Neighbor Selection and Hitting Probability in Small-World Graphs

Oskar Sandberg*

October 25, 2007

Abstract

Small-world graphs, which combine randomized and structured elements, are seen as prevalent in nature. Jon Kleinberg showed that in some graphs of this type it is possible to route, or navigate, between vertices in few steps even with very little knowledge of the graph itself.

In an attempt to understand how such graphs arrise we introduce a different criterion for graphs being navigable in this sense, relating the neighbor selection of a vertex with the hitting probability of routed walks. In several models starting from both discrete and continuous settings, this can be shown to lead to graphs with the desired properties. It also leads directly to an evolutionary model for the creation of similar graphs by the stepwise rewiring of the edges, and we conjecture, supported by simulations, that these too are navigable.

1 Introduction

1.1 Shortcut Graphs

Starting with the small-world model of Watts and Strogatz [22], rewired graphs have been the subject of much interest. Such graphs are constructed by taking a fixed graph, and randomly rewiring some portion of the edges. Later models of partially-random graphs have been created by taking a fixed base graph, and adding "long-range" edges between randomly selected vertices (see [18] [19]). The "small-world phenomenon", in this context, is that graphs with a high diameter (such as a simple lattice) attain a very low diameter with the addition of relatively few random edges.

Jon Kleinberg [11] studied such graphs, primarily ones starting from a two dimensional lattice, from an algorithmic perspective. He allowed for $O(n)$ long-range edges, and found that not only would this lead to a small

*Thanks to my advisors Olle Häggström and Devdatt Dubhashi, as well as Ian Clarke in discussion with whom some of these ideas were born, and Jon Kleinberg for taking the time to listen to and reflect upon my ideas.

diameter, but also that if the probability of two nodes having a long-range edge between them had the correct relation to the distance between them in the grid, the *greedy routing* pathlength between vertices was small as well. Greedy routing means, as the name implies, starting from one vertex and searching for another by always stepping to the neighbor that is closest to the destination. That the base graph is connected means that a non-overlapping greedy path always exists, so the question regards the utility of the long range contacts in shortening this path. Graphs where one can quickly route between two points using only local information at each step, as with greedy routing, are referred to as *navigable*.

Initially, we will stay in the one-dimensional translation invariant environment (that is, with the vertices arranged on a circle) for our work below. Later sections extend some of the results to other classes of graphs. In general, we will call graphs of the type discussed *shortcut graphs* and use the shorter term *shortcut* for the long-range edges.

1.2 Contribution

While Kleinberg's results are important and have been a catalyst for much study, it is not fully understood how the rather arbitrary and precise threshhold on the shortcut distribution might arise in practice. In this work, we present an alternative distributional requirement that associates the shortcut distribution with the hitting probabilities of queries under greedy routing. We study this in canonical case of a single loop, and in a wider setting of graphs induced the Voronoi tessellations of a Poisson process. We show that distributions that meet the criterion, which we call "balanced distributions" have $O(\log^2 n)$ mean routing times, similarly to the critical case in Kleinberg's model.

The relationship in this criterion naturally leads to a stepwise re-wiring algorithm for shortcut-graphs. The Markov chain on the set of possible shortcut configurations defined by this algorithm can easily be seen to have a stationary distribution with balanced marginals. While the previous results cannot be directly applied to this case, because the stationary distribution has dependencies between the shortcuts at nearby nodes, we argue through heuristics and simulation that these dependencies in fact work in our favor, and that graphs generated by our algorithm can be efficiently navigated.

1.3 Previous Work

The root of the recent work on navigable graphs are the papers by Jon Kleinberg [11] [1]. Further exposition is given in [2] [15] [16]. Continuum models similar to the ones discussed below have been introduced in [10] and [6], as well as, in a more practical context [14].

A very different algorithm that appears to produce navigable graphs has been independently proposed in [5], where it is tested by simulation. In [8] the emergence of navigable graphs is discussed in terms of a method for small-world construction without requiring an understanding of the geography, but the method developed is complicated an unnatural. An algorithm similar to that proposed below is present in Freenet [4] [3] [23] - the work below was in part inspired by attempts to place Freenet's algorithms en environments for conductive to analysis.

A recent survey of the field by Kleinberg is [13]. In the final section, he identifies the question of how small world graphs arise as one of the central question in the field.

2 Preliminaries

2.1 Decentralized Routing

The central problem in this area of research is that of *routing* through a graph with only limited knowledge of the graph itself. That is, given two vertices x and y in a (di)graph G, we want to find a path connecting x and y. In general, the combinatorial problems of finding such a path, and finding the shortest such path, are well understood problems involving $\Theta(n)$ and $\Theta(n^2)$ steps respectively. The question becomes more interesting if we allow some (but not all) information about the graph to be given when determining the path. In particular, we have a distance measure $d(x, y)$ between vertices given. With such a distance measure, one may define a *decentralized algorithm* (following Kleinberg [11]) as an algorithm which, in each step, uses only information about the distances between present points in the route and the destination to decide where to go next.

Definition 2.1. *A* decentralized algorithm *for finding a path from a point y to z in a graph G associated with a distance function $d : V \times V \mapsto \mathbb{R}$ is defined as follows.*

- *Let the $S_0 = \{y\}$.*
- *In step k, the algorithm chooses exactly one point in $N(S_{k-1})$ (the set of all neighbors in G of points in S_{k-1}) and appends this point to create S_k. The choice of x is a possibly random function of the subgraph of G induced by S_{k-1}, as well as the distance of all the vertices in $N(S_{k-1})$ to each-other and to z as given by d. In particular, it may not depend on the rest of G.*
- *The algorithm terminates in step k if $z \in S_k$.*

The definition is inspired by the small world experiments [17] where people were enlisted to forward a letter to a person strange to them through friend-to-friend links. The people in the experiment knew something about the final recipient (typically where he lived and his occupation), and so could compare how "close" acquaintances they considered sending the letter to were to him, but had no global knowledge of the social network itself.

For a decentralized algorithm to be able to perform better than a random walk, it is necessary that $d(x, y)$ in some way contains information about the structure of the graph. The extreme of this is where $d(x, y)$ is the graph metric implied by G, the minimal distance from x to y in G, which we denote $d_G(x, y)$. In this case routing is trivial: proceeding in each step to the neighbor which is closest to z will always calculate a minimal path. A more typical case that where $d(x, y)$ gives some, but not complete, information regarding where to go. In particular, we shall say that $d(x, y)$ is *adapted for routing* in a graph G, if for any z and $x \in V$, x has a neighbor y such that $d(y, z) < d(x, z)$. When such a distance measure exists, we can route to any point by always choosing such a y as the next step, though the path calculated may be far from optimal.

The common situation is to let H be a fixed graph, and G created by randomly augmenting H with links in order to create a semi-random graph. It is then trivially true that $d_H(x, y)$ is adapted for routing in G. The random edges need not be uniformly distributed, and indeed all the interesting cases arise when the probability of the presence of an edge in G depends on $d_H(x, y)$. Independence is usually assumed, however, so that presence of an added edge between x and z is $\ell(x, z)$ independent of which other edges are added.

Given such a random augmentation of edges, the question arises whether a decentralized algorithm can be found which efficiently routes through a family of graphs. In particular, for a family for finite graphs of bounded degree that are indexed by size, is there a decentralized algorithm such that the expected number of steps of a route between two points is asymptotically small (by which we typically mean that it grows at most poly-logarithmically with the size).

In Kleinberg's original work [11], the underlying graph was $\mathbb{Z}_n^2$ (the family of finite two dimensional grids) with edges between adjacent vertices, making the $d(x, y)$ the l^1 metric (Manhattan distance). He proved that poly-logarithmic routing was possible if $\ell(x, z) = K_n/x^\alpha$ with $\alpha = 2$ (K_n is the distributions normalizer), but impossible for all other values of α. Kleinberg's results also cover the same situation in $\mathbb{Z}_n^d$, in which case the single good value of α is exactly d. Similar analysis has been applied, among others, by Barriere et. al. [2] for thorough analysis of the directed loop, and Duchon et. al. [7] for a wider class of graph families. In all these cases (as well as in [12] [21] [14] [20]) it found that efficient routing is possible when

$$\ell(x, y) \propto \frac{1}{\mathrm{Vol}(B_x(d(x, y)))} \tag{1}$$

where $B_x(r) = \{z : d(x, z) \leq r\}$, or some slight variation thereof. (We will use this notation for the ball, as well as $S_x(r)$ for its boundary throughout the paper.)

Similarly, it turns out that in all these cases, the decentralized algorithm necessary is simply *greedy routing*, which means choosing in each step the unexplored neighbor of the previously explored nodes which is closest to the destination. When $d(x, y)$ is an underlying metric, greedy routing strictly approaches the target with each step and is always successful. The nature of the greedy walks through such graphs is the main emphasis of this paper.

The following is a very coarse, obvious, upper bound on the routing time:

Observation 2.2. *If a distance function $d : V \times V \mapsto \mathbb{R}$ is adapted for routing in a graph in $G = (V, E)$ then greedy routing from x to z takes at most $|\{v \in V : d(v, z) < d(x, z)\}|$ steps.*

2.2 Distribution and Hitting Probability

Consider an underlying graph $H = (V, E)$, which may be directed but must be connected in the sense that it contains a direction respecting path from any node to any other. Let $d(x, y)$ be the underlying metric implied by H, and let a random graph G be constructed by augmenting H with one random directed edge starting at each vertex. The edges added by the augmentation will be denoted at as $\gamma : V \mapsto V$. We call γ a *shortcut configuration*, and let $\Gamma = V \mapsto V$ be the set of all such. The general probability space over which we will work is $\Gamma \times V \times V$, where the two copies of V are from where the start and destination of queries are chosen. Let $\mathbf{P}$ be a probability measure on this set where the start and destination are chosen uniformly and independently of each other and the configuration chosen by some *shortcut distribution* $\ell(\gamma)$ which in the independent selection case is $\prod_{x \in V} \ell(x, \gamma(x))$.

On this space, we define $X_Z^Y(t)$ as a greedy walk in the graph from a uniformly chosen starting point $Y = X_z^Y(0)$ with a uniformly chosen destination Z. To make the greedy walk well defined, we dictate that ties are broken randomly (that is, if the m closest neighbors to the destination are equally far from it, one is selected uniformly at random.) Below, we will in particular be interested in the hitting probability of greedy walks with specific destinations. We define this formally as:

$$h(x, z) = \mathbf{P}(X_Z^Y(t) = x \text{ for some } t | Z = z) \tag{2}$$

If H is a transitive graph, and $\ell(x, y)$ is a function of $d(x, y)$ only (we call this *distance invariance*), then $h(x, z) = h(x - z, 0)$ for some distinguished node 0. Thus we will, without further loss of generality, consider the hitting probability as a function of one variable and write $h(x) = h(x, 0)$.

Our results concern relating $h(x)$ with the occurrence of shortcuts between nodes. Immediately, however, we can see that $h(x)$ gives us the expected length of a greedy path. Since such a path can hit each point only once, it follows that if T is the length of a greedy path from a random point to zero, then

$$T = \sum_{x \in V} \chi_{\{X_0^Y(t)=x \text{ for some } t\}}$$

whence it follows that:

$$E[T] = \sum_{x \in V} h(x). \tag{3}$$

We will call the expected greedy walk length $\tau = E[T]$.

3 Re-wiring Algorithm

Before proceeding to analyse our main model, we present the re-wiring algorithm wich motivates it. Running the algorithm modifies, in each step, the destinations of the shortcut edges of vertices in the graph in a random fashion. It is a steady-state algorithm in the sense that it neither creates nor destroys edges: it simply shifts the destinations of the single existing shortcut at each vertex.

In the sense that we propose a generative process which might explain why navigable graphs arise, this is similar to the celebrated preferential attachment model for power law graphs of Barabási and Albert. However, it is a not a growth model for the graph since the number of nodes and edges never changes, and is thus more similar to the model discussed in [9].

The proposed algorithm is as follows:

Algorithm 3.1. *Let (V, E_s) be the directed graph of shortcuts at time s. From each vertex there is exactly one edge. Let $0 < p < 1$. Then (V, E_{s+1}) is defined as follows.*

1. *Choose y_{s+1} and z_{s+1} uniformly from V.*
2. *If $y_{s+1} \neq z_{s+1}$, do a greedy walk from y_s to z_s along the lattice and the shortcuts of E_s. Let $x_0 = y_{s+1}, x_1, x_2, ..., x_t = z_{s+1}$ denote the points of this walk.*
3. *For each $x_0, x_1, ..., x_{t-1}$ independently with probability p replace its current shortcut with one to z_{s+1}.*

After a walk is made, E_{s+1} is the same as E_s, except that the shortcut from each node in walk $s + 1$ is with probability p replaced by an edge to the destination. In this way, the destination of each edge is a sample of the destinations of previous walks passing through it. We strongly believe that updating the shortcuts using this algorithm eventually results in a shortcut graph with greedy path-lengths of $O(\log^2 n)$. Though one can relate the

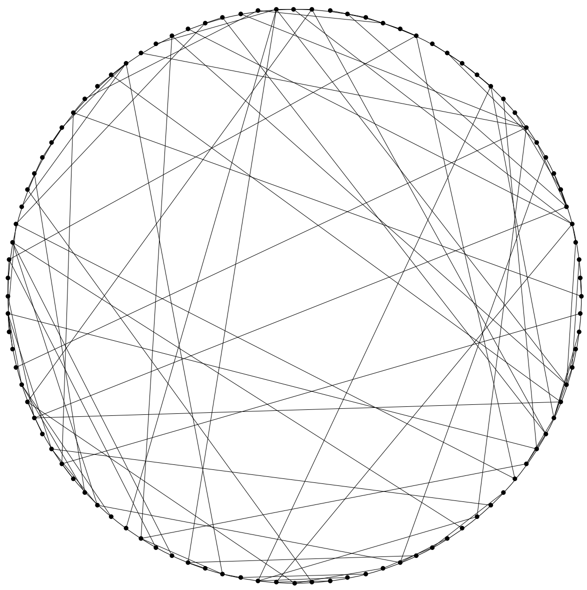

Figure 1: A shortcut graph generated by our algorithm ($n = 100$).

stationary regime of this algorithm to the balanced distributions (see below), a rigorous bound has not been proved.

The value of p is a parameter in the algorithm. It serves to disassociate the shortcut from a vertex with that of its neighbors. For this purpose, the lower the value of $p > 0$ the better, but very small values of p will also lead to slower sampling. It is hard to state an optimal value for p but there are simple heuristic arguments for why p should reasonably be on the order of one over the expected length of the greedy walks.

3.1 Markov Chain View

Each application of Algorithm 3.1 defines the transition of a Markov chain on the set of shortcut configurations, Γ. The Markov chain in question is defined on a finite (if large) state space. If it is irreducible and aperiodic, it thus converges a unique stationary distribution.

Proposition 3.2. *The Markov chain* $(E_s)_{s\geq 0}$ *is irreducible and aperiodic.*

Proof. Aperiodic: There is a positive probability that $y_s = z_s$ in which case nothing happens at step s.

Irreducible: We need to show that there is a positive probability of going from any shortcut configuration to any other in some finite number of steps. This follows directly if there is a positive probability that we can "re-point" the shortcut starting at a vertex x to point at a given target y without changing the rest of the graph. But the probability of this happening in a single iteration is:

$$\geq \frac{1}{n}\frac{1}{n}p(1-p)^{n-2} > 0.$$

□

Thus there does exist a unique stationary shortcut distribution, which assigns some positive probability to every configuration. The goal is to motivate that this distribution leads to short greedy walks.

Proposition 3.3. *Under unique stationary distribution of the Markov chain* $(E_s)_{s\geq 0}$

$$\ell(x,z) = \frac{h(x,z)}{\sum_{\xi=1}^{n-1} h(\xi)}$$

Proof. As selected by the algorithm, the shortcut from a vertex x at any time is simply a sample of the destination of the previous walks that x has seen. Under the stationary distribution this should not change with time, so

$$\ell(x,z) = \mathbf{P}(Z = z | X_Z^Y(t) = x \text{ for some } t).$$

Using Bayes' theorem, this can be seen as a statement relating ℓ to the hitting probability.

$$\begin{aligned} \ell(x,z) &= \mathbf{P}(Z = z | X_Z^Y(t) = x \text{ for some } t) \\ &= \frac{\mathbf{P}(X_Z^Y(t) = x \text{ for some } t | Z = z)\mathbf{P}(Z = z)}{\sum_{\xi \neq x} \mathbf{P}(X_Z^Y(t) = x \text{ for some } t | Z = \xi)\mathbf{P}(Z = \xi)} \end{aligned}$$

The first multiple in the numerator is the hitting probability $h(x,z)$. The formula then follows from the uniform distribution of Z and translation invariance. □

4 Balanced Shortcut Distributions

We use Proposition 3.3 as the starting point of our analysis, defining the class of all distributions meeting the same marginal property as follows.

Definition 4.1. *If a graph H with distance function $d(x,y)$ is randomly augmented such that:*

$$\ell(x,z) = \frac{h(x,z)}{\sum_{\xi=1}^{n-1} h(\xi)} = \frac{h(d(x,z))}{\tau} \tag{4}$$

where h is given by (2), then the joint distribution of shortcuts is called balanced*.*

We will show for several classes of graphs this relationship leads to navigable graphs, allowing for a characterization other than that of (1). Besides the relationship with Algorithm 3.1, balance is in some ways a natural requirement. The left hand side describes the distribution of destinations of walks that hit the point x, so our results simply say that a good choice of shortcuts is one that matches this.

Theorem 4.2. *For transitive graph H, there exists a balanced distribution which selects shortcuts independently at each node.*

Proof. Like before, we let $\ell(x, y)$ be the marginal probability that x has a shortcut to y. The joint distribution is simply the product over all vertices.

For a single walk toward a given z, we may view $X_z^Y(t)$, as Markov chain on the set of vertices, with some transition kernel $P_z(y, x)$. As above, we will set $z = 0$, and drop the index in the below calculations without loss of generality. The process hits every point except $z = 0$ at most once, and we can let this point be absorbing. The transition kernel P then consists of two mechanisms: either we step to x which is closer to 0 than y because it is the destination of the shortcut from y, or we step to one of y's neighbors in H because y's shortcut leads to somewhere from which it is further to 0 than y.

Let $N(x)$ be the set of neighbors of x in H, $L(x) = \{\xi \in V : d(\xi, 0) \geq d(x, 0), \xi \neq x\}$, the set of nodes at least as far as x from 0. Also, let $P(x) = \{\xi \in N(x) : d(\xi, 0) = d(x, 0) + 1\}$ (the set of "parent" nodes that can greedy route to x in H)and $C(x) = \{\xi \in N(x) : d(\xi, 0) = d(\xi, 0) = d(x, 0) - 1\}$ (the set of "child" nodes that x can route to). Then the transition kernel of the process described is:

$$P(y, x) = \begin{cases} 0 & \text{if } d(x, 0) \geq d(y, 0), x \notin C(y) \\ \ell(y, x) + \frac{1}{|C(y)|} \sum_{\xi \in L(y)} \ell(\xi) & \text{if } x \in C(y) \\ \ell(y, x) & \text{if } d(x, 0) < d(y, 0) - 1 \end{cases}$$

for $y \neq 0$. $P(0, x) = \chi_{\{x=0\}}$.

We can thus express the hitting probability for any $x \neq 0$ for a greedy walk as:

$$\begin{aligned} h(x) &= \sum_{\xi \in V : \xi \neq x} h(\xi) \ell(\xi, x) + \frac{1}{n-1} \\ &= \sum_{\xi : d(\xi, 0) > d(x, 0)} h(\xi) \ell(\xi, x) + \sum_{\xi \in P(x)} h(\xi) \sum_{\eta \in L(\xi)} \frac{\ell(\eta, \xi)}{|C(\xi)|} \\ &\quad + \frac{1}{n-1} \end{aligned} \tag{5}$$

Where the first two terms represent the probability that we enter x through a shortcut or a parent node in $P(x)$ respectively, and the last term is the probability that the walk starts at x.

Note that for any x (5) gives a recursive definition of $h(x)$ in terms of the distribution ℓ. Fix such a distribution ℓ'. From this we can thus calculate the hitting probabilities $h'(x)$, and define:

$$\ell''(x) = \frac{h'(x)}{\sum_{x \in V} h'(x)}.$$

The mapping of $\ell' \mapsto \ell''$ is continuous since $\sum_{x \in V} h'(x) > 1$ and maps the simplex of $n-1$ valued distributions into itself. Since the simplex is convex, Brouwer's fix-point theorem gives the existence of at least one fix-point ℓ^*, which is a balanced distribution. □

4.1 Almost Balanced Distributions

To increase the utility of the class distributions, one may weaken the requirements of (4) somewhat, and say that a shortcut distribution associated is *almost balanced* with respect to a base graph for some constant $a \in (0, 1]$ if

$$\ell(x, z) \geq \frac{ah(d(x, z))}{\tau}$$

for all ℓ in the family and all $x, z \in V$ of the graph to which ℓ is associated.

Since all the graphs dealt with are finite, all distributions with full support are in fact almost balanced for some $a > 0$, so on its own this definition becomes meaningless. For a family of base graphs, however, one could imagine constructing an associated family of distributions that are almost balanced for the same value of a. For reasons of clarity, we will use the stricter form of balance, but all the results carry over with only a constant multiplicative penalty (depending on a) to the performance, and minor modifications to the proofs.

5 The Directed Cycle

We let H be the directed cycle of n nodes, which will be numbered 0 through $n-1$ such that the edges are directed downward (modulo n). The implied distance function (which is not symmetric) is

$$d(x, y) = \begin{cases} x - y & \text{if } y \leq x \\ n - y + x & \text{otherwise.} \end{cases}$$

This environment is perhaps the most direct home to greedy routing, and has previously been the target of the thorough analysis by [2]. There exists exactly one point at each distance from 0, and greedy routing means selecting the shortcut if and only if its destination lies between 0 and the current position. Equation (5) here simplifies to:

$$h(x) = \sum_{\xi=x+1}^{n-1} h(\xi)\ell(\xi - x) + h(x+1) \sum_{\xi=x+2}^{n-1} \ell(\xi) + \frac{1}{n-1}.$$

To prove our result in this environment, we will need the following lemma:

Lemma 5.1. *If the shortcut configuration is chosen according to a distance invariant joint distribution, then $h(x)$ is non-increasing in x.*

Note that this lemma holds in greater generality than we need for the current section. Specifically, it does not depend on the shortcuts being chosen independently.

Proof. Let $I \subset \Gamma \times V$ be event consisting of all configurations and starting points such that a greedy walk for 0 hits the point $x+1$. Now we translate all the coordinates of this set down one coordinate (modulo n), and call the translated set J.

$$h(x+1) = \mathbf{P}(I) = \mathbf{P}(J)$$

by definition and distance invariance. However, every element in J corresponds to a starting point and shortcut configuration for which the greedy walk hits x. To see this, we pick a starting point y and configuration γ, such that $(\gamma, y) \in I$. This means that there is an integer m and a path $x_0, \ldots, x_m$ such that $x_0 = y$, $x_m = x+1$ and either

$$n - 1 \geq \gamma(x_i) > x_i \text{ and } x_{i+1} = x_i - 1$$

or

$$x_i > \gamma(x_i) \geq x + 1 \text{ and } x_{i+1} = \gamma(x_i)$$

for all $i = 0 \ldots m$. The corresponding configuration in J has a similar path $x'_0, \ldots, x'_m$ ($x'_i = x_i - 1$) where $x'_0 = y - 1$, $x'_m = x$ and either:

$$n - 2 \geq \gamma(x'_i) > x'_i \text{ and } x'_{i+1} = x'_i - 1$$

or

$$x'_i > \gamma(x'_i) \geq x \text{ and } x'_{i+1} = \gamma(x_i)'$$

for all $i = 0 \ldots m$. This means that starting in $y - 1$ will cause the greedy walk to hit x. (Note that not every configuration and starting point that cause greedy walks to hit x are necessarily in J, since $\gamma(x'_i)$ must be less than $n-2$ since rather than $n-1$ in the first line).

It now follows directly that

$$\mathbf{P}(J) \leq h(x).$$

□

We can now show that greedy routing here has the same order as the critical case in Kleinberg's model.

Theorem 5.2. *For every $n = 2^k$ with $k \geq 3$, the shortcut graph with shortcuts selected independently according to a balanced distribution has an expected greedy routing time*

$$\tau \leq 2k^2.$$

The proof method is similar those introduced by Kleinberg for the case (1) links, but the implicit definition of the shortcut distribution requires a somewhat more involved approach.

Proof. Assume that $\tau > 2k^2$. We will show that for k sufficiently large this always leads to a contradiction.

To start with, divide $\{1, \ldots, n-1\}$ into at most k disjoint phases. Each phase is a connected set of points, each successively further from the destination 0, and they are selected so that a greedy walk is expected to spend as many steps in each phase. Thus, the first phase is the interval $F_1 = \{1, \ldots, r_1\}$ where r_1 is the smallest number such that

$$\ell(F_1) = \sum_{\xi \in F_1} \ell(\xi) \geq 1/k$$

The second phase is defined similarly as the interval $\{r_1 + 1, \ldots, r_2\}$: again being the smallest such interval so that $\ell(F_2) \geq 1/k$. Let m be the total number of such intervals which can be formed, and let F_R denote remainder interval $\{r_m + 1, \ldots, n-1\}$, if necessary (let it be the empty set otherwise). By construction $\ell(F_R) < 1/k$ and the total number of phases, including F_R is $\leq k$.

Before proceeding, we need to bound how much ℓ of the different phases can deviate, since this will also tell us how much the expected number of steps in each phase can differ. From (4) and the assumed lower bound of τ, it follows that:

$$\ell(x) = \frac{h(x)}{\tau} \leq \frac{1}{2k^2}$$

for all x. This implies that $1/k \leq \ell(F_i) \leq 1/k + 1/(2k^2)$ for all $i \in \{1, \ldots, m\}$, and thus:

$$\ell(F_i) \leq \left(1 + \frac{1}{2k}\right) \ell(F_j) \tag{6}$$

for all $i, j \in \{1, \ldots, m\}$. It also gives $m \geq k^2/(k+1) - 1$.

Consider now $F_m = \{r_{m-1}+1, r_{m-1}+2 \ldots, r_m\}$ and let $L = \{0, 1, \ldots, r_{m-1}\}$. Our goal is to show that from any point in F_m there is a considerable probability of having a shortcut to L. We know that $r_m \leq n$. Assume that $r_{m-1} \geq r_m/2$. F_m then covers less than half of the distance from r_m to the target. In particular

$$\{r_m - F_m\} \subset L$$

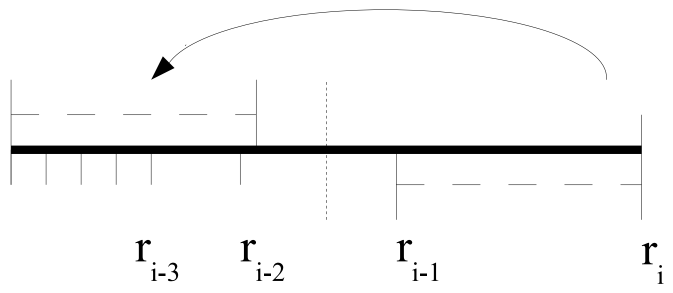

Figure 2: Illustration for the proof of Theorem 5.2. If a phase covers less then half of the “remaining ground”, then the a shortcut in the equivalent range takes us out of the phase.

Thus, if r_m has a shortcut with destination in $\{r_m - F_m\}$, any query which hits r_m will leave F_m in the next step. See Figure 2. We know that:

$$\ell(r_m, L) \geq \ell(r_m, \{r_m - F_m\}) = \ell(F_m) \geq 1/k.$$

Lemma 5.1 tells us that the probability of having a shortcut to L cannot decrease for points less than r_m, so for each vertex the query hits within F_m, there is an independent probability of $1/k$ of leaving F_m in the next step. This means that the expected number of steps the query can take in F_m is at most k.

The expected number of steps in a phase, $h(F_i) = \tau\ell(F_i)$, so by (6) it then holds that:

$$h(F_i) \leq (1 + 1/2k)h(F_m) \leq k + 1/2 \tag{7}$$

for all $i \in \{1, \ldots, m\}$ and also for F_R. There are at most k phases, so this implies that $\tau \leq k^2 + k/2$, which contradicts our assumption for all $k \geq 2$.

Thus the original assumption implies that $r_{m-1} \leq r_m/2 \leq n/2$. But by an identical argument for F_{m-1}, we can show that $r_{m-2} \leq r_{m-1}/2$. It follows by iteration that

$$r_i \leq \frac{1}{2^{m-i}} n.$$

and specifically:

$$r_1 \leq \frac{1}{2^{m-1}} n \leq 2^{\frac{k+2}{k+1}} \leq 4.$$

This means that F_1 contains at most 4 points, which means that $h(F_1) \leq 4$, and thus, again by (6), $h(F_i) \leq 5$ for all i. For $k \geq 3$ this contradicts the original assumption. The result follows. □

Note: It may seem strange that in the very last step we use that a constant is $O(k)$. In fact, this shows that the bound in the theorem could be strengthened somewhat, though it would have the same dominant order in k.

Theorem 5.2 gives us an alternate distributional criterion for attaining $O(\log^2 n)$ expected greedy path-lengths. Since Kleinberg showed that this cannot hold for many distributions, the balanced distributions must be "close" to the critical, harmonic decay. More specifically, drawing on the counterpoofs for navigability for other cases in [11], we can see that there can not exist $\delta > 0, \epsilon > 0$ and $N \in \mathbb{N}$ such that $\ell(\{1, 2, \ldots, n^\delta\}) \leq n^{-\epsilon}$ for the cycles of size $n \geq N$, as would be the case if the distributions' tails were dominated a power-law $(x^{-\alpha})$ decay with exponent $\alpha < 1$. Similarly, there cannot exist (possibly different) $\delta > 0, \epsilon > 0$ and $N \in \mathbb{N}$ such that $\ell(\{n^\delta, n^\delta + 1, \ldots, n-1\}) \leq n^{-\epsilon}$ for the cycles of size $n \geq N$, as would be the case if the distributions' tails were dominated by a power-law with exponent $s > 1$.

6 Delaunay Graphs

The small-world theory is not necessarily limited to cases where nodes are placed in a fixed grid. In the following section, we will let the vertices be points of a spacial Poisson process, and the distance function the euclidean metric. For simplicity, we will relax our requirements a little and let the graphs have degrees bounded in expectation, rather than uniformly.

Let S_r^d be the d-dimensional surface of a $d+1$ sphere with radius 1. We let $V = \{x_i\}$ be the N points of intensity $\lambda = n^d$ homogeneous Poisson process in this space. From this Poisson process we may construct the Voronoi tessellation, that is collection of cells $C(x_i)$ where:

$$C(x_i) = \{y \in \mathbb{R}_s^d : d(y, x_i) = \min_{z \in V} |z - x_i|\}.$$

$C(x_i)$ is that part of the space which is as close to x_i as any other point. Voronoi cells are closed convex polyhedrons that border other cells along each side, overlapping on Lebesgue measure zero set.

The tessellation induces a graph G with vertices V (known as the Delaunay graph) as follows, let $(x, y) \in E$ if and only if $C(x)$ and $C(y)$ intersect on an infinite number of points (this is a.s. equivalent to intersecting at one point). Intuitively this is the graph that connects a vertex x to all its neighbors in the tessellation. This graph is a natural base graph for greedy routing among the points:

Lemma 6.1. *Let $\{x_i\}$ be any point-set in S^d, and G its Delaunay graph. Then the euclidean metric $d(x, y) = |x - y|$ is adapted for routing in G.*

Proof. We must prove that $\forall x \neq z \in V$, there $\exists y \in V$ (which may be z) such that $(x, y) \in E$ and $|x - z| > |y - z|$. Consider the line xz. Let w be the first point as we more from x along xz such that $w \in C(y)$ for some $y \neq x$ (w is well-defined since the cells are compact).

It is clear that $w \in C(y)$ for some y such that x and y are connected in the Delaunay graph ($C(x)$ must border at least one cell that it meats at w). Clear $|y - w| = |y - w|$ since w is in both cells:

$$\begin{aligned} |y - z| &< |y - w| + |w - z| \\ &= |x - w| + |w - z| = |x - z| \end{aligned}$$

where the strict inequality follows from the fact that w is not on the line yz. □

Given this graph, we consider augmentations that allow for fast routing. A direct approach would be to connect a given node x to any other y with a probability depending on the distance between them, but this leads to complications regarding dependence between the progress at each step (though not insurmountable ones, see [6] for such an approach in a similar environment).

Instead, we augment the graph as follows. For each node $x \in V$, let $\{n_i(x)\}_{i=1}^{N(x)}$ be the points of a non-homogeneous Poisson process with given by the measure $\mu_x(A) = \ell_x(A \backslash C(x))$ for some *shortcut measure* ℓ on the Borel sets of S^d, and $\ell_x(A) = \ell(A - x)$.

We then augment G by adding a directed edge from x to y if $n_i(x) \in C(y)$ for any $i = 1, \ldots, N$.

Lemma 6.2. *If $x, z \in V$ and for some $i = 1, \ldots, N$ $|z - n_i(x)| \leq |z - x|/4$, then x has a shortcut $y \in V$ (which may be z) such that $|z - y| \leq |z - x|/2$.*

Proof. Let w be such an $n_i(x)$. With probability one it is in exactly one cell $C(y)$. If $y = z$ then x has a shortcut to z, or otherwise $|w - y| < |w - z|$. In the latter case,

$$|z - y| \leq |z - w| + |w - z| < 2|w - z| \leq |x - z|/2.$$

□

6.1 Kleinberg Augmentation

To motivate the model, we first show that that augmentation along the lines Kleinberg's model allows for a $\log^2(n)$ bound of the routing time. That is, as in 1

$$\ell(A) = \int_A \frac{dr}{\log n \text{Vol}(r)} \tag{8}$$

where $\text{Vol}(r)$ is the volume of a radius r ball in S^d.

Before proving a lower bound on the expected routing type, we need to ensure that we are not adding an unbounded number of edges.

Lemma 6.3. *The expected number of shortcuts added to each node under augmentation with intensity (8) is bounded by a constant.*

Proof. First note that $\mathbf{E}[\#\text{shortcuts added to } x] \leq E[N(x)]$. Now, let $R(x) = \inf |y - x| : y \in V, y \neq x$. If $R(x) = \delta$, then then all points within $\delta/2$ of x are in $C(x)$. Thus

$$\begin{aligned}
\mathbf{E}[N(x) \mid R(x) = \delta] &\leq \int_{S^d \quad B_{\delta/2}(x)} \frac{1}{\text{Vol}(x-y)} dy \\
&\leq \int_{\delta/2}^{1} \frac{1}{r \log s} dr = \frac{\log(2/\delta)}{\log n}
\end{aligned}$$

Hence, and since $\mathbf{E}[N(X) = \mid R(x) = y]$ is decreasing in y,

$$\begin{aligned}
\mathbf{E}[N(x)] &= \int_0^1 E[N(x) \mid R(x) = \delta] f_{R(x)}(\delta) d\delta \\
&\leq \mathbf{E}[N(x) \mid R(x) = 1/n] \mathbf{P}(R(x) \geq 1/n) \\
&\quad + \int_0^{1/n} \frac{\log(2/\delta)}{\log n} n^d S(\delta) e^{-n^d \text{Vol}(\delta)} d\delta \\
&\leq 2 + \frac{n^d S(1/n)}{\log n} \int_0^{1/n} \log(2/\delta) d\delta \\
&\leq c
\end{aligned}$$

where $S(\delta)$ is the surface area of a sphere of volume δ, and c is a constant independent of n. $\square$

The proof of the following theorem uses the by-now standard argument from [11].

Theorem 6.4. *For every n sufficiently large, the shortcut graph created by augmenting the Poisson-Delaunay graph with intensity (8) has an expected greedy routing time of $O(\log^2 n)$.*

Proof. Let the route currently be at the vertex x, such that $|x - z| = d > 1/n$. Let A be the event that $|n_i(x) - z| \leq d/4$ for some i.

$$\mathbf{P}(A) \geq \frac{\text{Vol}(d/4)}{\text{Vol}(3d/4) \log n} = \frac{c}{\log n}.$$

By Lemma 6.2, if such a $n_i(x)$ exists, then x has a neighbor within $d/2$ of z, and greedy routing at least halves the distance to z in the next step. If A fails to occur, then we know by Lemma 6.1 greedy routing will progress, and the event A is independent at each step. Thus the expected number of steps until halving the distance is $\log n$, which together with Lemma 2.2 proves the result. $\square$

6.2 Balanced Augmentation

In order to derive a result similar to Theorem 5.2 for the Delaunay setting, we will need to re-define the "balanced distribution" somewhat. In particular, we need will marginalize over the positions of the Poisson points.

Let the hitting measure of a $A \subset S^d$ be defined by:

$$h_z(A) = \mathbf{E}(\text{number of } t \text{ s.t. } X_Z(t) \in A \,|\, Z = z)$$

where $X_z(t)$ is the greedy routing process as above, and the existence of a point at z is included in the conditioning. Note that by the translation invariance of the construction, $h_Z(A) = h_0(A - z)$.

We call a distribution Poisson-balanced if

$$\ell(A) = \frac{h_0(A)}{\tau} \tag{9}$$

where $\tau = \mathbf{E}[\text{length of a greedy walk}] = h_z(S^d \backslash \{0\})$.

Lemma 6.5. *There exists a Poisson-balanced distribution.*

Proof. The proof is similar to the discrete case. A given shortcut measure ℓ' induces a hitting measure $h'_0(A)$, which in turn constructs a measure ℓ'' through (9). If we let L be the space of measures of total probability one on $S^d\{0\}$ equipped with the total variation metric

$$d(\mu, \nu) = \sup_{A \in \mathcal{B}(S^d\{0\})} |\mu(A) - \nu(A)|$$

then the mapping $\ell' \mapsto \ell''$ is a mapping from L to itself. L is continuous and compact, so it suffices to show that the mapping is continuous for us to apply Brouwer's fix-point theorem.

Since we know that $\tau > 1$, the second step of the mapping is certainly continuous. The first is also since the hitting probability depends only on a finite number of random variables with distribution depending on ℓ'. Formally:

Take $\epsilon > 0$ and any $m = n^d$. Let ℓ_1 and ℓ_2 be two shortcut measures. WLOG we assume that $\ell_2 \geq \ell_1$, and we let $d(\ell_1, \ell_2) \leq \epsilon'$, where

$$\epsilon' \leq \frac{\epsilon}{3m \max((e-1)n, \log(3m/\epsilon))}$$

We couple the routing processes X_0^1 and X_0^2 by letting them use the same set of Poisson process distributed vertices V, and the same starting point z. At each $x \in V$, we construct shortcuts $n_i(x)$ according to ℓ_1 which both processes may use, and then add an additional set of of shortcuts $n_i^2(x)$ according to $\ell_2 - \ell_1$ which only X_0^2 may use.

It follows that for any x, $|\{n_i^2(x)\}|$ is dominated by a Poi(ϵ') random variable, so

$$\mathbf{P}(|\{n_i^2(x)\}| > 0) \leq 1 - e^{-\epsilon'} \leq \epsilon'$$

Let the A be the event that any vertex x in V has $|\{n_i^2(x)\}| > 0$. Then

$$\begin{aligned}\mathbf{P}(A) &\leq \mathbf{P}(A \,|\, |V| \leq (e-1)m + q) + \mathbf{P}(|V| \leq (e-1)m + q) \\ &= ((e-1)m + q)\epsilon' + e^{-q} \\ &\leq \frac{\epsilon}{m}\end{aligned}$$

where the last step follows by setting $q = \log(3m/\epsilon)$.

Now let $H_1(A)$ and $H_2(A)$ be the number of points reached in A by the processes respectively. If the h_0^1 and h_0^2 are the respective hitting probabilities, then:

$$\begin{aligned}|h_0^1(A) - h_2(A)| &= \mathbf{E}[|H_1(A) - H_2(A)|] \\ &= E[|V|]\mathbf{P}(H_1(A) \neq H_2(A)) \leq \epsilon\end{aligned}$$

□

In order to the bound the routing time in this case, we will need the following geometrical fact.

Lemma 6.6. *There exists a $0 < q < 1$ such that if x and y are points in a S^d, such that $(3/4)\delta < |x - y| \leq \delta$, and $(3/4)\delta < r \leq \delta$, then the portion of the sphere $S_r(y)$ which lies inside $B_{(3/8)d}(x)$ is at least q. The constant q depends on d but not on δ and r.*

This follows from the fact that formulation is independent of scale. In one dimension $q = 1/2$ trivially, and in two it can easily be seen that it exceeds $1/8$, see Figure 3.

Theorem 6.7. *For every $n = \frac{4}{3}^k$ sufficiently large, the shortcut graph created by augmenting the Poisson-Delaunay graph with a Poisson-balanced distribution has an expected greedy routing time $\tau \leq \frac{k^2}{q}$.*

Proof. We let X_0^Y be a the routing process for zero, and define h_0 on $S^d \backslash \{0\}$ as above. We then divide $S^d \backslash \{0\}$ into k phases of the form $F_i = x \in S^d : r_{i-1} < |x| \leq r_i$ where $r_0 = 0$ and each subsequent r_i is defined so that:

$$h_0(F_i) = \frac{\tau}{k}.$$

For any phase F_i, assume that $r_{i-1} \geq \frac{3}{4} r_i$. Let x be a vertex in F_i: then q of each spherical "level" in $x + F_i$ lies in $L_i = B_0((3/8)r_i)$ by Lemma 6.6.

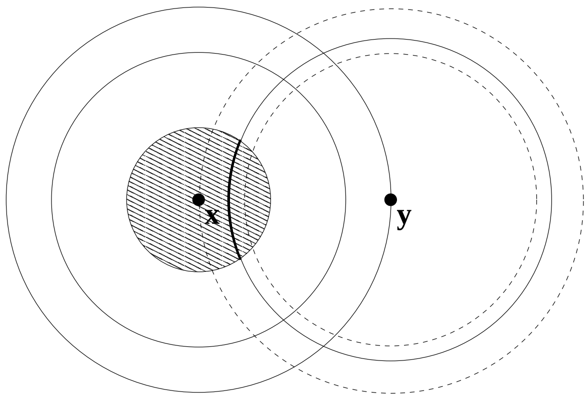

Figure 3: The circle around y intersects the ball around x in at least $1/8$ of its points.

By rotational invariance it follows that $\ell_x(L_i) = q\ell_x(x + F_i) = q\ell(F_i)$, so if A is the event x has a shortcut destination $n_i(x)$ closer than $r_{i-1}/2$

$$\mathbf{P}(A) = \ell_x(B_0((3/8)r_i)) \geq q\frac{h_0(F_i)}{\tau} = \frac{q}{k}.$$

By Lemma 6.2, if such a $n_i(x)$ exists, then x has a neighbor within $d/2$ of z, and greedy routing at least halves the distance to z in the next step. If A fails to occur, then we know by Lemma 6.1 greedy routing will progress, and the event A is independent at each step. Thus $h_0(F_i) \leq \frac{k}{q}$ whence $\tau \leq \frac{k^2}{q}$.

If, on other hand, for all i, $r_{i-1} \leq \frac{3}{4}r_i$, then

$$r_1 \leq \frac{3}{4}^{k-1} = \frac{4}{3n}.$$

Let N be the number of vertices in F_1. By adaptability:

$$h_0(F_1) \leq \mathbf{E}[N] = \frac{\mathrm{Vol}(r_1)}{n^d} = c$$

It follows that $\tau \leq ck$, so the result holds when $k > cq$. □

7 The Re-Wiring Algorithm, Again

Proposition 3.3 shows that under the stationary distribution of the re-wiring algorithm introduced above, the marginal shortcut distribution at each point is balanced, and it is tempting to apply Theorem 5.2. However, that theorem assumed that the shortcuts had been chosen independently at each vertex, which is not the case under Algorithm 3.1 which originally motivated the work. Showing that these dependencies do not negatively affect routing is an open problem, but which we discuss in general terms in this section.

There are two sources of dependencies between the shortcuts of neighboring vertices. Firstly, there is a chance that they sampled the destination of the same walk. When p is large, this dependency is substantial, and we see a highly detrimental effect even in the simulations. By using a small p, however, this dependence is muted. Another, more subtle dependence, has to do with the way the shortcuts of vertices around a vertex x may affect the destinations of the walks it sees. If $x+1$ has a shortcut to $x-10$, that will make it less likely for x see walks for places "beyond" $x-10$ since many such walks will have followed the shortcut at $x+1$, and thus skipping over x.

The first dependence, that of sampling from the same walk, can be handled by modifying the algorithm to make sure we do not sample more than once for each walk. Take $p \leq 1/n$ and once a walk is completed, we choose to update exactly one of its links with probability pw where w is the length of the walk. Which link to update is then chosen uniformly from the walk. This way, the probability a vertex updates its shortcut when hit by a walk is still always p, but we never sample two shortcuts from the same walk. The modified algorithm is less natural, but clearly a good approximation of the original for small p values. Although it is more complicated, it is easier harder to analyze, since it allows for the simplifying assumption that each edge is chosen from a different greedy walk.

The other dependencies are more complicated, and there is no easy way to modify the algorithm to remove them. However, it is worth noting that it is hard to see why these dependencies (unlike the first type) would be destructive for greedy routes. In fact, it makes sense that if x in our example gets few walks destined beyond $x-10$ because of the shortcut present at $x+1$, then it should also choose a shortcut to beyond $x-10$ with a smaller probability.

In the proof of Theorem 5.2 we use independence only to show that if the probability of having a shortcut out of a phase at the very furthest point is ρ, then the expected steps in the phase is bounded by $1/\rho$. There is little reason to believe this wouldn't hold under the algorithm, since if the link from the furthest point doesn't take us out the phase, it either goes to a point within the phase, or overshoots the destination. If it goes to a point within the phase, then we follow it, and the presence of that shortcut should not interfere with the shortcut from the destination. If it, on the other hand, overshoots, then by the above argument it should make it more likely that the following ones don't overshoot, giving a us a better than independent probability of leaving the phase.

Formalizing the requirements on the dependence, and proving that our stationary distribution indeed agrees with them, is the main open problem left to resolve about this work.

7.1 Computer Simulation

Simulations indicate that the algorithm gives results which scale as desired in the number of greedy steps, and that the distribution approximates $H_n/d(x,y)$.

The results in the directed one-dimensional grid can be seen in Figure 4. To get these results, the graph is started with no shortcuts, and then the algorithm is run $10n$ times to initialize the references. The value of $p = .10$ is used. The greedy distance is then measured as the average of 100,000 walks, each updating the graph according to the algorithm. The effect of running the algorithm, rather than freezing one configuration, seems to be lower the variance of the observed value.

The square root of the mean greedy distance increases linearly as the graph size increases exponentially, just as we would expect. In fact, as can be seen, our algorithm leads to better simulation results than choosing from Kleinberg's distribution. Doubling the graph size is found to increase the square route of the greedy distance by circa 0.41 when links are selected using our algorithm, compared to an increase of about 0.51 when Kleinberg's model is used. (In fact, in with Kleinberg's model we can use (5) to calculate numerically exact values for τ, allowing us to confirm this figure.)

In Figure 4 the marginal distribution of shortcut lengths is plotted. It is roughly harmonic in shape, except that it creates less links of length close to the size of the graph. This may be part of the reason why it is able to outperform Kleinberg's model: while Kleinberg's model is asymptotically correct, this algorithm takes into account finite size effects. (This reasoning is similar to that of the authors of [5]. Like them, we have no strong analytic arguments for why this should be the case, which makes it a tenuous argument at best.)

The algorithm has also been simulated to good effect using base graphs of higher dimensions. Figure 5 shows the mean greedy distance for two dimensional grids of increasing size. Here also, the algorithm creates configurations that seem to display square logarithmic growth, and which perform considerably better than explicit selection according to Kleinberg's model.

8 Conclusion

The study of navigable graphs is still in its infancy, but many interesting results have already been found, and the practical relevance to such fields as computer networks is beyond doubt. In this paper we have presented a different way of looking at the dynamics that cause graphs to be navigable, and we have presented an algorithm which may explain how navigable graphs arise naturally. The algorithm's simplicity also means that it can be useful in practice for generating graphs that can easily be searched, and important property for many structures on the Internet.

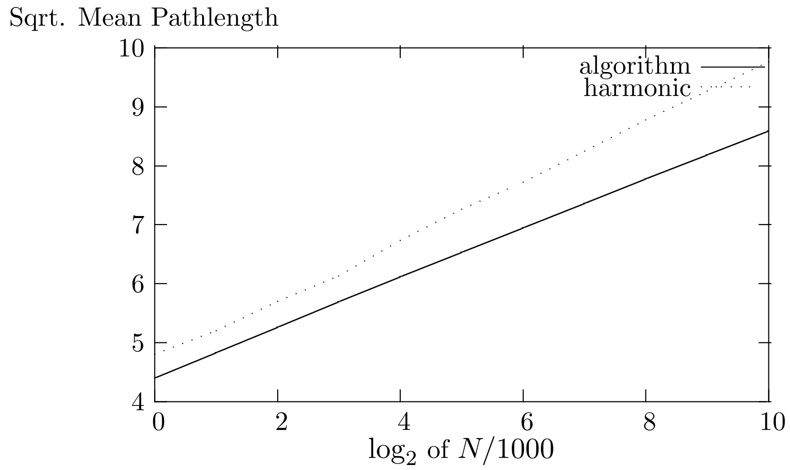

Figure 4: Data from the tables in Section 7.1 on the expected greedy walk length using our selection algorithm, compared to selection according the harmonic distribution.

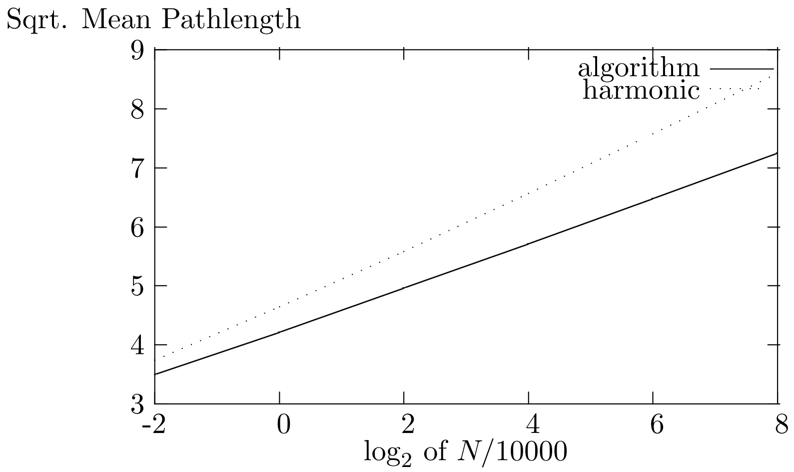

Figure 5: The expected greedy walk time of the selection algorithm, compared to selection according to harmonic distances, in a two dimensional base grid.

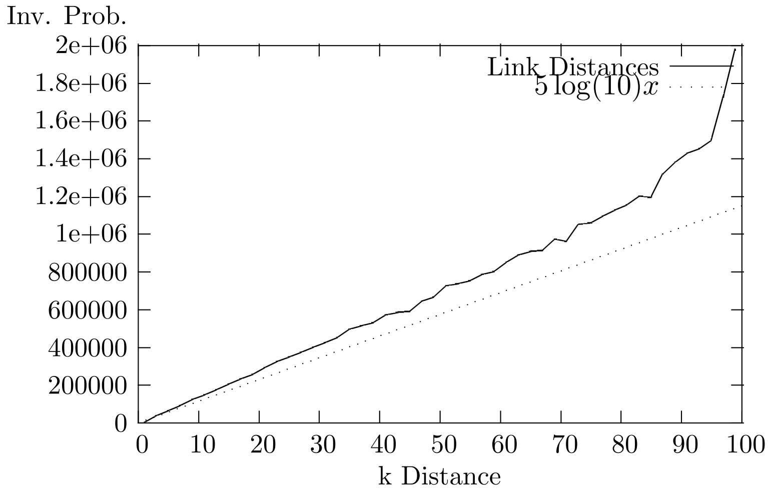

Figure 6: The inverse of distribution of shortcut distances, with $n = 100000$, $p = 0.10$. The straight line is the inverse of the harmonic distribution.

While many questions about these graphs in general, and our results in particular, remain unanswered, the prospects of going further with this work seem good. We are hopeful that these ideas will be fruitful, leading to further analysis of searching and routing in graphs of all kinds.

References

[1] Navigation in a small world. *Nature*, page 845, 2000.

[2] L. Barriere, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient routing in networks with long range contacts. In *Proceedings of the 15th International Symposium on Distributed Computing, DISC'01*, 2001.

[3] I. Clarke, T. Hong, S. Miller, O. Sandberg, and B. Wiley. Protecting free expression online with Freenet. *IEEE Internet Computing*, 6:40–49, 2002.

[4] I. Clarke, T. Hong, O. Sandberg, and B. Wiley. Freenet: A distributed anonymous information storage and retrieval system. In *Proc. of the ICSI Workshop on Design Issues in Anonymity and Unobservability*, pages 311–320, 2000.

[5] A. Clauset and C. Moore. How do networks become navigable? *Preprint*, 2003.

[6] M. Draief and A. Ganesh. Efficient routing in Poisson small-world networks. To appear in Journal of Applied Probability.

[7] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Could any graph be turned into a small world? *Theoretical Computer Science*, 355:96 – 103, 2006.

[8] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Towards small world emergence. In *Proceedings of 18th ACM Symposium on Parallelism in Algorithms and Architectures*, 2006.

[9] D. Eppstein and J. Y. Wang. A steady state model for graph power laws. In *2nd International Workshop on Web Dynamics*, May 2002.

[10] M. Franceschetti and R. Meester. Navigation in small world networks a scale-free continuum model. Technical Report UCB/ERL M03/33, EECS Department, University of California, Berkeley, 2003.

[11] J. Kleinberg. The small-world phenomenon: an algorithmic perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 2000.

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

[13] J. Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of the International Congress of Mathematicians (ICM)*, 2006.

[14] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geograph routing in social networks. In *Proceedings of the National Academy of Science*, volume 102, pages 11623–11628, 2005.

[15] G. Singh Manku. Know thy neighbor's neighbor: the power of lookahead in randomized p2p networks. In *Proceedings of the 36th ACM Symposium on Theory of Computing (STOC)*, 2004.

[16] C. Martel and V. Nguyen. Analyzing kleinberg's (and other) small-world models. In *PODC '04: Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, pages 179–188, New York, NY, USA, 2004. ACM Press.

[17] S. Milgram. The small world problem. *Psychology Today*, 1:61, 1961.

[18] M. Newman. Models of the small world: A review. *Journal of Statistical Physics*, 101:819–841, 2000.

[19] M. Newman and D. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263:341–346, 1999.

[20] J. Novak P. Raghavan A. Tomkins R. Kumar, D. Liben-Nowell. Theoretical analysis of geographic routing in social networks. Technical Report MIT-CSAIL-TR-2005-040, Computer Science and Artificial Intelligence Labratory, MIT, 2005.

[21] D.J. Watts, P. Dodds, and M. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

[22] D.J. Watts and S. Strogatz. Collective dynamics of small world networks. *Nature*, 393:440, 1998.

[23] H. Zhang, A. Goel, and R. Govindan. Using the small-world model to improve Freenet performance. In *Proc. IEEE Infocom*, 2002.