Adv. Appl. Prob. (SGSA) **40,** 1–21 (2008)
Printed in Northern Ireland

SHORT-LENGTH ROUTES IN LOW-COST NETWORKS VIA POISSON LINE PATTERNS

DAVID J. ALDOUS,* *University of California, Berkeley*
WILFRID S. KENDALL,** *The University of Warwick*

Abstract

In designing a network to link n points in a square of area n, we might be guided by the following two desiderata. First, the total network length should not be much greater than the length of the shortest network connecting all points. Second, the average route length (taken over source-destination pairs) should not be much greater than the average straight-line distance. How small can we make these two excesses? Speaking loosely, for a nondegenerate configuration, the total network length must be at least of order n and the average straight-line distance must be at least of order $n^{1/2}$, so it seems implausible that a single network might exist in which the excess over the first minimum is $o(n)$ and the excess over the second minimum is $o(n^{1/2})$. But in fact we can do better: for an arbitrary configuration, we can construct a network where the first excess is $o(n)$ and the second excess is almost as small as $O(\log n)$. The construction is conceptually simple and uses stochastic methods: over the minimum-length connected network (Steiner tree) superimpose a sparse stationary and isotropic Poisson line process. Together with a few additions (required for technical reasons), the mean values of the excess for the resulting random network satisfy the above asymptotics; hence, a standard application of the probabilistic method guarantees the existence of deterministic networks as required (speaking constructively, such networks can be constructed using simple rejection sampling). The key ingredient is a new result about the Poisson line process. Consider two points a distance r apart, and delete from the line process all lines which separate these two points. The resulting pattern of lines partitions the plane into cells; the cell containing the two points has mean boundary length approximately equal to $2r + \text{constant}(\log r)$. Turning to lower bounds, consider a sequence of networks in $[0, \sqrt{n}]^2$ satisfying a weak equidistribution assumption. We show that if the first excess is $O(n)$ then the second excess cannot be $o(\sqrt{\log n})$.

Keywords: Buffon argument; excess statistic; mark distribution; spatial network; Poisson line process; probabilistic method; ratio statistic; Slivynak theorem; Steiner tree; Vasershtein coupling; total variation distance

2000 Mathematics Subject Classification: Primary 60D05
Secondary 90B15

1. Introduction

We start with a counterintuitive observation and its motivation, which prompted us to probe more deeply into the underlying question.

Received 8 January 2007; revision received 14 November 2007.
* Postal address: Department of Statistics, 367 Evans Hall # 3860, University of California, Berkeley, CA 94720, USA.
Email address: aldous@stat.berkeley.edu
** Postal address: Department of Statistics, The University of Warwick, Coventry CV4 7AL, UK.
Email address: w.s.kendall@warwick.ac.uk

Consider n points ('cities', say) in a square of area n. Using the terminology of computer science, we are interested in both the *worst-case* setting where the points are located arbitrarily in the square, and the *average-case* setting where the points are random, independent, and uniformly distributed. Consider a connected network (a road network, say), made up of a finite number of straight line segments, linking these n points and perhaps other junction points. Recall that the minimum length connected network on a configuration of points $\boldsymbol{x}^n = \{x_1, \dots, x_n\}$ is the *Steiner tree* $\mathrm{ST}(\boldsymbol{x}^n)$.

It is well known and straightforward to prove that [9], [11] in both the worst case and the average case the (mean) total network length $\mathrm{len}(\mathrm{ST}(\boldsymbol{x}^n))$ grows as order $O(n)$. When designing a network, it is reasonable to regard the total network length as a 'cost'. A natural corresponding 'benefit' would be the existence (in some average sense) of short routes between points. Let $\ell(x_i, x_j)$ denote the route length (length of the shortest path) between points x_i and x_j in a given network, and let $|x_i - x_j|$ denote the Euclidean distance (so $\ell(x_i, x_j) \geq |x_i - x_j|$). A good network should satisfy the following informal criterion.

Criterion. (*The short routes property.*) Averaging over pairs (i, j) chosen uniformly at random, the route length $\ell(x_i, x_j)$ between points x_i and x_j is not much larger than the Euclidean distance $|x_i - x_j|$.

A first take on a statistic to measure this property for a connected network $G(\boldsymbol{x}^n)$ is the *ratio statistic*, based on averaging the ratios of network route lengths *versus* Euclidean distances. Consider a network $G(\boldsymbol{x}^n)$ to be the configuration of points $\boldsymbol{x}^n = \{x_1, \dots, x_n\}$ together with a collection of line segments which combine to connect every x_i to every other x_j.

Definition 1. (*Ratio statistic.*) Let $\mathrm{average}_{(i,j)}$ denote the average over all distinct pairs (i, j). Then

$$\mathrm{ratio}(G(\boldsymbol{x}^n)) = \underset{(i,j)}{\mathrm{average}} \frac{\ell(x_i, x_j)}{|x_i - x_j|} - 1 \geq 0.$$

Consider a network $G(\boldsymbol{x}^n)$ based on n uniform random points $\boldsymbol{x}^n \subset [0, \sqrt{n}]^2$, having (say) twice the total length of the Steiner tree. Initially, we speculated that in this case the expectation $\mathrm{E}[\mathrm{ratio}(G(\boldsymbol{x}^n))]$ would at best converge to some strictly positive constant as n tends to ∞. However, this intuition is wrong.

Counterintuitive observation. (See Section 5.3.) It is possible to construct networks over well-dispersed configurations whose total lengths are greater than the corresponding Steiner tree lengths by only an asymptotically negligible factor, but for which the ratio statistic converges to 0 as the total network length converges to ∞.

These considerations were originally motivated by analysis of real-world networks. Consider, for example, the 'core' part of the UK rail network; that part which links the 40 largest cities. Given a statistic R designed to capture the 'short routes' property, we can then consider how closely the observed value of R approaches optimality. Of course, the real network has evolved according to a complex historical process heavily influenced by topography; nevertheless, it is of interest to consider whether its value of R is close to the minimum possible value of R taken over all possible networks connecting the 40 cities, but of no greater total length.

We are then led to ask what statistic R might best capture the imprecisely expressed 'short routes' property, and our consideration of n cities in an idealised square $[0, \sqrt{n}]^2$ is designed to illuminate this question. The above counterintuitive observation can be interpreted as implying that the ratio statistic of Definition 1 is probably *not* a good choice of statistic, because we prove this observation by constructing networks which are approximately optimal by this criterion and yet are plainly rather different from many plausible real-world networks. What *is* a good choice of statistic will be discussed in a companion paper, along with some real-world examples.

Informally, the counterintuitive observation suggests that we can construct networks for configurations of n points which have total network length exceeding that of the Steiner tree by just $o(n)$, and such that the average excess of network distance over Euclidean distance is $o(n^{1/2})$ (bearing in mind that the average Euclidean distance for 'evenly spread out' configurations should be $O(n^{1/2})$). In fact, much more is true: whatever the configuration of n points in $[0, \sqrt{n}]^2$ (hence, even in 'worst case' scenarios), we can construct such networks with average excess of network distance over Euclidean distance barely more than $O(\log n)$. Thus, we can work on an additive rather than a multiplicative scale.

Definition 2. (*Excess average length for a network.*) The *excess route length* for a network $G(\boldsymbol{x}^n)$ is

$$\operatorname{excess}(G(\boldsymbol{x}^n)) = \operatorname*{average}_{(i,j)}(\ell(x_i, x_j) - |x_i - x_j|).$$

Theorem 1. (Upper bound on the minimum excess network length.) *For each n, let $\boldsymbol{x}^n$ be an arbitrary configuration of n points in a square of area n. The following asymptotics hold for large n.*

(a) *Let w_n tend to ∞. There exist networks $G(\boldsymbol{x}^n)$ connecting up the points such that*

(i) $\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\operatorname{ST}(\boldsymbol{x}^n)) = o(n)$;

(ii) $\operatorname{excess}(G(\boldsymbol{x}^n)) = o(w_n \log n)$.

(b) *Let $\varepsilon > 0$. There exist networks $G(\boldsymbol{x}^n)$ connecting up the points such that*

(i) $\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\operatorname{ST}(\boldsymbol{x}^n)) \leq \varepsilon n$;

(ii) $\operatorname{excess}(G(\boldsymbol{x}^n)) = O(\log n)$.

This result is proved in Sections 2 and 3. The idea is to build a hierarchical network. Details are given at the start of Section 3, but here is a sketch. At small scales, routes use the underlying Steiner tree. At large scales, routes use a sparse collection of randomly oriented lines (a realisation of a stationary and isotropic *Poisson line process*); this is the key ingredient that permits an excess of at most $o(w_n \log(n))$ and $O(\log(n))$ (Section 2), respectively. We believe that only these two scales are needed, but to simplify matters (so as to avoid nonelementary analysis of Steiner trees and geodesics in Poisson line networks) we introduce an intermediate scale consisting of a widely-spaced grid. Thus, a route from an originating city navigates through the Steiner tree to a grid line and then along the grid line to a line of the Poisson line process, and then navigates in the reverse sense down to the destination city. (For technical reasons, the discussion in Section 3 also introduces occasional small rectangles to permit circumnavigation around Steiner tree 'hotspots'.) The key ingredient in the analysis is a calculation concerning the Poisson line process, which has separate interest as a result in stochastic geometry (Theorem 4, below). Consider two points a distance r apart, and delete all lines from the line process which separate these two points. The resulting pattern of lines partitions the plane into cells; the cell containing the two points has mean boundary length which, for large r, is asymptotic to $2r + \text{constant}(\log r)$.

Note that randomness arises only through use of the Poisson line process to supply a relatively small number of long straight connections; the point pattern $\boldsymbol{x}^n$ is arbitrary. The probabilistic method may now be used to prove the existence of a nonrandom network satisfying the asymptotics described in Theorem 1, based on applying Markov's inequality to the expectations $\mathrm{E}[\mathrm{len}(G(\boldsymbol{x}^n))] - \mathrm{len}(\mathrm{ST}(\boldsymbol{x}^n)) = o(n)$, etc.

For *lower bounds*, it is necessary to impose some condition on the empirical distribution of the points in $\boldsymbol{x}^n$, since if all the points concentrate on a line then the excess is 0! We need a quantitative condition on equidistribution of points over a region, formalised via the following truncated *Vasershtein coupling* scheme.

Definition 3. (*Quantitative equidistribution condition.*) Let $\boldsymbol{x}^n$ for varying n form a sequence of configurations in the plane, let μ^n be a probability measure on the plane, and let $L_n > 0$. Say $\boldsymbol{x}^n$ *is* L_n*-equidistributed as* μ^n if there exists a coupling of random variables (X_n, Y_n) such that

(a) X_n has uniform distribution on the finite point set $\boldsymbol{x}^n$;

(b) Y_n has distribution μ^n;

(c) $\mathrm{E}\left[\min\left(1, \dfrac{|X_n - Y_n|}{L_n}\right)\right] \to 0$ as $n \to \infty$.

A sufficient condition for the following result is that $\boldsymbol{x}^n$ is L_n-equidistributed as the uniform distribution on the square of area n, for some $L_n = o(\sqrt{\log n})$. The purpose of introducing the *nonuniform* distribution μ^n in Definition 3 is to permit us to express Theorem 2, below, in terms of weaker and more local conditions; for example, a consequence of Theorem 2(b) is that we may replace the *uniform* reference distribution by any distribution μ on $[0, 1]^2$ with a continuous density component, rescaled to produce a distribution μ^n on $[0, n^{1/2}]^2$. In particular, the geometry of $[0, n^{1/2}]^2$ plays no role in this result.

We choose to express Definition 3 in stochastic terms purely for convenience of exposition. For example, arguments using the connection of total variation to coupling show that $\boldsymbol{x}^n$ is L_n-equidistributed as the uniform distribution on $[0, \sqrt{n}]^2$ if the following nonstochastic condition is satisfied: for some sequence of numbers $\lambda_n \to \infty$ with $\lambda_n / L_n \to 0$ and n/λ_n^2 being integral,

$$\frac{1}{n} \sum_{\text{box}} |\,\#(\boldsymbol{x}^n \cap \text{box}) - \lambda_n^2| \to 0,$$

with the sum being taken over n/λ_n^2 boxes partitioning $[0, \sqrt{n}]^2$ into cells of side length λ_n. Thus, a wide range of possible point patterns can be seen to be L_n-equidistributed in the above sense.

Theorem 2. (Lower bound on the minimum excess network length.) *Let* $\boldsymbol{x}^n$ *be a configuration of* n *points in a square* $[0, \sqrt{n}]^2$. *Let* $L_n = o(\sqrt{\log n})$. *Suppose that either*

(a) $\boldsymbol{x}^n$ *is* L_n*-equidistributed as the uniform distribution on the square of area* n*; or (more generally)*

(b) *for some fixed* ρ *and* ε*, there exists a subcollection* $\mathbf{y}^{k(n)}$ *of* $k(n)$ *points, all lying in a disk* D_n *of area* $\pi\rho n$*, such that* $k(n) > \pi\rho n\varepsilon$*, and such that* $\mathbf{y}^{k(n)}$ *is* L_n*-equidistributed as the uniform distribution on* D_n.

Let $G(\boldsymbol{x}^n)$ be a network based on the full collection of n points. If $\operatorname{len}(G(\boldsymbol{x}^n))/n$ *remains bounded as n tends to* ∞ *then*

$$\operatorname{excess}(G(\boldsymbol{x}^n)) = \Omega(\sqrt{\log n}). \tag{1}$$

(Here the Ω *notation implies that* $\liminf_{n\to\infty} \operatorname{excess}(G(\boldsymbol{x}^n))/\sqrt{\log n} > 0$.*)*

Configurations $\boldsymbol{x}^n$ produced by independent, uniform sampling from $[0, \sqrt{n}]^2$ satisfy the conditions of Theorem 2 (see Remark 2), but so will many other configurations exhibiting both clustering and repulsion. The proof of Theorem 2 is given in Section 4, and exploits a tension between the following two facts.

(a) An efficient route between x_i and x_j must run approximately parallel to the Euclidean geodesic and, hence, will tend to make almost orthogonal intersections with random segments perpendicular to this geodesic.

(b) On the other hand, the equidistribution condition means that two points x_i and x_j randomly chosen from the subcollection must be nearly independent, uniform draws from D_n, which permits the derivation of *upper bounds* on the probability of nearly orthogonal intersections of the form given in fact (a).

Finally, we might hope to improve the result by imposing a more restrictive assumption than the requirement that $\operatorname{len}(G(\boldsymbol{x}^n))/n$ remains bounded as n tends to ∞. This requirement is weaker than either of our two alternative assumptions on $\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n))$ in the upper bound (since $\operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n)) = O(n)$). However, we are unable to improve (1) under either of the two stronger assumptions.

2. The Poisson line process network

Our upper bound on minimal $\operatorname{excess}(G(\boldsymbol{x}^n))$ is based on a result from stochastic geometry (Theorem 4, below) which is of independent interest.

Recall that a Poisson line process in the plane $\mathbb{R}^2$ is constructed as a Poisson point process whose points lie in the space which parametrises the set of lines in the plane. We will consider only undirected lines, which will be parametrised by $(r, \theta) \in \mathbb{R} \times [0, \pi)$, where r is the signed distance from the line to a reference point and θ is the angle the line makes with a reference axis. A stationary and isotropic Poisson line process has intensity measure invariant under rotations and translations of $\mathbb{R}^2$: a stationary and isotropic Poisson line process Π of unit intensity is one for which the number of lines of Π hitting a unit segment has expectation 1 (further facts about Poisson line processes may be found in [10, Chapter 8]). We are interested in the cell containing two fixed points which is formed by the lines of Π that do not separate the two points, because this can be used as the efficient long-distance part of a network route between the two points (see Lemma 3, below). Theorem 4 establishes an asymptotic upper bound for the length of the mean cell perimeter in case of wide separation between the two points; we prepare for this by using a Buffon argument to derive an exact double-integral expression for the mean cell perimeter length.

Theorem 3. (Mean perimeter length.) *Let* Π *be a stationary and isotropic Poisson line process of unit intensity. Fix two points* v_i *and* v_j *which are a distance m apart. Delete the lines of* Π *which separate the two points* v_i *and* v_j. *The remaining line pattern partitions the plane: the cell* $\mathcal{C}(v_i, v_j)$ *containing the two fixed points has mean perimeter* $\mathrm{E}[\operatorname{len} \partial\mathcal{C}(v_i, v_j)] = 2m + J_m$,

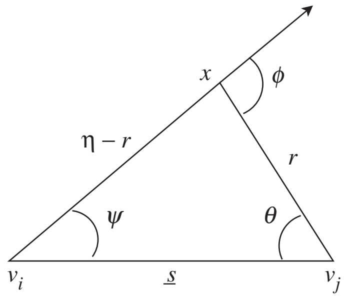

FIGURE 1: Definition of η and ϕ. Note that ϕ is the sum of the two interior angles ψ and θ.

where J_m is given by the double integral

$$J_m = \mathrm{E}[\operatorname{len} \partial \mathcal{C}(v_i, v_j)] - 2m = \frac{1}{2} \iint_{\mathbb{R}^2} (\phi - \sin\phi) \exp\left(-\frac{1}{2}(\eta - m)\right) \mathrm{d}x. \tag{2}$$

Here $\eta = \eta(x)$ is a sum of distances $|v_i - x| + |v_j - x|$, while $\phi = \phi(x)$ is the exterior angle at x of the triangle with vertices x, v_i, and v_j (see Figure 1).

Proof. This proof can be phrased in terms of measure-theoretic stochastic geometry, using the language of Palm distributions and Campbell measure. Since we deal only with constructions based on Poisson processes, we are able to adopt a less formal but more transparent exposition, for the sake of a wider readership.

Let $\underline{s}$ be the line segment of length m with endpoints v_i and v_j. The idea of the proof is to measure $\mathrm{E}[\operatorname{len} \partial \mathcal{C}(v_i, v_j)]$ by computing the expected number of hits on $\partial \mathcal{C}(v_i, v_j)$ made by an *independent*, homogeneous, isotropic Poisson line process $\widetilde{\Pi}$, again of unit intensity. Each hit corresponds to one of the points in the *intersection point process* $\mathcal{X} = \{\iota(\ell, \tilde{\ell}) \colon \ell \in \Pi, \ \tilde{\ell} \in \widetilde{\Pi}\}$, where

$$\iota(\ell, \tilde{\ell}) = \begin{cases} x & \text{if } \ell \cap \tilde{\ell} = \{x\}, \\ \text{undefined} & \text{if } \ell \text{ and } \tilde{\ell} \text{ are parallel.} \end{cases}$$

Note that with probability 1 the intersection point $\iota(\ell, \tilde{\ell})$ is defined for all $\ell \in \Pi$ and $\tilde{\ell} \in \widetilde{\Pi}$.

Not all intersection points $x \in \mathcal{X}$ correspond to hits on $\partial \mathcal{C}(v_i, v_j)$. The condition for $x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}$ to represent a hit on $\partial \mathcal{C}(v_i, v_j)$ is that ℓ should not hit $\underline{s}$ (for otherwise it cannot be involved in the construction of $\partial \mathcal{C}(v_i, v_j)$) and that x is not separated from $\underline{s}$ by any line from $\Pi \setminus \{\ell\}$. Recall that the Slivynak theorem [10, Section 4.4, Example 4.3] implies that $\Pi \setminus \{\ell\}$ conditional on $\ell \in \Pi$ is itself a homogeneous, isotropic, unit-rate Poisson line process. Consequently, under the condition that ℓ does not hit $\underline{s}$, the conditional probability of $x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}$ representing a hit on $\partial \mathcal{C}(v_i, v_j)$ is equal to the probability $p(x)$ of there being no line in Π which cuts both the segment from v_i to x and the segment from v_j to x.

A classic counting argument from stochastic geometry then reveals that

$$p(x) = \exp(-\tfrac{1}{2}(|v_i - x| + |v_j - x| - m)) = \exp(-\tfrac{1}{2}(\eta - m)).$$

Accordingly, if ν is the intensity of the point process $\mathcal{X}$ then we may compute the mean number of hits on $\partial\mathcal{C}(v_i, v_j)$ as

$$\iint_{\mathbb{R}^2} \nu\, \mathrm{P}[\ell \not\pitchfork \underline{s} \mid x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}] \exp\left(-\frac{1}{2}(\eta - m)\right) \mathrm{d}x$$
$$= 2m + \iint_{\mathbb{R}^2} \nu\, \mathrm{P}[\ell \not\pitchfork \underline{s},\ \tilde{\ell} \not\pitchfork \underline{s} \mid x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}] \exp\left(-\frac{1}{2}(\eta - m)\right) \mathrm{d}x.$$

Here '$\ell \not\pitchfork \underline{s}$' stands for 'the line ℓ does not hit $\underline{s}$'—noting that the conditioning in this context forces the Poisson line ℓ to pass through x but does not fix its orientation—and on the right-hand side the summand $2m$ corresponds to the fact that hits of $\widetilde{\Pi}$ on $\underline{s}$ count as automatic double hits on $\partial\mathcal{C}(v_i, v_j)$.

Condition on $x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}$ (which is to say, condition on there being Poisson lines $\ell \in \Pi$ and $\tilde{\ell} \in \widetilde{\Pi}$ both passing through x), and consider

(a) the angle ξ_1 made by ℓ with the line through v_i and x;

(b) the angle ξ_2 between ℓ and $\tilde{\ell}$.

By isotropy of Π, the random angle ξ_1 is uniform$(0, \pi)$. Conditional on ξ_1 and more generally on Π with an $\ell \in \Pi$ passing through x, the intersection of $\widetilde{\Pi}$ with ℓ is a Poisson point process on ℓ of unit intensity. Moreover, if the intersection points are marked with angles of intersection ξ_2 then the mark ξ_2 has mark density $\frac{1}{2}\sin\xi_2$ over $\xi_2 \in [0, \pi)$ (consider the length of the silhouette of a portion of ℓ viewed at angle ξ_2). Hence, the conditional distribution of ξ_2 for $x = \iota(\ell, \tilde{\ell})$ has density $\frac{1}{2}\sin\xi_2$ over $\xi_2 \in [0, \pi)$, and so we can compute (working with ξ_2 modulo π)

$$\mathrm{P}[\ell \not\pitchfork \underline{s},\ \tilde{\ell} \not\pitchfork \underline{s} \mid x = \iota(\ell, \tilde{\ell})] = \frac{1}{\pi}\int_0^{\phi}\left(\int_{-\xi_1}^{\phi-\xi_1} \frac{|\sin\xi_2|}{2}\,\mathrm{d}\xi_2\right)\mathrm{d}\xi_1$$
$$= \frac{\phi - \sin(\phi)}{\pi},$$

where $\phi = \theta + \psi$ is the exterior angle at x of the triangle formed by x, v_i, and v_j (see Figure 1).

Finally, the intensity ν of $\mathcal{X}$ can be computed as $\pi/2$ by, for example, computing the mean number of hits of the unit disk by Π, then computing the average length of the intersection of the disk with a line of Π conditional on that line hitting the disk. Thus,

$$J_m = \mathrm{E}[\mathrm{len}(\partial\mathcal{C}(v_i, v_j))] - 2m$$
$$= \nu \iint_{\mathbb{R}^2} \mathrm{P}[\ell \not\pitchfork \underline{s},\ \tilde{\ell} \not\pitchfork \underline{s} \mid x = \iota(\ell, \tilde{\ell}) \in \mathcal{X}] \exp\left(-\frac{1}{2}(\eta - m)\right) \mathrm{d}x$$
$$= \frac{1}{2}\iint_{\mathbb{R}^2} (\phi - \sin\phi) \exp\left(-\frac{1}{2}(\eta - m)\right) \mathrm{d}x,$$

as required.

We now state and prove the main result of this section: an $O(\log m)$ upper bound on the mean perimeter excess length J_m.

Theorem 4. (Asymptotic upper bound on the mean perimeter length.) *The mean perimeter excess length J_m is subject to the following asymptotic upper bound:*

$$J_m \leq O(\log m) \quad \text{as } m \to \infty.$$

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

Proof. Without loss of generality, place the points v_i and v_j at $(-m/2, 0)$ and $(m/2, 0)$, respectively. The double integral in (2) possesses mirror symmetry about each of the two axes, so we can write

$$\begin{aligned} J_m &= 2\iint_{[0,\infty)^2} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right)\mathrm{d}x \\ &= 2\int_0^{\pi/2}\int_0^{(m/2)\sec\theta} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right) r\,\mathrm{d}r\,\mathrm{d}\theta \\ &\quad + 2\int_{\pi/2}^{\pi}\int_0^{\infty} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right) r\,\mathrm{d}r\,\mathrm{d}\theta \end{aligned}$$

(using polar coordinates (r, θ) about the second point v_j located at $(m/2, 0)$). The integrand in the second summand is dominated by $\pi\exp(-r/2)r$, which is integrable over $(r, \theta) \in (0, \infty) \times (\pi/2, \pi)$. (In this region geometry shows that $\eta - m > r(1 - \cos\theta) \geq r$.) Thus, we can apply Lebesgue's dominated convergence theorem to deduce that the second summand is $O(1)$ as m tends to ∞; hence, it may be neglected.

In fact, we can also show that part of the first summand generates an $O(1)$ term: the dominated convergence theorem can be applied for any $\varepsilon \in (0, \pi/2]$ to show that

$$2\int_0^{\pi/2}\int_{\varepsilon}^{(m/2)\sec\theta} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right) r\,\mathrm{d}r\,\mathrm{d}\theta = O(1),$$

since the integrand is dominated by $\pi\exp(-(r/2)(1-\cos\theta))r$ over the region $(r, \theta) \in (0, \infty) \times (\varepsilon, \pi/2)$ (in this region geometry shows that $\eta - m > r(1 - \cos\theta) > r(1 - \cos\varepsilon)$). Thus, for fixed $\varepsilon \in (0, \pi/2)$ as m tends to ∞, we have the asymptotic expression

$$J_m = 2\int_0^{\varepsilon}\int_0^{(m/2)\sec\theta} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right) r\,\mathrm{d}r\,\mathrm{d}\theta + O(1),$$

where the implicit constant of the $O(1)$ term depends on the choice of $\varepsilon > 0$.

Now in the region where $0 < \theta < \varepsilon$ and $0 < r < (m/2)\sec\theta$ we know that $\phi < 2\theta < 2\varepsilon$, and moreover $\phi - \sin\phi$ is an increasing function of ϕ. Therefore, there exists a constant C_ε, converging to 0 as ε tends to 0, such that in this region

$$\phi - \sin\phi \leq 2\theta - \sin(2\theta) \leq \frac{C_\varepsilon}{8}\frac{(2\theta)^3}{6} \leq C_\varepsilon\frac{1 - \cos\theta}{3}\sin\theta.$$

Hence (as m tends to ∞ for fixed $\varepsilon > 0$),

$$\begin{aligned} &2\int_0^{\varepsilon}\int_0^{(m/2)\sec\theta} (\phi - \sin\phi)\exp\left(-\frac{1}{2}(\eta - m)\right) r\,\mathrm{d}r\,\mathrm{d}\theta \\ &\quad\leq \frac{2}{3}C_\varepsilon\int_0^{\varepsilon}\int_0^{(m/2)\sec\theta} (1 - \cos\theta)\sin\theta\exp\left(-\frac{r}{2}(1 - \cos\theta)\right) r\,\mathrm{d}r\,\mathrm{d}\theta \\ &\quad= \frac{8}{3}C_\varepsilon\int_0^{\varepsilon}\left(\int_0^{(m/4)(\sec\theta - 1)} s\mathrm{e}^{-s}\,\mathrm{d}s\right)\frac{\sin\theta\,\mathrm{d}\theta}{1 - \cos\theta} \qquad \left(\text{using } s = \frac{r}{2}(1 - \cos\theta)\right) \\ &\quad\leq \frac{8}{3}C_\varepsilon\int_0^{(m/4)(\sec\varepsilon - 1)}\left(\int_0^{v} s\mathrm{e}^{-s}\,\mathrm{d}s\right)\frac{1}{1 + 4v/m}\frac{\mathrm{d}v}{v} \qquad \left(\text{using } v = \frac{m}{4}(\sec\theta - 1)\right) \\ &\quad\leq \frac{8}{3}C_\varepsilon\log\left(\frac{m}{4}(\sec\varepsilon - 1)\right) + O(1). \end{aligned}$$

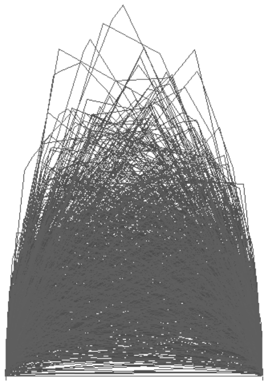

FIGURE 2: Simulation of semi-perimeters for 1000 independent cells for the unit-rate Poisson line process, with points located a distance 10^8 units apart. The figure is subject to vertical exaggeration: the y-axis is scaled at 10^4 times the x-axis. The empirical mean excess semi-perimeter is 27.63 with standard error ± 0.28, *versus* the predicted mean excess semi-perimeter, 27.5528 (using $o(1)$-asymptotics).

Remark 1. More careful analysis yields useful $o(1)$-asymptotics: in fact, it can be shown that, as m tends to ∞,

$$J_m = \tfrac{8}{3}(\log m + \gamma + \tfrac{5}{3}) + o(1),$$

where γ is the Euler–Mascheroni constant:

$$\gamma = \lim_{m\to\infty}\biggl(\biggl(\sum_1^m \frac{1}{r}\biggr) - \log m\biggr).$$

These $o(1)$-asymptotics show very good agreement with simulation: see, for example, the simulation reported in the legend of Figure 2.

3. A low-cost network with short routes

In this section we prove Theorem 1: for a given configuration $\boldsymbol{x}^n \subset [0, \sqrt{n}]^2$, we construct networks $G(x^n)$ for which both $\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\operatorname{ST}(\boldsymbol{x}^n))$ and $\operatorname{excess}(G(\boldsymbol{x}^n))$ are small. The network is constructed by augmenting the Steiner tree network $\operatorname{ST}(\boldsymbol{x}^n)$ in a hierarchical manner. The construction is stochastic: we construct a random augmentation for which the mean values of these excess values obey the desired asymptotics and then apply the probabilistic method to establish existence of the desired nonstochastic networks. Working from the largest scale downwards, we construct the following.

1. A stationary and isotropic Poisson line process Π of intensity η, where η will be small; note that this can be constructed from a unit intensity process by scaling by a magnification factor of $1/\eta$. A simple computation [10, Section 8.4] shows that the mean total length of the intersection of the resulting line pattern with $[0, \sqrt{n}]^2$ equals $\pi\eta n$.
2. A medium-scale rectangular grid with cell side length $s_n \sim (\log n)^{1/3}$. The total length

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

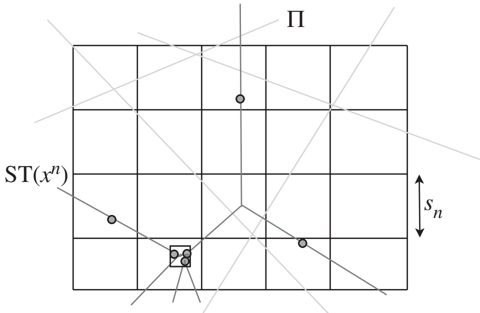

FIGURE 3: An illustration of the construction of a network to deliver an upper bound on the mean excess route length. Points are indicated by small circles. In this figure there is just one hot-spot cell.

of this grid in $[0, \sqrt{n}]^2$ is bounded above by

$$2\left(1+\frac{\sqrt{n}}{s_n}\right)\sqrt{n} = o(n).$$

3. The Steiner tree $\mathrm{ST}(\boldsymbol{x}^n)$.

4. A small number (at most $n/2$) of small hot-spot cells based on a small-scale rectangular grid with cell side length $t_n \sim 1/(\log n)^{1/6}$. A cell in this grid is described as a hot-spot cell if it contains two or more points. These hot-spot cells are used to bypass regions where the Steiner tree might become complicated and expensive in terms of network traversal. We add further small segments connecting each hot-spot cell perimeter to points within the hot-spot cell. The total length of these additions can be bounded by

$$4\frac{n}{2}t_n + n\frac{t_n}{2} = o(n).$$

Thus, the mean excess length of this augmented network is $o(n) + \pi\eta n$. The construction is illustrated in Figure 3. Note that we can choose s_n and t_n such that $n^{1/2}/s_n$ and s_n/t_n are integers, so that the small-scale lattice is a refinement of the medium-scale lattice, which itself refines the square $[0, \sqrt{n}]^2$.

3.1. Worst-case results for Steiner trees

We first record two elementary results on Steiner trees. The first result bounds the length of a Steiner tree in terms of the square root of the number of points (for the planar case).

Lemma 1. *Consider a configuration $\boldsymbol{x}^k$ of k points in a square of side r: there exists a constant C_1 not depending on k or r such that*

$$\mathrm{len}(\mathrm{ST}(\boldsymbol{x}^k)) \le C_1\sqrt{k}r.$$

Proof. See [9, Section 2.2].

The second result provides a local bound on the length contributed by a larger Steiner tree in a small square containing a fixed number of points.

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

Lemma 2. *Consider the Steiner tree* $\mathrm{ST}(\boldsymbol{x}^n)$ *for an arbitrary configuration* $\boldsymbol{x}^n$ *in the plane. Let G be the restriction of the network* $\mathrm{ST}(\boldsymbol{x}^n)$ *to a fixed open square of side length* t*. Suppose that k points* $x_1, \ldots, x_k$ *of the configuration* $\boldsymbol{x}^n$ *lie within the square. Then*

$$\mathrm{len}(G) \le t(4 + C_1\sqrt{k+1}).$$

Proof. Let $y_1, \ldots, y_m$ be the locations at which $\mathrm{ST}(\boldsymbol{x}^n)$ crosses into the interior of the square. (Note that $m = 0$ is possible if $\{x_1, \ldots, x_k\} = \boldsymbol{x}^n$: in this case choose y_1 arbitrarily from the perimeter of the square.) Then

$$\begin{aligned} \mathrm{len}(G) &\le \mathrm{len}(\mathrm{ST}(\{x_1, \ldots, x_k, y_1, \ldots, y_m\})) && \text{by minimality of } \mathrm{ST}(\boldsymbol{x}^n) \\ &\le \mathrm{len}(\mathrm{ST}(\{x_1, \ldots, x_k, y_1\})) + 4t && \text{using square perimeter} \\ &\le t(4 + C_1\sqrt{k+1}) && \text{using the previous lemma.} \end{aligned}$$

3.2. Route lengths in the medium-large network

The part of the construction involving the medium-scale grid and the Poisson line process is useful in variant problems, so we separate out the following estimate involving these ingredients.

Lemma 3. *Let* $n^{1/2}/s_n$ *be an integer. Consider the superposition of the rectangular grid with cell side length* s_n *and the Poisson line process of intensity* η*, intersected with the square* $[0, n^{1/2}]^2$*. Let* v_i *and* v_j *be vertices of the grid. Then*

$$\mathrm{E}[\textit{route length } v_i \textit{ to } v_j] \le |v_i - v_j| + C_2 \frac{1}{\eta} \log(\eta\sqrt{2n})$$

for an absolute constant C_2*.*

Proof. Let $\mathcal{C}(v_i, v_j)$ be the cell of Π containing v_i and v_j (having deleted lines from Π which separate v_i from v_j). Let $R(v_i, v_j)$ be the rectangle bounded by v_i and v_j; then, by convexity, the route length from v_i to v_j is bounded above by

$$\tfrac{1}{2}\,\mathrm{len}\,\partial(R(v_i, v_j) \cap \mathcal{C}(v_i, v_j)) \le \tfrac{1}{2}\,\mathrm{len}\,\partial\mathcal{C}(v_i, v_j),$$

whose mean value can be computed by recognising that the Poisson line process is a rescaled version of a homogeneous, isotropic, unit-rate Poisson line process. Hence, by scaling the asymptotic upper bound of Theorem 4 we have

$$\begin{aligned} \mathrm{E}\left[\frac{1}{2}\,\mathrm{len}\,\partial(R(v_i, v_j) \cap \mathcal{C}(v_i, v_j))\right] - |v_i - v_j| &\le O\left(\frac{1}{\eta}\log(\eta|v_i - v_j|)\right) \\ &= O\left(\frac{1}{\eta}\log(\eta\sqrt{2n})\right). \end{aligned}$$

3.3. Navigating the augmented network

We now explain how to move from points of $\boldsymbol{x}^n$ up to a vertex of the medium-scale grid.

Given $x_i \in \boldsymbol{x}^n$, if this is in one of the hot-spot cells then move to the perimeter of the hot-spot cell and thence to a suitable point of departure on the perimeter, with route length at most $\frac{5}{2}t_n$. Now move along the Steiner tree within the relevant medium-scale grid box to the box perimeter; however, bypass all hot-spot cells. There are $(s_n/t_n)^2 = ((\log n)^{1/3}(\log n)^{1/6})^2 = \log n$ small squares each of which involves a route length of either $2t_n$ (if the small square is a hot-spot box

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

which will be bypassed) or $t_n(4 + C_1\sqrt{2})$ (if not, by Lemma 2). Hence, the total trip to the medium-scale grid box perimeter (including emergence from the initial hot spot, if required) has length at most

$$\frac{5}{2}t_n + t_n(4 + C_1\sqrt{2}) \times \frac{s_n^2}{t_n^2} \sim \frac{5}{2}t_n + (4 + C_1\sqrt{2}) \times (\log n)^{5/6} = o(\log n).$$

Furthermore, the route length from perimeter to vertex of a medium-scale grid box is at most $\frac{1}{2}s_n \sim \frac{1}{2}(\log n)^{1/3} = o(\log n)$. So, for each x_i, there is a medium-scale grid vertex v_i for which the route length from x_i to v_i is $o(\log n)$. Combining with Lemma 3 and noting that the medium-scale grid geometry forces $|v_i - v_j| \leq |x_i - x_j| + 2(s_n/\sqrt{2})$, we find that

$$\mathrm{E}[\text{route length from } x_i \text{ to } x_j] - |x_i - x_j| \leq \sqrt{2}s_n + o(\log n) + C_2\frac{1}{\eta}\log(\eta\sqrt{2n}).$$

Averaging over the points of $\boldsymbol{x}^n$, it follows that the dominant contribution comes from the cell semi-perimeters, and indeed

$$\mathrm{E}[\operatorname{excess}(G(\boldsymbol{x}^n))] \leq O\left(\frac{1}{\eta}\log(\eta\sqrt{2n})\right),$$

at a cost in terms of network length which exceeds $\operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n))$ by a stochastic quantity of mean $\pi\eta n + o(n)$.

The two different results of Theorem 1 follow by choosing η to behave in two different ways:

(a) either $\eta \to 0$ and $\eta w_n \to \infty$; or

(b) $\eta = \varepsilon > 0$.

In either case we can apply the probabilistic method to exhibit existence of the required deterministic networks for cases (a) and (b) of Theorem 1. For example, in case (a) it is then the case that $\mathrm{E}[\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n))] \leq nc_n$ and $\mathrm{E}[\operatorname{excess}(G(\boldsymbol{x}^n))] \leq c_n w_n \log n$ for some $c_n \to 0$. But then, for any fixed n, we can apply Markov's inequality: $\mathrm{P}[\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n)) > 3nc_n] \leq \frac{1}{3}$ and $\mathrm{P}[\operatorname{excess}(G(\boldsymbol{x}^n)) > 3c_n w_n \log n] \leq \frac{1}{3}$. Hence, there is positive probability that the random network satisfies both $\operatorname{len}(G(\boldsymbol{x}^n)) - \operatorname{len}(\mathrm{ST}(\boldsymbol{x}^n)) \leq 3nc_n$ and $\operatorname{excess}(G(\boldsymbol{x}^n)) \leq 3c_n w_n \log n$; hence, such a network exists for each n.

We can view these applications of Markov's inequality as indicating a simple rejection sampling algorithm to be used to generate the required sequence of networks.

4. A lower bound on the average-excess route length

In this section we prove Theorem 2. The proof is divided into four parts. Firstly (Subsection 4.1), we show how to reduce the problem to an analogous case in which the excess is computed for two random points drawn independently and uniformly from the whole disk D_n of area $\pi\rho n$ given in condition (b) of the theorem. Then (Subsection 4.2) we show that the network geodesic must run almost parallel to the Euclidean geodesic if the excess is small. On the other hand (Subsection 4.3), we can use the uniformity of the two random points to control the extent to which network segments can run both close to and nearly parallel to the Euclidean geodesic. Finally (Subsection 4.4), we use the opposing estimates of Subsections 4.2 and 4.3 to derive a proof of the theorem using the method of contradiction.

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

4.1. Reduction to the case of a pair of uniformly random points

First we indicate how condition (a) of Theorem 2 implies condition (b). Under condition (a), we can use the coupling between X_n and Y_n to show that $\#\{\boldsymbol{x}^n \cap D_n\}/n \to \pi\rho$: therefore, for large n, the number of points in $\boldsymbol{x}^n \cap D_n$ is approximately $\pi\rho n$. On the other hand, the same coupling can be used to bound the total variation distance between the two conditional distributions $\mathcal{L}(Y_n \mid X_n \in D_n)$ and $\mathcal{L}(Y_n \mid Y_n \in D_n) = \text{uniform}(D_n)$, and to show that this bound tends to 0. We can then use rejection sampling techniques to couple $\mathcal{L}(Y_n \mid X_n \in D_n)$ and $\text{uniform}(D_n)$ so that the truncated Vasershtein distance tends to 0 as n tends to ∞; as the distance is a metric, we can combine this coupling with the (conditioned) coupling of $\mathcal{L}(X_n \mid X_n \in D_n)$ and $\mathcal{L}(Y_n \mid X_n \in D_n)$ to obtain a coupling which satisfies condition (b).

We now note that it is sufficient to consider the analogous result for a configuration $\boldsymbol{x}^n$ of n points in the disk D_n. For then we can apply the result to the lesser configuration $\mathbf{y}^{k(n)}$ (for $k(n)$ as given in condition (b) of Theorem 2) and obtain

$$\text{excess}(G(\mathbf{y}^{k(n)})) = \Omega(\sqrt{\log k(n)}) = \Omega(\sqrt{\log \pi\rho n\varepsilon}) = \Omega(\sqrt{\log n}),$$

while

$$\begin{aligned}\text{excess}(G(\mathbf{y}^{k(n)})) &= \frac{n(n-1)}{k(n)(k(n)-1)}\,\text{excess}(G(\boldsymbol{x}^n)) \\ &\le \frac{1}{\pi\rho\varepsilon(\pi\rho\varepsilon - 1/n)}\,\text{excess}(G(\boldsymbol{x}^n)),\end{aligned}$$

from which Theorem 2 follows.

We therefore consider $\boldsymbol{x}^n \subset D_n$ being L_n-equidistributed as the uniform distribution on D_n. So, by definition, there is a coupling (X_1, Y_1) (here we omit dependence on n), where X_1 has uniform distribution on $\boldsymbol{x}^n$, Y_1 has uniform distribution on D_n, and

$$\Delta_n = \mathrm{E}\left[\min\left(1, \frac{|X_1 - Y_1|}{L_n}\right)\right] \to 0 \quad \text{as } n \to \infty.$$

Write (X_2, Y_2) for an independent copy of (X_1, Y_1). In the definition of *excess* it makes no asymptotic difference if we allow $j = i$ in $\text{average}_{(i,j)}$, so we may take

$$\text{excess}(G(\boldsymbol{x}^n)) = \mathrm{E}[\ell(X_1, X_2) - |X_1 - X_2|].$$

Set

$$A_n = [|Y_1 - X_1| \le L_n] \cap [|Y_2 - X_2| \le L_n], \tag{3}$$

so that, by Markov's inequality,

$$\mathrm{P}[A_n] \ge 1 - 2\Delta_n. \tag{4}$$

Define $\ell(Y_1, Y_2)$ by supposing that Y_i is plumbed into the network using a connection by a *temporary* line segment with endpoints Y_i and X_i. A direct computation shows that, on the event A_n,

$$\begin{aligned}\ell(Y_1, Y_2) - |Y_1 - Y_2| &\le (\ell(X_1, X_2) + |X_1 - Y_1| + |X_2 - Y_2|) \\ &\quad - (|X_1 - X_2| - |X_1 - Y_1| - |X_2 - Y_2|) \\ &\le \ell(X_1, X_2) - |X_1 - X_2| + 4L_n.\end{aligned}$$

Consequently,

$$\mathrm{E}[\ell(Y_1, Y_2) - |Y_1 - Y_2|;\ A_n] \le \text{excess}(G(\boldsymbol{x}^n)) + 4L_n. \tag{5}$$

By hypothesis, $L_n = o(\sqrt{\log n})$, and so the proof of Theorem 2 reduces to showing that the left-hand side (the excess for two random points chosen uniformly in the disk) is $\Omega(\sqrt{\log n})$.

4.2. Near-parallelism for the case of small excess

We now substantiate our previous remark that the network geodesic must run almost parallel to the Euclidean geodesic if the excess is small.

It is convenient to situate the disk D_n in the complex plane $\mathbb{C}$ so as to have a compact notation for rotations. For $t > 0$, we define Z_t and Φ by

$$\begin{aligned} e^{i\Phi} &= \frac{Y_2 - Y_1}{|Y_2 - Y_1|}, \\ Z_t &= Y_1 + t \times e^{i\Phi}. \end{aligned} \tag{6}$$

Let $\gamma \colon [0, \ell(Y_1, Y_2)] \to \mathbb{C}$ be the unit-speed network geodesic running from Y_1 to Y_2 (using the temporary plumbing to move from Y_1 to X_1 and then again from Y_2 to X_2). Then (bearing in mind that $|\gamma'(t)| = 1$)

$$\ell(Y_1, Y_2) = \int_0^{\ell(Y_1,Y_2)} |\gamma'(s)| \, ds \geq \int_0^{|Y_1 - Y_2|} |\gamma'(\tau(t))| \tau'(t) \, dt,$$

where $\tau(t)$ is the first time s at which $\langle \gamma(s) - Y_1, e^{i\Phi} \rangle = t$. (Note that our networks are formed from finite collections of line segments. Hence, τ' will be defined and finite, save perhaps at a finite number of times.) This and the following constructions are illustrated in Figure 4.

Defining $\theta(t)$ by $\sec\theta(t) = \tau'(t)$, and using $\sec\theta \geq 1 + \frac{1}{2}\theta^2$, we deduce that

$$\ell(Y_1, Y_2) \geq |Y_1 - Y_2| + \frac{1}{2} \int_0^{|Y_1 - Y_2|} \theta(t)^2 \, dt. \tag{7}$$

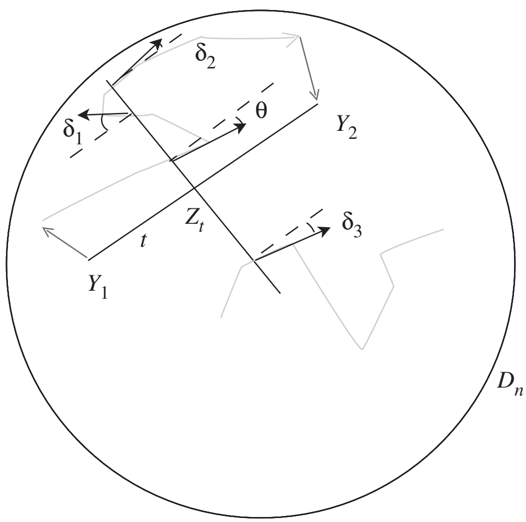

FIGURE 4: An illustration of the construction of Y_1, Y_2, and Z_t. The angles $\theta(t)$ and $\delta_1, \delta_2, \ldots$ are computed using the angles of incidence of network segments on the perpendicular running through Z_t; $\Upsilon_{t,\chi}$ is the minimum of absolute values of all such angles of points of intersection within $\sqrt{2t\chi + \chi^2}$ of Z_t.

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

Furthermore, we can use Pythagoras and the geodesic property of the Euclidean line segments to show the following. Let $H(t)$ be the maximum $|r|$ for which, for some s,

$$\gamma(s) = Z_t + \mathrm{i}r\mathrm{e}^{\mathrm{i}\Phi}.$$

If the excess for the network geodesic from Y_1 to Y_2 is bounded above by $\ell(Y_1, Y_2) - |Y_1 - Y_2| \leq \chi$ then $H(t) \leq \sqrt{2t\chi + \chi^2}$.

Let $\Upsilon_{t,\chi}$ be the smallest $|\delta|$ such that some network segment intersects the perpendicular $\{Z_t + \mathrm{i}r\mathrm{e}^{\mathrm{i}\Phi} : r \in \mathbb{R}\}$ at an angle of $\pi/2 + \delta$ and at a distance of at most $\sqrt{2t\chi + \chi^2}$ from Z_t (thus, δ is the angle of incidence of this network segment on the perpendicular). If $\ell(Y_1, Y_2) - |Y_1 - Y_2| \leq \chi$ and $|Y_1 - Y_2| \geq \kappa\sqrt{\rho n}$, we can use (7) to deduce that

$$\ell(Y_1, Y_2) - |Y_1 - Y_2| \geq \frac{1}{2}\int_0^{\kappa\sqrt{\rho n}} \Upsilon_{t,\chi}^2 \,\mathrm{d}t - \frac{1}{2}\left(\frac{\pi^2}{4}\right)(|X_1 - Y_1| + |X_2 - Y_2|).$$

(The second summand allows for the temporary plumbing in of connections X_1Y_1 and X_2Y_2, for which the angle $\theta(t) \in (0, \pi/2)$ is not controlled by permanent network segments.) So, introduce the event

$$B_{\kappa,\chi} = [\ell(Y_1, Y_2) - |Y_1 - Y_2| \leq \chi,\ |Y_1 - Y_2| \geq \kappa\sqrt{\rho n}],$$

and from (3) recall that the event $A_n = \bigcap_{i=1}^{2}[|Y_i - X_i| \leq L_n]$. Taking expectations, we deduce that

$$\mathrm{E}[\ell(Y_1, Y_2) - |Y_1 - Y_2|;\ B_{\kappa,\chi} \cap A_n] \geq \frac{1}{2}\int_0^{\kappa\sqrt{\rho n}} \mathrm{E}[\Upsilon_{t,\chi}^2;\ B_{\kappa,\chi} \cap A_n]\,\mathrm{d}t - \frac{\pi^2}{4}L_n.$$

Using integration by parts to replace the expectation by a probability,

$$\begin{aligned}
&\mathrm{E}[\ell(Y_1, Y_2) - |Y_1 - Y_2|;\ B_{\kappa,\chi} \cap A_n] + \frac{\pi^2}{4}L_n \\
&\quad\geq \int_0^{\kappa\sqrt{\rho n}}\int_0^{\infty} \mathrm{P}[[\Upsilon_{t,\chi} > u] \cap B_{\kappa,\chi} \cap A_n]u\,\mathrm{d}u\,\mathrm{d}t \\
&\quad= \int_0^{\kappa\sqrt{\rho n}}\int_0^{\infty} (\mathrm{P}[B_{\kappa,\chi} \cap A_n] - \mathrm{P}[[\Upsilon_{t,\chi} \leq u] \cap B_{\kappa,\chi} \cap A_n])u\,\mathrm{d}u\,\mathrm{d}t \\
&\quad\geq \int_0^{\kappa\sqrt{\rho n}}\int_0^{\infty} \max(\mathrm{P}[B_{\kappa,\chi} \cap A_n] - \mathrm{P}[\Upsilon_{t,\chi} \leq u], 0)u\,\mathrm{d}u\,\mathrm{d}t. \qquad (8)
\end{aligned}$$

Also note that, from the definitions of $B_{\kappa,\chi}$ and A_n, using (4), (5), and Markov's inequality,

$$\begin{aligned}
1 - \mathrm{P}[B_{\kappa,\chi} \cap A_n] &= 1 - \mathrm{P}[A_n] + \mathrm{P}[A_n \setminus B_{\kappa,\chi}] \\
&\leq 2\Delta_n + \mathrm{P}[|Y_1 - Y_2| < \kappa\sqrt{\rho n}] + \frac{\mathrm{excess}(G(\boldsymbol{x}^n)) + 4L_n}{\chi}. \qquad (9)
\end{aligned}$$

To make progress, we now need to find an upper bound for $\mathrm{P}[\Upsilon_{t,\chi} \leq u]$, and this is the subject of the next section.

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

4.3. Upper bounds using uniform random variables

Firstly, we compute an upper bound on the joint density of the quantities Z_t and Φ from the previous section, illustrated in Figure 5.

Lemma 4. *Suppose that Y_1 and Y_2 are independent, uniformly distributed random points in a disk D_n of radius $\sqrt{\rho n}$ and centre 0 in the complex plane $\mathbb{C}$. With Z_t and Φ defined as in (6), the joint density of Z_t and Φ is given over $\mathbb{C} \times [0, 2\pi)$ by*

$$\mathbf{1}[z - te_\phi \in D_n] \frac{(t + s(z, \phi))^2}{2\pi^2 \rho^2 n^2} \, \mathrm{d}z \, \mathrm{d}\phi, \tag{10}$$

where $e_\phi = \mathrm{e}^{\mathrm{i}\phi}$ is the unit vector making angle ϕ with a reference x-axis, and $s(z, \phi)$ is the distance from z to the disk boundary ∂D_n in the direction ϕ (thus, in particular, $z + s(z, \phi)e_\phi$ is on the disk boundary).

Proof. Express the joint density for Y_1 and Y_2 as a product of a uniform density over D_n for Y_1 and polar coordinates (r, ϕ) about Y_1 for Y_2:

$$\mathbf{1}[y_1 \in D_n] \frac{\mathrm{d}y_1}{\pi \rho n} \, \mathbf{1}[y_1 + r\mathrm{e}^{\mathrm{i}\phi} \in D_n] \frac{r \, \mathrm{d}r \, \mathrm{d}\phi}{\pi \rho n}.$$

Obtain the result by integrating out the r variable and transforming the y_1 variable to z by $z = y_1 + t\mathrm{e}^{\mathrm{i}\phi}$.

Corollary 1. *The density for Z_t and Φ $\pmod{\pi}$ is*

$$f(z, \phi) = \left(\mathbf{1}[z - te_\phi \in D_n] \frac{(t + s(z, \phi))^2}{2} + \mathbf{1}[z + te_\phi \in D_n] \frac{(t + s(z, \pi + \phi))^2}{2} \right) \times \mathbf{1}[0 \le \phi < \pi] \frac{\mathrm{d}z \, \mathrm{d}\phi}{\pi^2 \rho^2 n^2}, \tag{11}$$

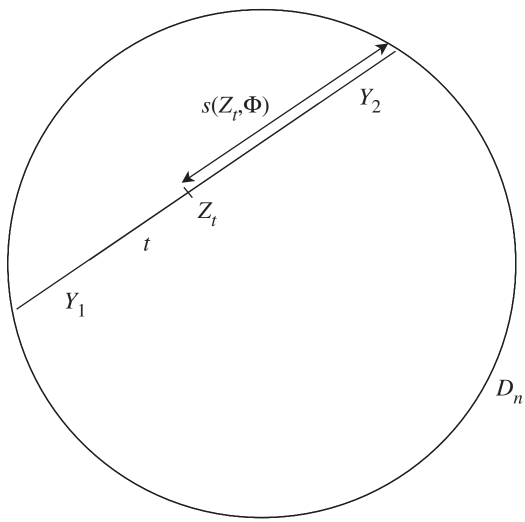

Figure 5: An illustration of the construction in Lemma 4.

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

with an upper bound

$$f(z,\phi) \leq 4 \times \mathbf{1}[0 \leq \phi < \pi] \frac{\mathrm{d}z\,\mathrm{d}\phi}{\pi^2 \rho n}. \tag{12}$$

Proof. Equation (11) follows immediately from adding the two expressions from (10) for $\phi \pmod{\pi}$. The upper bound follows by noting that the maximum will occur

1. when $z - te_\phi$ runs along a diameter as t varies;
2. furthermore when one of $z \pm te_\phi$ lies on the disk boundary; and
3. furthermore when $z = 0$ is located at the centre of the disk (so $t = s(z, \pm\phi) = \sqrt{\rho n}$).

Now consider the line segment $\mathcal{S}_{t,\chi}$ centred at Z_t, with endpoints given by the pair

$$\pm \mathrm{i}\sqrt{2t\chi + \chi^2}\mathrm{e}^{\mathrm{i}\Phi},$$

and consider the rose-of-directions empirical measure of angles made by intersections of network edges with this segment:

$$\mathcal{R}_{t,\chi}(A) = \#\{\text{network intersections on } \mathcal{S}_{t,\chi} \text{ with angle of incidence lying in } A\}$$

(here angles are measured modulo π, and $A \subseteq [0,\pi)$). We may apply a Buffon-type argument to bound $\mathrm{E}[\mathcal{R}_{t,\chi}(A)]$ using inequality (12). Consider the contribution to the expectation from a fixed line segment of the network of length ℓ: the result of disintegrating the integral expression for this according to the value of ϕ is an integral of $f(z,\phi)$ with respect to z over a region formed by intersecting the disk with a parallelogram of base side length ℓ and height $2\sqrt{2t\chi + \chi^2}\sin\alpha$ (here the angle α depends implicitly on ϕ and z). Of course the integral vanishes if $\phi \notin A$. Thus, inequality (12) yields a bound

$$\mathrm{E}[\mathcal{R}_{t,\chi}(A)] \leq \frac{4}{\pi^2 \rho n} \int_{G(\boldsymbol{x}^n)} \int_A 2\sqrt{2t\chi + \chi^2} \sin\alpha \,\mathrm{d}\alpha\,\mathrm{d}z.$$

For constant χ, the event $[\Upsilon_{t,\chi} \leq u]$ is the event $[\mathcal{R}_{t,\chi}(\pi/2 - u, \pi/2 + u) \geq 1]$, and so

$$\mathrm{P}[\Upsilon_{t,\chi} \leq u] \leq \mathrm{E}\left[\mathcal{R}_{t,\chi}\left(\frac{\pi}{2} - u, \frac{\pi}{2} + u\right)\right] \leq \frac{16}{\pi^2 \rho} \frac{\mathrm{len}(G(\boldsymbol{x}^n))}{n} \sqrt{2t\chi + \chi^2}\, u. \tag{13}$$

4.4. Calculations

We have assembled the ingredients for the proof of Theorem 2, and so now we can perform the calculations to obtain a quantitative lower bound.

We proceed by contradiction. Suppose that $\mathrm{excess}(G(\boldsymbol{x}^n)) = o(\sqrt{\log n})$. Inspecting (9) we see that we can choose $\chi = \chi_n = o(\sqrt{\log n})$ and some small $\kappa > 0$ such that, for all sufficiently large n,

$$\mathrm{P}[B_{\kappa,\chi} \cap A_n] \geq 2^{-1/3}.$$

So, we can combine (5) and (8) (and the fact that $\pi^2/4 < 3$) to obtain

$$\mathrm{excess}(G(\boldsymbol{x}^n)) + 7L_n \geq \int_0^{\kappa\sqrt{\rho n}} \int_0^\infty \max(2^{-1/3} - \mathrm{P}[\Upsilon_{t,\chi} \leq u], 0) u \,\mathrm{d}u\,\mathrm{d}t.$$

By (13) and the hypothesis of Theorem 2, there exists a constant B such that

$$\mathrm{P}[\Upsilon_{t,\chi} \leq u] \leq \sqrt{\frac{B}{12}} \sqrt{2t\chi + \chi^2}\, u.$$

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

Applying the formula $\int_0^\infty \max(0, \alpha - \beta u)u \, du = \alpha^3/6\beta^2$ we see that

$$\text{excess}(G(\boldsymbol{x}^n)) + 7L_n \geq \frac{1}{B}\int_0^{\kappa\sqrt{\rho n}} \frac{1}{2t\chi + \chi^2} \, dt = \frac{\log(\kappa\sqrt{\rho n} + \chi/2) - \log(\chi/2)}{2\chi B}. \quad (14)$$

Recall that this holds under the assumption that $\chi_n = o(\sqrt{\log n})$ and that $\kappa > 0$ is constant. We are given that $L_n = o(\sqrt{\log n})$, and we have supposed, for the purposes of contradiction, that $\text{excess}(G(\boldsymbol{x}^n)) = o(\sqrt{\log n})$. But then (14) takes the form

$$o(\sqrt{\log n}) \geq \frac{\Omega(\log n)}{o(\sqrt{\log n})},$$

which is impossible. We deduce that we must have $\text{excess}(G(\boldsymbol{x}^n)) = \Omega(\sqrt{\log n})$.

5. Closing remarks and supplements

5.1. Spatial network design

Within the realm of spatial network design, the closest work we know is that of Gastner and Newman [1], who considered the similar notion of a *distribution network* for transporting material from one central vertex to all other vertices. They gave a simulation study (their Figure 2) of a certain algorithm on random points, and made the following comment.

> Thus, it appears to be possible to grow networks that cost only a little more than the [minimum-length] network, but which have far less circuitous routes.

Our Theorem 1 provides a strong formalisation of this idea.

An algorithm for minimizing excess for a given length is described in [8], where results for a 39 point configuration are shown. But neither this nor [1] has led to the study of n tending to ∞ asymptotics.

5.2. Fractal structure of the Steiner tree on random points

A longstanding topic of interest in statistical physics is that of the continuum limits of various discrete two-dimensional self-avoiding walks arising in probability models, for example,

- uniform self-avoiding walks on the lattice;
- paths within uniform spanning trees in the lattice;
- paths within minimum spanning trees in the lattice.

This study has recently been complemented by spectacular successes of rigorous theory [5]. It is conjectured that routes in Steiner trees on random points have similar fractal properties [7]: route length between points a distance n apart should grow as n^γ for some $\gamma > 1$. However, as our construction shows, such results have little relevance to spatial network design.

5.3. The counterintuitive observation

The counterintuitive observation following Definition 1 follows quickly from the work of Theorem 1. Suppose that the configuration $\boldsymbol{x}^n$ is well dispersed, in the weak sense that, for some $\gamma \in (0, 1)$, we find the number of point pairs within $n^{\gamma/2}$ of each other is $o(\binom{n}{2}n^{\gamma-1})$ (certainly this is the case for most patterns generated by uniform random sampling from $[0, \sqrt{n}]^2$). Consider a network $G(\boldsymbol{x}^n)$ produced by augmenting the Steiner tree according to the construction in the proof of Theorem 1. Using the properties of this construction, the following can be shown:

$$\begin{aligned}\mathrm{E}[\mathrm{ratio}(G(\boldsymbol{x}^n))] &= \mathrm{E}\bigg[\underset{(i,j)}{\mathrm{average}}\,\frac{\ell(x_i, x_j)}{|x_i - x_j|} - 1\bigg]\\ &\le (\mathrm{constant})o(n^{\gamma-1}) + (1 - o(n^{\gamma-1}))\bigg(\frac{O(\log\sqrt{2n})}{n^{\gamma/2}}\bigg)\\ &\le O\bigg(\max\bigg(\frac{1}{n^{1-\gamma}}, \frac{\log n}{n^{\gamma/2}}\bigg)\bigg).\end{aligned}$$

5.4. Derandomisation

Theorem 1 is a purely deterministic assertion, though our proof used randomisation (supplied by the Poisson line process). It seems intuitively plausible that we could give a purely deterministic proof, say by replacing the Poisson line process with a suitable sparse set of deterministically positioned lines having a dense set of orientations.

5.5. Quantifying equidistribution

The classical equidistribution property, which states that the empirical distribution of

$$\{n^{-1/2}x_i^n,\ 1 \le i \le n\}$$

converges weakly to the uniform distribution on $[0, 1]^2$, is equivalent (by a straightforward argument) to the property that $\boldsymbol{x}^n$ is L_n-equidistributed as the uniform distribution on the square of area n, for *some* $L_n = o(n^{1/2})$. Replacing one sequence of L_n by another slower-growing sequence makes equidistribution a stronger assumption, and so our assumption in Theorem 2(a) (equidistribution for some $L_n = o(\log^{1/2} n)$) is stronger than the classical equidistribution property. Indeed Theorem 2 fails under the classical equidistribution property, as the following example shows.

Example 1. Let $L_n = n^\gamma$ for some $\gamma \in (\frac{3}{8}, \frac{1}{2})$. There exist networks $G(\boldsymbol{x}^n)$ which are L_n-equidistributed as the uniform distribution on the square of area n, for which $\mathrm{len}(G(\boldsymbol{x}^n)) = o(n)$ whilst $\mathrm{excess}(G(\boldsymbol{x}^n)) \to 0$.

For example, partition $[0, n^{1/2}]^2$ into subsquares of side length $L_n/\log n$, construct the complete graph on all centres of such subsquares, allocate the n points evenly amongst subsquares and position them arbitrarily close to the centres.

As is apparent from the nonstochastic condition implying L_n-equidistribution, there is a wide variety of configurations satisfying L_n-equidistribution. Here we consider the particular case of independent uniform sampling, and show that this generates an L_n-equidistributed sequence of configurations.

Remark 2. Sample the configuration $\boldsymbol{x}^n$ independently and uniformly from $[0, \sqrt{n}]^2$. Let L_n tend to ∞, perhaps arbitrarily slowly. Then the probability that the configuration $\boldsymbol{x}^n$ is L_n-equidistributed with the uniform distribution converges to 1. This follows by dividing $[0, \sqrt{n}]^2$ into cells of side length asymptotic to $L_n/\sqrt{2}$, by conditioning on $\boldsymbol{x}^n$, and by 'blurring' the points of $\boldsymbol{x}^n$ by replacing each point $x \in \boldsymbol{x}^n$ by an independent draw taken uniformly from the cell containing x. Then a uniform random draw $\tilde{Y}_n$ of one of the blurred points can be coupled to lie within L_n of a uniform random draw X_n from the finite configuration $\boldsymbol{x}^n$. A simple argument using the binomial distribution then shows that the total variation distance between

https://doi.org/10.1239/aap/1208358883 Published online by Cambridge University Press

$\tilde{Y}_n$ and uniform$([0, \sqrt{n}]^2)$ tends to 0; it follows that X_n can be coupled to a uniform$([0, \sqrt{n}]^2)$ random variable Y_n so that

$$\mathrm{E}\left[\min\left(1, \frac{|X_n - Y_n|}{L_n}\right) \,\middle|\, \boldsymbol{x}^n\right] \to 0,$$

where the convergence takes place in probability.

5.6. Poisson line process networks

Remark 1 indicates that more can be said about the mean semi-perimeter

$$\tfrac{1}{2}\,\mathrm{E}[\mathrm{len}(\partial\mathcal{C}(v_i, v_j))],$$

and this will be returned to in later work. For example, consider the network formed entirely from a Poisson line pattern. If the pattern is conditioned to contain points v_i and v_j then the perimeter $\partial\mathcal{C}(v_i, v_j)$ will be close to providing a genuine network geodesic.

Note that questions about $\mathcal{C}(v_i, v_j)$ bear a family resemblance to the D. G. Kendall conjecture about the asymptotic shape of large cells in a Poisson line pattern (see D. G. Kendall's foreword to [10]). However, $\mathcal{C}(v_i, v_j)$ is the result of a very explicit conditioning and, hence, explicit and rather complete answers can be obtained by direct methods, in contrast to the striking work on resolving the conjecture about large cells [2], [3], [4], [6].

5.7. An open question

In the random points model we can pose a more precise question. Over choices of network G subject to the constraint

$$\mathrm{E}[\mathrm{len}(G(\boldsymbol{x}^n)) - \mathrm{len}(\mathrm{ST}(\boldsymbol{x}^n))] = o(n),$$

or the constraint

$$\mathrm{E}[\mathrm{len}(G(\boldsymbol{x}^n))] = O(n),$$

what is the minimum value of $\mathrm{E}[\mathrm{excess}(G(\boldsymbol{x}^n))]$? Our results pin down this minimum value, in the latter case to the range $[\Omega(\sqrt{\log n}), O(\log n)]$ and in the former case to the range $[\Omega(\sqrt{\log n}), o(w_n \log n)]$. But it remains an open question what should be the exact order of magnitude.

Acknowledgement

The research of the first author was supported by N.S.F. Grant DMS-0203062.

References

[1] Gastner, M. T. and Newman, M. E. J. (2006). Shape and efficiency in spatial distribution networks. *J. Statist. Mech. Theory Experiment* **2006,** P01015.

[2] Hug, D., Reitzner, M. and Schneider, R. (2004). The limit shape of the zero cell in a stationary Poisson hyperplane tessellation. *Ann. Prob.* **32,** 1140–1167.

[3] Kovalenko, I. N. (1997). A proof of a conjecture of David Kendall on the shape of random polygons of large area. *Kibernet. Sistem. Anal.* **1997,** 187 (in Russian). English translation: *Cybernet. Systems Anal.* **33,** 461–467.

[4] Kovalenko, I. N. (1999). A simplified proof of a conjecture of D. G. Kendall concerning shapes of random polygons. *J. Appl. Math. Stoch. Anal.* **12,** 301–310.

[5] Lawler, G. F., Schramm, O. and Werner, W. (2004). Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Ann. Prob.* **32,** 939–995.

[6] Miles, R. E. (1995). A heuristic proof of a long-standing conjecture of D. G. Kendall concerning the shapes of certain large random polygons. *Adv. Appl. Prob.* **27,** 397–417.

[7] READ, N. (2005). Minimum spanning trees and random resistor networks in d dimensions. *Phys. Rev. E* **72,** 036114.
[8] SCHWEITZER, F., EBELING, W., ROSE, H. AND WEISS, O. (1998). Optimization of road networks using evolutionary strategies. *Evolutionary Comput.* **5,** 419–438.
[9] STEELE, J. M. (1997). *Probability Theory and Combinatorial Optimization* (CBMS-NSF Regional Conf. Ser. Appl. Math. **69**). Society for Industrial and Applied Mathematics, Philadelphia, PA.
[10] STOYAN, D., KENDALL, W. S. AND MECKE, J. (1995). *Stochastic Geometry and Its Applications*, 2nd edn. John Wiley, Chichester.
[11] YUKICH, J. E. (1998). *Probability Theory of Classical Euclidean Optimization Problems* (Lecture Notes Math. **1675**). Springer, Berlin.