Analyzing Kleinberg's (and other) Small-world Models*

Chip Martel
Computer Science Dept.
University of California
Davis, CA 95616
martel@cs.ucdavis.edu

Van Nguyen
Computer Science Dept.
University of California
Davis, CA 95616
nguyenvk@cs.ucdavis.edu

ABSTRACT

We analyze the properties of Small-World networks, where links are much more likely to connect "neighbor nodes" than distant nodes. In particular, our analysis provides new results for Kleinberg's Small-World model and its extensions. Kleinberg adds a number of directed long-range random links to an $n \times n$ lattice network (vertices as nodes of a grid, undirected edges between any two adjacent nodes). Links have a non-uniform distribution that favors arcs to close nodes over more distant ones. He shows that the following phenomenon occurs: between any two nodes a path with expected length $O(\log^2 n)$ can be found using a simple greedy algorithm which has no global knowledge of long-range links.

We show that Kleinberg's analysis is tight: his algorithm achieves $\theta(\log^2 n)$ delivery time. Moreover, we show that the expected diameter of the graph is $\theta(\log n)$, a $\log n$ factor smaller. We also extend our results to the general k-dimensional model. Our diameter results extend traditional work on the diameter of random graphs which largely focuses on uniformly distributed arcs. Using a little additional knowledge of the graph, we show that we can find shorter paths: with expected length $O(\log^{3/2} n)$ in the basic 2-dimensional model and $O(\log^{1+1/k} n)$ in the general k-dimensional model (for $k \geq 1$).

Finally, we suggest a general approach to analyzing a broader class of random graphs with non-uniform edge probabilities. Thus we show expected $\theta(\log n)$ diameter results for higher dimensional grids, as well as settings with less uniform base structures: where links can be missing, where the probability can vary at different nodes, or where grid-related factors (e.g. the use of lattice distance) has a weaker role or is dismissed, and constraints (such as the uniformness of degree distribution) are relaxed.

*This work was supported by NSF grant CCR-85961

PODC'04, July 25–28, 2004, St. Johns, Newfoundland, Canada.

Categories and Subject Descriptors

F.2 [**Analysis of Algorithms and Problem Complexity**]: Miscellaneous; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph Algorithms, Network Problems*; C.2.2 [**Network Protocols**]: *Routing Protocols*

General Terms

Algorithms, Performance, Theory

Keywords

Small-world network, random graphs, routing, diameter

1. INTRODUCTION

Small-world networks (SWN) have been an active and common topic in many disciplines, including the social and natural sciences. These networks possess a striking property, the so called small-world phenomenon, also often spoken of as "six degrees of separation" (between any two people in the United States). Milgram discovered this in his pioneering work in the 1960's [21], and the more recent work by Dodds et al. suggests its still true [10]. Since many real networks exhibit small-world properties, a number of network models have been proposed as a framework to study this phenomenon. Recently, Kleinberg [14], building on the work of Watts and Strogatz [22], proposed a family of SWNs to study another compelling aspect of Milgram's original findings: a greedy algorithm using only local information can construct short paths.

As Kleinberg has commented, it is striking that short paths not only exist but can be found with limited knowledge of the global network. Algorithmic results in this area improve our understanding of many practical network structures and bring in potential applications related to routing problems in the Internet and peer-to-peer networks. Thus, finding models with this feature of SWNs, with an emphasis on algorithmic aspects, is well motivated. Kleinberg developed an interesting model of this, and it has generated considerable followup work.

Kleinberg's basic model uses a two-dimensional grid as a base with long-range random links added between any two nodes u and v with a probability proportional to $d^{-2}(u, v)$, the inverse square of the lattice distance between u and v. In the basic model, from each node there is an undirected *local* link to each of its four grid neighbors and one directed *long-range* random link. In this setting Kleinberg shows that a simple greedy algorithm using only local information

finds routes between any source and destination using only $O(\log^2 n)$ expected links [14].

Kleinberg leaves two important issues open in the analysis of routing in his model. We complete the analysis in this paper and then extend our techniques to a broader range of settings. First we show that the $O(\log^2 n)$ expected time analysis is tight (thus except for pairs which are quite close, Kleinberg's algorithm uses expected $\Theta(\log^2 n)$ links).

Our second main result shows that the expected diameter of this graph is $\Theta(\log n)$. This extends traditional work on the diameter of random graphs which largely focuses on uniformly distributed arcs [7]. This diameter result shows that an algorithm with global knowledge of the random links can improve on Kleinberg's decentralized algorithm by a log factor. We then give an intermediate algorithm which uses some additional, fairly local, information to improve the expected time to $O(\log^{3/2} n)$ in the basic 2-dimensional model and $O(\log^{1+1/k} n)$ in the k-dimensional model (for $k \geq 1$).

In addition, we develop techniques for analyzing random graphs with non-uniform arc distributions. We are able to characterize properties which lead to small diameter graphs, and use these to prove $O(\log n)$ expected bounds on the diameter of several settings: k-dimensional versions of Kleinberg's model; settings where arc probabilities are proportional to $d^{-r}(u, v)$ for $0 \leq r < 2$; also for some non-grid settings and where nodes have differing degree or arc distributions.

Since real-world networks often are better modeled by random networks with non-uniform arc distributions, our analysis results may help create and analyze more accurate network models.

The structure of this paper. Section 3 presents definitions and supporting facts on Kleinberg's Small-world setting. Section 4 discusses our bound on Kleinberg's delivery time and introduces our alternative algorithm. Section 5 describes our diameter results for Kleinberg's model and more general settings. In Section 6 we summarize our results and suggest additional open problems.

2. RELATED WORK

There has been considerable work on the small-world phenomenon. See [15] for early surveys and [14] for a more recent account on modeling small-world networks. Before Kleinberg's model, Watts and Strogatz [22] proposed a refined model by randomly rewiring the edges of a ring lattice each with a probability parameter p. Watts and Strogatz observed that for small p the model reflects many practical small-world networks with small typical path length and non-negligible clustering coefficient.

Applications have been found using Kleinberg's small-world model or the ideas behind it, such as decentralized search protocols in peer-to-peer systems [20, 25], and gossip protocols for spreading information in a communication network [11]. See also [13] for a generalization that encompasses both lattice-based and tree-based ("taxonomic" or "hierarchical") small-world networks, and [1] for a practical approach to measure the diameter of the World Wide Web using a simulation model based on the power law distribution.

The diameter of random graphs is a classic problem [5, 6, 7, 8] but most results use uniformly distributed arcs. Bollobas and Chung [6], study a graph model very similar to Watts and Strogatz in [22] with the nodes of a cycle (or a "ring") randomly matched to form additional long-range links. The closest diameter work with non-uniform arc probabilities is by Benjamini and Berger who study the diameter of 1-D long-range percolation graphs [4] and Coppersmith et al. who extend this to k-D grids [9]. As in Kleinberg's model, a grid with (undirected) local links is augmented by undirected long-range random links whose probability is inversely related to their distance. Both papers prove diameter results which show how the diameter changes as the arc probability parameters change. Note that in contrast to Kleinberg's model, the added links are undirected, and the out-degree of a node is not fixed. Thus the analysis techniques here are rather different than those to analyze Kleinberg's and related models. Our $\Theta(\log n)$ diameter bound improves on the best prior result of $O(\log n \log\log^2 n)$ due to Lebhar and Schabanel's work on finding short paths [16].

There have also been several recent papers which analyze greedy routing in other small-world like networks. Barriere et al. [3] give matching upper and lower bounds for greedy routing in a ring network augmented by random links. They also give a nice general framework for analyzing greedy routing, and prove a polylog upper bound on greedy routing for a broad class of what they call "Long-range contact" graphs. Our results have some similarities, though we focus on diameter issues and use mostly different techniques. Aspnes et al. [2] and Manku and Manku et al. [18, 17] look at using small-world type networks for routing in peer-to-peer systems. Manku et al. compares deterministic and randomized structures, as well as considering the effect of one-lookahead. Aspnes et al. analyze expected performance with additional links and look at the effect of edge failures. They produce some useful results for doing analysis though their focus is on the one dimensional setting. Recently, Fraigniaud et al. [12] study greedy routing and also show that with some extra knowledge about the graph it is possible to route using $O(\log^{1+1/k} n)$ expected steps in the general k-dimensional model (for $k \geq 1$). However, they achieve this with an oblivious greedy algorithm, while ours requires keeping additional state information. In addition, they show this path length is the best possible for a range of greedy routing algorithms (even ones which have complete global knowledge). They give a general lower-bound on greedy routing which also shows that Kleinberg's $O(\log^2 n)$ analysis is tight for greedy routing in the basic model.

3. BASIC FACTS AND DEFINITIONS

We now present our notation and basic facts on Kleinberg's small world setting. Let V denote the set of all nodes, the size of which is n^2. We use $\mathcal{K}(n, p, q)$ to refer to the class of all random graphs based on Kleinberg's model: an $n \times n$ grid, undirected local links (from a node) to all nodes within distance p, and q long range directed links/node such that the probability of a link from u to v is proportional to $d^{-2}(u, v)$. We also use $\mathcal{K}^*(n, p, q)$ for a similar class, defined with respect to the lattice distance with wrap-around: $d(u, v) = min\{|k-i|, n-|k-i|\}+min\{|l-j|, n-|l-j|\}$. For simplicity, we usually show our main results for $\mathcal{K}^*(n, p, q)$ but most of our work can be easily extended to $\mathcal{K}(n, p, q)$.

Let $B_l(u)$ denote the set of all the nodes within lattice distance l from u; we call this a ball of radius l and center u (actually shaped as a diamond in a two dimensional grid). Define $b_l(u)$ as the number of nodes at distance l from u, i.e. nodes on the 'surface' of $B_l(u)$.

First, we consider some basic facts for the simple models with $p = q = 1$, which, we denote by $\mathcal{K}^*$ and $\mathcal{K}$. For any two nodes u and v, let $p(u, v)$ denote the probability that there is a random link from u to v. Since $p(u, v) \sim d^{-2}(u, v)$, we have $p(u, v) = d^{-2}(u, v)/c_u$, where c_u is the inverse normalized coefficient, a constant depending on n and the position of u:

$$c_u = \sum_{\forall v \neq u} d^{-2}(u, v) = \sum_{j=1}^{2n-2} b_j(u) j^{-2}.$$

Note that we always have $p(u, v) = p(v, u)$ if the graph is from $\mathcal{K}^*$, and so c_u is the same for all u; call this value c^*. It is easy to see that $b_j(u) = \theta(j)$, so c_u can be approximated by a harmonic sum. The following fact is trivially implied from Kleinberg's analysis in [14].

FACT 1. *For graphs from $\mathcal{K}^*$ or $\mathcal{K}$, the inverse normalized coefficient $c_u = \theta(\log n)$ and for any two distinct nodes u and v, $p(u, v) = \Omega((n^2 \log n)^{-1})$. Especially, $c^* = 4 \ln n + O(1)$*[1].

3.1 Links into or out of a ball

Our analysis of the expected diameter (section 5) builds on a standard approach. We consider the expected length of a path from an arbitrary start s to a destination t; for simplicity, we start with $\mathcal{K}^*(n, 1, 4)$ for a basic framework then extend our analysis to $\mathcal{K}^*(n, 1, 1)$ and others. For $\mathcal{K}^*(n, 1, 4)$, using only the long-range links consider all nodes at distance one from s, then all new nodes these can reach (so at distance two), and so on. If these sets grow exponentially in size, we will quickly reach a large subset of the nodes in the graph. We prove two important facts which help to analyze the growth rate.

For a given set Q of nodes, we construct a set R, a collection of nodes not in Q but reachable by a random link from Q; we also consider a reverse scenario where R contains nodes with a random link *into* Q. We focus on lower bounding the ratio $|R|/|Q|$, which is supported by these two elementary experiments: to see if a random link from the center of a ball goes out of this ball and to see if there is a random link from outside the ball which goes to its center. Note that we have non-uniform probabilities, so the ratio $|R|/|Q|$ depends on the shape of Q as well as its size; we also have directed links, so the analysis on incoming links differs from outgoing links [2].

FACT 2. *On a graph from $\mathcal{K}^*$, given any positive $\theta < 1$, any integer $1 \leq l \leq n^\theta$, for n large enough: i) the probability that a random link from a given node u goes to a node outside of $B_l(u)$ is greater than $1 - \theta - o(1)$; ii) the probability that there is a random link to u from a node outside of $B_l(u)$ is greater than $1 - e^{\theta + o(1) - 1}$* (i.e. almost $1 - e^{\theta - 1}$).

We are most interested in values of θ just above 0.5 since then $|B_l(u)| = \Omega(n \log n)$. Thus, a link going out has probability $\lessapprox 0.5$ ($1 - \theta$) and in is $\lessapprox 0.39$ ($1 - e^{\theta - 1}$).

PROOF. For i) let E be the event that u has a random link going to a node $v \notin B_l(u)$. We have

$$\Pr[\overline{E}] = \sum_{v \in B_l(u)} d^{-2}(u, v)/c^* \leq \sum_{j=1}^{l} b_j(u)(j^{-2})/c^*.$$

Since $b_j(u) \leq 4j$ we have

$$\sum_{j=1}^{l} b_j(u)(j^{-2}) \leq 4 \sum_{j=1}^{l} j^{-1} \leq 4 \ln(3l) \leq 4 \ln(3n^\theta).$$

With $c^* = 4 \ln n + O(1)$ we have

$$\Pr[\overline{E}] \leq \frac{4 \ln(3n^\theta)}{4 \ln n + O(1)} \leq \frac{\theta \ln n + \ln 3}{\ln n + O(1)} = \theta + o(1)$$

Thus $\Pr[E] \geq 1 - \theta - o(1)$.

For ii) let F be the event that there is a random link coming to u from a node v outside $B_l(u)$; thus,

$$\Pr[\overline{F}] = \prod_{v \notin B_l(u)} (1 - p(v, u)) \leq \prod_{v \notin B_l(u)} e^{-p(v,u)} = e^{-\sum_{v \notin B_l(u)} p(v,u)}$$

Here we use the well known fact $e^x \geq 1 + x$ to obtain $e^{-p(u,v)} \geq 1 - p(u, v)$. Since $p(u, v) = p(v, u)$,

$$\Pr[\overline{F}] \leq e^{-\sum_{v \notin B_l(u)} p(u,v)} = e^{-\Pr[E]} \leq e^{\theta + o(1) - 1}$$

So $\Pr[F] = 1 - Pr[\overline{F}] \geq 1 - e^{\theta + o(1) - 1}$. □

In a similar manner, we have an equivalent result for $\mathcal{K}$ as follows. See [19] for the proof of this.

FACT 3. *On a graph from $\mathcal{K}$, given any positive $\theta < 1$ and integer $1 \leq l \leq n^\theta$, for n large enough: i) the probability that a random link from a given node u goes to a node outside of $B_l(u)$ is greater than $\frac{1-\theta}{1+3\theta} + o(1)$; ii) the probability that there is a random link to u from a node outside of $B_l(u)$ is greater than $1 - e^{-(1-\theta)/4 + o(1)}$.*

3.2 Extended models

As a natural generalization, we can extend all these concepts and basic facts to the classes of graphs based on a k-dimensional grid for $k \geq 1$ by updating the distribution rule of the random links: $p(u, v) \sim d^{-k}(u, v)$ instead. Let $\mathcal{K}^*(k, n, p, q)$ and $\mathcal{K}(k, n, p, q)$ denote such classes of graphs based on this k-dimensional grid (with size n in each dimension) with and without wrap-around, respectively. In $\mathcal{K}(k, n, p, q)$, for a node u close to the edge of the grid, a ball centered at u may not be a 'full' ball but it is *easy to extend all our results with balls to cover such a case.*

It is relatively easy to prove that $b(j) = \theta(j^{k-1})$ [3]. So, $b_j(u) j^{-k} = \theta(j^{-1})$, which means that we can still use harmonic sums to bound the likes of $\sum b_j(u) j^{-k}$ as before. Thus, we still have $c_u = \theta(\log n)$. More importantly, we can generalize the facts on "links into or out of a ball" as follows (for simplicity, we do not claim more exact bounds).

FACT 4. *On a graph from $\mathcal{K}(k, n, p, 1)$ or $\mathcal{K}^*(k, n, p, 1)$, for any given [4] positive $\theta < 0.6$, there exist positive constants ξ_1 and ξ_2 such that for n large enough and $l \leq n^\theta$:*
i) the probability that there is a random link from a node u to a node outside of $B_l(u)$ is $\geq \xi_1$;
ii) the probability that there is a random link (with a given label) to u from a node outside of $B_l(u)$ is $\geq \xi_2$.

[1] Since in $\mathcal{K}^*$, $b_j(u) = 4j$ if $j \leq \lceil n/2 \rceil$ and 0 otherwise.
[2] In $\mathcal{K}^*|\mathcal{K}$, each node has exactly one outgoing random link, but the number of incoming links varies.
[3] Based on counting the number of ways to choose k positive numbers such that they sum to a given positive number (j).
[4] The fact holds for any $0 < \theta < 1$ but we use $\theta < .6$ since we only need θ about 0.5 in our later analysis).

It is also easy to extend this fact for arbitrary $q \geq 1$. Each node now has q independent random links where each of them can be labeled from 1 to q. Fact 3 is then, concerned with a link specified by a given label and an arbitrary center node u.

3.3 Links to a spherical surface

We now study another probabilistic experiment in the general models $\mathcal{K}(k,n,p,q)$ and $\mathcal{K}^*(k,n,p,q)$, namely, if a random link from a given node u goes to the surface of a given ball B, where u is outside of B. We denote the probability of this by $P(u,S_B)$, where S_B is the set of nodes on the surface of ball B. Also define distance $d(u,B)$ between node u and ball B as the minimum lattice distance from u to a node on B's surface. Our lower bound in section 4 will be based on this experiment.

FACT 5. *For any $k \geq 1$, there exists a constant $\hat{c}$ such that for any ball $B = B_l(v)$ and node u outside B on a graph from $\mathcal{K}^*|\mathcal{K}(k,n,p,q)$: $P(u,S_B) \leq \frac{\hat{c}}{m \log n}$, where $m = d(u,B)$. Also, if $k \geq 2$ and $l \leq m$ then $P(u,S_B) \leq \frac{\hat{c}l}{m^2 \log n}$.*

Note that for the first (main) part of the fact, the bound $(\frac{\hat{c}}{m \log n})$ depends on the distance from u to the ball only, and thus, is independent of the size of the ball. Intuitively, if we think of this probability as a measure of 'attractive force' (to u), the force generated by S_B is not stronger than the joint force generated by m nodes at about distance m from u. That is, a small fraction of nodes in S_B, which are at the 'pole' closest to u, generates a dominant term for $P(u,S_B)$.

A rigorous proof for the first part is complicated, so we leave the proof to the appendix. Meanwhile, the second part is easy: $P(u,S_B) = \sum_{w \in S_B} p(u,w) = O(\frac{l}{m^2 \log n})$ with $p(u,w) \leq \frac{1}{m^k log n}$ and $|S_B| = b(l) = O(l^{k-1}) \leq l \times O(m^{k-2})$.

4. DECENTRALIZED ROUTING

Kleinberg proved that the expected number of links used to route from a source s to a destination t (*Kleinberg's delivery time*) by the greedy algorithm is $O(\log^2 n)$ for the two-dimensional case [14][5]. We will prove that indeed the greedy algorithm takes expected $\Omega(\log^2 n)$ *delivery time for the general k-dimensional model* and suggest other variants with better delivery time. Let $\mathcal{GA}$ denote the greedy algorithm. For a node v we call $N(v)$ the next node found by $\mathcal{GA}$ when we are at v (initially, $v = s$) and seeking a path to the destination node t. Define random variable $\delta(v) = d(v,t) - d(N(v),t)$, which characterizes the speed that $\mathcal{GA}$ converges to t. The following lemma, which is actually a re-statement of fact 5, establishes a probabilistic bound for this speed.

LEMMA 6. *For each $k \geq 1$ (the dimension), there exists a constant $\hat{c}$ such that, for a graph from $\mathcal{K}(k,n,p,q)$ or $\mathcal{K}^*(k,n,p,q)$, any two nodes v and t, and any integer $1 < m < d(v,t)$, $\Pr[\delta(v) = m] \leq \frac{\hat{c}}{mlogn} \times min\{1, \frac{d(v,t)-m}{m}\}$.*

The following theorem shows that for the majority of $s-t$ pairs, *Kleinberg's delivery time* can not be $o(log^2 n)$, therefore we have the tight bound of $\theta(log^2 n)$ for the greedy algorithm.

[5]It is straightforward to generalize his proof for the general k-dimensional model

THEOREM 7. *For any constant $c_1 > 0$, there exists a constant $c_2 > 0$ such that, for any two nodes s and t on a graph from $\mathcal{K}^*|\mathcal{K}(k,n,1,1)$ with $d(s,t) > c_1 n$,* Kleinberg's delivery time *w. r. t. endpoints s and t is greater than $c_2 log^2 n$ with probability at least 0.5, when n is large enough* [6].

PROOF. Consider a series of random trials, where, given a node v, we find $N(v)$ the node for the next trial. Define random variable $X_v = \frac{d(v,t)}{d(N(v),t)}$, which reflects the ratio between the distance to t before and after such a trial. Let $\{v_1 = s, v_2, \ldots, v_k\}$ be the nodes on a run of $\mathcal{GA}$ after k such steps (thus $v_{i+1} = N(v_i)$). $\prod_{i=1}^{k-1} X_{v_i}$ reflects the ratio between the distance at the initial and at the current state. To make this ratio always well-defined, if $N(v) = t$ we define $X_v = d(v,t)$ and if $v = t$ we define $X_v = 1$.

It is clear that $1 \leq \prod_{i=1}^{k-1} X_{v_i} \leq d(s,t)$ and we need to analyze the expected k to have $\prod_{i=1}^{k-1} X_{v_i} = d(s,t)$ (indicating that $d(v_k,t) \leq 1$, so $\mathcal{GA}$ is about to finish), or $\sum Z_{v_i} = ln(d(s,t))$ where $Z_v = lnX_v$. From lemma 6, we can prove that $E[Z_v] < c/\log n$ for some constant c (proved below), if $d(v,t) \geq \ln n$. To get rid of the condition, $d(v_i,t) > \ln n$, we modify $\mathcal{GA}$ a bit, so whenever we get to a node v within a distance $\ln n$ from t (i.e. $d(v,t) \leq \ln n$), set $N(v) = t$ and redefine $Z_v = 0$ (this will not weaken our proof since when we get there we can simply walk to t by at most $\lceil lnn \rceil$ local links). Thus we always have $E[Z_{v_i}] \leq c/lnn, \forall i = 1..k$.

Consider $Z^k = Z_{v_1} + Z_{v_2} + \ldots + Z_{v_k}$. If we reach t with less than k steps, or say, $v_j = t$ for some $j < k$, then as mentioned before, we have $Z_{v_j} = Z_{v_{j+1}} = \ldots = Z_{v_{k-1}} = 0$. We now show that

$$Pr[Z^k < M] \geq 0.5 \tag{1}$$

where $k = \lfloor \frac{ln^2 n}{4c} \rfloor$ and $M = ln(\frac{d(s,t)}{\ln n}) \geq \ln n + \ln c_1 - \ln(\ln n)$.

Since $E[Z_{v_i}] < c/\ln n$, we have

$$E[Z^k] \leq k \times \frac{c}{\ln n} \leq \frac{lnn}{4} < 0.5M$$

for n large enough. Thus, $Pr[Z^k \geq M] < 0.5$ since otherwise $E[Z^k] \geq 0.5M$. Hence, $Pr[Z^k < M] \geq 0.5$. From (1), choose $c_2 = \lfloor \frac{1}{4c} \rfloor$ then at least half of the time, $\mathcal{GA}$ can not finish with less than $c_2 log^2 n$ steps; the theorem follows.

Now, we show that there exists a constant c such that $E[Z_v] \leq c/\ln n$ if $d = d(v,t) \geq \ln n$. Define the function $f(i) = \ln(\frac{d}{d-i})$; note that $Z_v = \ln(\frac{d}{d-\delta(v)})$. We need to show that $\sum_{i=1}^{d} f(i) \Pr[\delta(v) = i] \leq c/\ln n$ for some constant c. First we have

$$\sum_{i=1}^{d/2} f(i) \Pr[\delta(v) = i] \leq \frac{1}{d-1} + \frac{c'}{\ln n} \sum_{i=d/2}^{d-1} 1/i \leq \frac{c'+1}{\ln n}$$

for some constant c'. Note that we have $\Pr[\delta(v) = i] \leq \frac{c'}{i \ln n}$ from lemma 6 and $f(i) \leq \frac{i}{(d-i)}$ from the common fact that $ln(1+x) < x, \forall x > -1$. Similarly, we have $\sum_{i=d/2}^{d} f(i) \Pr[\delta(v) = i] \leq \frac{c''}{\ln n}$ for some constant c'', using $\Pr[\delta(v) = i] \leq \frac{c''(d-i)}{i^2 \ln n}$ from lemma 6. Now, choose $c = c' + c'' + 1$. □

[6]Note that we can actually make the probability above an arbitrarily high constant (< 1), but we use 0.5 for simplicity.

4.1 Algorithms for improved delivery time

We now consider variants of Kleinberg's greedy algorithm which use additional knowledge of the graph to improve expected path length. Our basic algorithm operates on graphs from class $\mathcal{K}^*(n,1,1)$ but it can be easily extended to more general classes. We assume that each node u knows the long-range links of the $\log n$ neighbor nodes closest to u in the grid. Call W_u, the set of these neighbors and their long range contacts, the view at u. During a *basic step* of routing, if u is the current message router, we go next to v, the node in W_u closest to t (we may need to follow several local links and, possibly one final random link to v). Once we reach v, it becomes the current node.

In Kleinberg's algorithm W_u includes only the nodes incident to links from u (at most 5 in a two-dimensional grid) and a routing decision is made at each node on the path. In our algorithm, once we make a decision for a basic step (at the initial message router) we follow the sub-path to the next node where a decision is made. When we are at an intermediate node in a basic step, we follow local links (up, down, left or right) except for the final link which may be a random long-range link). We can describe this sub-path using only $O(\log\log n)$ extra bits to specify the last node on the sub-path reached by local links only. Alternately, we can use $2\sqrt{\log n}$ bits to describe the sequence of local moves (two bits per local move).

We now adapt Kleinberg's analysis to show that *the expected number of links used in this algorithm is* $O(\log^{3/2} n)$. We say that node u is in *phase* i if $2^i < d(u,t) \leq 2^{i+1}$. Kleinberg proved that if u is in phase i, and v is $u's$ long range contact, the probability that v is in some phase $j < i$ is proportional to $1/\log n$ (for $i > \log\log n$). Let X_i be a random variable recording the number of basic steps our algorithm executes from nodes in phase i.

LEMMA 8. *For $i > \log\log n$, $E[X_i] < c$ for a constant c.*

PROOF. Each basic step considers at least $\log n/2$ new long-range links (since at most half of the current message holder's set W_u overlaps all previous sets), and by Kleinberg's result has at least a constant probability of finding a link to a node within distance 2^i of t. □

THEOREM 9. *Our improved algorithm visits expected $O(\log^{3/2} n)$ nodes in $\mathcal{K}^*(n,1,1)$.*

PROOF. Using the lemma above, the expected number of basic steps till we get within distance $\log n$ is $O(\log n)$ (constant per phase for $\log n$ phases). Each basic step visits at most $\sqrt{\log n}$ nodes (for a 2-D grid), so the total length is $O(\log^{3/2} n)$. □

More careful work (using tail inequalities) gives a slightly stronger result: a longest path found has length $O(\log^{3/2} n)$ with overwhelming probability. The underlying intuition is to think of doing a chain of basic steps until we have about $\log n$ which advance to a new phase.

If we use the same algorithm on classes $\mathcal{K}^*/\mathcal{K}(n,p,q)$ where $p,q \geq 1$, it is relatively easy to show a better bound of $O(\frac{\log^{3/2} n}{p\sqrt{q}})$. So, the bound is just $O(\log n)$ if for instance, $p = \Omega(\log^{1/2} n)$ and q is a constant, or p is a constant and $q = \Omega(\log n)$. It is also easy to extend our algorithm to $\mathcal{K}^*/\mathcal{K}(k,n,p,q)$. *For $p = 1, q = 1$, we obtain the upper bound $O(\log^{1+1/k} n)$*, which is close to the ideal $O(\log n)$ when k is large.

A related approach is analyzed in [12] who show that the $O(\log^{1+1/k} n)$ bound for k-dimensions can be achieved by an oblivious algorithm. They also show that this bound is tight for a range of greedy algorithms (even with much more information).

5. EXPECTED DIAMETER

5.1 Basic approach

For simplicity we will first consider Kleinberg's basic 2-D grid, but with four long-range links per node. We show that the expected diameter of a graph from $\mathcal{K}^*(n,1,4)$ is $\Theta(\log n)$ and then extend our approach to other and more general classes of graphs.

For a source s and a destination $t \neq s$ chosen arbitrarily from V, we will show that there is an $O(\log n)$ length path from s to t with overwhelming probability. The basic idea is to construct two sets of nodes S and T which are each of size $\Theta(n\log n)$ such that: all the nodes in S can be reached from s by a directed path of length $O(\log n)$ of long-range links; all the nodes in T have a directed path of length $O(\log n)$ to t using long-range links. It is then simple to show that with overwhelming probability, there is a directed arc from a node in S to a node in T, thus there is an $O(\log n)$ length path from s to t. The bulk of the proof is showing that appropriate sets S and T exist with high probability.

Define $\chi(u)$ with $u \in V$ as the set of nodes v such that (u,v) is a long-range link; also $\chi(A) = \bigcup_{u\in A}\chi(u)$, for any $A \subset V$. Our basic idea is to 'grow two trees' with roots from s and t to create S and T. We construct a chain of disjoint subsets $\{S_k\}_{k=0}^{\mu}$ with $S_0 = B_r(s)$ where $r = r_0\sqrt{\log n}$ for some constant r_0 [7]. Let $C_k = \bigcup_{i=0}^{k} S_i$ and define $S_{k+1} = \chi(S_k) - C_k$. Thus, S_{k+1} is built by iteratively applying χ on elements of S_k and taking only 'fresh' nodes, which have not been in any preceding subsets.

Later we construct a similar subset chain $\{T_k\}_{k=0}^{\nu}$ to create the $t-$tree. The probability of success (finding sets S,T which connect) can be made as close to one as needed by choosing r_0 (and thus S_0) sufficiently large. Thus we get an $O(\log n)$ bound on the expected diameter of a graph in $\mathcal{K}^*(n,1,4)$.

A worst-case analysis with a ball experiment.

We will show that the subset chain almost surely grows exponentially in size: there is a constant $\gamma > 1$ such that $\Pr[|S_{k+1}|/|S_k| > \gamma]$ is almost one when $c\log n < |S_k| < \alpha n\log n$ for suitable constants c and α. Our goal is to finally form a set S_μ such that $|S_\mu| \geq \alpha n\log n$ and $\mu = O(\log n)$. For a fixed γ, consider the following *S-construction process*: starting with S_0 we successively create $S_1, S_2, \ldots$ until either we get a set S_μ with $|S_\mu| \geq \alpha n\log n$ (we succeed) or we have some $|S_{k+1}|/|S_k| \leq \gamma$ and we fail. As soon as either case happens we stop the process. Note that if we succeed we do so in $O(\log n)$ $\chi(u)$ steps, so $\mu = O(\log n)$ and S_μ is our desired set S. The following analysis shows that our S-construction succeeds with high probability.

The main concern is how many *fresh* nodes we get from $\chi(S_k)$ to include in S_{k+1}. We fix an order to scan S_k and for each $u \in S_k$, a node $v \in \chi(u)$ is fresh if it has not been included in a subset G, which contains the current tree (the union of the current S_{k+1} and all the preceding S_i, $i \leq k$).

[7] Thus, there are $\theta(\log n)$ nodes in S_0 all reachable from s by $O(\sqrt{\log n})$ local links.

To analyze this we make the following crucial observation. We consider this experiment $\mathcal{E}(u, G)$: we generate a random link [8] from u and *consider if* $u \notin G$; let $X(u, G)$ denote the indicator random variable of this happening. We do this experiment 4 times for each $u \in S_k$ and $|S_{k+1}|$ will be the sum of these $4|S_k|$ random variables $X(u, G)$; note that these variables are not identical since G keeps growing larger. Let $\mathcal{G}_k$ denote the whole process. Note that $\Pr[X(u, G) = 1]$ depends on both the size and *shape* of G. Thus we ask: *if we move the elements in G around, how would we minimize* $\Pr[X(u, G) = 1]$*?* The nature of the inverse second power distribution makes it clear: this can be done by moving elements of G closer to u; in fact, $\Pr[X(u, G) = 1]$ is minimized (for a fixed size for G) *when G is like a ball with u as its center.*

Thus we can lower bound $|S_{k+1}|/|S_k|$ using the following worst-case setting for selecting a link from u. Let $C = \Theta(n \log n)$ be the maximum size set C_j we can ever have in our S-construction process. Let H be a ball with center s and size C (or the next larger size to make a full ball). Since H is chosen such that $|H| \geq |C_{k+1}|$ (i.e. always $\geq |G|$) and has the worst possible shape, $\Pr[X(u, G) = 1] \geq \Pr[X(s, H) = 1]$ for any pair u, G used in our S-construction process. Since by fact 2, $E[X(s, H)] > \beta$ for any fixed $\beta < 0.5$ when n is large enough, we conclude:

FACT 10. *For any fixed* $\beta < 0.5$*,* $E[X(u, G)] > \beta$ *for each* u, G *pair in our S-construction.*

We now show that each new set is successfully constructed with high probability:

LEMMA 11. *For any* $\eta > 0$*, and* $\gamma \in (1, 2)$ *we can choose a constant* $\hat{c}$ *such that, for* $k \geq 0$*, if* $|S_k| \geq \hat{c} \ln n$ *and* $|C_{k+1}| = O(n \log n)$ *then* $\Pr[|S_{k+1}|/|S_k| > \gamma] = 1 - O(n^{-\eta})$ *for n large enough.*

PROOF. Let $m = |S_k|$, so $|S_{k+1}|$ is the sum of $4m$ $X(u, G)$ indicator R.V's. By fact 10, we treat this as the sum of $4m$ independent Bernoulli random variables each with expectation $\geq \beta$. Applying Chernoff's inequality, for $0 < \delta < 1$,

$$\Pr[|S_{k+1}| \leq (4m\beta)(1 - \delta)] \leq e^{-2m\beta\delta^2} \leq n^{-2\hat{c}\beta\delta^2} \leq n^{-\eta}$$

for a proper choice of $\hat{c}$. Thus, $\Pr[|S_{k+1}|/|S_k| > \gamma] = 1 - O(n^{-\eta})$ for $\gamma = 4\beta(1 - \delta)$.

Indeed we can always choose $\delta > 0$ small enough so that $\gamma = 4\beta(1 - \delta)$ can be arbitrarily close to 4β, and thus, arbitrarily close to 2 as $\beta = 0.5 - o(1)$. Once a value from $(1, 2)$ is chosen for γ (and so $0 < \delta < 1$ and $0 < \beta < 0.5$ are chosen accordingly), *given any* $\eta > 0$, we can choose $\hat{c}$ such that when $m \geq \hat{c} \log n$ we have $\Pr[|S_{k+1}|/|S_k| > \gamma] = 1 - O(n^{-\eta})$. □

In fact, by some simple computation, we can see that $\hat{c} > \eta/(1 - \frac{\gamma}{2})^2$ is a sufficient condition to find such $\hat{c}$. So, if we have $\eta = 6$ ([9]), $\gamma = 1.8$ (by $\beta = 0.48$ and $\delta = 0.0625$ for example) then we can choose $\hat{c} = 601$.

Thus, given fixed constants η and γ we can choose $\hat{c}$ as above, so if $|S_0| \geq \hat{c} \log n$, we can expect the cardinality series $\{|S_i|\}$ to grow exponentially before reaching or exceeding the threshold $\alpha n log n$ during the first $g + 1$ terms, where $g = \lceil log_\gamma(\alpha n/\hat{c}) \rceil = \theta(\log n)$. By iterating lemma 11 appropriately (in each step of growing S_{i+1} from S_i when $|S_i| < \alpha n log n$), we can show that almost surely $|S_\mu| \geq \alpha n log n$ for some $\mu \leq g = \theta(\log n)$. Note that S_μ contains only 'fresh' nodes such that we have not yet considered their random links. We now prove that we almost surely get a large s-tree within $O(\log n)$ steps.

LEMMA 12. *For any given node* $s \in V$*, any* $\theta > 0$ *and any* $\alpha > 0$*, by constructing s's subset chain as above, we will obtain subset* S_μ *where any node in* S_μ *can be reached from s by an* $O(\log n)$ *path and* $\Pr[|S_\mu| \geq \alpha n log n] = 1 - O(n^{-\theta})$.

PROOF. Consider a successful S-construction where we have a succession of steps where $|S_{k+1}|/|S_k| > \gamma$ until we reach size $\alpha n \log n$. From lemma 11 we can pick $\eta > \theta$ so that $\Pr[|S_{k+1}|/|S_k| > \gamma] > 1 - O(n^{-\eta})$ for each step.

The probability that the S-construction process succeeds is greater than the probability we get g consecutive successful growth steps with $g = \lceil log_\gamma(\alpha n/\hat{c}) \rceil = \theta(\log n)$ Thus $\Pr[S\text{-construction process succeeds}] \geq \left(1 - c_1 n^{-\eta}\right)^g \geq 1 - g c_1 n^{-\eta}$ for some constant $c_1 > 0$ (using a basic calculus fact: $(1 + x)^n \geq 1 + nx$ for any $x > -1$ and $n \geq 1$). Since $\theta < \eta$ and , $g = O(\log n)$, it is easy to see that $g c_1 n^{-\eta} = O(n^{-\theta})$ and hence, we create the desired final set with probability $\geq 1 - O(n^{-\theta})$. □

Another ball experiment to analyze 't-tree'.

We now consider a tree of nodes with *paths to* t. We now use a function $\hat{\chi}$ which, given an input node u, outputs the nodes with a random link *to* u. As before, we construct a subset chain $\{T_k\}_0^\nu$ by having $T_{k+1} = \hat{\chi}(T_k) - \hat{C}_k$ where $\hat{C}_k = \bigcup_{i=0}^k T_i$. Thus, we include into T_{k+1} all the fresh nodes which have a random link to any node in T_k. We can use much the same approach as before, but with some modifications and additional details. We still use a state-variable G to denote the set of all nodes in the tree we have reached so far: $G = \hat{C}_k$ if we have just finished the first k subsets, otherwise G is the union of $\hat{C}_k$ and the developing T_{k+1}. Let $\overline{G} = V - G$, the set of nodes not in the tree yet. Also, for each random link from each node $u \in V$ we assign a label $i = 1..4$.

We now look closer at process $\hat{\mathcal{G}}_k$, the growing step of T_{k+1} from T_k, which is more complicated than the counterpart in the s-tree. Let $\hat{\mathcal{E}}(u, i, G)$ denote an experiment which has each node $w \in \overline{G}$ look at its random link labeled i, and if this link hits u we add w to G (for simplicity we may also use $\hat{\mathcal{E}}_u$ instead). Process $\hat{\mathcal{G}}_k$ simply repeats the $\hat{\mathcal{E}}(u, i, G)$ experiment [10] for each node $u \in T_k$ and for each $i = 1..4$. As with the $s-$tree, we now consider an artificial experiment which should be 'outperformed' by the real $\hat{\mathcal{G}}_k$ but is easier to analyze. Let H be a ball centered at u with size at least $\hat{C}$, the maximum size we allow a set to get in our T-construction process. Let $\hat{X}(u, H)$ be a random variable which is one if at least one random link from a node in $V - H$ hits u, otherwise it is zero. In our T-construction process, let $\hat{X}(u, i, G)$ be an indicator random variable recording whether some node in $\bar{G}$ has a link with label i to u. We now argue that $Pr[\hat{X}(u, i, G) = 1] > .39$ (the bound on a link into a ball in Fact 2).

[8] Using inverse second power distribution

[9] Later, in our final step, we use this value to prove the two trees, s-tree and t-tree, connect with overwhelming probability

[10] Note that G is changing as we discover new fresh nodes and we can fix the order to scan T_k initially.

We compare the real and artificial processes above. Looking closely at each real experiment $\hat{\mathcal{E}}_u$, observe that it is slightly preconditioned by the results of the previous ones. A random link from a node $w \in \overline{G}$ must not go to a node in $\hat{C}_{k-1}$, because if w 'hits' a node $v \in \hat{C}_{k-1}$, then w would have been included in G earlier during $\hat{\mathcal{E}}_v$. In fact, this precondition makes this random link from w more likely to hit u. Thus in the artificial process, each link is less likely to hit u than in the real process (as before, a ball is the worst shape for G, and we start with $|H|$ at least the maximum size for G). Thus, $Pr[\hat{X}(u,i,G) = 1] > Pr[$a node outside H has a link to u$] > .39$. We can then use an argument analogous to lemma 11: the size of a new set T_{k+1} is lower bounded by a sum of indicator variables, and we lower bound the probability each of these variables is one. Thus we can pick a value of $\hat{\gamma} > 1$ such that for any $\hat{\eta} > 0$, $Pr[|T_{k+1}|/|T_k| > \hat{\gamma}] > 1 - O(n^{-\hat{\eta}})$ (again assuming we start with a large enough initial set T_0 of size $\theta(\log n)$). Once we have this, the same argument as in lemma 12 shows that the entire T-construction process succeeds with high probability.

FACT 13. *For any given node $t \in V$, any $\theta > 0$ and any $\alpha > 0$, if we construct t's subset chains as above, we will obtain a subset T_ν such that t can be reached from each node in T_ν by a path of length $O(\log n)$ and $Pr[|T_\nu| > \alpha n \log n] = 1 - O(n^{-\theta})$.*

The tight bound. Using fact 13, it is not hard to see that there almost surely exists two such subsets S_μ and T_ν with our desired properties. We now show they either intersect or S_μ is one-link separate from T_ν, i.e. there exists $u \in S_\mu$ such that $\chi(u) \cap T_\nu \neq \emptyset$. This means, there almost surely exists an $O(\log n)$ path from s to t. Moreover, by choosing appropriate constants θ, γ and α, we can make this probability arbitrarily close to one and as a result, we have the following theorem.

THEOREM 14. *The expected diameter of the graphs in $\mathcal{K}^*(n,1,4)$ is $\theta(\log n)$*

PROOF. We will show that for any pair of nodes s,t there is an $s-t$ path of length $O(\log n)$ with probability at least $1 - O(n^{-6})$, since there are fewer than n^4 pairs of vertices in the grid and each has a path of length at most $2n$, this shows the expected diameter is $O(\log n)$. Since the graph is constant degree it must have diameter $\Omega(\log n)$ so the result follows. We now show that the two trees almost surely both exist and are within one-link of each other.

Suppose that we construct set S_μ and then construct set T_ν. If at any time in the t-tree construction process we include a node in any of our S_i sets we have the desired short path and are done. To analyze the probability both construction processes succeed, note that it is not straightforward to use the prior two lemmas with $\theta = 6$ and conclude this probability is $\geq (1 - O(n^{-6}))^2$, since when we are constructing T_ν, we have already been conditioned by the existence of s's subset chain. To fix this, let the state-variable subset G, which includes all the nodes already in the t-tree, be larger, containing all the nodes in the completed $s-$tree and the developing $t-$tree. Since the number of nodes in the $s-$tree is $O(nlogn)$, we still have $|G| = O(nlogn)$, so all the results for t's subset chain still apply. Thus $Pr[$both succeed$] \geq (1 - O(n^{-6}))^2$, which is also $1 - O(n^{-6})$ again using $(1+x)^a \geq 1 + ax$ when $x > -1, a \geq 1$.

Given an arbitrary node $u \in S_\mu$, let p denote the probability that u 'misses' T_ν, i.e. none of the 4 random links from u goes to any node in T_ν. From fact 1, the probability of a random link from u to any other node is at least $\epsilon = C_u(2n)^{-2} = \Omega(n^{-2}log^{-1}n)$, therefore $p \leq (1 - \epsilon|T_\nu|)^4$, but $|T_\nu| \geq \alpha nlogn$, so there exists a constant c_1 such that $p \leq (1 - \alpha c_1 n^{-1})^4 \leq e^{-4\alpha c_1 n^{-1}}$. Now the probability that all the nodes in S_μ 'miss' T_ν is at most

$$(e^{-4\alpha c_1 n^{-1}})^{\alpha nlogn} \leq e^{-4\alpha^2 c_1 logn} = n^{-4\alpha^2 c}$$

where $c = c_1 loge$. Thus, if we let E be the event that the two trees intersect or are separated by one link, $\Pr[E] \geq 1 - n^{-4\alpha^2 c}$ and when we choose α large enough:

$$\Pr[E] \geq 1 - O(n^{-6})$$

□

With a bit more care, we also show that the expected diameter is smaller than $3\log n$ (see [19] for details). Based on the approach we have used above to analyze the expected diameter of graphs with 4 random links out from each node, we now study variations which also have expected diameter $\theta(\log n)$.

The diameter of $\mathcal{K}^*(n,1,1)$. We now show that we only need each node to have at least one out-going random link. The key idea is to modify the subset chain construction so that for a chosen node u, $\chi(u)$ still has at least 4 outputs. These "super nodes" are formed by combining adjacent nodes (and their random links).

COROLLARY 15. *The expected diameter of the graphs in $\mathcal{K}^*(n,1,1)$is $\theta(\log n)$*

PROOF SKETCH. Given a graph $G \in \mathcal{K}^*(n,1,1)$ we can partition it into 2 by 2 blocks and contract each node into a supernode to obtain G' (so there is an edge (i,j) in G' iff there is an edge (u,v) in G where u was contracted to i and v to j where grid edges map to grid edges and long-range edges to long-range edges). Note that G' is almost a graph from problem $\mathcal{K}^*(n/2,1,4)$ but the probabilities for long-range edges are slightly different (due to rounding effects). However, it is easy to see that facts 1 and 2 still apply to this setting, so the diameter proof applies to G' and the diameter of G differs from G' by a constant factor. □

5.2 Extensions

The work presented in subsection 5.1 can be used as a framework to analyze the diameter problem on many other lattice-based settings, where additional random links are independently added from each node. Grouping adjacent nodes to form 'super nodes' (with enough out-going random links) has added some flexibility to our approach. For constructing subset chains, when we need to make a node u a super node, we can collect the random links of many of u's neighbors and make a virtual re-assignment of these links to u, i.e. $\chi(u)$ contains the nodes incident to these links. In other words, we only need to maintain the principle of always having enough out-going random links from any small neighborhood with a determined size in order to form a super node. In fact, *we can relax many conditions in Kleinberg's original model yet we can still analyze the new graphs using our approach.* Hence, we can include more practical graphs: distribution of links are less uniform, some nodes

may have many long-range contacts some may have none, and local links may be broken or missing.

The issue of missing local links may affect the view of a grid, however, for now we still assume that the locations of all the vertices form a geographic grid; thus we can still use lattice distance to determine link probabilities. We now suggest the following conditions for a lattice-based setting, wherein our current approach could be applicable though, these conditions alone will not always be sufficient. To adapt this approach to such a new setting, the key is to suitably formulate underlying properties, analogies of facts 2 and others, with respect to the new scenario.

I. The setting is rich enough in local links so that we can construct starting subsets with size $\Omega(\log n)$).

II. There exists a mechanism for producing super nodes so that we can get enough long-range random links per super node. We suggest the following proposition, called "*Sufficiency of random links everywhere*", but other similar ones can be used instead: *for any constant $C > 1$, there exists a constant L such that, for n large enough, for any node u, there almost surely exists a set of nodes with at least C outgoing random links and reachable from u by no more than L grid links* [11].

III. An analogue to fact 2: there exist positive constants ξ_1 and ξ_2 such that for any positive $\theta < 0.6$, for n large enough, for any node u with an out-going random link, the probability that this random link goes to a node outside of $B_l(u)$, where $l = n^\theta$, is greater than ξ_1. Also, the probability that there is a random link to u from a node outside of $B_l(u)$ is greater than ξ_2.

These conditions can be stated either as facts to be proved in a new setting or as assumptions. In order to have the two trees grow exponentially before they connect (after $O(\log n)$ steps of growing), we need to find proper C (and hence, L) for *II* such that $C\xi_1 > 1$ and $C\xi_2 > 1$ when $\theta > 0.5$. We now present a few special cases and illustrate the use of these conditions. A more general approach will be mentioned next.

The diameter of $\mathcal{K}(n, p, q)$ where $p, q \geq 1$. It is easy to see that condition *I* and *II* are automatically met since the local links are 'full' and condition *III* is met using fact 3 ('links out of a ball' for $\mathcal{K}$, the non-wraparound grid model). Similarly as in section 5.1 we use fact 3 with $\theta = 0.5 + o(1)$ and hence obtain $\xi_1 = 0.2 + o(1)$ and $\xi_2 = 1 - e^{0.125+0(1)}$ (roughly, $0.117 + o(1)$). Thus, we need to specify constant C so that $C\xi_1 > 1$ and $C\xi_2 > 1$, which help to establish the exponential growth of the two trees. Therefore any $C \geq 9$ will work and if we choose $L = 2$ we can take $C = 9..13$ (since $|B_2(.)| = 1 + 4 + 8 = 13$). Thus, when we want to make a node u a super node we need to 'collect' the random long-range contacts of u's neighbors within distance 2 and assign them to $\chi(u)$ (similarly for $\hat{\chi}(u)$). If u is too close to the edges of the grid, we can simply drop it as the fraction of such u is negligible (note that we do not need to care about this case with $u = s$ or t, since we only need a big enough initial S_0 or T_0, which is always trivial in this setting). Then we can continue similarly as before to show that the diameter of $\mathcal{K}(n, p, q)$, for $p, q \geq 1$, is $\theta(logn)$.

[11] A similar condition is also needed with respect to the destination node, and for simplicity, constants C and L can be chosen to be the same in both cases. These conditions also reflect a sufficient ratio of random links per node.

Higher dimensions: we now consider the diameter of $\mathcal{K}(k, n, p, q)$ where $k, p, q \geq 1$. Our prior results leave us with a simple task. It is again clear that conditions *I* and *II* are met. Fact 4 meets condition *III*, albeit we can not give exact values for the constant parameters as above. The model's connectedness and sufficiency of random links (everywhere) makes it as easy as above to find C and L such that $C\xi_1 > 1$ and $C\xi_2 > 1$ for any given positive constants ξ_1 and ξ_2. Our approach is then applicable as before which results in the following theorem, solving the general diameter problem in Kleinberg's model.

THEOREM 16. *For any $k, p, q \geq 1$, the expected diameter of the graphs from $\mathcal{K}(k, n, p, q)$ is $\theta(\log n)$.*

Other probability distributions: $0 \leq r < 2$. For a 2-D grid, when $0 \leq r < 2$ there is less of a bias in favor of links to closer nodes. Thus the worst case probability of a link existing is larger, so the Fact 1 lower bound on $p(u, v)$ still holds, and it is also easy to redo Fact 2 for this setting (intuitively, a link is more likely to leave or enter a ball when the distribution is more uniform). Thus our diameter proof easily applies in this setting as well, so the diameter is still $\Theta(\log n)$.

The case of lacking local links. We mention the classes $\mathcal{K}(k, n, 0, q)$ and $\mathcal{K}^*(k, n, 0, q)$ with $q > 0$, a special case when we do not assume the existence of local links. Note that this does not rule out the possibility of two adjacent nodes being connected by a random link. We show that (see [19]), if q is a constant then no matter how large q is, a graph in $\mathcal{K}^*(n, 0, q)$ is not strongly connected with probability tending to one when n goes to infinity. Moreover, the graph is even not semi-connected if we consider all the directed links as undirected links.

On a more abstract model. We have defined and analyzed a wider class of small-world graphs, which include the ones considered in this paper and others which are significantly different, e.g. the ones mentioned in [13]. These graphs are small-worlds with respect to graph diameter as well as Kleinberg's delivery time. Our main ideas are to abstract the key features of the $\mathcal{K}(k, n, p, q)$ setting which support the exponential growth property as in lemma 11. We start with an abstract model where a graph has fixed edges and for each vertex we choose independently one or more random outgoing edges based on a probability distribution.

6. CONCLUSION AND OPEN PROBLEMS

We have extended the analysis of Kleinberg's small-world model. We proved a tight $\theta(log^2 n)$ bound for Kleinberg's routing scheme and $\theta(\log n)$ for the expected diameter in this model. We also demonstrate an improved routing algorithm, using additional, fairly local, information. Finally, we suggest and demonstrate a more general approach to analyze other extensions of Kleinberg's small-world model. We have created (but do not present here) an abstract framework to define and analyze a class of small-world graphs, using only a few abstract features of Kleinberg's grid setting.

However, there are still many interesting problems. Prior results are mainly concerned with uniform-degree graphs but we also want to extend our framework to better handle non-uniform degrees such as power-law graphs (e.g. the Internet topology). Another possible extension is to relax some of

our framework criteria, e.g. "sufficiency of random links everywhere" can be relaxed to sufficiency of random links with a high probability. Greedy routing seems an obvious choice for the settings we have considered, however, backtracking or look-ahead would be useful in some other contexts with more network information available. For example, for networks with non-uniform degrees, a routing decision could be a function of two variables: the distance from the considered node to the destination and the prospect of reaching from this node's neighborhood to an outside node. Finally, we wish to extend our diameter result to Kleinberg's network using the inverse r^{th} power distribution for $r > 2$, especially when r is close to 2.

7. REFERENCES

[1] R. Albert, H. Jeong, A.-L. Barabasi, "The diameter of the World Wide Web" Nature 401, 130 (1999).

[2] J. Aspnes, Z. Diamadi, G. Shah, " Fault-tolerant Routing in Peer-to-peer Systems", Proc. 22nd ACM Symp. on Princ. of Dist. Computing, 2002, p. 223-232

[3] L. Barriere, P. Fraigniaud, E. Kranakis, D. Krizanc, "Efficient Routing in Networks with Long Range Contacts", Proc. 15th Intr. Symp. on Dist. Computing, DISC 01, 2001, p. 270-284

[4] I. Benjamini and N. Berger, "The Diameter of a Long-Range Percolation Clusters on Finite Cycles" Random Structures and Algorithms, 19,2:102-111 (2001).

[5] B. Bollobas, "The Diameter of Random Graphs", IEEE Trans. Inform. Theory 36 (1990), no. 2, 285-288.

[6] B. Bollobas, F.R.K. Chung, "The diameter of a cycle plus a random matching", SIAM J. Discrete Math. 1,328 (1988)

[7] B. Bollobas, *Random Graphs* (Academic Press, London, 1985)

[8] F. Chung and L. Lu, "The Diameter of Random Sparse Graphs", Advances in Applied Math. 26 (2001), 257-279.

[9] D. Coppersmith, D. Gamarnik, and M. Sviridenko "The diameter of a long- range percolation graph", Random Structures and Algorithms, 21,1:1-13 (2002).

[10] P. Dodds, R. Muhamad, and D. Watts, "An experimental study of search in global social networks", Science, 301:827-829, 2003.

[11] D. Kemper, J. Kleinberg, and A. Demers, "Spatial gossip and resource location protocols", in Proc. ACM Symp. Theory of Computing, 163-172, 2001.

[12] P. Fraigniaud, C. Gavoille and Christophe Paul, "Eclecticism Shrinks the World", Technical Report LRI-1376, University Paris-Sud, Nov. 2003. Also, to appear in PODC 2004.

[13] J. Kleinberg, "Small-World Phenomena and the Dynamics of Information", in *Advances in Neural Information Processing Systems (NIPS)* 14, 2001.

[14] J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective", in Proc. 32nd ACM Symp. Theory of Computing, 163-170, 2000.

[15] M. Kochen, Ed., *The Small-World* (Ablex, Norwood, 1989)

[16] E. Lebhar and N. Schabanel, "Almost optimal decentralized routing in long-range contact networks", to appear in ICALP 2004, 2004.

[17] G. S. Manku and M. Naor and U. Wieder, "Know Thy Neighbour's Neighbour: The Role of Lookahead in Randomized P2P Networks", Proc. 36th ACM Symposium on Theory of Computing (STOC 2004).

[18] G. Manku, "Routing Networks for Distributed Hash Tables", Proc. 22nd ACM Symposium on Principles of Distributed Computing, PODC 2003, p. 133-142, June 2003

[19] Charles Martel and Van Nguyen, "The complexity of message delivery in Kleinberg's Small-world Model", Tech-Report CSE-2003-23, CS-UCDavis, 2003.

[20] D. Malkhi, M. Naor, D. Ratajczak, "Viceroy: A Scalable and Dynamic Emulation of the Butterfly", Proc. of the 21st ACM Symposium on Principles of Distributed Computing, 2002

[21] S. Milgram, "The small world problem" Psychology Today 1,61 (1967).

[22] D. Watts and S. Strogatz, "Collective Dynamics of small-world networks" Nature 393,440 (1998).

[23] Xu, J., Kumar, A., and Yu, X., "On the Fundamental Tradeoffs between Routing Table Size and Network Diameter in Peer-to-Peer Networks", accepted to IEEE Journal on Selected Areas in Communications

[24] S. Yook, H. Jeong, A.-L. Barabasi, "Modeling the Internet's large-scale topology", Proceedings of the National Academy of Sciences, 99:13382-13386, 2002.

[25] H. Zhang, A. Goel and R. Govidan, "Using the Small-World Model to Improve Freenet Performance" Proc. IEEE Infocom, 2002.

APPENDIX

Proof of fact 5.

To evaluate "the attractive force" generated by S_B, we partition it into simpler subsets, which are hyper-intervals defined by linear systems. The number of these subsets can be easily upper bounded by a constant depending on k only, hence we just need to show an $O(\frac{1}{mlogn})$ contribution from each subset. For instance, consider a two-dimensional space with u as the origin. S_B actually has a diamond shape and these subsets are linear segments (intervals) on this diamond perimeter, separated by the 4 vertices and the axis. In a general k-dimensional space, each subset of nodes $(x_1, x_2, \ldots, x_k)$ can be defined as:

$$s_1x_1+s_2x_2+\ldots+s_kx_k = a, s_i \in \{-1, 1\}, x_i \in [a_i, b_i], i = 1..k$$

where a, a_i, b_i, s_i are integer constants depending on l, m and k such that $a > 0, a_i \times b_i \geq 0$. To bound the "attractive force" of each subset as such, we again partition the subset into u-centric spherical surfaces where all nodes generate the same force and the number of nodes (integer coordinates) can be counted easily.

We now give a proof for fact 5 for $\mathcal{K}(k, n, p, q)$ but it is easy to extend our proof for $\mathcal{K}^*(k, n, p, q)$. Our proof is based on claims 17 and 18 below. Once these claims are verified, fact 5 immediately follows. Let node v and integer $l > 0$ be the center and radius of ball B. Let $m = d(u, B)$ and $d = d(u, v)$; clearly $d = m + l$. Let $\mathcal{Z}^+ = \{i \in \mathcal{Z} : i \geq 0\}$. Assume that the nodes on the grid are the integer points (wrapped by a hypercube with size n) in the space $\mathcal{R}^k$. For simplicity, we also assume the origin of $\mathcal{R}^k$ is at u (i.e. u has all coordinates zero) and $v(d_1, d_2, \ldots, d_k)$ is in the 'positive

angle' relative to u, i.e. $v \in \mathcal{Z}^{+^k}$. Also call $N(\mathcal{L})$ the number of integer nodes within an object $\mathcal{L}$.

CLAIM 17. *Consider a linear hyper-multilateral (LHM) object $\mathcal{L}$ which is defined by a linear system of the following form*

$$s_1x_1 + s_2x_2 + \ldots + s_kx_k = a \quad (2)$$
$$s_i \in \{-1, 1\}, x_i \in [a_i, b_i], i = 1..k$$
$$a > 0, a_i \times b_i \geq 0$$

where a, a_i, b_i, s_i are constant integers. For $u = (0, 0, \ldots, 0)$, if $|x_1| + |x_2| + \ldots + |x_k| \geq m$ then $P(u, \mathcal{L}) \leq \frac{c}{mlogn}$ for some constant c depending on k only.

CLAIM 18. *S_B can be decomposed into at most C LHM objects (determined by an instance of (2) above), where C is a constant dependent on k only.*

For $k = 1$, a ball is an interval, so a LHM is simply an endpoint of an interval. Similarly, a LHM is an edge of a diamond when $k = 2$, a face of a cube when $k = 3$, and so on. With claims 17 and 18, fact 5 follows immediately. We now prove claim 17, which is also supported by the following simple fact, directly implied from section 3.2.

FACT 19. *(i) There exist positive constants c_1 and c_2 such that for a k-D grid of size n, $c_1 \leq b_j(u)/j^{k-1} \leq c_2, \forall u \in V, \forall j = 1..n$.*

(ii) for $\mathcal{K}^*(k, n, p, q)$ or $\mathcal{K}(k, n, p, q)$, the inverse normalized coefficient $c_u = \theta(logn)$ and for any two distinct nodes u and v, $p(u, v) = \Omega((n^k logn)^{-1})$.

See [19] for the proof of fact 19.

PROOF OF CLAIM 17. Using equation (2), we have either both a_i and b_i positive or negative. For simplicity, assume $0 \leq a_i < b_i$ for the rest of the proof; it is easy to extend our proof for the other cases. By a proper permutation of indexes we can make the characteristic equation become $x_1 + \ldots + x_j - x_{j+1} - \ldots - x_k = a$ for some integer j in $1..k$. Obviously this preserves $P(u, \mathcal{L})$. Also note that $x_1 + x_2 + \cdots + x_k \geq m$ since $x_i \geq a_i \geq 0, \forall i$.

We now consider the special case of $j = k$, i.e. $x_1 + \ldots + x_k = a$. Thus, all points in $\mathcal{L}$ are at distance a from u. From fact 19(i), $N(\mathcal{L}) \leq b_a(u) \leq c_1 a^{k-1}$ for some constant c_1. Therefore, from fact 19(ii) above, $P(u, \mathcal{L}) \leq c_1 a^{k-1} \times \frac{c_2}{a^k logn} \leq \frac{c}{mlogn}$ for some constants c_2 and $c = c_1c_2$ (note that $a > m$).

Now consider the case of $j < k$. Set $x = x_1 + \ldots + x_j$ and $y = x_{j+1} + \ldots + x_k$ we then have $x - y = a$. Clearly, for each value $\xi = x + y$, there is an unique pair of x, y. Let $\mathcal{L}_\xi = \mathcal{L} \cap B_\xi(u)$, i.e. the subset of nodes in $\mathcal{L}$ with $x + y = \xi$. Let Z_1 be the $\mathcal{Z}^j$ space of $x_1, \ldots, x_j$ and Z_2 be the $\mathcal{Z}^{k-j}$ space of $x_{j+1}, \ldots, x_k$. We now project $\mathcal{L}_\xi$ onto Z_1 to obtain an image $\mathcal{I}_1$, which has characteristic equation $x_1 + \ldots x_j = x$, and onto Z_2 to obtain an image $\mathcal{I}_2$, which has characteristic equation $x_{j+1} + \ldots x_k = y$. Again by fact 19(i), we have $N(\mathcal{I}_1) = O(x^{j-1})$ and $N(\mathcal{I}_2) = O(y^{k-j-1})$. Thus, clearly $N(\mathcal{L}_\xi) = N(\mathcal{I}_1) \times N(\mathcal{I}_2) \leq c_3 x^{j-1} y^{k-j-1} \leq c_3 \xi^{k-2}$ for some constant c_3.

Therefore, from fact 19(ii), $P(u, \mathcal{L}_\xi) \leq c_3 \xi^{k-2} \times \frac{c_4}{\xi^k logn} \leq \frac{c_3 c_4}{\xi^2 logn}$ for some constant c_4. Now, summing $P(u, \mathcal{L}_\xi)$ over all possible values of $\xi \geq m$, we have $P(u, \mathcal{L}) < \sum_{\xi=m}^{\infty} \frac{c_3 c_4}{\xi^2 logn}$. Note that $\sum_{i=m}^{\infty} 1/i^2 \leq 1/m^2 + \sum_{i=m}^{\infty} \frac{1}{i(i+1)} \leq 1/m^2 + \sum_{i=m}^{\infty} (\frac{1}{i} - \frac{1}{i+1}) \leq 1/m^2 + 1/m$; thus $P(u, \mathcal{L}) \leq \frac{c}{mlogn}$ for $c = 2c_3c_4$. □

PROOF OF CLAIM 18. We need to show that S_B can be decomposed into at most C LHM objects (determined by an instance of (2) above), where C is a constant dependent on k only. Clearly, a node $W(x_1, x_2, \ldots, x_k) \in \mathcal{Z}^k$ belongs to S_B, the surface of $B_l(u)$, if and only if $|x_1 - d_1| + |x_2 - d_2| + \ldots + |x_k - d_k| = l$. Note that $|x_1| + \ldots + |x_k|$ is at least $m = d(u, B)$.

Thus, S_B is composed of 2^k faces, each of which is on a hyper-plane determined by an equation of the form $\pm x_1 \pm x_2 \pm \ldots \pm x_k = a$, where integer constant a may vary for different hyper planes. (There are 2^k combinations of these $+/-$ so we have 2^k such hyper planes.) More specifically, such a face can be determined by a linear system similar to (2) but without $a_ib_i \geq 0$, where a, a_i, b_i are integer constants completely determined by $d_1, \ldots, d_k, n$ and l (for instance is it not hard to see that $a = l + s_1d_1 + s_2d_2 + \ldots + s_kd_k$). Call such a linear system a LES.

We can think of S_B as being decomposed into 'pieces'. Let us further decompose this as follows. For each axis-hyper plane $x_i = 0$, we divide the pieces, being 'cut through' by this hyper plane, each into two 'smaller pieces': one 'above' and one 'below' this hyper plane. For example, if a piece has LES with $a_i < 0 < b_i$ for some $i = 1..k$, we then split this object into two, each of which is described by the same LES with additional $x_i \in [0, b_i]$ for the 'the above' and $x_i \in [a_i, 0]$ for the 'the below'. We repeat this process until all pieces are inside a 'right angle': none of them are 'cut through' by any axis-hyper plane. Now each obtained object is described by an instance of (2) and the number of them is less than $2^k \times 2^k = 2^{2k}$. □