Depth of Field and Cautious-Greedy Routing in Social Networks*

David Barbella
Department of Computer Science
Carleton College
barbelld@carleton.edu

George Kachergis†
Department of Psychology & Brain Sciences
Indiana University
gkacherg@indiana.edu

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

Anna Sallstrom
Department of Computer Science
Carleton College
sallstra@carleton.edu

Ben Sowell†
Department of Computer Science
Cornell University
sowell@cs.cornell.edu

Abstract

Social networks support efficient decentralized search: people can collectively construct short paths to a specified target in the network. *Rank-based friendship*—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—is an empirically validated model of acquaintanceship that provably results in efficient decentralized search via greedy routing, even in networks with variable population densities. In this paper, we introduce *cautious-greedy routing*, a variant of greedy that avoids taking large jumps unless they make substantial progress towards the target. Our main result is that cautious-greedy routing finds a path of short expected length from an arbitrary source to a randomly chosen target, independent of the population densities. To quantify the expected length of the path, we define the *depth of field* of a metric space, a new quantity that intuitively measures the "width" of directions that leave a point in the space. Our main result is that cautious-greedy routing finds a path of expected length $O(\log^2 n)$ in n-person networks that have aspect ratio polynomial in n, bounded doubling dimension, and bounded depth of field. Specifically, in k-dimensional grids under Manhattan distance with arbitrary population densities, the $O(\log^2 n)$ expected path length that we achieve with the cautious-greedy algorithm improves the best previous bound of $O(\log^3 n)$ with greedy routing.

1 Introduction

As large-scale datasets of social interactions have become widespread, computer scientists have begun to explore *social networks*, graphs in which nodes representing people are connected by edges

*A preliminary version of this paper appears in *Proceedings of the 18th Annual International Symposium on Algorithms and Computation (ISAAC'07)*, December 2007, pages 574–586. Comments are welcome. This work was supported in part by NSF grant CCF-0728779 and by grants from Carleton College.

†Work performed in part while at Carleton College.

representing friendships. Social networks are a particularly appealing domain for interdisciplinary application of computational thinking: a graph-theoretic, algorithmic approach can lend interesting insight to questions traditionally in the domain of the social sciences. The well-known *small-world phenomenon*—first observed explicitly by Stanley Milgram's ingenious experiments [21], in which people demonstrated an ability to collectively construct short chains of acquaintances to pass a message to a specified target personin the social network—is a marked example [15]. A key observation of Jon Kleinberg [12, 13], made some thirty years after Milgram's original experiment, is that these results are at heart algorithmic: people are able to use some sort of distributed algorithm to construct short paths through the social network, with each node having only limited, local information about the friendships in the network. Kleinberg gave a simple model that suffices to produce a navigable small world. Start from a k-dimensional grid of people, with $k = \Theta(1)$, where we view proximity in the grid as corresponding to similarity in geographic location, occupation, or some other attribute. Connect each person to $2k$ *local neighbors*, her immediate neighbors in the grid. Also endow each person u with one *long-range link*, chosen randomly so that $\texttt{Pr}[u \to v] \propto d(u,v)^{-\beta}$, where $d(u,v)$ is the Manhattan distance between u and v and $\beta \geq 0$ is a parameter of the model. (The presence of additional long-range links does not substantially affect the results.) Kleinberg studied *greedy routing*—to route a message from s to t, person s forwards the message to the friend of s who is closest in lattice distance to t—and showed that this simple algorithm finds paths of expected length $O(\log^2 n)$ in n-person networks when $\beta = k$, and of length $n^{\Omega(1)}$ when $\beta \neq k$. A $\Omega(\log^2 n)$ bound on the expected length of the path found by greedy routing when $\beta = k$ has also been shown [4, 20].

These results have subsequently been extended and adapted to situations in which the underlying network is not a grid, but is instead, for example, a tree [14] or a network that has low treewidth [8], bounded growth rate [6, 7], or low doubling dimension [24]. But an important feature present in the real world is lacking in almost all of these models: whatever one chooses as the space in which to measure similarity, people are not uniformly distributed among the points in this space(geographic locations, occupations, etc.). The *group-structure* model [14] can handle differential population density, as can *rank-based friendship* [16, 18]. We focus on the latter, which has been shown empirically to be a close match for friendship patterns in a real large-scale online social network [18]. The model is still based on an underlying similarity measure, but an arbitrary number of people can live at each point. Define the *rank* of a person v with respect to u as the number of people who live at least as close to u as v does. Under rank-based friendship, we generate long-range links probabilistically using rank instead of distance: a person u chooses a long-range link v so that $\texttt{Pr}[u \to v]$ is proportional to the rank of v with respect to u. (It is worth noting that rank-based friendship implicitly handles the dimensionality of the underlying similarity measure: in a k-dimensional grid with $\Theta(1)$ population at each point, we have $\mathsf{rank}_u(v) = \Theta(d(u,v)^k)$, matching the navigable distribution that explicitly involves k in Kleinberg's distance-based setting.)

In joint work of Ravi Kumar, Andrew Tomkins, and the third author [16], greedy routing was shown to perform well in rank-based social networks, in the following sense. In a $\Theta(1)$-dimensional grid with arbitrary positive population at every point and with total population n, the greedy algorithm finds a path of expected length $O(\log^3 n)$ from an arbitrary source to the point of a target chosen uniformly from the population. In fact, these results were shown in a much more general setting [16]: if similarity is measured by proximity in an arbitrary metric space, then Greedy finds a path of expected length $O(\log n \cdot \log^2 \Delta \cdot 2^{O(\alpha)})$ to a randomly chosen target, where n is the population size; Δ is the distance between maximally distant people; and α is the doubling dimension of the metric space, a combinatorial measure of the implicit dimensionality of the space. Formally, the doubling dimension of a metric space is the smallest α such that every ball of radius r can be covered by 2^α balls of radius $r/2$(see [2, 10, 24], e.g.).

Our contributions. The theorem for rank-based networks with variable population densities restricted to k-dimension grids under Manhattan distance says that GREEDY finds a path of expected length $O(\log^3 n)$ to a randomly chosen target. This result is weaker than Kleinberg's theorem in two ways: there is an additional $O(\log n)$ factor in the path length; and there is "in expectation for a random target" in place of "for any target." (These negatives are, of course, counterbalanced by the increased generality of the model, which can handle essentially arbitrary population distributions.)

In this paper, we improve the upper bound on expected path length to $O(\log^2 n)$ in k-dimensional grids with arbitrary population densities, closing the gap with Kleinberg's analysis of the uniform-population case. We achieve this bound with the *cautious-greedy algorithm*, a variation on greedy routing that we introduce. Intuitively, the greedy algorithm can get into trouble in variable-density networks as follows. Imagine a person s who only lives near points with unit population. Suppose that s has a friend u so that $d(u,t) = d(s,t) - \varepsilon$ but the jump from s to u "overshoots" the target t. If there is a city with massive population near u (but farther from t than u is), then s may be making a mistake by choosing u as the next step in the chain: because u is in the "shadow" of the city, then we have $\texttt{Pr}[s \to t] \gg \texttt{Pr}[u \to t]$. Indeed, the same is true from u for all points close to t, and showing that GREEDY is making progress at every step is difficult. (This difficulty seems intuitive: if s and t live 60 and 25 miles due west of New York City, respectively, then overshooting t from s by jumping to a friend in Manhattan seems like a bad idea, because the chain is now stuck inside New York's "basin of attraction.") CAUTIOUSGREEDY differs from GREEDY in that it conservatively takes local links *unless* there is a long-range link that halves the distance to the target, in which case it follows that link. We are able to show that, in rank-based social networks with arbitrary population distributions derived from k-dimensional grids under Manhattan distance, the expected length of the CAUTIOUSGREEDY(s,t) path for a randomly chosen target t is $O(\log^2 n)$. We also show that this bound is tight in the 1-dimensional case.

The result for k-dimensional grids is a corollary of our main theorem, which shows that cautious-greedy routing performs well in expectation for social networks with similarity measures derived from arbitrary metric spaces. As usual, the performance of the algorithm depends on certain properties of the metric space, including the aspect ratio Δ and the doubling dimension α. We also uncover a new quantity $\rho \in (0,1]$ characterizing metric spaces, intuitively measuring the smallest "width" of each direction leaving a point in the space, that is crucial to our analysis. By analogy to the corresponding concept in photography, we call ρ the *depth of field* of the metric space. More precisely, the depth of field is the minimum over points s and t of the ratio $r(s,t)/d(s,t)$, where $r(s,t)$ is the maximum radius of a ball B around t so that a shortest path from s to every point in B has the same first step. In networks with large depth of field (e.g., Manhattan distance in k-dimensional grids), we can prove much better routing bounds than in networks with small depth of field (e.g., Euclidean distance in k-dimensional grids). Our main result is that CAUTIOUSGREEDY finds a path of expected length $O(\log n \cdot \log \Delta \cdot (c\rho)^{-\alpha})$ for constant c, improving by a $O(\log \Delta)$ factor the bound from [16] in networks with bounded ρ and α.

Other related work. Although GREEDY is the most commonly analyzed decentralized social-network routing algorithm, there has been significant work on other algorithms as well. Typically—and in contrast to CAUTIOUSGREEDY, which still uses only completely local information in constructing the path—these algorithms endow individuals with additional structural information about the network, such as awareness of friend's friends (e.g., [9, 17, 19, 20, 24]). Other studies of the navigability of social networks and, in particular, good local-information algorithms for navigation have been performed, largely focusing on simulations and empirical studies of real networks (e.g., [3, 5, 23, 25]). There has also been recent relevant work in metric embeddings in which the underlying metric space can be simplified without large distortion of distances between nodes that are close together [1] and in designing peer-to-peer systems where node distributions are nonuniform in keyspace [11].

2 Depth of Field in Metric Spaces

Consider a metric space $\mathcal{M} = \langle X, d \rangle$. For convenience, throughout the paper we scale every metric space so that $\min_{x \neq y} d(x, y) = 1$. We first mention a few standard notions that we use throughout:

- Let $\Delta := \max_{x,y \in X} d(x, y)$ denote the *aspect ratio* of the metric space.
- Let $B_r(x) := \{y : d(x, y) < r\}$ denote the open radius-r ball around $x \in X$.
- Let α denote the *doubling dimension* of the metric space: α is the smallest value such that, for every $r > 0$, every set $Y \subseteq X$ of radius $2r$ can be covered by at most 2^α subsets of X, each of radius r.

We will need to develop a notion that quantifies the following intuition: if a point s is far away from a point t, then making a small step closer to t from s should bring one closer to points near t as well. We will define the *depth of field* of the metric space to quantify the notion of "near t." (Our use of this term is inspired by the same term in photography: for a camera focused on a point t, "depth of field" refers to the range of points around t that are also in focus. A large depth of field means that many points near t are also in focus.) Let $r > 0$ denote the largest radius so that some point u is on a shortest path from s to *every* point in the ball $B_r(t)$. The farther apart s and t are, the larger one would expect r to be; thus, we will be concerned with the ratio between r and $d(s, t)$.

Definition 2.1 (Depth of field). *For arbitrary points $s, t, u \in X$ with $s \notin \{t, u\}$, define $r_u(s, t)$ as the maximum value such that*

$$\forall z \in X \qquad d(z, t) < r_u(s, t) \implies d(s, z) = d(s, u) + d(u, z).$$

We define the depth of field *of s and t as $\rho(s, t) := \max_{u \neq s} r_u(s, t)/d(s, t)$. The* depth of field *of the metric space $\mathcal{M}$ is $\rho(\mathcal{M}) := \min_{s \neq t} \rho(s, t)$.*

Lemma 2.2. *For any metric space $\mathcal{M}$, we have $0 < \rho(\mathcal{M}) \leq 1$.*

Proof. Consider arbitrary distinct s, t. For sufficiently small r, the ball $B_r(t)$ is just $\{t\}$. So $r_t(s, t) > 0$, and $\rho(s, t) > 0$. On the other hand, we have $s \in B_r(t)$ for $r > d(s, t)$. Thus $r_u(s, t) \leq d(s, t)$ for every u. Thus $0 < \rho(s, t) \leq 1$. ∎

One can give similar intuitive descriptions of the depth of field and the doubling dimension of a metric space: they both aim to quantify the number of distinct directions emanating from a point in the metric space. However, they measure "number of distinct directions" in different ways. Informally, doubling dimension counts something like the number of directions one can go from s; depth of field counts something like the "width" of the narrowest of these directions.

Denote depth of field and doubling dimension by ρ and α, respectively. In what follows, we will find algorithms whose running times depend exponentially on α and $1/\rho$, so we are most interested in metric spaces where both α and $1/\rho$ are bounded. To clarify the relationship between these quantities, we note metric spaces where one or the other or both of these quantities is/are small:

- a k-dimensional grid under Manhattan distance has $\rho = 1/k$ and $\alpha = \Theta(k)$.
- a 2-by-n grid under Euclidean distance has $\rho = 1/n$ and $\alpha = \Theta(1)$.
- a metric space in which distances are given by shortest paths in an n-node star graph (a tree with $n-1$ leaves) has $\rho = 1$ and $\alpha = \log n$.

3 Social Networks, Rank-based Friendship, and Routing Algorithms

A *social network* is a directed graph $\langle P, E\rangle$, where a node represents a person and an edge denotes a friendship between its endpoints. Let $\Gamma(u)$ denote the friends of $u \in P$. Our general framework consists of some attributes of the people in P, with friendships derived in some way from these attributes. (We use terminology suggestive of a geographic interpretation, but other attributes are equally valid in this setting.) The attributes will be globally known to all people, but the friendships will be known only to the people involved. As a formal encoding of this framework, we consider the following structures throughout:

- a finite set L of *points*;
- a *distance metric* $d : L \times L \to \mathbb{R}^{\geq 0}$ on the points (so $\langle L, d\rangle$ is a metric space);
- a finite set P of *people*; and
- a *location function* $\mathsf{loc} : P \to L$ so that $\mathsf{loc}(u)$ is the point where $u \in P$ lives.

We will permit ourselves to write an element u of P in contexts where an element of L is expected, with the understanding that u is shorthand for $\mathsf{loc}(u)$. For $\ell \in L$ or $L' \subseteq L$, let $\mathsf{pop}(\ell) := |\{u \in P : \mathsf{loc}(u) = \ell\}|$ and $\mathsf{pop}(L') := \sum_{\ell \in L'} \mathsf{pop}(\ell)$ denote the *population* of a point or set of points. Write $n := \mathsf{pop}(L) = |P|$ for the total population. We will also impose a condition called *neighbor connectivity* [16] on the social networks that we consider. For every $p, q \in P$ with $\mathsf{loc}(q) \neq \mathsf{loc}(p)$, we require that p have a friend in some location ℓ such that $d(\mathsf{loc}(p), \ell) + d(\ell, \mathsf{loc}(q)) = d(\mathsf{loc}(p), \mathsf{loc}(q))$. (Formally, if G_d is the minimal graph on L where shortest paths correspond to the metric d, then p must have a friend in every neighbor of $\mathsf{loc}(p)$ in G_d.) Among other things, this guarantees that, for people $s, t \in P$ with $\mathsf{loc}(s) \neq \mathsf{loc}(t)$, person s has a friend u with $d(s, t) > d(u, t)$.

3.1 Rank-based Friendships

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, so $\mathsf{rank}_u(v) := |\{w \in P : d(u, w) < d(u, v)\}|$. For concreteness, we will specify that $\mathsf{rank}_u(u) := 1$ for every person $u \in P$. Ties in distance are broken using a canonical ordering on P, so for any person $u \in P$ and any rank $i \in \{1, \ldots, n\}$, 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 $\mathtt{Pr}[u \text{ links to } v] \propto 1/\mathsf{rank}_u(v)$. Let $H_n = \Theta(\log n)$ denote the nth harmonic number. By normalization, we have the following:

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

We endow each person with ≥ 1 rank-based friendship, chosen according to (1), along with an arbitrary set of *local neighbors* satisfying neighbor connectivity.

3.2 Decentralized Routing Algorithms

Given *source* and *target* individuals $s, t \in P$, a *routing algorithm* seeks a path $\sigma = \langle u_0, u_1, \ldots, u_k \rangle$ from $s = u_0$ to u_k with $\mathsf{loc}(u_k) = \mathsf{loc}(t)$ in the graph $\langle P, E \rangle$. (We do not model routing within points in this paper.) A *decentralized* algorithm computes the next step u_{i+1} from the current person u_i without taking the entire graph $\langle P, E \rangle$ as input: only the edges in E incident to u_i are available to the algorithm. (Full information about L, d, P, and loc is available to the algorithm; thus, for example, a person s can compute $r_u(s,t)$ for any u, t in the sense of Def. 2.1.) We focus on two particular decentralized algorithms. Under the well-studied *greedy algorithm*, the current person u_i simply chooses her friend who is closest to t as the next step in the path. We also introduce and analyze the *cautious-greedy algorithm*, which is a conservative variant of Greedy that refuses to take long jumps that do not make significant progress (specifically, by halving the distance) to the target, instead opting for a "safe" local link.

CautiousGreedy(s,t):
(halving step.) Let $u := \mathrm{argmin}_{u \in \Gamma(s)} d(u,t)$. If $d(u,t) \leq \frac{d(s,t)}{2}$, forward to u. *(nonhalving step.)* Else forward to $\mathrm{argmax}_{u \in \Gamma(s)} r_u(s,t)$ in the sense of Def. 2.1.

Lemma 3.1. *Suppose the network is neighbor connected and let ρ be the depth of field of $\langle L, d \rangle$. If* CautiousGreedy(s,t) *takes a nonhalving step from s to w, then $d(s,z) = d(s,w) + d(w,z)$ for every z such that $d(z,t) < \rho \cdot d(s,t)$.* □

Proof. Let v^* be the person so that $r_{v^*}(s,t)$ is maximized, so that $\rho \leq \rho(s,t) = r_{v^*}(s,t)/d(s,t)$ and, for every z such that $d(z,t) < r_{v^*}(s,t)$, we have

$$d(s,z) = d(s,v^*) + d(v^*,z). \tag{2}$$

Let G_d denote the minimal graph where shortest paths correspond to the metric d, and let $\Gamma_G(s)$ denote the neighbors of s in G_d. By definition of neighbor connectivity, there is a local link from s to every neighbor of s in G_d. In particular, there is a link from s to a $u \in \Gamma_G(s)$ so that

$$d(s,v^*) = d(s,u) + d(u,v^*), \tag{3}$$

because shortest paths in G_d correspond to distances under d. We claim, then, that $r_u(s,t) \geq r_{v^*}(s,t)$. For every z such that $d(z,t) < r_{v^*}(s,t)$:

$$\begin{aligned}
d(s,z) &= d(s,v^*) + d(v^*,z) && \text{by (2)} \\
&= d(s,u) + d(u,v^*) + d(v^*,z) && \text{by (3)} \\
&\geq d(s,u) + d(u,z) && d(u,v^*) + d(v^*,z) \geq d(u,z) \text{ by triangle inequality} \\
&\geq d(s,z). && d(s,u) + d(u,z) \geq d(s,z) \text{ by triangle inequality}
\end{aligned}$$

Because $d(s,z) \geq d(s,u) + d(u,z) \geq d(s,z)$ for every z such that $d(z,t) < r_{v^*}(s,t)$, we in fact have that $d(s,z) = d(s,u) + d(u,z)$ for every such z, and therefore $r_u(s,t) \geq r_{v^*}(s,t)$. By the assumption of neighbor connectivity, then, person s has a link to a person u with $r_u(s,t) \geq r_{v^*}(s,t) \geq \rho \cdot d(s,t)$. Therefore the person w to whom s forwards must have $r_w(s,t) \geq r_u(s,t) \geq \rho \cdot d(s,t)$. □

As a corollary, we note that both Greedy and CautiousGreedy make strict progress towards their targets in every step. Among other things, this fact guarantees that every person is encountered only once by a run of the algorithm. Thus we can invoke the principle of deferred decisions in our analysis (see [22]): we proceed as if the long-range links of each person are generated only once the algorithm encounters that person.

4 An Upper Bound for Cautious Greedy Routing

Consider fixed L, d, P, loc as defined previously. Let ρ and α, respectively, denote the depth of field and doubling dimension of the metric space $\langle L, d \rangle$. We will be interested in the length of the path found by CautiousGreedy(s,t), where $s \in P$ is arbitrary and $t \in P$ is chosen uniformly at random from the population. Most of our effort will be focused on analyzing the number of steps required to halve the distance to the target—or, more precisely, to reach a person u such that $d(u,t) < 2^{i-1}$ from a source s when t is chosen randomly from $\{t : d(s,t) < 2^i\}$.

Halving the distance to the target

Fix a source person s and some integer $i \in \mathbb{Z}^{\geq 1}$. Let $B := B_{2^i}(s)$ denote the ball of radius 2^i around s. It will turn out to be handy to fix notation for $D := B_{2^i}(s) - \{z : d(s,z) \leq 2^{i-1}\}$, the "donut" formed by removing the (closed) ball of radius 2^{i-1} centered at s from B.

We will use a cover of B by a small (in terms of α and $1/\rho$) set of balls of radius $\rho \cdot 2^{i-2}$. This radius is chosen carefully to ensure the following: whenever we take a nonhalving step towards target t from a node u with $d(u,t) \geq 2^{i-1}$, we have stepped along a shortest path to every node in the ball containing t that is included in the cover.

Lemma 4.1. *There is a set $\mathcal{Q} \subseteq B$ where $|\mathcal{Q}| \leq (8/\rho)^\alpha$ and $\bigcup_{\ell \in \mathcal{Q}} B_{\rho \cdot 2^{i-2}}(\ell) \supseteq B$.*

Proof sketch. By repeatedly applying the definition of doubling dimension, we find that we can cover B with a set of at most $2^{(3+\log_2(1/\rho))\alpha} = (8/\rho)^\alpha$ balls of radius at most $\rho \cdot 2^{i-2}$ that cover B. We let $\mathcal{Q}$ denote the set of their centers. □

Throughout this section, fix $\mathcal{Q}$ to denote this set, so that, for every $u \in B$, there exists a point $q \in \mathcal{Q}$ such that $d(u,q) \leq \rho \cdot 2^{i-2}$. Write $q(u) := \operatorname{argmin}_{q \in \mathcal{Q}} d(q,u)$, and write $C(u) := B \cap B_{\rho \cdot 2^{i-2}}(q(u))$. Notice that $u \in C(u)$.

Lemma 4.2. *Let $t \in B$ be arbitrary. Suppose* CautiousGreedy(s,t) *arrives at node u with $d(u,t) > 2^{i-1}$. Let $v \in C(t)$ be arbitrary. Then $B_{d(u,v)}(u) \subseteq B$.*

Proof by induction on the number of steps taken by CautiousGreedy(s,t). If no steps have been taken by CautiousGreedy(s,t), then the claim is trivial, because $u = s$ and $C(t) \subseteq B$. Otherwise, suppose that u was reached by a step from a node w. Notice that every step that CautiousGreedy(s,t) has taken before reaching u must be nonhalving: otherwise, there was a halving step from some u^* with $d(s,t) \geq d(u^*,t)$ to a neighbor v^* with $d(v^*,t) \leq d(u^*,t)/2$; but then u would not satisfy the conditions of the lemma, as $d(v^*,t) \leq d(u^*,t)/2 \leq d(s,t)/2 \leq 2^{i-1}$. Observe that $d(v,t) < \rho \cdot 2^{i-1}$:

$$d(v,t) \leq d(v,q(t)) + d(q(t),t) < (\rho \cdot 2^{i-2}) + (\rho \cdot 2^{i-2}) = \rho \cdot 2^{i-1}.$$

by the triangle inequality and the fact that $v, t \in C(t) \subseteq B_{\rho \cdot 2^{i-2}}(q(t))$ by the definition of v and $C(t)$. We also have that $d(w,t) > d(u,t) > 2^{i-1}$ by assumption. Therefore, because we took a nonhalving step from w to u, Lemma 3.1 implies that $d(w,v) = d(w,u) + d(u,v)$, because $d(v,t) < \rho \cdot 2^{i-1}$ implies $d(v,t) < \rho \cdot d(w,t)$. To complete the proof of the lemma, consider a generic point x. We show that $x \in B_{d(u,v)}(u) \implies x \in B$:

$$\begin{aligned}
x \in B_{d(u,v)}(u) &\iff d(x,u) < d(u,v) \\
&\iff d(x,u) < d(w,v) - d(w,u) && (d(w,v) = d(w,u) + d(u,v) \text{ as above}) \\
&\implies d(x,w) < d(w,v) && (d(x,w) \leq d(x,u) + d(u,w) \text{ by the triangle inequality}) \\
&\iff x \in B_{d(w,v)}(w) \\
&\implies x \in B && \text{(inductive hypothesis).}
\end{aligned}$$

Thus $B_{d(u,v)}(u) \subseteq B$, as desired. □

For an arbitrary target $t \in B$, let X_t be a random variable denoting the number of steps that CautiousGreedy(s,t) takes before it reaches a person u with $d(u,t) \leq 2^{i-1}$. Note that, for $t \in B - D$, we have $X_t = 0$: the node s itself satisfies the desired condition for u. For $t \in D$, it suffices to reach a node $u \in C(t)$: we have

$$d(u,t) \leq d(u,q(t)) + d(q(t),t) < \rho \cdot 2^{i-2} + \rho \cdot 2^{i-2} = \rho \cdot 2^{i-1} \leq 2^{i-1}$$

by the triangle inequality, the definition of $C(t)$ and $q(t)$ and the fact that $t \in C(t)$, and Lemma 2.2. Note CautiousGreedy(s,t) will take a halving step to follow a link to any node in $C(t)$ if $d(s,t) > 2^{i-1}$, as $d(u,t) < 2^{i-2} < d(s,t)/2$; thus it suffices to compute the probability that a node has a link to $C(t)$.

Lemma 4.3. *For $t \in D$, we have $\mathtt{E}[X_t] \leq H_n \cdot \mathsf{pop}(B)/\mathsf{pop}(C(t))$.*

Proof. Suppose CautiousGreedy has generated a partial path to t and that u is the last node on that path. If $d(u,t) \leq 2^{i-1}$, we are done, so consider a node u such that $2^{i-1} < d(u,t) \leq d(s,t) \leq 2^i$, as every step of CautiousGreedy moves closer to the target t. The probability that a rank-based link from such a node u goes into $C(t)$ is

$$\mathtt{Pr}[u \to C(t)] = \sum_{v \in C(t)} \frac{1}{H_n \cdot \mathsf{rank}_u(v)} \geq \sum_{v \in C(t)} \frac{1}{H_n \cdot \mathsf{pop}(B)} = \frac{\mathsf{pop}(C(t))}{H_n \cdot \mathsf{pop}(B)} \tag{4}$$

by (1) and Lemma 4.2. Thus at every step while CautiousGreedy is farther than 2^{i-1} from the target, the probability of the current node having a rank-based link to $C(t)$ is given by (4). Thus the expected number of steps until CautiousGreedy either reaches a point within distance 2^{i-1} of t through nonhalving steps or reaches a person in $C(t)$ via a halving step is no larger than the expected waiting time for success in geometric random process with success probability $\mathsf{pop}(C(t))/(H_n \cdot \mathsf{pop}(B))$. The claim follows. □

Lemma 4.4. *Let t be a target chosen uniformly at random from B. Then the expected length of* CautiousGreedy(s,t) *before it arrives at some node u with $d(u,t) \leq 2^{i-1}$ is at most $|\mathscr{Q}| \cdot H_n$, where the expectation is taken both over the random construction of the network and over the random choice of t.*

Proof. Let X be a random variable denoting the length of the path found by CautiousGreedy(s,t) before it arrives at a node u with $d(u,t) \leq 2^{i-1}$ when t is drawn uniformly at random from B. Then

$$\begin{aligned}
\mathtt{E}[X] &= \mathtt{E}[\,X \mid t \in D] \cdot \mathtt{Pr}[t \in D \,|\, t \in B] \\
&\qquad + \mathtt{E}[\,X \mid t \in B - D] \cdot \mathtt{Pr}[t \in B - D \,|\, t \in B] \\
&= \mathtt{E}[\,X \mid t \in D] \cdot \mathtt{Pr}[t \in D \,|\, t \in B] && (t \in B - D \text{ is done in zero steps}) \\
&= \left[\sum_{t \in D} \mathtt{E}[X_t] \cdot \frac{1}{\mathsf{pop}(D)}\right] \cdot \frac{\mathsf{pop}(D)}{\mathsf{pop}(B)} \\
&\leq \sum_{t \in D} \frac{H_n \cdot \mathsf{pop}(B)}{\mathsf{pop}(C(t))} \cdot \frac{1}{\mathsf{pop}(D)} \cdot \frac{\mathsf{pop}(D)}{\mathsf{pop}(B)} && \text{(Lemma 4.3)} \\
&= H_n \sum_{t \in D} \frac{1}{\mathsf{pop}(C(t))} \\
&= H_n \sum_{q \in \mathscr{Q}} \frac{\mathsf{pop}(C(q))}{\mathsf{pop}(C(q))} && \left(\tfrac{1}{\mathsf{pop}(C(q))} \text{ is summed once for each person in } C(q)\right) \\
&= H_n \cdot |\mathscr{Q}|.
\end{aligned}$$

□

Reaching the target

Lemma 4.4 establishes that, for a fixed source s and a fixed i, the expected number of steps for CautiousGreedy to get to within distance 2^{i-1} of a target chosen uniformly at random from the ball of radius 2^i around s is small as long as $|\mathscr{Q}|$ is small. By repeated invocations of this lemma, we can establish a polylogarithmic upper bound on the expected length of the path found by CautiousGreedy:

Theorem 4.5. *Fix a neighbor-connected rank-based social network with population size n, depth of field ρ, doubling dimension α, and aspect ratio Δ. Let s be an arbitrary source person, and let t be a target person chosen uniformly at random from the population. Then the expected length of the* CautiousGreedy(s,t) *path from s to $\mathsf{loc}(t)$ is $O(\log n \cdot \log \Delta \cdot (8/\rho)^\alpha)$, where the expectation is taken over both the random construction of the network and the random choice of t.*

Proof. Let $Y_{u,i}$ be a random variable that denotes the number of people that CautiousGreedy(u,t) encounters before it reaches a person within distance 2^{i-1} of a target t chosen uniformly at random from $B_{2^i}(u)$. For any u and i, by Lemma 4.4, we have $\mathtt{E}[Y_{u,i}] \leq |\mathscr{Q}| \cdot H_n \leq (8/\rho)^\alpha \cdot H_n$. Choose a target t uniformly at random from the population. For any u and any i, conditioned on $d(u,t) < 2^i$, the target t is a uniformly chosen person from $B_{2^i}(u)$. Thus, starting from any source node u_i, conditioned on the distance to the target being at most 2^i, the expected number of steps before CautiousGreedy reaches a node within distance 2^{i-1} of the target is $O(\log n \cdot (8/\rho)^\alpha)$. The total length of the $\textsc{CautiousGreedy}(s,t)$ is at most the number of steps required to reduce the distance to the target from Δ down to $1/2$—i.e., $O(\log \Delta)$ iterations of this process. By linearity of expectation and the above bound, the theorem follows. $\square$

5 A Tight Lower Bound on Cautious Greedy

We have shown that $\mathtt{E}_t[|\textsc{CautiousGreedy}(s,t)|] = O(\log^2 n)$ in networks with constant doubling dimension, constant depth of field, and aspect ratio that is polynomial in the population size. In this section, we exhibit a network with $\alpha = \Theta(1)$, $1/\rho = \Theta(1)$, and $\Delta = n$ such that the expected length of the path $\textsc{CautiousGreedy}(s,t)$ for a randomly chosen t is $\Omega(\log^2 n)$. Our results rely heavily on a lower bound proven by Martel and Nguyen [20] on $\textsc{Greedy}$: in a k-dimensional grid with $\mathtt{Pr}[u \to v] \propto d(u,v)^{-k}$, for any source–target pair $\langle s,t\rangle$ with $d(s,t) > cn$, the expected length of $\textsc{Greedy}(s,t)$ is $\Omega(\log^2 n)$, where a dependence on the constant c is hidden by the $\Omega(\cdot)$. Because rank-based link probabilities differ by only a constant factor from $\mathtt{Pr}[s \to t] \propto d(s,t)^{-k}$ in uniform-population grids [16], we need only connect $\textsc{CautiousGreedy}$ to $\textsc{Greedy}$ to derive a lower bound.

Consider a uniform-population rank-based social network $\mathcal{R}^n_{uniform}$ where the underlying metric space is a ring (that is, we take $L = \{0, 1, \ldots, n\}$ and $d(i,j) = \min(|i-j|, |j-i|)$ and exactly one person living at each point). For a target person t, write $g_s(t) := \mathtt{E}[|\textsc{Greedy}(s,t)|]$ and $c_s(t) := \mathtt{E}[|\textsc{CautiousGreedy}(s,t)|]$ to denote the expected length of the paths found by the two algorithms to a particular target t. (Here the expectation is taken only over the random choices of the rank-based friendships.)

Lemma 5.1. *In $\mathcal{R}^n_{uniform}$, (i) if $n/4 \geq d(s',t) \geq d(s,t)$, then $g_{s'}(t) \geq g_s(t)$; and (ii) if $d(s,t) \leq n/4$, then $c_s(t) \geq g_s(t)$.* $\square$

(The proofs, both by induction on $d(s,t)$, are omitted due to space constraints; the proof of claim (ii) relies on claim (i).)

Theorem 5.2. *In $\mathcal{R}^n_{uniform}$, for any source s and a target t chosen uniformly at random from the population,* $\textsc{CautiousGreedy}(s,t)$ *has expected length $\Omega(\log^2 n)$.*

Proof. For any constant $c \in (0, 1/4)$, there are $\Omega(n)$ people t with $cn < d(s,t) \leq n/4$. Thus with $\Omega(1)$ probability, the random target t satisfies $cn < d(s,t) \leq n/4$. For any such target t, the expected length of $\textsc{CautiousGreedy}(s,t)$ satisfies $c_s(t) \geq g_s(t) = \Omega(\log^2 n)$ by Lemma 5.1(ii) and the aforementioned theorem of Martel and Nguyen [20], trivially adapted to handle the differences between the distance-based and rank-based models, which can affect probabilities by a constant factor. For a constant fraction of the choices of t, then, we have an $\Omega(\log^2 n)$ bound on the expected length of $\textsc{CautiousGreedy}(s,t)$. $\square$

6 Future Directions

We have shown that, in rank-based networks, cautious-greedy routing performs well, in expectation for a randomly chosen target, as long as the underlying metric space has small aspect ratio, small doubling dimension, and large depth of field. In particular, we have been able to improve by a $O(\log \Delta)$ factor the results on $\textsc{Greedy}$ [16]. But two natural questions remain:

- Is the expected length of the path found by GREEDY to a randomly chosen target also short, say $O(\log^2 n)$ in a $\Theta(1)$-dimensional grid under Manhattan distance? (Or can CAUTIOUSGREEDY be much better than GREEDY?)

- Do either GREEDY or CAUTIOUSGREEDY achieve short paths—say of length $O(\log^{\Theta(1)}(n + \Delta) \cdot f(\rho, \alpha))$—for an *arbitrary* target (in expectation only over the random construction of the network)?

In fact, these questions appear to be intimately connected. In Sect. 1, we gave an example in which GREEDY appears to hurt itself by taking a long-range link that brings it very close to a "distracting" point of high population, which attracts a large fraction of the rank-based links from people that GREEDY subsequently encounters. But it is not too hard to see that GREEDY "escapes" from the shadow of a distracting point in polylog(n) steps, because there is a reasonable probability of increasing one's distance from the distraction by a factor of 3/2 at any step. However, it is an open question as to whether some adversarial construction of a set of distracting points of various sizes might cause a particular target to be hard to reach efficiently. This question appears to be closely related to the question of whether the distracting points that GREEDY may encounter in fact substantially slow down its performance. (For example, a natural analogue to Lemma 4.2 does not appear to hold for GREEDY unless the ball B is expanded, as was done in previous analysis [16], which had the carryover effect of the extra logarithmic factor.)

It would also be interesting to better understand the role of the depth of field of the metric space. It is known, for example, that for doubling dimension $\alpha = \omega(\log \log n)$, there is no decentralized routing algorithm that achieves polylogarithmic routing time for all pairs of nodes [10]. The interaction between doubling dimension and depth of field is an interesting direction for further study—for example, is there a similar lower bound in networks with low doubling dimension but very small depth of field?

7 Acknowledgements

Thanks to Esteban Arcaute, Seth Gilbert, Ravi Kumar, Jeff Ondich, Andrew Tomkins, and Sergei Vassilvitskii for helpful conversations.

References

[1] Ittai Abraham, Yair Bartal, and Ofer Neiman. Local embeddings of metric spaces. In *Symposium on Theory of Computing*, pages 631–640, June 2007.

[2] Ittai Abraham, Cyril Gavoille, Andrew Goldberg, and Dahlia Malkhi. Routing in networks with low doubling dimension. In *Intl. Conf. on Distributed Computing Systems*, July 2006.

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

[4] Lali Barrière, Pierre Fraigniaud, Evangelos Kranakis, and Danny Krizanc. Efficient routing in networks with long range contacts. In *Intl. Symp. on Distributed Computing*, pages 270–284, October 2001.

[5] 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.

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

[7] Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar, and Nicolas Schabanel. Towards small world emergence. In *Symposium Parallel Algorithms and Archictectures*, pages 225–232, August 2006.

[8] Pierre Fraigniaud. Greedy routing in tree-decomposed graphs. In *European Symposium on Algorithms*, pages 791–802, October 2005.

[9] Pierre Fraigniaud, Cyril Gavoille, and Christophe Paul. Eclecticism shrinks even small worlds. In *Principles of Distributed Computing*, pages 169–178, July 2004.

[10] Pierre Fraigniaud, Emmanuelle Lebhar, and Zvi Lotker. A doubling dimension threshold $\Theta(\log\log n)$ for augmented graph navigability. In *European Symposium on Algorithms*, September 2006.

[11] Sarunas Girdzijauskas, Anwitaman Datta, and Karl Aberer. On small world graphs in non-uniformly distributed key spaces. In *International Conference on Data Engineering*, page 1187, 2005.

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

[13] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Symposium on Theory of Computing*, pages 163–170, June 2000.

[14] Jon Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems*, pages 431–438, December 2001.

[15] Jon Kleinberg. Complex networks and decentralized search algorithms. In *International Congress of Mathematicians*, August 2006.

[16] Ravi Kumar, David Liben-Nowell, and Andrew Tomkins. Navigating low-dimensional and hierarchical population networks. In *European Symposium on Algorithms*, pages 480–491, September 2006.

[17] Emmanuelle Lebhar and Nicolas Schabanel. Close to optimal decentralized routing in long-range contact networks. In *Intl. Colloq. on Automata, Languages, and Programming*, pages 894–905, July 2004.

[18] 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.

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

[20] Chip Martel and Van Nguyen. Analyzing Kleinberg's (and other) small-world models. In *Principles of Distributed Computing*, pages 179–188, July 2004.

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

[22] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge U. Press, 1995.

[23] Ozgur Şimşek and David Jensen. Decentralized search in networks using homophily and degree disparity. In *Intl. Joint Conference on Artificial Intelligence*, pages 304–310, August 2005.

[24] Aleksandrs Slivkins. Distance estimation and object location via rings of neighbors. In *Principles of Distributed Computing*, pages 41–50, July 2005.

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