Navigating Low-Dimensional and Hierarchical Population Networks*

Ravi Kumar
Yahoo! Research
ravikumar@yahoo-inc.com

David Liben-Nowell
Department of Computer Science
Carleton College
dlibenno@carleton.edu

Andrew Tomkins
Yahoo! Research
atomkins@yahoo-inc.com

Abstract

Social networks are *navigable small worlds*, in which two arbitrary people are likely connected by a short path of intermediate friends that can be found by a "decentralized" routing algorithm using only local information. We develop a model of social networks based on an arbitrary metric space of points, with population density varying across the points. We consider *rank-based friendships*, where the probability that person u befriends person v is inversely proportional to the number of people who are closer to u than v is. Our main result is that greedy routing can find a short path (of expected polylogarithmic length) from an arbitrary source to a randomly chosen target, independent of the population densities, as long as the doubling dimension of the metric space of locations is low. We also show that greedy routing finds short paths with good probability in tree-based metrics with varying population distributions.

1 Introduction

The last few years have witnessed increased interest in measuring, modeling, and exploiting *social networks*—collections of people connected by edges representing acquaintance, friendship, or other social relationships. Numerous internet startups have arisen predicating that one's social network requires the same careful husbandry as one's credit rating or investment portfolio. A common focus of scientific studies of social networks is the *small-world phenomenon*, the observation that most pairs of people are connected through short chains of friends. A remarkable experiment of Stanley Milgram [24] in the 1960s empirically validated this hypothesis, showing that two typical people in the United States were connected by a chain of acquaintances with an average length of six, thereby introducing the concept of "six degrees of separation" into popular culture. It is surprising that short paths exist, but it is remarkable that members of the network are able to discover these short paths with only information about their local neighborhood and some scant information about the destination [16]. Thus, Milgram's experiment suggests not only that social networks have small diameter but also that they admit efficient "decentralized" search.

*A preliminary version of this paper appears in *Proceedings of the 14th Annual European Symposium on Algorithms (ESA'06)*, September 2006, pages 480–491. Comments are welcome.

There have been a number of recent models of social networks that attempt to include an account of observed social-network properties like the small-world phenomenon. Watts and Strogatz [31] proposed a model consisting of two superimposed sets of edges, a structured set of edges and a smaller number of random edges. The former is meant to capture "typical" social friendships created, say, by geographic proximity. The latter is meant to capture long-range connections; these edges reduce the network's diameter, but do not explain navigability. Kleinberg [16, 15] proposed a simple model that suffices to produce a navigable small world. The underlying network is a k-dimensional grid, and each person occupies a unique grid location. As in the Watts–Strogatz model, the network has two kinds of links. Each person is connected via short-range links to her immediate neighbors in the grid, and she has one long-range link, chosen randomly so that the probability that a person u befriends a person v is proportional to $d(u,v)^{-\alpha}$, where $d(u,v)$ is the lattice distance between u and v, and $\alpha \geq 0$ is a parameter of the model. Kleinberg studied *greedy routing*—to route a message from s to t, person s sends the message to the neighbor of s who is closest in lattice distance to t—and showed that with high probability this simple algorithm finds paths of length polylogarithmic in the population size if and only if $\alpha = k$.

As with most models of complex real-world phenomena, Kleinberg's mathematically appealing model makes certain simplifying assumptions. In particular, it postulates that each grid point is occupied by a single individual, and hence the grid exhibits uniform population density—a significant deviation from most real-world populations. This issue has been addressed by two subsequent models, both designed to handle nonuniform population distributions without compromising analytical tractability. The first is Kleinberg's *group-structure* model, based on set systems [17]. The second is *rank-based friendship*, which we proposed in joint work with Jasmine Novak and Prabhakar Raghavan [21]. We showed empirically that the geographic distribution of friendships in the LiveJournal blogging community closely matches rank-based friendship when the population is modeled in two-dimensional Euclidean space, and we proved that short paths can be found in two-dimensional grids by greedy routing [21, 19].

In this paper, we focus on rigorous analysis of rank-based friendship in a wide variety of social-network settings. For intuition on this model, consider two people u and v who live 500 meters apart. In rural Minnesota, say, u and v are probably next-door neighbors, and are very likely to know each other; in Manhattan, there may be more than 10,000 people who live closer to u than v does, and u and v have probably never met. This discrepancy suggests why distance alone is insufficient as the basis for a model of real friendships. Instead, we model long-range links using a notion of "rank": the *rank* of a person v with respect to u is the number of people who live at least as close to u as v does. Thus, the closest candidate friend to u has rank 1, the next one has rank 2, and so forth. In rank-based friendship, the probability that u befriends v is inversely proportional to the rank of v with respect to u. (Intuitively, to be befriended by u, person v will have to compete with all of the more "convenient" candidate friends for u, i.e., all people w who live closer to u than v does.) An important feature of this model is that it implicitly accounts for the dimensionality of the space in which people live. For example, in a k-dimensional grid of uniform population density, for any k, the rank-based formulation induces exactly the unique value of α proven by Kleinberg to generate a navigable network.

Our contributions. In this paper, we study the properties of rank-based friendship. Our main theorem shows that greedy routing discovers short paths for arbitrary (not necessarily uniform) population densities on any metric space of low doubling dimension. We extend the main theorem in two directions. First, we present a recursive formulation of a population in which it is possible to route a message to the city of Manhattan, then to the appropriate block of the city, and finally to the floor of the building where the target person lives. We show that the theorem holds even for recursive structures of polynomial depth. Next, we analyze greedy routing and rank-based friendship under tree metrics, which have been proposed to capture non-geographic proximity between individuals in a social network [17, 30]. (One can naturally model the proximity of two hobbies or occupations, e.g., through a tree.) Through a more detailed analysis, we are able to remove the notion of "local neighbors" entirely in this case.

In Sect. 2, we present background material on metric spaces, population networks, and routing algorithms, and we define a notion of a social structure, which can be formed from an arbitrary metric space. Points in the metric space correspond to locations, at which many people may reside. In Sect. 3, we formally define rank-based friendship, and we construct social networks with long-range links generated via rank-based friendship. We also define a general notion of short-range links based on the metric space. In Sect. 4, we present our main result on greedy routing in these social networks. Specifically, we show that greedy routing finds a path from an arbitrary source person to the location of a target person chosen uniformly at random from the population. The expected length of this path is polylogarithmic in the size of the population and exponential in the doubling dimension of the metric space. Thus, if the underlying metric space has low doubling dimension—like a constant-dimensional grid—greedy routing yields expected polylogarithmic paths. We then turn to our two extensions. In Sect. 5, we describe recursive population networks and analyze greedy routing in these networks; due to potentially polynomial-depth leaves in the tree of locations, we must adopt a more complex measure of progress towards the target. Finally, in Sect. 6, we investigate tree social structures, in which the points of the population network are leaves of a tree.

Other related work. There have been several relevant extensions to Kleinberg's original model, which we review here. In k-dimensional grids, there has also been considerable work on upper and lower bounds for the diameter and the length of the greedy path (e.g., [23, 26, 4]), and partially decentralized algorithms other than greedy routing have also been considered [8, 20, 22, 23, 27]. Kleinberg has extended his model to tree-based structures and group structures [17]. Fraigniaud analyzes circumstances under which a graph can be augmented to be navigable, and gives a positive answer in the case of bounded treewidth or bounded chordality [9]. Analysis of navigability in (uniform-population) networks of low doubling dimension has been performed by Duchon et al. [7] and Slivkins [28]. Broadly speaking, these papers give stronger bounds on network navigability than the present work, but are limited to uniform populations; the treatment of nonuniform population distributions is the major contribution of this paper.

In previous work with J. Novak and P. Raghavan, we defined rank-based friendship and referenced a technical report that includes a theorem regarding population networks on two-dimensional grids [21, 19]. The current paper gives a significantly more general analysis of rank-based friendship, and subsumes the particular theorem contained in that technical report. Notions similar to rank-based friendships can also be found in geometric data structures [5, 10].

The question of searching in social networks was also considered by Adamic et al. [1, 2, 3] and Kim et al. [13], and Milgram's experiment was replicated in a larger-scale email setting by Dodds et al. [6]. For a comprehensive treatment of social networks, good sources include the book by Wasserman and Faust [29] and a recent survey by Kleinberg [18].

2 Preliminaries

Background on metric spaces. Let $\langle X, d\rangle$ be a metric space. Denote by $B_r(x) := \{y \in X : d(x,y) < r\}$ the open radius-r ball around point $x \in X$. Define the *aspect ratio* as $\Delta := \max_{x,y \in X} d(x,y) / \min_{x,y \in X, x \neq y} d(x,y)$. The *doubling dimension* of $\langle X, d\rangle$ is the smallest α such that, for every $r > 0$ and every $Y \subseteq X$ of diameter $2r$, Y can be covered by at most 2^α subsets of diameter r.

Social structures and population networks. A *social network* is a graph $\langle P, E\rangle$, where a node represents a person and an edge represents a friendship between its endpoints. Edges are directed to allow nonreciprocal friendships. Let $\Gamma(u)$ denote the out-neighbors of $u \in P$.

A *social structure* is a quadruple $\langle L, d, P, \mathsf{loc}\rangle$, where L is a finite set of *points*; $d : L \times L \to \mathbb{R}^{\geq 0}$ is a *distance metric* on the points (so $\langle L, d\rangle$ is a metric space); P is an ordered finite set of *people*; and $\mathsf{loc} : P \to L$ is the *location function*, which maps people to the point in which they live. For convenience, we assume that d is scaled so that $\min_{\ell,\ell' \in L, \ell \neq \ell'} d(\ell, \ell') = 1$. Extend d to $d : (P \cup L) \times (P \cup L) \to \mathbb{R}^{\geq 0}$ where $d(u, \cdot) := d(\mathsf{loc}(u), \cdot)$ and $d(\cdot, v) := d(\cdot, \mathsf{loc}(v))$ for all people $u, v \in P$. We use the ordering on P to break ties in comparing distances: for people $u, v, v' \in P$, write $d(u,v) < d(u,v')$ as shorthand for $\langle d(u,v), v\rangle \prec_{lexicographic} \langle d(u,v'), v'\rangle$, where the ordering on the second component is given by the ordering on P. This tie-breaking role is the only purpose of the ordering on people.

A *population network* is a quintuple $\langle L, d, P, \mathsf{loc}, E\rangle$ where $\langle L, d, P, \mathsf{loc}\rangle$ is a social structure and $E \subseteq P \times P$ is a set of *friendships*. (Thus $\langle P, E\rangle$ is a social network.) Let $\mathsf{pop}(\ell) := |\{u \in P : \mathsf{loc}(u) = \ell\}|$ denote the *population* of $\ell \in L$. Extend pop so that $\mathsf{pop}(L') := \sum_{\ell \in L'} \mathsf{pop}(\ell)$ for a subset $L' \subseteq L$ of the points. Write $n := \mathsf{pop}(L) = |P|$ for the total population. Let $\mathsf{dens} : L \to [0,1]$ be a probability distribution denoting the *population density* of each point $\ell \in L$, so that $\mathsf{dens}(\ell) := \mathsf{pop}(\ell)/n$. As before, we extend dens to allow us to write $\mathsf{dens}(L') := \sum_{\ell \in L'} \mathsf{dens}(\ell)$ for $L' \subseteq L$.

Routing algorithms. Given a population network $\langle L, d, P, \mathsf{loc}, E\rangle$, a *source individual* $s \in P$, and a *target individual* $t \in P$, a *routing algorithm* seeks a path $\rho = \langle u_0, u_1, \ldots, u_k\rangle$ from $s = u_0$ to $t = u_k$ in the graph $\langle P, E\rangle$.

We are interested in routing algorithms that compute the next step u_{i+1} from the current person u_i without taking the entire graph $\langle P, E\rangle$ as input. The algorithm is *decentralized* if, when computing the next step u_{i+1} in the path, the only information used is u_i, t, the social structure $\langle L, d, P, \mathsf{loc}\rangle$, and the set of neighbors $\Gamma(u_i)$ of the current node u_i. (That is, the edges in E excluding those incident to u_i are not available as input to the decentralized algorithm.) In this paper, we focus on one particular decentralized algorithm: the *greedy algorithm* Greedy selects $u_{i+1} := \mathrm{argmin}_{v \in \Gamma(u_i)} d(v, t)$.

3 Rank-based friendship

For two people $u, v \in P$, the *rank of v with respect to u* is the number of people $w \in P$ who are closer to u than v is. Formally, this quantity is given by $\mathsf{rank}_u(v) := |\{w \in P : d(u,w) < d(u,v)\}|$, where we break ties in distance from person $u \in P$ using the linear ordering on P so that, for any $i \in \{1, \ldots, n\}$ and any person $u \in P$, there is exactly one person v such that $\mathsf{rank}_u(v) = i$.

A *rank-based* friendship for a person $u \in P$ is one generated as follows: a friend v is chosen randomly for u according to the probability distribution $\mathtt{Pr}[u \text{ links to } v] \propto 1/\mathsf{rank}_u(v)$. For any person $u \in P$, we have $\sum_v 1/\mathsf{rank}_u(v) = \sum_{i=1}^{n} 1/i = H_n$, the nth harmonic number. Therefore, by normalizing, we have

$$\mathtt{Pr}[\text{a particular rank-based link from } u \text{ links to } v] = 1/(H_n \cdot \mathsf{rank}_u(v)). \tag{1}$$

Up to constant factors, this rank-based formulation gives the same link probabilities as Kleinberg's distance-based model for a uniform-population k-dimensional mesh. Thus Kleinberg's results [16] immediately imply that rank-based friendship produces a navigable grid for a uniformly distributed population:

Theorem 1. *Let $\langle L, d, P, \mathsf{loc} \rangle$ be a social structure where L is a k-dimensional mesh for $k = \Theta(1)$, d is the Manhattan (L_1) distance, and we have a uniform population P in which exactly one person lives at each point on the grid. Endow each person in the network with $2k$ "local" friends (the immediate neighbors in each cardinal direction) and one "long-range" friend, chosen according to rank-based friendship. Then, with high probability, the length of the* Greedy *path from any $s \in L$ to any $t \in L$ is $O(\log^2 n)$.*

In this paper, we will consider networks with more complicated metrics on the points. To do so, we will need a generalization of the $2k$ "local" neighbors from Theorem 1. For a social structure $\langle L, d, P, \mathsf{loc} \rangle$, construct a population network $\langle L, d, P, \mathsf{loc}, E \rangle$ by generating friendships as follows:

- Endow each person $p \in P$ with δ rank-based links, chosen according to (1).
- Endow each person $p \in P$ with "local neighbors," as follows. Let $G = \langle L, E_G \rangle$ be a graph where shortest paths correspond to the metric d—i.e., the shortest ℓ-to-ℓ' path in G has length $d(\ell, \ell')$. For $\ell \in L$, let $\Gamma_G(\ell)$ be the neighbors of ℓ in G. For every person $p \in P$ with $\mathsf{loc}(p) = \ell$ and for every $\ell' \in \Gamma_G(\ell)$, choose an arbitrary q such that $\mathsf{loc}(q) = \ell'$ and add the edge $\langle p, q \rangle$ to E.

We refer to a network satisfying the latter condition as a *neighbor-connected* network. Neighbor connectivity ensures that, for any s and any t, the first step taken by $\mathsf{Greedy}(s, t)$ will be to a person u such that $d(u, t) < d(s, t)$. Among other things, this condition guarantees that every person encountered by Greedy is encountered only once. Thus we can invoke the Principle of Deferred Decisions in our analysis (see [25]): we proceed as if the long-range links of each person are generated only once the greedy algorithm encounters that person. Furthermore, the greedy algorithm never gets "stuck"; a person u fails to link to a person v such that $d(v, t) < d(u, t)$ only if $\mathsf{loc}(u) = \mathsf{loc}(t)$. (Notice also that neighbor connectivity requires that every point has strictly positive population.)

4 Routing in networks with low doubling dimension

Let $\langle L, d, P, \mathsf{loc} \rangle$ be an arbitrary social structure, where $n := |P|$. Let α and Δ, respectively, be the doubling dimension and aspect ratio of $\langle L, d \rangle$. We derive a neighbor-connected degree-δ rank-based population network $\langle L, d, P, \mathsf{loc}, E \rangle$ by endowing each person $p \in P$ with "local" neighbors as required to achieve neighbor connectivity and δ rank-based friends. In this section, we show that greedy routing finds a short path to a target location whenever α is small.

Lemma 2 (Greedy quickly (in expectation) halves distance to target). *For arbitrary $s \in P$ and $t \in P$ chosen uniformly at random from P, the expected number of rank-based links examined before* $\mathsf{Greedy}(s,t)$ *reaches a person in $B_{d(s,t)/2}(\mathsf{loc}(t))$ is $O(\log n \cdot \log \Delta \cdot 2^{O(\alpha)})$, where the expectation is taken over both the random construction of the network and the random choice of t.*

Proof sketch. An *r-net*, for any $r > 0$, is a set $S \subseteq L$ such that (i) for all $x \in L$, there is some $s \in S$ with $d(x,s) < r$; and (ii) for all distinct $s, s' \in S$, we have $d(s,s') \geq r$. An r-net can be greedily constructed for any $r > 0$. Let $\mathscr{R} := \{1, 2, 4, \ldots, 2^{2+\lceil \log \Delta \rceil}\}$. For every $r \in \mathscr{R}$, we define a set of balls of radius r, where the set $\mathscr{C}_r$ of ball centers forms an $(r/2)$-net. Let r_t denote the minimum $r \in \mathscr{R}$ such that $s, t \in B_r(s')$ for some $s' \in \mathscr{C}_r$. We show that $2r_t \geq d(s,t)/2 \geq r_t/8$; thus it will suffice to show that the expected number of links examined before $\mathsf{Greedy}(s,t)$ lands in $B_{r_t/8}(t) \subseteq B_{d(s,t)/2}(t)$ is $O(\log n \cdot \log \Delta \cdot 2^{O(\alpha)})$.

Suppose that $\mathsf{Greedy}(s,t)$ has generated a partial path from s, where the last element of the path so far is some person $u \in P$. Each step taken by Greedy decreases the distance from the current point to the target t, so we have that $d(u,t) \leq d(s,t) \leq 2r_t$.

We refer to a link from u as *good$_t$* if it connects u to any person living in the ball $B_{r_t/8}(t)$. Let $\beta_{u,t}$ denote the probability that a particular link from u is good$_t$. We show that there is a point $z_t \in \mathscr{C}_{16r_t}$ such that $B_{8r_t}(t) \subseteq B_{16r_t}(z_t)$, and that $\beta_{u,t} \geq \mathsf{pop}(B_{r_t/8}(t))/(\mathsf{pop}(B_{16r_t}(z_t)) \cdot H_n)$, independent of u. Define $\beta_t := \mathsf{pop}(B_{r_t/8}(t))/(\mathsf{pop}(B_{16r_t}(z_t)) \cdot H_n)$. Thus the probability that a particular link from u is good$_t$ is at least β_t for every person u along the Greedy path, and is independent at each step. Therefore, the expected number of links examined by Greedy before we reach a good$_t$ link (or t itself) is at most $1/\beta_t$, where the expectation is taken over the random construction of the network.

We now examine the expected value of $1/\beta_t$ when t is chosen uniformly at random from the population. We show that

$$\begin{aligned}
\mathsf{E}_t[1/\beta_t] &\leq H_n \cdot \sum\nolimits_{x \in L} \mathsf{dens}(x) \cdot \mathsf{dens}(B_{16r_x}(z_x))/\mathsf{dens}(B_{r_x/16}(z'_x)) \\
&\leq H_n \cdot \sum_{r \in \mathscr{R}, z \in \mathscr{C}_{16r}} \quad \sum_{z' \in \mathscr{C}_{r/16} : z' \in B_{16r}(z)} \frac{\mathsf{dens}(B_{16r}(z))}{\mathsf{dens}(B_{r/16}(z'))} \sum_{x \in B_{r/16}(z')} \mathsf{dens}(x),
\end{aligned}$$

where the second line follows by reindexing the summation to be over radii and ball centers from the appropriate r-nets rather than over target locations x. From this, we obtain

$$\mathsf{E}_t[1/\beta_t] \leq H_n \cdot \sum\nolimits_{r \in \mathscr{R}} \sum\nolimits_{z \in \mathscr{C}_{16r}} \mathsf{dens}(B_{16r}(z)) \cdot |\{z' \in \mathscr{C}_{r/16} : z' \in B_{16r}(z)\}|.$$

Using properties of r-nets, we are able to show that the inner summation is upper bounded by $2^{O(\alpha)}$, independent of r. Thus, the expectation is upper bounded by $H_n \cdot |\mathscr{R}| \cdot 2^{O(\alpha)}$, which is $O(\log n \cdot \log \Delta \cdot 2^{O(\alpha)})$ by definition of $\mathscr{R}$. □

Theorem 3. *Let $\langle L, d, P, \mathsf{loc}, E \rangle$ be a neighbor-connected degree-δ rank-based population network. Let $s \in P$ be arbitrary, and let $t \in P$ be chosen uniformly at random. Then the expected length of the* $\mathsf{Greedy}(s,t)$ *path from s to* $\mathsf{loc}(t)$ *is $O(\max\{\log \Delta, \log n \cdot \log^2 \Delta \cdot 2^{O(\alpha)}/\delta\})$, where the expectation is taken both over the random construction of the network and over the random choice of t.*

As a corollary, in the k-dimensional mesh under L_1 distance, where each person has δ rank-based friends and $2k$ local friends, for an arbitrary source s and a uniformly chosen target t, the expected length of $\mathsf{Greedy}(s, \mathsf{loc}(t))$ is $O(\log^3 n \cdot 2^{O(k)}/\delta)$, which is just $O(\log^3 n)$ when $\delta = \Omega(1)$ and $k = O(1)$.

5 Recursive population networks

In this section, we describe a recursive model of population networks that allows higher resolution of location, and that allows the routing of messages to an *individual*, rather than just to that individual's city or town.

Recursive social structures. In the model described previously, a point $\ell \in L$ represents a collection of collocated individuals. Here, we extend the model so that each $\ell \in L$ represents either a single individual or a substructure refining distances between ℓ's inhabitants.

A *recursive social structure (RSS)* is the following: we have a social structure consisting of people living at various points, with a distance function describing the separation between points. For each point ℓ in which strictly more than one person lives, we have, recursively, a social structure for the people living in point ℓ. Formally, an RSS σ on a nonempty set P of people is given as follows:

- If $|P| = 1$, then σ is simply the lone individual in P.
- If $|P| \geq 2$, then $\sigma = \langle L, d, P, \mathsf{loc}, M \rangle$, where $\langle L, d, P, \mathsf{loc} \rangle$ is a social structure with $|L| \geq 2$ and $\mathsf{pop}(\ell) \geq 1$ for every $\ell \in L$, and, for every $\ell \in L$, we have that $M(\ell) = \sigma_\ell$ is an RSS on the set of people $P_\ell := \{u \in P : \mathsf{loc}(u) = \ell\}$.

For an RSS σ, define a tree $\mathscr{T}(\sigma)$ of social structures, where each social structure $\langle L, d, P, \mathsf{loc} \rangle$ contained in σ has "child structures" for each point $\ell \in L$ with $|P_\ell| \geq 2$. The leaves of the tree are the points with a single resident. Let $\mathscr{M}(\sigma)$ denote the internal nodes in $\mathscr{T}(\sigma)$. For $N \in \mathscr{T}(\sigma)$, let $\mathsf{depth}(N)$ denote the depth of N in the tree $\mathscr{T}(\sigma)$, and let $\mathsf{depth}(\mathscr{T}(\sigma))$ denote the depth of the deepest leaf in $\mathscr{T}(\sigma)$. (The root of $\mathscr{T}(\sigma)$ has depth one.)

For $u \in P$, let $N_u = u$ denote the leaf of $\mathscr{T}(\sigma)$ where u is the lone person. For a structure $N \in \mathscr{M}(\sigma)$, we write $u \in N$ to denote that N_u is in the subtree of $\mathscr{T}(\sigma)$ rooted at N—i.e., that $u \in P_N$ where $N = \langle L_N, d_N, P_N, \mathsf{loc}_N \rangle$. Write $\mathsf{depth}(u) := \mathsf{depth}(N_u)$, and for any $1 \leq i \leq \mathsf{depth}(u)$, write $\mathsf{structure}_i(u)$ to denote the unique structure at depth i in $\mathscr{T}(\sigma)$ such that $u \in \mathsf{structure}_i(u)$. Finally, for two individuals $u, v \in P$, let $\mathsf{LCA}(u, v)$ denote the least common ancestor of u and v in $\mathscr{T}(\sigma)$—i.e., the smallest-population structure N in $\mathscr{T}(\sigma)$ such that $u, v \in N$.

From an RSS σ on a set P of people, we derive a (standard) social structure $\mathscr{S}(\sigma)$, where the distances between people are derived from σ. Because the leaves of $\mathscr{T}(\sigma)$ are just the people of P, there will be a unique location in $\mathscr{S}(\sigma)$ for each person of P. To derive distances $d_\sigma(u, v)$ in σ, we consider only the coarsest-resolution structure N in which u and v live in distinct points. Formally, let $N := \mathsf{LCA}(u, v)$, where $N = \langle L_N, d_N, P_N, \mathsf{loc}_N \rangle$. (Note that $u, v \in P_N$ and that $\mathsf{loc}_N(u) \neq \mathsf{loc}_N(v)$.) We define $d(u, v) := \langle -\mathsf{depth}(N), d_N(\mathsf{loc}_N(u), \mathsf{loc}_N(v)) \rangle$, and we use standard lexicographic ordering on pairs to compare distances.

Recursive population networks. Given an RSS σ on a set P of people, we can generate a *recursive population network (RPN)* $\rho = \langle \sigma, E \rangle$ by endowing the people of P with friendships. Let $d = d_\sigma$ be the derived distance function as described above. (We will abuse notation and write $\mathscr{T}(\rho) := \mathscr{T}(\sigma)$, etc.) In a degree-$\delta$ rank-based RPN, we endow each person in P with δ long-range links, chosen according to (1). We assume that ties in distance are broken randomly for the purposes of generating rank-based friendships.

As before, we introduce local neighbors to guarantee (minute) progress from any source to any target t. (The condition is similar to the one introduced in Section 4, but slightly more complicated.)

Let $h : P \to \mathbb{R}$ be a function assigning a "social height" to the people in the network. Consider any $N \in \mathscr{T}(\sigma)$ where $N = \langle L_N, d_N, P_N, \mathsf{loc}_N \rangle$, and let P_ℓ denote the set of people living in point $\ell \in L_N$. For a person $p \in P_\ell$, consider the following conditions:

1. Person p has a local link to a person $q \in P_\ell$ so that $h(q) > h(p)$.
2. Suppose that the metric d_N on L_N is a shortest-path metric in a graph $G = \langle L_N, E_N \rangle$. For every point ℓ' for which the edge $\langle \ell, \ell' \rangle \in E_N$, there exists a q where $\mathsf{loc}_N(q) = \ell'$ such that p has a local link to person q.

If every person in an RPN ρ satisfies one of these two conditions for every structure in $\mathscr{M}(\sigma)$, then we say that ρ is *neighbor connected*.

We also add a tie-breaking rule to Greedy using the social-height function. Suppose that source s seeks a greedy path to a target $t \neq s$, and there is no friend u of s such that $d(u,t) < d(s,t)$. (Thus s cannot fall into Case 2 of the definition of neighbor connectivity for the structure $N = \mathsf{LCA}(s,t)$.) The next step in the Greedy path is a neighbor u of s such that $d(s,t) = d(u,t)$ and $h(u) > h(s)$. This tie-breaking rule guarantees that Greedy can "lift" itself out of a substructure to reach a target in a different structure.

For *non-local* ties in distance—i.e., s has two distinct friends u, v such that $d(u,t) = d(v,t) < d(s,t)$—we assume that ties are broken uniformly at random.

Routing on rank-based RPNs. We will prove that Greedy finds short paths in expectation in any neighbor-connected rank-based RPN derived from an RSS σ as long as the maximum doubling dimension of $N \in \mathscr{M}(\sigma)$ is small.

Notice the following fact, which follows immediately by definition of $d = d_\sigma$: for any $u \in P$ and any depth $i \leq \mathsf{depth}(u)$, all people in $\mathsf{structure}_i(u)$ are closer to u than any person outside $\mathsf{structure}_i(u)$ is to u. An immediate consequence of this fact is that the path found by Greedy aiming for a target t will never leave $\mathsf{structure}_i(t)$ once it enters this subtree.

The expected time required to reach a target t drawn uniformly from the population P is bounded by $O(\max\{\log\Delta, \log n \cdot \log^2\Delta \cdot 2^{O(\alpha)}/\delta\} \cdot \mathsf{depth}(\mathscr{T}(\sigma)))$, by Theorem 3: in expectation we reach the target point in any particular structure in $O(\max\{\log\Delta, \log n \cdot \log^2\Delta \cdot 2^{O(\alpha)}/\delta\})$ steps, and we must find the correct point $\mathsf{depth}(t)$ times before we have arrived at the target person herself. In the following, we remove the dependence on $\mathsf{depth}(\mathscr{T}(\sigma))$.

Theorem 4. *Let ρ be an arbitrary degree-δ rank-based neighbor-connected RPN with $n = |P|$ people, maximum doubling dimension α, and maximum aspect ratio Δ. For an arbitrary source person $s \in P$ and a target person $t \in P$ chosen uniformly at random from P, we have that the expected length of the* Greedy *path from s to t is* $O(\max\{\log\Delta, \log^2\Delta \cdot \log n \cdot 2^{O(\alpha)}/\delta\} \cdot \min\{\mathsf{depth}(\mathscr{T}(\rho)), \log n\})$.

Proof sketch. Our proof proceeds by showing that within a polylogarithmic number of steps we will reduce by a factor of two the number of people closer to the target than the current person on the greedy path is. Let $N_{\mathsf{LCA}} := \mathsf{LCA}(s,t)$, and let $P_{\mathsf{LCA}} := \mathsf{pop}(N_{\mathsf{LCA}})$ be its population. In the structure N_{LCA}, we begin at some point ℓ_s and we wish to reach some point ℓ_t. There are two cases to consider. If $\mathsf{pop}(\ell_t) \leq |P_{\mathsf{LCA}}|/2$ (i.e., the subpopulation containing the target is not too big), then simply reaching ℓ_t as per Theorem 3 constitutes considerable progress towards the target. If $\mathsf{pop}(\ell_t) > |P_{\mathsf{LCA}}|/2$, then any node encountered on the Greedy path has a probability $\Omega(1/H_n)$ of linking to one of the $|P_{\mathsf{LCA}}|/2$ people closest to t. Thus in $O(\log n)$ steps with high probability we reach one of the $|P_{\mathsf{LCA}}|/2$ people closest to t, which is also considerable progress towards the target.

In either case, we have reduced by a factor of two the number of people closer to the target than the current person on the greedy path; a logarithmic number of repetitions of this process will find the target individual herself.

To formalize the above intuitive argument, consider running Greedy starting from person s until the completion of the following two-phase operation:

Phase 1 ("Halfway there"): Run Greedy starting from s until we reach a person v such that either (i) $v \in \mathsf{structure}_{\mathsf{depth}(t)-1}(t)$—i.e., the structure that directly contains the target t—or (ii) $\mathsf{rank}_t(v) \leq \mathsf{pop}(\mathsf{LCA}(s,t))/2$.

Phase 2 ("One level deeper"): Run Greedy starting from v until we reach a person w such that either $w = t$ or $\mathsf{depth}(\mathsf{LCA}(w,t)) > \mathsf{depth}(\mathsf{LCA}(v,t))$.

We show the following:

- After Phase 2 has ended, either $w = t$ or $\mathsf{pop}(\mathsf{LCA}(w,t)) \leq \mathsf{pop}(\mathsf{LCA}(s,t))/2$.
- The expected number of steps before we complete a single two-phase operation is $O(\max\{\log \Delta, \log^2 \Delta \cdot \log n \cdot 2^{O(\alpha)}/\delta\})$.

Thus after a logarithmic number of repetitions of the two-phase process—or $\mathsf{depth}(\mathscr{T}(\rho))$ repetitions, if that quantity is smaller—we reach the target t. □

6 Routing in trees

We now turn to *tree social structures* $\langle L, d, P, \mathsf{loc}\rangle$, where the elements of L are the leaves of a k-ary tree T. We abuse notation and also let T denote the nodes of this tree. Let $T[r]$ denote the subtree of T rooted at $r \in T$. We restrict d so that, for every point $x \in L$ and every node r that is an ancestor of x in T, the point x is closer to every node in $T[r]$ than it is to any node outside of $T[r]$. The population P consists of an arbitrary set of n people, and $\mathsf{loc} : P \to L$ is an arbitrary location function. In particular, we do not impose the condition that $\mathsf{pop}(\ell)$ be strictly positive for every $\ell \in L$; we can simply treat zero-population leaves as not appearing in the tree. If each person in a k-ary tree social structure is endowed with δ edges chosen according to rank-based friendship, then we refer to the resulting population network as a *rank-based δ-degree k-ary tree population network*. Proofs of the following are omitted due to space constraints.

Lemma 5. *Fix an arbitrary $s \in P$. Fix any internal node $r \in T$ such that $\mathsf{loc}(s) \in T[r]$. Choose $t \in P$ uniformly at random from $\{t : \mathsf{loc}(t) \in T[r]\}$. Let r_t denote the child of r such that $\mathsf{loc}(t) \in T[r_t]$. Then, with probability at least $1 - \frac{(k-1)\cdot H_n}{e\cdot\delta}$, within one step the path from s to t found by Greedy reaches $T[r_t]$.*

Theorem 6. *Let $\langle L, d, P, \mathsf{loc}, E\rangle$ be a rank-based δ-degree k-ary tree population network. Fix an arbitrary $\eta \geq 1$. If the degree δ satisfies $\delta \geq \eta \cdot k \cdot H_n \cdot \mathsf{depth}(T)/e$, then the following holds with probability at least $1 - 1/\eta$: for an arbitrary source person $s \in P$ and a target person $t \in P$ chosen uniformly at random from P, the Greedy path from s to $\mathsf{loc}(t)$ has length at most $\mathsf{depth}(T)$.*

As a corollary, consider a rank-based population network derived from a binary tree with depth $O(\log^k n)$ and with degree $\delta = \Omega(\eta \cdot \log^{k+1} n)$. Then with probability at least $1 - 1/\eta$, for arbitrary $s \in P$ and uniformly chosen $t \in P$, the length of the Greedy path from s to $\mathsf{loc}(t)$ has length $O(\log^k n)$.

7 Discussion and future work

Here we highlight some interesting open questions for future study, focusing on the model described in Sections 4 and 5. We have shown that $\mathsf{E}_t[|\mathsf{Greedy}(s,t)|] = \text{polylog}(|P|)$ for any $s \in P$ in rank-based networks. In contrast, Kleinberg has shown that, for uniform populations, with high probability, $\mathsf{Greedy}(s,t)$ has polylogarithmic length for any s and for any t when link probabilities are chosen according to the correct distance-based distribution. There may be population distributions for which the "for all t" condition cannot be achieved in our context, perhaps if there is a recluse who is very unlikely to be reached by long-range links. It remains open whether Greedy finds a short expected path for *any* target.

We use the assumption that there are no empty locations in our network to guarantee that Greedy never gets "stuck" at a person u without a local neighbor closer to the target than u herself is. Investigating the limitations of Greedy in a model with zero-population locations (like lakes and deserts in the real world) is an intriguing direction, and would eliminate the most unrealistic limitation in our model. Geographic routing via local-information algorithms in general, and geographic routing around obstacles in particular, has been previously considered in the wireless-networking community [11, 12, 14]. It is an interesting question as to whether these results, where there is typically a technologically inspired threshold on the geographic distance that a message can traverse in a single hop, can be adapted to the social-network setting.

A number of partially decentralized algorithms (e.g., [8, 20, 22, 23, 27]) have been shown to outperform Greedy theoretically or experimentally; it would be interesting to analyze them in rank-based networks. More generally, our results can be viewed as extending Kleinberg's theorem to a dimension-independent model that allows varying population density (and one that holds in real networks [21]). There have been some recent theoretical results extending and refining Kleinberg's result—for example, considering routing on other types of underlying graphs [28, 7, 9], among other results [4, 23, 26]—and we might hope to be able to make analogous improvements to our results.

Acknowledgements. Thanks to David Barbella, Erik Demaine, George Kachergis, David Karger, Jon Kleinberg, Danny Krizanc, Jasmine Novak, Prabhakar Raghavan, Anna Sallstrom, and Ben Sowell. Part of the second author's work wasperformed while he was visiting Yahoo! Research and was partially supported by Carleton College.

References

[1] Lada A. Adamic and Eytan Adar. How to search a social network. *Social Networks*, 27(3):187–203, July 2005.

[2] Lada A. Adamic, Rajan M. Lukose, and Bernardo A. Huberman. Local search in unstructured networks. In *Handbook of Graphs and Networks*. Wiley-VCH, 2002.

[3] Lada A. Adamic, Rajan M. Lukose, Amit R. Puniyani, and Bernardo A. Huberman. Search in power-law networks. *Physical Review Letters E*, 64(046135), 2001.

[4] Lali Barrière, Pierre Fraigniaud, Evangelos Kranakis, and Danny Krizanc. Efficient routing in networks with long range contacts. In *Proceedings of the International Conference on Distributed Computing (DISC)*, pages 270–284, October 2001.

[5] Erik D. Demaine, John Iacono, and Stefan Langerman. Proximate point searching. *Computational Geometry: Theory and Applications*, 28(1):29–40, May 2004.

[6] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. An experimental study of search in global social networks. *Science*, 301:827–829, 8 August 2003.

[7] Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar, and Nicolas Schabanel. Could any graph be turned into a small world? *Theoretical Computer Science*, 355(1):96–103, 2006.

[8] P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In *Proceedings of the Symposium on Principles of Distributed Computing (PODC)*, pages 169–178, 2004.

[9] Pierre Fraigniaud. Greedy routing in tree-decomposed graphs. In *Proceedings of the European Symposium on Algorithms (ESA)*, pages 791–802, 2005.

[10] John Iacono and Stefan Langerman. Proximate planar point location. In *Proceedings of the Symposium on Computational Geometry (SoCG)*, pages 220–226, June 2003.

[11] Brad Karp. *Geographic Routing for Wireless Networks*. PhD thesis, Harvard, October 2000.

[12] Brad Karp and H. T. Kung. GPSR: Greedy perimeter stateless routing for wireless networks. In *Proceedings of the International Conference on Mobile Computing and Networking (Mobicom)*, pages 243–254, August 2000.

[13] B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong. Path finding strategies in scale-free networks. *Physical Review Letters E*, 65(027103), 2002.

[14] Young-Jin Kim, Ramesh Govindan, Brad Karp, and Scott Shenker. Geographic routing made practical. In *Proceedings of the Symposium on Networked Systems Design and Implementation (NSDI)*, May 2005.

[15] Jon Kleinberg. Navigation in a small world. *Nature*, 406:845, 24 August 2000.

[16] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the Symposium on the Theory of Computing (STOC)*, volume 14, 2000.

[17] Jon Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing (NIPS)*, volume 14, 2001.

[18] Jon Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of the International Congress of Mathematicians*, 2006.

[19] Ravi Kumar, David Liben-Nowell, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. Theoretical analysis of geographic routing in social networks. Technical Report MIT-CSAIL-TR-2005-040.

[20] E. Lebhar and N. Schabanel. Close to optimal decentralized routing in long-range contact networks. In *Proceedings of the International Colloquium on Automata, Languages, and Programming (ICALP)*, 2004.

[21] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, August 2005.

[22] Gurmeet Singh Manku, Moni Naor, and Udi Wieder. Know thy neighbor's neighbor: the power of lookahead in randomized P2P networks. In *Proceedings of the Symposium on the Theory of Computing (STOC)*, pages 54–63, June 2004.

[23] Chip Martel and Van Nguyen. Analyzing Kleinberg's (and other) small-world models. In *Proceedings of the Symposium on Principles of Distributed Computing (PODC)*, pages 179–188, 2004.

[24] Stanley Milgram. The small world problem. *Psychology Today*, 1:61–67, May 1967.

[25] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge Univ. Press, 1995.

[26] V. Nguyen and C. Martel. Analyzing and characterizing small-world graphs. In *Proceedings of the Symposium on Discrete Algorithms (SODA)*, pages 311–320, 2005.

[27] O. Şimşek and David Jensen. A probabilistic framework for decentralized search in networks. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[28] Aleksandrs Slivkins. Distance estimation and object location via rings of neighbors. In *Proceedings of the Symposium on Principles of Distributed Computing (PODC)*, July 2005.

[29] Stanley Wasserman and Katherine Faust. *Social Network Analysis*. Cambridge Univ. Press, 1994.

[30] Duncan J. Watts, Peter Sheridan Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 17 May 2002.

[31] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.