Efficient Routing in Networks with Long Range Contacts (Extended Abstract)*

Lali Barrière[1], Pierre Fraigniaud[2], Evangelos Kranakis[3], and Danny Krizanc[4]

[1] Dept. de Matemàtica Aplicada i Telemàtica, Universitat Politècnica de Catalunya. Email: lali@mat.upc.es

[2] CNRS, LRI, Université Paris-Sud, France. Email: pierre@lri.fr

[3] School of Computer Science, Carleton University, Canada. Email: kranakis@scs.carleton.ca Research supported in part by NSERC (Natural Sciences and Engineering Research Council of Canada).

[4] Department of Mathematics, Wesleyan University, USA. Email: dkrizanc@caucus.cs.wesleyan.edu

Abstract. We investigate the notion of Long Range Contact graphs. Roughly speaking, such a graph is defined by (1) an underlying network topology G, and (2) one (or possibly more) extra link connecting every node u to a "long distance" neighbor, called the long range contact of u. This extra link represents the *a priori* knowledge that a node has about far nodes and is set up randomly according to some probability distributions p. To illustrate the claim that Long Range Contact graphs are a good model for the small world phenomenon, we study greedy routing in these graphs. Greedy routing is the distributed routing protocol in which a node u makes use of its long range contact to progress toward a target, if this contact is closer to the target, than the other neighbors. We give upper and lower bounds on greedy routing on the n-node ring C_n augmented with links chosen using the r-harmonic distributions. In particular, we show a tight $\Theta(\log^2 n)$-bound for the expected number of steps required for routing in C_n augmented using the 1-harmonic distribution. Hence, our study shows that the model of Kleinberg [11] can be simplified by using the ring rather than the mesh while preserving the main features of the model. Our study also demonstrates the significant difference (in term of both diameter and routing) between the ring augmented with long range contacts chosen with the harmonic distribution and the ring augmented with a random matching as introduced by Bollobas and Chung [3]. Finally, using epimorphisms of a graph onto another, for any network G, we show how to define a probability distribution p and study the performance of greedy routing in G augmented with p. For appropriate embeddings (if they exist), this performance turns out to be $O(\log^2 n)$.

* Part of this work was done while the first and third author were visiting LRI at Orsay (additional support from Université Paris-Sud), and while the first and second authors were visiting Carleton University. Additional support from Carleton University, NATO, and MITACS (Mathematics of Information Technology and Complex Systems) grants.

1 Introduction

The small-world phenomenon arises from rather anecdotal experience that has been witnessed in many large interconnected systems: it is a phenomenon that formalizes the paradoxical ability of an entity in the system to be only a few "degrees" of separation away from any other entity in the system. This paradoxical occurrence of the small-world phenomenon has been backed by statistical data of reachability and has several instantiations in the scientific literature from sociology to the web. It has become the subject of investigation in popular as well as artistic culture (see [7, 8, 16]).

To understand this phenomenon studies have been made that include the introduction of two graph theoretic models: *relational* graphs and *spatial* graphs. In relational graphs the probability of the vertices becoming connected depends only upon preexisting connections [3, 5, 16, 17]. In spatial graphs, the corresponding probability is a function of the vertices [11, 16]. In recent years the web has been the focus of investigations. Here researchers have investigated power-laws, i.e., the probability that a node has degree k is given by k^{-c}, for some constant $c > 0$; this implies that nodes with low degree are the most numerous and the probability of nodes with given degree k decreases as k increases proportionately with k^{-c} [1, 4, 6, 12]. All these studies show that random graphs $\mathcal{G}_{n,p}$ as defined by Erdös and Rényi, are not good models for the small world phenomenon, because they have a large diameter when the average degree is small [2].

In this paper, we study the notion of Long Range Contact graphs. Let $G = (V, E)$ be a network on n vertices. Consider a probabilistic mapping p on the vertices of G such that $\sum_{v \in V} p(u, v) = 1$, for all $u \in V$. I.e., each node $u \in V$ has an associated probability distribution $p(u, \cdot)$. Given G and p, the Long Range Contact graph (G, p) is a directed graph defined on the same set of vertices, such that every node u has $\deg_G(u) + 1$ out-neighbors, that is its $\deg_G(u)$ neighbors in G, plus one additional out-neighbor chosen at random according to p. This latter neighbor is called the *long range contact* of u. The probabilistic mapping p, i.e., the probability distributions $p(u, \cdot)$'s, reflect "vague knowledge" available at the nodes about the possible status and location of a desired information located at some node of the network.

In small world graphs, not only have the nodes a few degrees of separation, but these nodes are able (or expected) to find reasonably short routes between them. Therefore, the following two parameters have been the source of much research: (1) The *diameter* of (G, p), i.e., the maximum distance between any two nodes in the augmented graph; and (2) The performance of *greedy routing* in (G, p), i.e., routing from a source s to a target t is executed by selecting, at each intermediate node u, the next node as the neighbor of u (including its long range contact) which is closer (in the graph G) to the target t.

These two parameters depend first on the probability distribution to select a long range contact and second on the underlying topology of the graph. To be a good candidate to abstract small world phenomenon, a graph model must insure that both the diameter, and the number of greedy routing steps, be small. In this paper, we study the model in which G is the ring C_n, and p is the harmonic distribution.

Related research Among the previously cited papers, two are strongly connected to this paper. Bollobas and Chung [3] have studied the diameter of a ring plus a random matching, selected uniformly among all possible matchings. They have shown that the resulting augmented ring has a diameter $\Theta(\log n)$ with a probability tending to 1 as n goes to infinity. However, the performance of greedy routing can be very bad in this model. Indeed, Kleinberg [11] has shown that the ring augmented with long range contacts chosen uniformly at random offers very bad properties in term of routing ($\Omega(\sqrt{n})$ lower bound for the expected number of steps). As an attempt to model the small world phenomenon, Kleinberg has therefore proposed to use the 2-dimensional square grid augmented with long range contacts chosen according to the 2-harmonic distribution. He showed that, in this model, greedy routing performs in $O(\log^2 n)$ expected number of steps. Moreover he showed that this is optimal in the sense that for $r \neq 2$ any distributed routing algorithm based on the r-harmonic distribution has an $n^{\Omega(1)}$ lower bound on the expected number of steps. He concluded that the grid with the 2-harmonic distribution is a good model for the small world phenomenon.

Results of the paper Motivated by the research of Bollobas and Chung, we have investigated the augmented ring. Motivated by the research of Kleinberg, we have investigated *r-harmonic* mappings p_r, $r \geq 0$, defined as follows. Given two nodes u and v, the probability for u to have v as long range contact is given by $p_r(u,v) = \frac{d(u,v)^{-r}}{\sum_{w \neq u} d(u,w)^{-r}}$, where $d(\cdot,\cdot)$ is the distance function in the network. The *uniform* distribution (which is obtained for $r = 0$), i.e., $p(i,j) = 1/n$, and the *Zipf* distribution [18] (which is obtained for $r = 1 - \log .80/\log .20$), are two examples of harmonic distributions. We have performed an exhaustive study of the performances of greedy routing in the ring augmented with harmonic long range contacts, for all $r \geq 0$. Table 1 summarizes our results.

One important result in this table is the tight $\Theta(\log^2 n)$-bound for the expected number of steps of greedy routing in the ring augmented with long range contact chosen using the 1-harmonic distribution. The upper bound $O(\log^2 n)$ shows that the simple ring can perform as well as the square mesh, and hence provides a simpler model for the small world phenomenon. The lower bound $\Omega(\log^2 n)$, as well as the other lower bounds for $r \neq 1$, show that greedy routing cannot perform faster than $\log^2 n$ steps in any ring augmented with an harmonic distribution. It seems to be a challenging task to prove or disprove the existence of a distribution allowing greedy routing to perform faster in the ring, the square grid, or even the k-dimensional mesh, $k \geq 3$.

As a last contribution, we show how to extend the results of the ring to any network G, by using epimorphisms of a graph onto another. In particular,

r-Harmonic Distribution	Lower Bound	Reference	Upper Bound	Reference
$0 \leq r < 1$	$\Omega(n^{\frac{1-r}{2-r}})$	Theorem 4	$O(n^{1-r})$	Theorem 2
$r = 1$	$\Omega(\log^2 n)$	Theorem 5	$O(\log^2 n)$	Theorem 1
$1 < r < 2$	$\Omega(n^{\frac{r-1}{r}})$	Theorem 4	$O(n^{r-1})$	Theorem 1
$r = 2$	$\Omega(\sqrt{n})$	Theorem 4	$O(\frac{n \log\log n}{\log n})$	Theorem 3
$2 < r$	$\Omega(n^{\frac{r-1}{r}})$	Theorem 4	$O(n)$	Trivial

Table 1. Expected number of steps of greedy routing in the ring augmented with long range contacts chosen according to the r-harmonic distribution.

we show how to define a probabilistic mapping p and study the performance of greedy routing in (G, p). For appropriate embeddings this performance turns out to be $O(\log^2 n)$.

2 Preliminary Results

For the purpose of simplification of the presentation, all our results are formally proven for the *directed* ring, i.e., the digraph in which nodes are labeled from 0 to n, and where node i has node $i+1$ as out-neighbor, and $i-1$ as in-neighbor (unless specified otherwise, all operations are performed modulo $n+1$). In each case, the result in the undirected ring differs by a constant factor only. We denote by R_{n+1} the directed ring of $n+1$ nodes.

The r-harmonic random variable H_r, with values in $\{1, \ldots, n\}$ has the probability distribution defined by $\mathbf{Pr}(\{H_r = k\}) = \frac{k^{-r}}{H_n^{(r)}}$, where $H_n^{(r)} = \sum_{i=1}^{n} i^{-r}$ is the *r-harmonic number* of order n. Therefore, if R_{n+1} is augmented using the r-harmonic mapping p_r, then, given two nodes i and j, the probability for i to have j as long range contact in (R_{n+1}, p_r) is given by $p_r(i,j) = \frac{((j-i) \bmod n+1)^{-r}}{H_n^{(r)}}$. This formula can be made more explicit by noticing that the harmonic numbers satisfy the following identities.

Lemma 1. *The r-harmonic number of order n is*

$$H_n^{(r)} = \begin{cases} \frac{1}{1-r} n^{1-r} + O(1) & \text{if } r < 1; \\ \log n + O(1) & \text{if } r = 1; \\ O(1) & \text{if } r > 1. \end{cases}$$

The next lemma shows thresholds in the behavior of the harmonic distributions. Not surprisingly, these thresholds are those appearing in Table 1.

Lemma 2. *The expected value of H_r is*

$$\mathbf{E}(H_r) = \begin{cases} \Theta(n) & \text{if } 0 \le r < 1 \\ \Theta(n/\log n) & \text{if } r = 1 \\ \Theta(n^{r-1}) & \text{if } 1 < r < 2 \\ \Theta(1/\log n) & \text{if } r = 2 \\ \Theta(1) & \text{if } 2 < r. \end{cases}$$

For our analysis of greedy routing in R_{n+1}, we will always assume that the source node is 0, and the target node is n. It is indeed easy to observe that this is a worst case, as far as greedy routing is concerned. Indeed, the probability for a node to have a long range contact at distance d on the ring decreases as d increases. Therefore the farther a source is from a target, the larger is the expected number of steps to route from that source to that target.

A very naive interpretation of Lemma 2 would be to derive that, e.g., greedy routing in the ring augmented with the 1-harmonic distribution performs in $O(\log n)$ expected number of steps. This reasoning fails because the expected gain of using long range contacts decreases as one gets closer to the destination (as long range contacts may lead farther away from that destination than one currently is). The following clarifies that point. Given a node $s \in \{0, \dots, n-1\}$, greedy routing defines a random variable J_s as the length of the "jump" performed at s toward the target n. It satisfies: $J_s = H_r$if $H_r \le n - s$, and -1 otherwise. One can easily show the following.

Lemma 3. *For $k \le n - s$, we have*

$$\mathbf{Pr}(\{J_s = k\} = \begin{cases} \mathbf{Pr}(\{H_r = k\}) + \mathbf{Pr}(\{H_r > n - s\}) & \text{if } k = 1 \\ \mathbf{Pr}(\{H_r = k\}) & \text{if } 1 < k \le n - s. \end{cases}$$

And the expected value of the jump J_s at node s in (R_{n+1}, p_r) is:

$$\mathbf{E}(J_s) = \begin{cases} \Theta((n-s)^{2-r}/n^{1-r}) & \text{if } r < 1 \\ \Theta((n-s)/\log n) & \text{if } r = 1 \\ \Theta((n-s)^{2-r}) & \text{if } 1 < r < 2 \\ \Theta(\log(n-s)) & \text{if } r = 2 \\ \Theta(1) & \text{if } r > 2. \end{cases}$$

3 Upper Bounds

We begin with general considerations which apply to arbitrary networks. Then we will refine these concepts for the specific case of the ring. For each vertex u of $G = (V, E)$, and each real number $r > 0$, define the ball $B_r^G(u)$ of radius r around u as the set of vertices at distance at most r from u. (If the graph used is clear from the context we will omit the superscript G from $B_r^G(u)$ and write $B_r(u)$.) For any set S of vertices of (G, p) and any vertex $u \in V$ define

$p[u \to S] = \sum_{v \in S} p(u,v)$. Here we are trying to quantify the weight that a node u gives to a contact in S in the sense that $p[u \to S]$ is the probability that a node u has a long-range contact in the set S.

Definition 1. *Let G be a graph, p a probabilistic mapping on G, $c > 1$ a constant, and f a function. The pair (G,p) is called an (f,c)-Long Range Contact graph if for any pair (u,t) of vertices of G at distance at most d we have that $p[u \to B_{d/c}(t)] \geq \frac{1}{f(d)}$.*

Lemma 4. *Let $G = (V,E)$ be a graph of diameter D. If (G,p) is an (f,c)-Long Range Contact graph then greedy routing in (G,p) performs in $O\Big(\sum_{i=1}^{\log_c D} f(D/c^i)\Big)$ expected number of steps.*

Proof. What is the probability, for a given node u at distance at most d from the target t, that the long range contact selected is at a distance at most d/c from the target? By definition, this is equal to $p[u \to B_{d/c}(t)]$. Moreover, by the geometric distribution, the expected number of trials to guarantee success is $1/p[u \to B_{d/c}(t)]$. When a trial fails, we make a move towards the target by going to a neighbor along a shortest path from the current node to the target. The next trial is therefore performed at a node still at distance at most d from t. It follows from Definition 1 that the expected number of trials to get a contact in $B_{d/c}(t)$ is at most $\frac{1}{p[u \to B_{d/c}(t)]} \leq f(d)$. This implies that after at most $f(d)$ expected number of routing steps from u, we enter $B_{d/c}(t)$. Iterating this we conclude that the expected number of steps for routing is at most $O\Big(\sum_{i=1}^{\log_c D} f(D/c^i)\Big)$.

Using specific probabilistic mappings we can simplify our analysis.

Definition 2. *A probabilistic mapping p on a graph G is* distance-invariant *if $p(u,v)$ depends only on the distance $d(u,v)$. A distance-invariant mapping is called* non-increasing *if it is a non-increasing function of the distance.*

To simplify notation we use the same symbol to denote the resulting mapping, namely $p(u,v) = p(d(u,v))$. We can prove the following result.

Lemma 5. *If p is a non-increasing distant-invariant mapping on the graph G then for all vertices u,t with $d(u,t) \leq d$ and all constants $c > 0$, we have that $p[u \to B_{d/c}(t)) \geq p\Big((c+1)d/c\Big) \cdot |B_{d/c}(t)|$.*

Proof. Let v be a node in $B_{d/c}(t)$. For any node u, $d(u,v) \leq d(u,t) + d(t,v) \leq d + d/c = (c+1)d/c$. It follows that

$$p[u \to B_{d/c}(t)] = \sum_{v \in B_{d/c}(t)} p(u,v)$$

$$= \sum_{v \in B_{d/c}(t)} p(d(u,v)) \text{ since } p \text{ is distance invariant}$$
$$\geq \sum_{v \in B_{d/c}(t)} p((c+1)d/c) \text{ since } p \text{ is non increasing}$$
$$= p((c+1)d/c) \cdot |B_{d/c}(t)|,$$

which completes the proof of the lemma.

As a direct consequence of Lemma 5, and by definition of (f,c)-Long Range Contact graphs, we obtain the following result.

Lemma 6. *Consider a graph G and a non-increasing distance-invariant mapping p. Then, for any $c > 1$, the pair (G,p) is an (f,c)-Long Range Contact graph where the function $f(d)$ is defined by $f(d) = \frac{1}{p((c+1)d/c) \cdot \min_{t \in V} |B_{d/c}(t)|}$.*

Theorem 1. *The expected number of steps for greedy routing on R_{n+1} is*

$$\begin{cases} O(\log^2 n) & \textit{if } r = 1 \\ O(n^{r-1}) & \textit{if } 1 < r < 2. \end{cases}$$

Proof. The r-harmonic mapping p_r on a graph G is a non-increasing distance-invariant mapping. From Lemma 6, (G, p_1) is a Long Range Contact graph with $f(d) \simeq 1/\log n$. The $O(\log^2 n)$ bound then results from the application of Lemma 4. Similarly, for $r > 1$, $G(,p_r)$ is a Long Range Contact graph with $f(d) \simeq d^{r-1}$ (cf. Lemma 1). The result then follows by application of Lemma 4.

In order to obtain non trivial upper bounds when either $r < 1$ or $r \geq 2$ we can use the method of probabilistic recurrences. First we recall the following discussion from [14] (Theorem 1.3, page 15). Let $g(x)$ be a monotone non-decreasing function from positive reals to positive reals. Consider a particle starting from position 0 and moving along the discrete line segment from 0 to n and whose position changes in discrete time intervals. If the particle is currently at position s it moves to position $s + X$ where X is a random variable ranging over the integers $1, \ldots, n-s$ such that $\mathbf{E}[X] \geq g(n-s)$. The following result due to Karp, Upfal and Widgerson was first stated in [10] (see also [9] for additional information on probabilistic recurrences):

Lemma 7. (Karp, Upfal, Widgerson [10]) *Let T be the random variable denoting the number of steps in which the particle reaches the position n. Then $\mathbf{E}(T) \leq \int_1^n dx/g(x)$.*

We can use Lemma 7 to analyze greedy routing when $r < 1$. More precisely, we can prove the following result.

Theorem 2. *The expected number of steps for greedy routing on R_{n+1} using r-harmonic distributions with $0 \leq r < 1$ is $O(n^{1-r})$.*

Proof. Greedy routing is similar to the motion of the particle described above. By Lemma 3, if the particle is in position s then the expected length of a jump is $\Theta((n-s)^{2-r}/n^{1-r})$. If we let $g(x) = \Theta(x^{2-r}/n^{1-r})$ then Lemma 7 is applicable and we obtain that the expected number of steps of greedy routing is at most

$$\int_1^n \frac{dx}{x^{2-r}/n^{1-r}} = \frac{n^{1-r}}{1-r} - \frac{1}{1-r}.$$

The Lemma on probabilistic recurrences can also be used for analysing greedy routing when using 2-harmonic distributions.

Theorem 3. *The expected number of steps for greedy routing on R_{n+1} using 2-harmonic distribution is $O(\frac{n \log\log n}{\log n})$.*

Proof. Combining Lemmas 3 and 7, we can show that up to a constant the expected number of steps of greedy routing is at most $\int_2^n \frac{dx}{\log x}$. This is easily seen to be in $O(\frac{n \log\log n}{\log n})$.

4 Lower Bounds

The proof of the following result is based on a proof in [11].

Lemma 8. *Let p be any distance-invariant mapping on R_{n+1}. Assume that there exists d and D, and ϵ, $0 < \epsilon < 1$, such that such that one of the two following conditions holds:*

1. *$d \geq D$ and $D \cdot \sum_{i=1}^{d} p(i) \leq \epsilon$;*
2. *$d \cdot D < n$ and $D \cdot \sum_{i>d} p(i) \leq \epsilon$.*

Then the expected number of steps of greedy routing is at least $(1-\epsilon)D$.

Proof. First we prove the lemma under condition 1. Let B denote the ball of R_{n+1} centered at n and radius d, i.e., $B = \{n-d, \ldots, n-1, n\}$. Recall that we consider greedy routing from 0 to n. Consider the events:

•E: In at most D steps we reach n.
•E': In at most D steps we reach a node that has a long range contact to a node in B.
•E'_i: In step i we reach a node that has a long range contact to a node in B.

Let X be the random variable which counts the number of steps to reach n from 0. In view of condition 1 we have that

$$\mathbf{Pr}(E') = \mathbf{Pr}(\cup_{i=1}^{D} E'_i) \leq \sum_{i=1}^{D} \mathbf{Pr}(E'_i) \leq D \cdot p[0 \to B] \leq D \cdot \sum_{i=1}^{d} p(i) \leq \epsilon.$$

It follows that

$$\mathbf{Pr}(\overline{E'}) = 1 - \mathbf{Pr}(E') \geq 1 - \epsilon. \tag{1}$$

Since $d \geq D$, $E \subseteq E'$, and hence $\overline{E'} \subseteq \overline{E}$. It follows that $\mathbf{Pr}(E|\overline{E'}) = 0$. Using this and Inequality 1 we can show that

$$\mathbf{E}[X] = \sum_k k \cdot \mathbf{Pr}(\{X = k\}) \geq \sum_k k \cdot \mathbf{Pr}(\{X = k\} \cap \overline{E'}) = \mathbf{Pr}(\overline{E'}) \cdot \mathbf{E}[X|\overline{E'}] \geq (1-\epsilon)D.$$

This proves the first part of the lemma. Next we prove the lemma under condition 2. Consider the events

•E: In at most D steps we reach n.
•E': In at most D steps, we reach a node u_0 that has a long range contact to a node $u_0^+ \neq n$ such that $d(u_0, u_0^+) > d$.
•E'_i: In step i, we reach a node u_0 that has a long range contact to a node $u_0^+ \neq n$ such that $d(u_0, u_0^+) > d$.

Again, let X be the random variable which counts the number of steps to reach n from 0. For every node u, let u^+ be the long range contact of u. Using Condition 2 of the lemma, we obtain

$$\mathbf{Pr}(E') = \mathbf{Pr}(\cup_{i=1}^{D} E'_i) \leq \sum_{i=1}^{D} \mathbf{Pr}(E'_i) \leq D \cdot \mathbf{Pr}(\{d(u, u^+) > d\}) = D \cdot \sum_{i > d} p(i) \leq \epsilon.$$

Since $dD < n$, $E \subseteq E'$, and hence $\overline{E'} \subseteq \overline{E}$. Therefore, $\mathbf{E}[X] \geq \mathbf{Pr}(\overline{E'}) \cdot \mathbf{E}[X|\overline{E'}] \geq (1 - \epsilon)D$. This completes the proof of the lemma.

Theorem 4. *The expected number of steps for greedy routing on R_{n+1} under the r-harmonic distribution is bounded from below by (up to a constant):*

$$\begin{cases} n^{\frac{1-r}{2-r}} & \text{if } r < 1 \\ n^{\frac{r-1}{r}} & \text{if } 1 < r \end{cases}$$

Proof. The cumulative distributions of the r-harmonics random variable H_r are given (up to a multiplicative constant) by the formulas

$$\mathbf{Pr}(\{H_r \leq k\}) \approx \begin{cases} (k/n)^{1-r} & \text{if } r < 1; \\ 1 - k^{1-r} & \text{if } r > 1. \end{cases}$$

When $r < 1$ we apply condition 1 of Lemma 8 with $d = D = n^{\frac{1-r}{2-r}}$, and $\epsilon = 1/2$. When $r > 1$ we apply condition 2 of Lemma 8 with $d = n^{\frac{1}{r}}$, $D = n^{\frac{r-1}{r}}$, and $\epsilon = 1/2$.

In the specific case $r = 1$, one can prove the optimality of Theorem 1 for the 1-harmonic distribution.

Theorem 5. *The expected number of steps of greedy routing using the 1-harmonic distribution is at least* $\Omega(\log^2 n)$.

Proof. Let H be a 1-harmonic random variable in $\{1,\dots,n\}$, i.e., $\mathbf{Pr}(\{H = i\}) = 1/(i \cdot H_n)$ where $H_n = \sum_{i=1}^{n} 1/i = \Theta(\log n)$. For any s, $0 \le s \le n-1$, the jump at node s is $J_s = \begin{cases} H \text{ if } H \le n-s; \\ 1 \;\; \text{otherwise.} \end{cases}$ Greedy routing from 0 to n constructs a sequence $s_0 = 0, s_1, s_2, \dots$ such that $s_{i+1} = s_i + J_{s_i}$. From Lemma 3, we have, for any $k \in \{1,\dots,n-s\}$,

$$\mathbf{Pr}(\{J_s = k\}) = \begin{cases} \mathbf{Pr}(\{H = k\}) & \text{if } 1 < k \le n-s; \\ \mathbf{Pr}(\{H = 1\}) + \mathbf{Pr}(\{H > n-s\}) & \text{otherwise.} \end{cases} \quad (2)$$

and

$$\mathbf{E}(J_s) = \mathbf{Pr}(\{H > n-s\}) + \frac{n-s}{H_n} \le 1 + \frac{n-s}{H_n}. \quad (3)$$

For $0 \le i \le \lfloor \log_2 n \rfloor$, let $n_i = n \cdot (1 - 1/2^i)$ and $I_i = [n_i, n_{i+1})$. Let $i > 0$, $s \in I_{i-1}$, and E_s be the event that the long range contact of s is in $[n_{i+1}, n]$ (i.e., greedy routing from s to n "jumps" over I_i). We have $\mathbf{Pr}(E_s) = \mathbf{Pr}(\{J_s \ge n_{i+1} - s\})$, and thus, thanks to Equation 2, $\mathbf{Pr}(E_s) = \sum_{k=n_{i+1}-s}^{n-s} \mathbf{Pr}(\{H = k\}) \simeq \frac{1}{\log n} \log\left(\frac{n-s}{n_{i+1}-s}\right) = \frac{1}{\log n} \log\left(1 + \frac{n-n_{i+1}}{n_{i+1}-s}\right)$. For $s \in I_{i-1}$, we have $2 - \log 3 \le \log\left(1 + \frac{n-n_{i+1}}{n_{i+1}-s}\right) \le 1$. As a consequence,

$$\mathbf{Pr}(E_s) = \Theta\Big(\frac{1}{\log n}\Big). \quad (4)$$

Let K be the random variable defined as the number of consecutive first intervals containing at least one node s_i, while performing greedy routing from 0 to n. More precisely, if greedy routing constructs the sequence $s_0 = 0, s_1, s_2, \dots$, then $K = \min\{j \;:\; s_i \notin I_j, \forall i\} - 1$. From Equation 4,

$$\mathbf{Pr}(\{K = k\}) = \Theta\Big(\frac{(1 - 1/\log n)^k}{\log n}\Big).$$

By using $\ln(1+x) \sim x$ when x is small, easy calculations show that $\mathbf{E}(K) = \Theta(\log n)$. Let us now concentrate on the time it takes to traverse an interval I_i, $i \le K$. Let

$$t_i = \min\{s_j \;:\; s_j \in I_i\} \text{ and } t'_i = \max\{s_j \;:\; s_j \in I_i\}.$$

Then let $\Delta_i = t_i - n_i$ and $\Delta'_i = n_{i+1} - t'_i$. If $t_i = s_j$ and $s_{j-1} \in I_\ell$, then $\Delta_i \le J_{n_\ell} = J_{n(1-1/2^\ell)}$ and thus, thanks to Equation 3,

$$\mathbf{E}(\Delta_i) \le \mathbf{E}(J_{n(1-1/2^\ell)}) \le 1 + \frac{n}{2^\ell H_n}.$$

Similarly,

$$\mathbf{E}(\Delta'_i) \le \mathbf{E}(J_{n(1-1/2^i)}) \le 1 + \frac{n}{2^i H_n}.$$

Therefore, if $i \leq K$, we get $\ell = i - 1$, and thus

$$\mathbf{E}(\Delta_i) \leq 1 + \frac{4|I_i|}{H_n} \text{ and } \mathbf{E}(\Delta'_i) \leq 1 + \frac{2|I_i|}{H_n}. \tag{5}$$

Let $D_i = t'_i - t_i$. We have $D_i = (n_{i+1} - n_i) - (\Delta_i + \Delta'_i)$, and thus from Equation 5,

$$\mathbf{E}(D_i) \geq |I_i|(1 - 6/H_n). \tag{6}$$

In the interval I_i, the long range contacts are at distance at most $J_{n_i} = J_{n(1-1/2^i)}$. Let $X^{(i)} = J_{n(1-1/2^i)}$, and let N_i be the stopping time for $X^{(i)}$, that is

$$N_i = \min\{k | \sum_{j=1}^{k} X^{(i)} \geq D_i\}.$$

From Equation 6, we have $\mathbf{E}(\sum_{j=1}^{N_i} X^{(i)}) \geq \mathbf{E}(D_i) \geq |I_i|(1-6/H_n)$. On the other hand, by Wald's Equation (see [15] (Corollary 6.2.3)), we have $\mathbf{E}(\sum_{j=1}^{N_i} X^{(i)}) = \mathbf{E}(N_i) \cdot \mathbf{E}(X^{(i)})$. Therefore, from Equation 3, we get

$$\mathbf{E}(N_i) \geq \frac{|I_i|(1 - 6/H_n)}{1 + 2|I_i|/H_n} = \Omega(\log n).$$

To summarize, the expected number of consecutive intervals I_i traversed by the greedy routing is $\Omega(\log n)$, and the expected number of steps to traverse each of these intervals is $\Omega(\log n)$. Therefore the expected number of steps of greedy routing is at least $\Omega(\log^2 n)$.

It is an open problem whether or not the lower bound of Theorem 5 is valid under any distance invariant distribution on the ring R_{n+1}. However we note the following general result which is an immediate corollary of Lemma 8.

Corollary 1. *Let p be any non-increasing distance-invariant mapping on R_{n+1} and $D < n/4$ an integer such that*

$$\min\left\{\sum_{i=1}^{O(D)} p(i), \sum_{i=\Omega(n/D)}^{n} p(i)\right\} \leq O\left(\frac{1}{D}\right).$$

Then the expected number of steps of greedy routing is in $\Omega(D)$.

5 Long Range Contact Graphs

In this section, we show how to generalize the results obtained on the ring to arbitrary graphs. More precisely, we consider the issue of how to produce an appropriate probabilistic mapping p on an arbitrary graph G so that routing can be done in a small number of steps in (G, p). We begin with the class of k-dimensional tori.

Kleinberg [11] considers the two dimensional grid. We can generalize his result in the following manner. Consider the k-dimensional torus T_q^k: with $n = q^k$ vertices, i.e., q vertices per dimension and $k \geq 1$. It is clear that balls of radius d have size $\Theta(d^k)$, and spheres of radius d have size $\Theta(d^{k-1})$. Moreover the diameter is $D = \Theta(n^{1/k})$. Let us consider the r-harmonic distribution on the graph T_q^k. For the r-harmonic distribution we have

$$p(d) = \frac{d^{-r}}{\sum_{i=1}^{D} i^{-r}|S_i(t)|} = \Theta\left(\frac{d^{-r}}{\sum_{i=1}^{q} i^{-1} \cdot i^{k-r}}\right). \tag{7}$$

Equation (7) indicates that we should select $r = k$. In this case we obtain that $p(d) = \Theta(d^{-k}/\log q)$. In particular, using Lemma 6, (T_q^k, p) becomes an (f, c)-Long Range Contact graph, where

$$\begin{aligned} f(d) &= 1/(p(3d/2) \cdot |B_{d/2}(t)|) \\ &= 1/(\Theta((3d/2)^{-k}/\log q) \cdot (d/2)^{-k}) \\ &= \Theta\left(\frac{3^k}{k} \log n\right). \end{aligned}$$

Since the diameter of the graph T_q^k is $D = \Theta(n^{1/k})$ we can use Lemma 4 to obtain the following result.

Lemma 9. *Let T_q^k be the k-dimensional torus of dimension $k \geq 1$ and $n = q^k$ nodes, and let p_k be the k-harmonic mapping. Then (T_q^k, p_k) is an $(f, 2)$-Long Range Contact graph, where $f(d) = \Theta(\frac{3^k}{k} \log n)$. Moreover, greedy routing in (T_q^k, p) performs in $O\left(\frac{3^k}{k^2} \log^2 n\right)$ expected number of steps.*

It follows from Lemma 9 that greedy routing can be performed in $O(\log^2 n)$ expected number of steps in the k-dimensional torus T_q^k, where k is constant and the probabilistic mapping is defined as before. Let us now present a tool to extend results on a greedy routing in a graph G to other graphs G'. First, we recall the notion of an epimorphism.

Definition 3. *Consider two graphs $G = (V, E)$ and $G' = (V', E')$. An* epimorphism *of G onto G' is an onto mapping $\phi : V \rightarrow V'$ such that $\{u, v\} \in E \Rightarrow \{\phi(u), \phi(v)\} \in E'$, for all vertices $u, v \in V$.*

Note that, if ϕ is an epimorphism, then $d_{G'}(\phi(u), \phi(v)) \leq d_G(u, v)$ for every u and v. Next we define the notion of *distance maintaining* epimorphism.

Definition 4. *Let α be a positive constant. An epimorphism ϕ from the graph $G = (V, E)$ onto the graph $G' = (V', E')$ is called* α-distance maintaining *if for all $u, v \in V$, $d_G(u, v) \leq \alpha \cdot d_{G'}(\phi(u), \phi(v))$. The epimorphism ϕ is called* distance maintaining *if it is* α-distance maintaining *for some positive constant α.*

It is not hard to see that if p is a probabilistic mapping on the vertices of G then p' is a probabilistic mapping on the vertices of G', where

$$p'(u',v') = \frac{1}{|\phi^{-1}(u')|} \sum_{\substack{u\in\phi^{-1}(u')\\ v\in\phi^{-1}(v')}} p(u,v). \tag{8}$$

Lemma 10. *Assume that there is an α-distance maintaining epimorphism ϕ from G onto G'. Let (G,p) be an $(f,\alpha c)$-Long Range Contact graph. Then (G',p') is an (f',c)-Long Range Contact graph, where p' is defined in Equation 8 and $f'(d) = f(\alpha d)\cdot \max_{u'\in V'} |\phi^{-1}(u')|$.*

Proof. Let ϕ be a distance maintaining epimorphism from G onto G'. First of all observe that for any t, t' such that $\phi(t) = t'$ we have that

$$B_{d'}^{G'}(t') = \{v' : d_{G'}(v',t') \le d'\} = \{\phi(v) : d_{G'}(\phi(v),\phi(t)) \le d'\}.$$

Therefore, $B_{d'}^{G'}(t') \supseteq \phi(\{v : d_G(v,t) \le d'/c'\})$ from the definition of epimorphism. Hence $B_{d'}^{G'}(t') \supseteq \phi(B_{d'}^{G}(t))$. It follows that

$$B_{d'}^{G'}(t') \supseteq \bigcup_{t\in\phi^{-1}(t')} \phi(B_{d'}^{G}(t)). \tag{9}$$

Let $u', t' \in V'$ be vertices such that $d_{G'}(u',t') \le d'$. From the definition of epimorphism, there exist vertices $u_0, t_0 \in V$ such that $\phi(u_0) = u'$, $\phi(t_0) = t'$. Then from the definition of distance maintaining, we have $d_G(u_0,t_0) \le \alpha\cdot d_{G'}(u',t') \le \alpha\cdot d'$. We have

$$\begin{aligned} p'[u' \to B_{d'/c}^{G'}(t')] &= \sum_{v'\in B_{d'/c}^{G'}(t')} p'(u',v') \\ &= \frac{1}{|\phi^{-1}(u')|} \sum_{v'\in B_{d'/c}^{G'}(t')} \sum_{u\in\phi^{-1}(u')} \sum_{v\in\phi^{-1}(v')} p(u,v) \\ &\ge \frac{1}{|\phi^{-1}(u')|} \sum_{v'\in B_{d'/c}^{G'}(t')} \sum_{v\in\phi^{-1}(v')} p(u_0,v) \\ &= \frac{1}{|\phi^{-1}(u')|} \sum_{v\in\phi^{-1}(B_{d'/c}^{G'}(t'))} p(u_0,v). \end{aligned}$$

Therefore, from Inequality 9, we get

$$p'[u' \to B_{d'/c}^{G'}(t')] \ge \frac{1}{|\phi^{-1}(u')|} \sum_{v\in\bigcup_{t\in\phi^{-1}(t')} B_{d'/c}^{G}(t)} p(u_0,v).$$

If follows that

$$
\begin{aligned}
p'[u' \to B^{G'}_{d'/c}(t')] &\geq \frac{1}{|\phi^{-1}(u')|}\, p[u_0 \to \bigcup_{t\in\phi^{-1}(t')} B^{G}_{d'/c}(t)] \\
&\geq \frac{1}{|\phi^{-1}(u')|}\, p[u_0 \to B^{G}_{d'/c}(t_0)] \\
&\geq \frac{1}{|\phi^{-1}(u')|}\, p[u_0 \to B^{G}_{(\alpha d')/(\alpha c)}(t_0)] \\
&\geq \frac{1}{|\phi^{-1}(u')|}\, \frac{1}{f(\alpha d')} \\
&\geq 1/\Big(f(\alpha d') \cdot \max_{v'\in V'} |\phi^{-1}(v')|\Big) \\
&\geq 1/f'(d').
\end{aligned}
$$

This completes the proof of the Lemma.

Lemma 10 enables us to define new distributions on graphs.

Theorem 6. *Let $G = (V, E)$ be any graph such that there is a distance maintaining epimorphism ϕ from a k-dimensional torus of size $O(n)$ onto G. Further assume that $\max_{v\in V} |\phi^{-1}(v)| = O(1)$. Then there is a probabilistic mapping p on G such that greedy routing in (G, p) performs in $O(\frac{3^k}{k} \log^2 n)$ expected number of steps.*

Proof. From Lemma 9, (T^k_q, p_k) is an $(f, 2)$-Long Range Contact graph, where $f(d) = \Theta(\frac{3^k}{k} \log n)$. By application of Lemma 8, the probability p' defined in Equation 8 is such that (G, p') is an $(f', 2)$-Long Range Contact graph where $f'(d) \leq \beta \cdot f(\alpha d)$ for some constants α and β. That is $f'(d) = O(\frac{3^k}{k} \log n)$. It follows from Lemma 4 that greedy routing in (G, p') performs in $O(\frac{3^k}{k} \log n)$ expected number of steps.

6 Conclusion and Open Problems

In this paper we have studied the performance of greedy routing in the ring augmented with long range contacts chosen using r-harmonic distributions. We have also shown how to extend our results to arbitrary networks via appropriate mappings of multidimensional tori onto the network. Under certain conditions it is shown that greedy routing performs quite efficiently, i.e., $O(\log^2 n)$ expected number of steps. In particular, the ring augmented with the 1-harmonic distribution provides a simple model for the small world phenomenon.

Several interesting problems remain. For a general network, can we define probabilistic mappings for which greedy routing has better performance? Is our

$\Omega(\log^2 n)$ lower bound on the ring valid for all distance invariant mappings (not just the r-harmonic) on the n-node ring? Similar questions apply to any multidimensional torus. We note that in this paper we emphasized greedy routing, in the sense that nodes forward messages to their neighbors which are closer to the destination. An interesting open problem is to study the resulting tradeoff between memory (required at the nodes of the network) and type of routing being used.

References

1. W. Aiello, F. Chung, and L. Lu, "A Random Graph Model for Massive Graphs", in 32nd Annual ACM Symposium on Theory of Computing (STOC), 2000.
2. B. Bollobas, "Random graphs", Academic Press, 1985.
3. B. Bollobas and F. Chung, "The Diameter of a Cycle Plus a Random Matching", SIAM Journal on Discrete Mathematics 1:328-333, 1988.
4. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, "Graph Structure in the Web", in 9th International World Wide Web Conference, Amsterdam, May 15-19, 2000.
5. F. Chung and M. Garey, "Diameter Bounds for Altered Graphs", Journal of Graph Theory, 8:511-534, 1984.
6. M. Faloutsos, P. Faloutsos, C. Faloutsos, "On Power-Law Relationships of the Internet Topology", in ACM annual Technical Conference of the Special Interest Group on Data Communication (SIGCOMM), 1999.
7. B. Hayes, "Graph Theory in Practice: Part I", American Scientist, January-February, Vol. 88, No. 1, 2000.
8. B. Hayes, "Graph Theory in Practice: Part II", American Scientist, March-April, Vol. 88, No. 2, 2000.
9. R. Karp, "Probabilistic Recurrence Relations", in Proceedings of 23rd Annual Symposium on Theory of Computing, pages 190-197, 1991.
10. R. Karp, E. Upfal, and A. Wigderson, "Constructing a Perfect Matching is in Random NC", Combinatorica, 6:35-48, 1986.
11. J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective", in 32nd Annual ACM Symposium on Theory of Computing (STOC), 2000.
12. J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The Web as a Graph: Measurements, Models, and Methods", in 5th International Conference on Computing and Combinatorics, Tokyo, July 26-28, 1999.
13. D. Knuth, "The Art of Computer Programming: Sorting and Searching", Vol. 3, 2nd edition, Addison-Wesley, 1998.
14. R. Motwani and P. Raghavan, "Randomized Algorithms", Cambridge University Press, 1995.
15. S. Ross, "Stochastic Processes", 2nd ed., John Wiley and Sons, 1996.
16. D. Watts, "Small World: The Dynamics of Networks between Order and Randomness", Princeton University Press, 1999.
17. D. Watts and S. Strogatz, "Collective Dynamics of Small-World Networks", Nature 393:440-442, 1998.
18. G. K. Zipf, "Human Behavior and the Principle of Least Effort", Addison Wesley, Cambridge MA, 1949, Reprint New York, Hefner, 1972.