the number of green-yellow adjacent pairs where neither vertex is in Ξ^k or in $\square_k$. Then there exist $q > 0$ and $\alpha > 0$ such that*

$$\mathbb{P}(c'(\Xi^k) < \alpha k) \to 0 \text{ as } k \to \infty.$$

Outline proof. Re-color the vertices in the small subsquare to become all the same color, and apply Lemma 2.2 to the new configuration. We omit details.

2.2 Proof of Theorem 1.2

Take large integers k and m, and set $n = km$. Consider the $n \times n$ square $[0,n]^2$ in the plane. Write $\Sigma_{m,k}$ for the index set of the natural partition of the square $[0,n]^2$ into k^2 subsquares σ of side m – call these the *natural* subsquares. The set $\Sigma_{m,k}$ is isomorphic to the $k \times k$ vertex grid G_k. In particular a subsquare $\square$ of G_k, say with $s \times s$ vertices, corresponds to a subsquare $\boxplus$ of the square $[0,n]^2$, with side sm, consisting of s^2 natural subsquares.

Write q_m for the probability that a realization of a rate-1 Poisson point process on a $m \times m$ square is not *balanced*, in the sense of Lemma 2.1. By the weak law of large numbers for the Poisson distribution,

$$q_m \to 0 \text{ as } m \to \infty. \tag{2.12}$$

Given a rate-1 Poisson point process, the collection of not-balanced subsquares can be identified with the random subset Ξ^k in Corollary 2.3 with $q = q_m$. In the bounds below we assume that m and k are sufficiently large, independently,

Now consider a rate-1 Poisson point process on the whole plane and a tree-network connecting them. Write N for the number of Poisson points in the square $[0,n]^2$. Consider the subtree spanned by all the Poisson points in that square. This subtree will typically extend outside the square. But there will exist a centroid, in the sense of a vertex v_* (maybe a Steiner point, and maybe outside the square) such that, writing $B_1, B_2, \ldots$ for the sets of Poisson points within the square that are in the different branches from v_*, the largest such set has size at most $N/2$. It is then always possible to merge (if necessary) these sets into a bipartition $\{B, B^c\}$ of the points in the square such that $N/3 \le |B| \le N/2$. The key observation is that the path from any $v \in B$ to any $v' \in B^c$ must go via v_*. We will use this to prove the following key result, from which Theorem 1.2 will follow quite easily.

Lemma 2.4. *There exists $\beta_0 > 0$ such that, with probability $\to 1$ as $m, k \to \infty$, there are at least $\beta_0 m^4 k$ pairs of points from B and B^c within straight-line distance $2^{1/2}m$ but whose route-length is at least $0.001mk$.*

Proof. Color red the points in B, and color blue the points in B^c. The "probability" parts of the argument are the following easy consequences of the law of large numbers. Outside an event of probability $\to 0$ as $m, k \to \infty$:

$$\text{the total numbers of blue and of red points each exceed} 0.33m^2k^2; \tag{2.13}$$

the total number of points not in balanced natural subsquares is at most $m^2k^2\psi(m)$,

$$\text{where } \psi(m) \downarrow 0 \text{ as } m \to \infty. \tag{2.14}$$

The remainder of the argument is deterministic, and the precise numerical constants are not important.

Write $\square$ for a subsquare of G_k with $0.001k \times 0.001k$ vertices, and write $\boxplus$ for the corrresponding square of side $0.001n$ within $[0,n]^2$. The essential issue is to find a lower bound for

$$N_{\text{pair}} := \min_{\boxplus} N_{\text{pair}}(\boxplus), \quad \text{where}$$

$$N_{\text{pair}}(\boxplus) := \text{ number of blue-red point pairs outside } \boxplus$$

$$\text{and at distance at most } 2^{1/2}m \text{ apart.}$$

Write b_σ for the number of blue points in the natural subsquare σ. Recall that a balanced natural subsquare must have between $0.98m^2$ and $1.02m^2$ points – call this the *size condition*. Amongst balanced natural subsquares, σ, consider

- the number S for which $0.09m^2 \le b_\sigma \le 0.89m^2$,
- the number $S^<$ for which $b_\sigma < 0.09m^2$,
- the number $S^>$ for which $b_\sigma > 0.89m^2$.

If a balanced natural subsquare σ has $0.09m^2 \le b_\sigma \le 0.89m^2$ then, by the size condition, there are at least $0.98m^2 - b_\sigma$ red points, and so at least $0.08m^4$ blue-red pairs in σ. So we immediately have

$$N_{\text{pair}} \ge 0.08m^4 S^*, \text{ where } S^* := \min_\square S(\square) \text{ and}$$

$$S(\square) := \text{ number of balanced natural subsquares outside } \square \text{ for which}$$

$$0.09m^2 \le b_\sigma \le 0.89m^2.$$

where (as above) $\square$ is a subsquare of G_k with $0.001k \times 0.001k$ vertices. A given subsquare $\square$ cannot intersect more than $0.000001k^2$ natural subsquares, and so

$$N_{\text{pair}} \ge 0.08m^4(S - 0.000001k^2). \tag{2.15}$$

If S is indeed of order k^2 then this inequality gives all we require (see (2.17) below) but the key issue is to analyze the case where S is small. We can lower bound the total number N_{blue} of blue points by (2.13) and upper bound it by (2.14) and the definitions of $(S, S^<, S^>)$: this gives

$$0.33m^2k^2 \le 0.09m^2 S^< + 0.89m^2 S + 1.02m^2 S^> + m^2k^2\psi(m).$$

Using $S^< + S + S^> \le k^2$ to eliminate the $S^<$ term, this rearranges to

$$0.24m^2k^2 \le 0.8m^2 S + 0.93m^2 S^> + m^2k^2\psi(m).$$

With the corresponding inequality arising from counting red points, we obtain that for m sufficiently large

$$\text{if } S \le 0.005k^2 \text{ then } \min(S^<, S^>) \ge k^2/4. \tag{2.16}$$

Color a natural subsquare σ yellow if $b_\sigma < 0.1m^2$, or green otherwise. In the case $\min(S^<, S^>) \ge k^2/4$ we can apply Corollary 2.3 provided m is sufficiently large (recall (2.12)), to conclude that (outside an event of probability $\to 0$ as $k \to \infty$) the number of adjacent balanced green-yellow pairs is at least αk, and these can be taken to avoid any choice of $\square$ corresponding to a choice of $\boxplus$. A given subsquare can be in at most 4 such green-yellow pairs, so we can find $\alpha k/4$ disjoint pairs. By Lemma 2.1, within each such pair there exist at least $0.088m^4$ blue-red pairs and so in this case

$$N_{\text{pair}} \ge 0.088m^4 \times \alpha k/4 := \beta_0 m^4 k \tag{2.17}$$

for some constant β. In the opposite case, that is by (2.16) if $S > 0.005k^2$, inequality (2.15) gives an essentially larger bound, so we may assume (2.17).

Apply the bound (2.17) where the given subsquare $\boxplus$ is a square of side $0.001mk$ centered near the tree centroid: we obtain the conclusion of Lemma 2.4. $\qquad\square$

Completing the proof of Theorem 1.2. Take expectation in Lemma 2.4 and re-write in terms of

$$r := 2^{1/2}m, \quad d := 0.001mk = 0.001n$$

as follows.

(∗) There are constants $\beta_1 > 0$ and $r_0, \rho_0 < \infty$ such that, for $r \geq r_0$ and $d/r \geq \rho_0$, and for any invariant tree model, the mean number of pairs of Poisson points within $[0, n]^2$ at distance $\leq r$ apart and with route-length $\geq d$ is at least $\beta_1 dr^3$.

Let $\chi(r, d)$ be the probability, in a given model of an invariant tree-network over the Poisson points, that between two typical Poisson points at distance $\leq r$ the route-length is $\geq d$. The mean total number of pairs within $[0, n]^2$ at distance $\leq r$ apart is bounded above by $\frac{1}{2}n^2\pi r^2$. So the mean number of such pairs with route-length $\geq d$ is bounded above by $\frac{1}{2}n^2\pi r^2\chi(r, d)$. This holds in particular when $d = 0.001n$, and now comparing the upper and lower bound we find

$$\chi(r, d) \geq \beta_2 \tfrac{r}{d}; \quad r \geq r_0, d/r \geq \rho_0 \tag{2.18}$$

for a constant β_2. In the notation of Theorem 1.2 we have $\chi(r, d) = \mathbb{P}(D_r \geq d)$, and this inequality is equivalent to the form stated in Theorem 1.2.

3 Remarks

3.1 The Euclidean MST

Consider the *random geometric graph* $\mathcal{G}(r_0)$ whose vertices form the rate-1 Poisson point process and whose edges link all pairs of points at Euclidean distance at most r_0. Write $N(v, r_0)$ for the number of vertices in the component $C(v, r_0)$ of $\mathcal{G}(r_0)$ containing a typical vertex v. It is well known [16] that

$$\text{for sufficiently small } r_0, \text{ all moments of } N(v, r_0) \text{ are finite.} \tag{3.1}$$

Consider now the Euclidean MST over the Poisson process. Fix r_0 satisfying (3.1). By considering Prim's algorithm started from a vertex of $C(v, r_0)$, we see that the algorithm first constructs a spanning tree within he component $C(v, r_0)$, using edges of length $\leq r_0$, before using any edge of length $> r_0$ needed to escape that component. In other words, the restriction of the MST to $C(v, r_0)$ is a spanning tree within $C(v, r_0)$. So the route length in the MST between v and another vertex v' at distance $\leq r_0$ is at most $r_0 N(v, r_0)$. We can now use a size-biasing argument to show that, for r_0 satisfying (3.1),

$$\mathbb{E}D_{r_0} := \frac{\int_0^{r_0} 2\pi r\rho(r)dr}{\pi r_0^2} < \infty \tag{3.2}$$

which is essentially saying that $\rho(r_0) < \infty$. Consider a component $C_{n,m}$ of $\mathcal{G}(r_0)$ with $n \geq 2$ vertices and some number m of pairs of vertices with inter-pair distance $\leq r_0$. Note $n - 1 \leq m \leq \binom{n}{2}$. Such components appear at some rate $\beta(n, m)$ per unit area. In the entire Poisson process, the rate of pairs at inter-pair distance $\leq r_0$ is

$$\mathbf{rate}(r_0) = \tfrac{1}{2}\pi r_0^2 = \sum_{n \geq 2} \sum_{m=n-1}^{\binom{n}{2}} m\beta(n, m). \tag{3.3}$$

The route length between such a pair within $C_{n,m}$ is at most nr_0, so the contribution to "sum of route lengths between such pairs in a component" from one such $C_{n,m}$ is at most mnr_0. So the contribution to that sum "per unit area" from such components is at

most $r_0 m n \beta(n, m)$. Writing (N_0, M_0) for the values (n, m) in a typical (uniformly chosen) component of $\mathcal{G}(r_0)$, we have

$$\mathbb{E}D_{r_0} \leq r_0 \frac{\mathbb{E}[M_0 N_0 \beta(N_0, M_0)]}{\mathbf{rate}(r_0)}.$$

Using (3.3) we have $\beta(n, m) \leq \mathbf{rate}(r_0)$ and we also know $M_0 \leq \binom{N_0}{2}$, so

$$\mathbb{E}D_{r_0} \leq r_0 \mathbb{E}[N_0^3]/2.$$

Now $N = N(v, r_0)$ at (3.1) is the size-biased transform of N_0, that is $\mathbb{P}(N = n) = n\mathbb{P}(N_0 = n)/\mathbb{E}[N_0]$. So $\mathbb{E}[N^3] \geq \mathbb{E}[N_0^3]$ and (3.2) is established.

3.2 Constructing invariant tree-networks

Some examples are given in [14]; here is our general discussion. Take an arbitrary tree-network linking m^2 independent uniform random vertices in the continuum square $[0, m]^2$, and write ℓ_m for the expectation of the average (over vertices) length of the edge from the vertex toward the centroid. Randomly re-center, that is translate the plane as $(x, y) \to (x - U, y - V)$ for (U, V) uniform on $[0, m]^2$, and then apply a uniform random rotation. A sequence of such networks with ℓ_m bounded as $m \to \infty$ is tight in the natural "local weak convergence" topology, and any subsequential weak limit network has invariant distribution. This very general construction suggests that the class of invariant tree-networks should be very rich. But there are two issues.

In general the weak limit structure is guaranteed to be a forest with infinite tree-components, but is not guaranteed to be a single tree. The planar MST limit is known to be a tree [7] but the proof heavily exploits its explicit structure; there seem to be no useful general methods for proving that a construction via local weak convergence gives a limit tree. To illustrate a more algorithmic construction, consider "Poisson rain" on the plane – rate 1 per unit area per unit time over time $0 < t \leq 1$. The construction rule "each arriving point is a child of the nearest existing point" gives a genealogical tree studied in [2]. Representing the parent-child relation by drawing a line segment, the network is not a tree because such lines may cross, but instead one can draw just the part of the segment from the child to the existing network, and the analysis in [2] implies this will be a tree. Presumably other rules for connecting arriving points to the existing network within this Poisson rain framework will also yield invariant trees.

The second issue is illustrated by the notion of minimal (shortest length) Steiner tree. In the finite random setting this is a.s. unique. The local weak convergence scheme produces limit random forests attaining the minimum length-per-unit area possible, but (even if one could prove that limits are trees) it is not clear how to prove there is an a.s. *unique* limit tree attaining that minimum.

More abstractly, infinite trees arising as local weak limits are *unimodular*: general theory for unimodular trees at the graph-theoretic level is given in [11, 8] but is not specifically adapted to the spatial setting.

3.3 Outline of possible counter-examples to the natural conjecture

Take $r_i \to \infty$ very fast and $\delta_i \to 0$ very fast. Draw a line segment between the Poisson point pairs which are at some distance in $\cup_i [r_i, r_i + \delta_i]$. One can arrange that the density of intersections of these lines is arbitrarily small. Break the (rare) circuits. Then assign random arrival times and use a "Poisson rain" construction as in the section above. In this way it might be possible to construct an invariant tree-network such that $\rho(r) < \infty$ for $r \in \cup_i [r_i, r_i + \delta_i]$.

3.4 Possible generalizations

For Theorem 1.2, can one replace the Poisson point process by a more general point process satisfying some spatial mixing condition? The *centroid argument* outlined below the statement of Theorem 1.2 seems heuristically to require very little beyond invariance. A more sophisticated proof of the Poisson case might allow such generalizations, and this might be a good starting project for a student wishing to engage percolation theory.

3.5 General spatial networks

For general (i.e. non-tree) invariant networks over Poisson points, the quantity

$$\rho(r) := \text{ mean route length between two Poisson points at distance } r$$

at (1.1) is a natural object of study. From [1, 13] we know that under very weak assumptions (which roughly correspond to "not a tree"), not only is $\rho(r) < \infty$ for all r, but also (by subadditivity, heuristically) there exists the limit

$$\lim_{r \to \infty} r^{-1}\rho(r) := \rho^* < \infty.$$

That is, average route lengths are asymptotically linear in straight-line distance. But quantitative analytic study of $\rho(r)$ or ρ^* seems very difficult even in simple-to-describe network models.

For reasons explained in [6], it is not always wise to use ρ^* as a summary statistic for efficiency at providing short routes. Instead, in [6] we recommend the statistic $\sup_r r^{-1}\rho(r)$ to ensure that the network provides short routes *on all scales*. This line of thought also motivates study of exactly self-similar networks (so $r^{-1}\rho(r)$ is constant) on the continuum plane [3].

References

[1] David Aldous. Which Connected Spatial Networks on Random Points have Linear Route-Lengths? *ArXiv e-prints*, November 2009.

[2] David Aldous. Random partitions of the plane via Poissonian coloring and a self-similar process of coalescing planar partitions. *Ann. Probab.*, 46(4):2000–2037, 2018. MR-3813984

[3] David Aldous and Karthik Ganesan. True scale-invariant random spatial networks. *Proc. Natl. Acad. Sci. USA*, 110(22):8782–8785, 2013. MR-3082274

[4] David Aldous and Tamar Lando. The stretch-length tradeoff in geometric networks: average case and worst case study. *Math. Proc. Cambridge Philos. Soc.*, 159(1):125–151, 2015. MR-3349335

[5] David Aldous and Wilfrid S. Kendall. Short-length routes in low-cost networks via Poisson line patterns. *Adv. in Appl. Probab.*, 40(1):1–21, 2008. MR-2411811

[6] David Aldous and Julian Shun. Connected spatial networks over random points and a route-length statistic. *Statist. Sci.*, 25(3):275–288, 2010. MR-2791668

[7] Kenneth S. Alexander. Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.*, 23(1):87–104, 1995. MR-1330762

[8] Francois Baccelli, Mir-Omid Haji-Mirsadeghi, and Ali Khezeli. Eternal family trees and dynamics on unimodular random graphs. In *Unimodularity in randomly generated graphs*, volume 719 of *Contemp. Math.*, pages 85–127. Amer. Math. Soc., Providence, RI, 2018. MR-3880014

[9] M. Barthélemy. *Morphogenesis of Spatial Networks*. Springer, 2018. MR-3752147

[10] Itai Benjamini, Russell Lyons, Yuval Peres, and Oded Schramm. Uniform spanning forests. *Ann. Probab.*, 29(1):1–65, 2001. MR-1825141

[11] Itai Benjamini, Russell Lyons, and Oded Schramm. Unimodular random trees. *Ergodic Theory Dynam. Systems*, 35(2):359–373, 2015. MR-3316916

[12] Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, third edition, 2013. MR-3236788

[13] C. Hirsch, D. Neuhäuser, C. Gloaguen, and V. Schmidt. First passage percolation on random geometric graphs and an application to shortest-path trees. *Adv. in Appl. Probab.*, 47(2):328–354, 2015. MR-3360380

[14] Alexander Holroyd and Yuval Peres. Trees and matchings from point processes. *Electron. Comm. Probab.*, 8:17–27, 2003. MR-1961286

[15] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. MR-3616205

[16] Ronald Meester and Rahul Roy. *Continuum percolation*, volume 119 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. MR-1409145