



Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2005-040
MIT-LCS-TR-990

June 3, 2005

Theoretical Analysis of Geographic Routing in
Social Networks

Ravi Kumar, David Liben-Nowell, Jasmine Novak,
Prabhakar Raghavan, Andrew Tomkins

Theoretical Analysis of Geographic Routing in Social Networks

Ravi Kumar* David Liben-Nowell† Jasmine Novak* Prabhakar Raghavan‡
Andrew Tomkins*

Abstract

We introduce a formal model for geographic social networks, and introduce the notion of *rank-based friendship*, in which the probability that a person v is a friend of a person u is inversely proportional to the number of people w who live closer to u than v does. We then prove our main theorem, showing that rank-based friendship is a sufficient explanation of the navigability of any geographic social network that adheres to it.

1 A Model of Population Networks

There are two key features that we wish to incorporate into our social-network model: *geography* and *population density*. We will first describe a very general abstract model; in later sections we examine a concrete grid-based instantiation of it.

Definition 1.1 (Population network) *A population network is a 5-tuple $\langle L, d, P, \text{loc}, E \rangle$ where*

- L is a finite set of locations $(\ell, s, t, x, y, z, \dots)$;
- $d : L \times L \rightarrow \mathbb{R}^+$ is an arbitrary distance function on the locations;
- P is a finite ordered set of people (u, v, w, \dots) ;
- $\text{loc} : P \rightarrow L$ is the location function, which maps people to the location in which they live; and
- $E \subseteq P \times P$ is the set of friendships between people in the network.

The ordering on people is required only to break ties when comparing distances between two people. Let $\mathcal{P}(L)$ denote the power set of L .

Let $\text{pop} : L \rightarrow \mathbb{Z}^+$ denote the *population* of each point on L , where $\text{pop}(\ell) := |\{u \in P : \text{loc}(u) = \ell\}|$. We overload notation, and let $\text{pop} : \mathcal{P}(L) \rightarrow \mathbb{Z}^+$ denote the population of a subset of the locations, so that $\text{pop}(L') := \sum_{\ell \in L'} \text{pop}(\ell)$. We will write $n := \text{pop}(L) = |P|$ to denote the total population, and $m := |L|$ to denote the total number of locations in the network.

*IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120. ravi, jnovak, tomkins@almaden.ibm.com.

†MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), 32 Vassar Street, Cambridge, MA 02139. dln@theory.lcs.mit.edu. The majority of this work was performed while the author was visiting the IBM Almaden Research Center.

‡Verity, Inc., 892 Ross Drive, Sunnyvale, CA 94089. pragh@verity.com

Let $\text{density} : L \rightarrow [0, 1]$ be a probability distribution denoting the *population density* of each location $\ell \in L$, so that $\text{density}(\ell) := \text{pop}(\ell)/n$. We similarly extend $\text{density} : \mathcal{P}(L) \rightarrow [0, 1]$ so that $\text{density}(L') = \sum_{\ell \in L'} \text{density}(\ell)$. Thus $\text{density}(L) = 1$.

We extend the distance function to accept both locations and people in its arguments, so that we have the function $d : (P \cup L) \times (P \cup L) \rightarrow \mathbb{R}^+$ where $d(u, \cdot) := d(\text{loc}(u), \cdot)$ and $d(\cdot, v) := d(\cdot, \text{loc}(v))$ for all people $u, v \in P$.

When comparing the distances between people, we will use the ordering on people to break ties. For people $u, v, v' \in P$, we will write $d(u, v) < d(u, v')$ as shorthand for $\langle d(u, v), v \rangle \prec_{\text{lex}} \langle d(u, v'), v' \rangle$. This tie-breaking role is the only purpose of the ordering on people.

2 Rank-Based Friendship

Following the navigable-small-world model of Kleinberg [1, 2], each person in the network will be endowed with one *long-range link*. We diverge from the model of Kleinberg in the definition of our long-range links. Instead of distance, the fundamental quantity upon which we base our model of long-range links is *rank*:

Definition 2.1 (Rank) For two people $u, v \in P$, the rank of v with respect to u is defined as

$$\text{rank}_u(v) := |\{w \in P : d(u, w) < d(u, v)\}|.$$

Note that by breaking ties in distance consistently according to the ordering on P , for any $i \in \{1, \dots, n\}$ and any person $u \in P$, there is exactly one person v such that $\text{rank}_u(v) = i$. We now define a model for generating a *rank-based* social network using this notion:

Definition 2.2 (Rank-based friendship) For each person u in the network, we generate one long-range link from u , where

$$\Pr[u \text{ links to } v] \propto \frac{1}{\text{rank}_u(v)}.$$

Intuitively, one justification for rank-based friendship is the following: person v will have to compete with all of the more “convenient” candidate friends for person u , i.e., all people w who live closer to u than v does. Note that, for any person u , we have $\sum_v 1/\text{rank}_u(v) = \sum_{i=1}^n 1/i = H_n$, the n th harmonic number. Therefore, by normalizing, we have that

$$\Pr[u \text{ links to } v] = \frac{1}{H_n \cdot \text{rank}_u(v)}. \tag{1}$$

Under rank-based friendship, the probability of a link from u to v depends only on the number of people within distance $d(u, v)$ of u , and not on the geographic distance itself.

One important feature of this model is that it is *independent* of the dimensionality of the space in which people live. For example, in the k -dimensional grid with uniform population density and the L_1 distance on locations, we have that $|\{w : d(u, w) \leq \delta\}| \propto \delta^k$, so the probability that person u links to person v is proportional to $d(u, v)^{-k}$. That is, the rank of a person v with respect to a person u satisfies $\text{rank}_u(v) \approx d(u, v)^k$. Thus, although this model has been defined without explicitly embedding the locations in a metric space, our rank-based formulation gives essentially the same long-distance link probabilities as Kleinberg’s model for a uniform-population k -dimensional mesh.

3 Rank-Based Friendship on Meshes

In the following, our interest will lie in population networks that are formed from meshes with arbitrary population densities. Let $L := \{1, \dots, q\}^k$ denote the points on the k -dimensional mesh, with length q on each side. We write $x = \langle x_1, \dots, x_k \rangle$ for a location $x \in L$. We will consider Manhattan distance (L_1 distance) on the mesh, so that $d(\langle x_1, \dots, x_k \rangle, \langle y_1, \dots, y_k \rangle) := \sum_{i=1}^k |x_i - y_i|$. The only restriction that we impose on the people in the network is that $\text{pop}(\ell) > 0$ for every $\ell \in L$ —that is, there are no ghost towns with zero population. This assumption will allow us to avoid the issue of disconnected sets in what follows.

Thus a *mesh population network* is fully specified by the dimensionality k , the side length q , the population P (with an ordering to break ties in interpersonal distances), the friendship set E , and the location function $\text{loc} : P \rightarrow \{1, \dots, q\}^k$, where for every location $\ell \in \{1, \dots, q\}^k$, there exists a person $u_\ell \in P$ such that $\text{loc}(u_\ell) = \ell$.

Following the Kleinberg’s model of navigable small worlds, we include *local links* in E for each person in the network. For now, we assume that each person u at location $\ell^{(u)} = \text{loc}(u)$ in the network has a local link to some person at the mesh point in each cardinal direction from $\ell^{(u)}$, i.e., to some person at each of the $2k$ points $\langle \ell_1^{(u)}, \dots, \ell_{i-1}^{(u)}, \ell_i^{(u)} \pm 1, \ell_{i+1}^{(u)}, \dots, \ell_k^{(u)} \rangle$, for any coordinate $i \in \{1, \dots, k\}$. Thus, for any two people u and v , there exists a path of length at most qk between them, and, more specifically, the geographically greedy algorithm will find a path of length no longer than qk .

In a *rank-based mesh population network*, we add one long-range link to E per person in P , where that link is chosen probabilistically by rank, according to (1).

4 The Two-Dimensional Grid

For concreteness, we focus on the two-dimensional grid, where we have $L := \{1, \dots, q\} \times \{1, \dots, q\}$, and thus $m = |L| = q^2$. We may think of the two-dimensional grid as representing the intersection of integral lines of longitude and latitude, for example.

In this section, we will show that the geographically greedy algorithm on the two-dimensional grid produces paths that are on average very short—more precisely, that the expected length of the path found by the geographically greedy algorithm is bounded by $O(\log^3 n)$ when the target is chosen randomly from the population P . Formally, the geographically greedy algorithm *GeoGreedy* proceeds as follows: given a target t and a current message-holder u , person u examines her set of friends, and forwards the message to the friend v of u who is geographically closest to the target t . First, we need a few definitions:

Definition 4.1 (L_1 -Ball) For any location $x \in L$ and for any radius $r \geq 0$, let

$$B_r(x) = \{y \in L : d(x, y) \leq r\} = \{y \in L : |x_1 - y_1| + |x_2 - y_2| \leq r\}$$

denote the L_1 -ball of radius r centered at location x .

We consider an exponentially growing set $\mathcal{R} := \{2^i : i \in \{1, 2, 4, \dots, 128 \lceil \log q \rceil\}\}$ of ball radii, and we place a series of increasingly fine-grained collections of balls that cover the grid:

Definition 4.2 (Covering Radius- r Ball Centers) For any radius $r \in \mathcal{R}$, let the set

$$\mathcal{C}_r := \{z \in L : 2z_1 \bmod r = 2z_2 \bmod r = 0\}$$

be the set of locations z such that z_i/r is half-integral for $i \in \{1, 2\}$.

For each radius $r \in \mathcal{R}$, we will consider the set of radius- r balls centered at each of the locations in \mathcal{C}_r . We begin with a few simple facts about these L_1 -balls:

Fact 4.3 (Only a small number of balls in \mathcal{C}_r overlap) For each each radius $r \in \mathcal{R}$:

1. For each location $x \in L$, we have that $|\{z \in \mathcal{C}_r : d(z, x) \leq r\}| \leq 25$.
2. For each location $z \in \mathcal{C}_r$, we have that $|\{z' \in \mathcal{C}_{r/2} : B_{r/2}(z') \cap B_r(z) \neq \emptyset\}| \leq 169$.

Proof. For the first claim, note that if $|z_1 - x_1| > r$ or if $|z_2 - x_2| > r$, then $d(z, x) > r$, and z is not an element of the set of relevance. Thus every $z \in \mathcal{C}_r$ such that $d(z, x) \leq r$ must fall into the range $\langle x_1 \pm r, x_2 \pm r \rangle$. There are at most five half-integral values of z/r that can fall into the range $[b, b + 2r]$ for any b , so there are at most twenty-five total points $z \in \mathcal{C}_r$ that satisfy $d(x, z) \leq r$.

For the second claim, notice that any ball of radius $r/2$ that has a nonempty intersection with $B_r(z)$ must have its center at a point z' such that $d(z, z') \leq 3r/2$. Thus the only $z' \in \mathcal{C}_{r/2}$ that could be in the set of relevance must have $z'_i \in [z_i - 3r/2, z_i + 3r/2]$ for $i \in \{1, 2\}$ and have $2z'_i/(r/2)$ be half-integral. As in the first claim, the number of half-integral values of $2z'/r$ that can fall into the range $[b, b + 3r]$ is at most thirteen for any b . Thus there can be at most 169 total points $z' \in \mathcal{C}_{r/2}$ so that $B_{r/2}(z') \cap B_r(z) \neq \emptyset$. \square

Fact 4.4 (Relation between balls centered in L and in \mathcal{C}_r) For each location $x \in L$ and for each radius $r \in \mathcal{R}$:

1. There exists a location $z \in \mathcal{C}_r$ such that $B_{r/2}(x) \subseteq B_r(z)$.
2. There exists a location $z' \in \mathcal{C}_{r/2}$ such that $B_{r/2}(z') \subseteq B_r(x)$ and $x \in B_{r/2}(z')$.

Proof. For the first claim, let $z \in \mathcal{C}_r$ be the closest point to x in \mathcal{C}_r . Note that $x_1 \in [z_1 - r/4, z_1 + r/4]$; otherwise x would be strictly closer to either $\langle z_1 - r/2, z_2 \rangle \in \mathcal{C}_r$ or $\langle z_1 + r/2, z_2 \rangle \in \mathcal{C}_r$. Similarly we have $x_2 \in [z_2 - r/4, z_2 + r/4]$. Therefore we have $d(x, z) = \sum_{i \in \{1, 2\}} |x_i - z_i| \leq r/2$. Let $y \in B_{r/2}(x)$ be arbitrary. Then by the triangle inequality we have $d(z, y) \leq d(z, x) + d(x, y) \leq r/2 + r/2 = r$. Thus we have $y \in B_r(z)$, which proves the claim.

For the second claim, let $z' \in \mathcal{C}_{r/2}$ be the closest point to x in $\mathcal{C}_{r/2}$. By the same argument as above, we have $d(x, z') \leq r/4$. Immediately we have $x \in B_{r/2}(z')$. Let $y \in B_{r/2}(z')$ be arbitrary. Then $d(x, y) \leq d(x, z') + d(z', y) \leq r/4 + r/2 < r$, and $y \in B_r(x)$, which proves the claim. \square

Let x and y be two arbitrary locations in L . In what follows, we will use the size of the smallest ball in $\bigcup_{r \in \mathcal{R}} \{B_r(z) : z \in \mathcal{C}_r\}$ that includes both x and y as a ceiling-like proxy for $d(x, y)$, and as the measure of progress towards the target. We will also need a large ball from $\{B_r(z) : z \in \mathcal{C}_r\}$ that includes both x and y and also includes a large ball centered at y .

Definition 4.5 (Minimum enclosing-ball radius) For any two locations $x, y \in L$, let $\text{mebr}(x, y)$ (“minimum enclosing-ball radius”) denote the minimum $r \in \mathcal{R}$ such that, for some $z \in \mathcal{C}_r$, we have $x, y \in B_r(z)$.

Fact 4.6 (Relating distance and minimum enclosing-ball radius) For any $x, y \in L$, let $r := \text{mebr}(x, y)$. Then we have $2r \geq d(x, y) \geq r/4$.

Proof. Let $z \in \mathcal{C}_r$ be such that $x, y \in B_r(z)$, and note that, by definition, there is no $z' \in \mathcal{C}_{r/2}$ such that $x, y \in B_r(z')$. The first direction is easy: by the triangle inequality, we have that $d(x, y) \leq d(x, z) + d(z, y) \leq r + r = 2r$. For the other direction, suppose for a contradiction that $d(x, y) \leq r/4$. Let $z^* \in \mathcal{C}_{r/2}$ be such that $B_{r/4}(x) \subseteq B_{r/2}(z^*)$, as guaranteed by Fact 1. But then we have $x, y \in B_{r/4}(x)$ because $d(x, y) \leq r/4$, which implies that $x, y \in B_{r/2}(z^*)$, which in turn contradicts the minimality of r . \square

Thus, in the path from any source $s \in L$ to any target $t \in L$ found by GeoGreedy, the path will always remain inside the ball $B_{d(s,t)}(t) \subseteq B_{2 \cdot \text{mebr}(s,t)}(t)$.

Definition 4.7 (Sixteenfold enclosing ball) Let $x, y \in L$ be an arbitrary pair of locations, and let $r = \text{mebr}(x, y)$. Let $\text{sebc}(y, r)$ (“sixteenfold enclosing ball center”) denote the location $z_{y,r}^* \in \mathcal{C}_{16r}$ such that $B_{8r}(y) \subseteq B_{16r}(z_{y,r}^*)$ whose existence is guaranteed by Fact 1.

Lemma 4.8 (Relationship between ball population and rank) Let $s, t \in L$ be an arbitrary source/target pair of locations. Let $r = \text{mebr}(s, t)$, and let $z^* = \text{sebc}(t, r)$. Let $x, y \in L$ be arbitrary locations such that $x \in B_{2r}(t)$ and $y \in B_{r/8}(t)$, and let $u, v \in P$ be arbitrary people such that $\text{loc}(u) = x$ and $\text{loc}(v) = y$. Then $\text{rank}_u(v) \leq \text{pop}(B_{16r}(z^*))$.

Proof. First, we note

$$\begin{aligned} d(x, y) &\leq d(x, t) + d(t, y) && \text{triangle inequality} \\ &\leq 2r + r/8 && \text{assumptions that } x \in B_{2r}(t) \text{ and } y \in B_{r/8}(t) \\ &= 17r/8. \end{aligned} \tag{2}$$

We now claim the following:

$$\text{for any location } \ell \in L, \text{ if } d(x, \ell) \leq d(x, y), \text{ then } d(z^*, \ell) \leq 16r. \tag{3}$$

To prove (3), let ℓ be an arbitrary location so that $d(x, \ell) \leq d(x, y)$. Then we have

$$\begin{aligned} d(t, \ell) &\leq d(t, y) + d(y, x) + d(x, \ell) && \text{triangle inequality} \\ &\leq d(t, y) + d(y, x) + d(x, y) && \text{assumption that } d(x, \ell) \leq d(x, y) \\ &\leq r/8 + d(y, x) + d(x, y) && \text{assumption that } y \in B_{r/8}(t) \\ &\leq r/8 + 17r/8 + 17r/8 && (2) \\ &= 35r/8. \end{aligned}$$

Then, we have that $\ell \in B_{35r/8}(t) \subseteq B_{8r}(t) \subseteq B_{16r}(z^*)$ by the definition of $z^* = \text{sebc}(t, r)$, which proves (3). Now, by definition of rank, we have that

$$\begin{aligned} \text{rank}_u(v) &\leq |\{w \in P : d(u, w) \leq d(u, v)\}| \\ &= \sum_{\ell \in L: d(x, \ell) \leq d(x, y)} \text{pop}(\ell) \\ &= \text{pop}(\{\ell \in L : d(x, \ell) \leq d(x, y)\}) \\ &\leq \text{pop}(\{\ell \in L : d(\ell, z^*) \leq 16r\}) \\ &= \text{pop}(B_{16r}(z^*)) \end{aligned}$$

where the second inequality follows from (3). \square

We are now ready to prove the main technical result of this section, namely that the geographically greedy algorithm will halve the distance from the source to the target in a polylogarithmic expected number of steps, for a randomly chosen target person.

Lemma 4.9 (GeoGreedy halves distance in expected polylogarithmic steps) *Let $s \in L$ be an arbitrary source location, and let $t \in L$ be a randomly chosen target location, according to the distribution density (\cdot) . Then the expected number of steps before the geographically greedy algorithm started from location s reaches a point in $B_{d(s,t)/2}(t)$ is $O(\log n \log m) = O(\log^2 n)$, where the expectation is taken over the random choice of t .*

Proof. Let $r_t := \text{mebr}(s, t)$, and let $z_t := \text{sebc}(t, r_t)$ so that

$$z_t \in \mathcal{C}_{16r_t} \text{ and } B_{8r_t}(t) \subseteq B_{16r_t}(z_t). \quad (4)$$

Let z'_t be the location whose existence is guaranteed by Fact 2 such that

$$z'_t \in \mathcal{C}_{r_t/16} \text{ and } B_{r_t/16}(z'_t) \subseteq B_{r_t/8}(t) \text{ and } t \in B_{r_t/16}(z'_t). \quad (5)$$

Putting together (4) and (5), we have the following two facts:

$$B_{r_t/16}(z'_t) \subseteq B_{r_t/8}(t) \subseteq B_{8r_t}(t) \subseteq B_{16r_t}(z_t) \quad (6)$$

$$t \in B_{r_t/16}(z'_t). \quad (7)$$

By Fact 4.6, we know that $d(s, t)/2 \geq r_t/8$. Thus it will suffice to show that the expected number of steps before GeoGreedy started from location s lands in $B_{r_t/8}(t) \subseteq B_{d(s,t)/2}(t)$ is $O(\log n \log m)$.

Suppose that we start GeoGreedy at the source s , and the current point on the path found by the algorithm is some person $u \in P$, at location $x_u = \text{loc}(u)$. By definition, every step of GeoGreedy decreases the distance from the current location to the target t , so we have that

$$d(x_u, t) \leq d(s, t) \leq 2r_t. \quad (8)$$

We refer to a person u as *good* if there exists a long-range link from that person to any person living in the ball $B_{r_t/8}(t)$. Let $\alpha_{u,t}$ denote the probability that a person $u \in P$ living at location $x_u = \text{loc}(u) \in L$ is good. Then

$$\alpha_{u,t} = \sum_{v:\text{loc}(v) \in B_{r_t/8}(t)} \frac{1}{\text{rank}_u(v) \cdot H_n} \geq \sum_{v:\text{loc}(v) \in B_{r_t/8}(t)} \frac{1}{\text{pop}(B_{16r_t}(z_t)) \cdot H_n} = \frac{\text{pop}(B_{r_t/8}(t))}{\text{pop}(B_{16r_t}(z_t)) \cdot H_n}$$

by the definition of good, by Lemma 4.8 (which applies by (8)), and by the definition of $\text{pop}(\cdot)$. Noting that the lower bound on $\alpha_{u,t}$ is independent of u , we write

$$\alpha_t := \frac{\text{pop}(B_{r_t/8}(t))}{\text{pop}(B_{16r_t}(z_t)) \cdot H_n} \leq \alpha_{u,t}.$$

Thus the probability that u is good is at least α_t for every person u along the GeoGreedy path. Furthermore, each step of the algorithm brings us to a new node never seen before by the algorithm because the distance to t is strictly decreasing until we reach node t . Thus the probability of finding a good long-range link is independent at each step of the algorithm until it terminates. Therefore, the expected number of steps before we reach a good person (or t itself) is at most $1/\alpha_t$.

We now examine the expected value of $1/\alpha_t$ for a target location $t \in L$ chosen according to the distribution density(\cdot):

$$\begin{aligned}
\mathbf{E}_{t \in L \sim \text{density}(\cdot)}[1/\alpha_t] &= \sum_t \text{density}(t) \cdot \frac{1}{\alpha_t} \\
&= \sum_t \text{density}(t) \cdot \frac{\text{pop}(B_{16r_t}(z_t)) \cdot H_n}{\text{pop}(B_{r_t/8}(t))} \\
&= H_n \cdot \sum_t \text{density}(t) \cdot \frac{\text{density}(B_{16r_t}(z_t))}{\text{density}(B_{r_t/8}(t))} \\
&\leq H_n \cdot \sum_t \text{density}(t) \cdot \frac{\text{density}(B_{16r_t}(z_t))}{\text{density}(B_{r_t/16}(z'_t))}.
\end{aligned}$$

The equalities follow the definition of expectation, the definition of α_t , and from the fact that $\text{density}(\cdot) = \text{pop}(\cdot)/n$. The inequality follows from the definition of z'_t in (5), using the fact that $B_{r_t/16}(z'_t) \subseteq B_{r_t/8}(t)$ and the monotonicity of $\text{density}(\cdot)$.

We now reindex the summation to be over radii and ball centers rather than over targets t . Recall that $z_t \in \mathcal{C}_{16r_t}$ and $z'_t \in \mathcal{C}_{r_t/16}$, and that $B_{r_t/16}(z'_t) \subseteq B_{16r_t}(z_t)$ by (6), and therefore that $z'_t \in B_{16r_t}(z_t)$. Thus, we have that

$$\begin{aligned}
\mathbf{E}_{t \in L \sim \text{density}(\cdot)}[1/\alpha_t] &\leq H_n \cdot \sum_t \text{density}(t) \cdot \frac{\text{density}(B_{16r_t}(z_t))}{\text{density}(B_{r_t/16}(z'_t))} \\
&\leq H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \sum_{z' \in \mathcal{C}_{r/16}: z' \in B_{16r}(z)} \frac{\text{density}(B_{16r}(z))}{\text{density}(B_{r/16}(z'))} \sum_{t: z'_t = z'} \text{density}(t) \\
&\leq H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \sum_{z' \in \mathcal{C}_{r/16}: z' \in B_{16r}(z)} \frac{\text{density}(B_{16r}(z))}{\text{density}(B_{r/16}(z'))} \sum_{t \in B_{r/16}(z')} \text{density}(t)
\end{aligned}$$

where the last inequality follows from (7). But then

$$\begin{aligned}
\mathbf{E}_{t \in L \sim \text{density}(\cdot)}[1/\alpha_t] &\leq H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \sum_{z' \in \mathcal{C}_{r/16}: z' \in B_{16r}(z)} \frac{\text{density}(B_{16r}(z))}{\text{density}(B_{r/16}(z'))} \sum_{t \in B_{r/16}(z')} \text{density}(t) \\
&= H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \sum_{z' \in \mathcal{C}_{r/16}: z' \in B_{16r}(z)} \frac{\text{density}(B_{16r}(z))}{\text{density}(B_{r/16}(z'))} \cdot \text{density}(B_{r/16}(z')) \\
&= H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \sum_{z' \in \mathcal{C}_{r/16}: z' \in B_{16r}(z)} \text{density}(B_{16r}(z)) \\
&= H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \text{density}(B_{16r}(z)) \cdot |\{z' \in \mathcal{C}_{r/16} : z' \in B_{16r}(z)\}|.
\end{aligned}$$

Now we are almost done: by applying Fact 2 a constant number of times, we have that

$$|\{z' \in \mathcal{C}_{r/16} : z' \in B_{16r}(z)\}| = O(1). \quad (9)$$

Furthermore, we have $\sum_{z \in \mathcal{C}_{16r}} \text{density}(B_{16r}(z)) \leq 25$: by Fact 1, there are at most twenty-five balls in \mathcal{C}_r that include any particular location, so we are simply summing a probability distribution

with some “double counting,” but counting each point at most twenty-five times. Thus we have

$$\begin{aligned}
\mathbb{E}_{t \in L \sim \text{density}(\cdot)}[1/\alpha_t] &\leq H_n \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \text{density}(B_{16r}(z)) \cdot |\{z' \in \mathcal{C}_{r/16} : z' \in B_{16r}(z)\}| \\
&\leq H_n \cdot O(1) \cdot \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{C}_{16r}} \text{density}(B_{16r}(z)) \\
&\leq H_n \cdot O(1) \cdot \sum_{r \in \mathcal{R}} 25 \\
&= H_n \cdot O(1) \cdot 25 \cdot |\mathcal{R}| = H_n \cdot O(\log q) = O(\log n \log m).
\end{aligned}$$

because $|\mathcal{R}| = \Theta(\log q) = \Theta(\log m)$. □

In the case of uniform population density, the value of $\alpha_{u,t} = \Omega(1/\log n)$ is independent of s and t , and the greedy algorithm finds an s - t path of length $O(\log^2 n)$ with high probability [1, 2].

Theorem 4.10 (GeoGreedy finds short paths in all 2-D meshes) *For any 2-dimensional mesh population network with n people and m locations, the expected length of the search path found by GeoGreedy from an arbitrarily chosen source location s and a uniformly chosen target t is $O(\log n \log^2 m) = O(\log^3 n)$.*

Proof. Immediate by inductive application of Lemma 4.8: the expected number of hops required before moving to a node s' with $d(s', t) \leq d(s, t)/2$ or t itself is $O(\log n \log m)$; by repeating this process $O(\log(\max_{s,t} d(s, t))) = O(\log qk) = O(\log q^k) = O(\log m)$ times, we must arrive at the target node t itself. □

References

- [1] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Symposium on the Theory of Computation*, volume 32, 2000.
- [2] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406:845, 24 August 2000.