Distrib. Comput. (2006)
DOI 10.1007/s00446-005-0137-4

SPECIAL ISSUE PODC 04

Pierre Fraigniaud · Cyril Gavoille · Christophe Paul

Eclecticism shrinks even small worlds

Submitted: 28 October 2004 / Accepted: 4 July 2005 / Published online: 3 February 2006

Abstract We consider small world graphs as defined by Kleinberg (2000), i.e., graphs obtained from a d-dimensional mesh by adding links chosen at random according to the d-harmonic distribution. In these graphs, greedy routing performs in $O(\log^2 n)$ expected number of steps. We introduce *indirect-greedy* routing. We show that giving $O(\log^2 n)$ bits of topological awareness per node enables indirect-greedy routing to perform in $O(\log^{1+1/d} n)$ expected number of steps in d-dimensional augmented meshes. We also show that, independently of the amount of topological awareness given to the nodes, indirect-greedy routing performs in $\Omega(\log^{1+1/d} n)$ expected number of steps. In particular, augmenting the topological awareness above this optimum of $O(\log^2 n)$ bits would drastically decrease the performance of indirect-greedy routing.

Our model demonstrates that the efficiency of indirect-greedy routing is sensitive to the "world's dimension," in the sense that high dimensional worlds enjoy faster greedy routing than low dimensional ones. This could not be observed in Kleinberg's routing. In addition to bringing new light to Milgram's experiment, our protocol presents several desirable properties. In particular, it is totally *oblivious*, i.e., there is no header modification along the path from the source to the target, and the routing decision depends only on the target, and on information stored locally at each node.

A preliminary version of this paper appeared in the proceedings of the 23rd ACM Symposium on Principles of Distributed Computing (PODC), St. Johns, Newfoundland, Canada, July 25–28, 2004.

P. Fraigniaud (✉)
CNRS, LRI, Paris-Sud
E-mail: pierre@lri.fr

C. Gavoille
University of Bordeaux, LaBRI, Bordeaux
E-mail: gavoille@labri.fr

C. Paul
CNRS, LIRMM, Montpellier
E-mail: paul@lirmm.fr

Keywords Small world graphs · Routing · Milgram's experiment

1 Introduction

We consider small world graphs as defined by Kleinberg [7], i.e., graphs obtained from a d-dimensional mesh, for some fixed $d \geq 1$, by adding *long-range* links chosen at random according to the d-harmonic distribution, i.e., the probability that x chooses y as long-range contact is $h(x, y) = 1/(Z_x \cdot \text{dist}(x, y)^d)$ where dist() is the Manhattan distance in the mesh (i.e., the distance in the L_1 metric), and Z_x is a normalizing coefficient (cf. Sects. 5.1 and 5.2 for more details). This model aims at giving formal support to the "six degrees of separation" between individuals experienced by Milgram [14], and recently reproduced by Dodds, Muhamad, and Watts [5] (see also [1]). In a social context, professional as well as leisure occupation, citizenship, geography, ethnicity, and religiousness are all intrinsic dimensions of the human multi-dimensional world, playing different roles with possibly different impact degrees [6]. Each of these dimensions should be used as an independent criterion for routing in the social graph. In this context, one would thus expect that the more criteria used the more efficient the routing should be. Surprisingly however, Kleinberg's model does not reflect this fact, in the sense that greedy routing has the same performance whether the number of mesh dimensions considered is one, two, or more. Indeed, Kleinberg has shown that greedy routing in the n-node d-dimensional mesh augmented with long-range links chosen according to the d-harmonic distribution performs in $O(\log^2 n)$ expected number of steps, i.e., independently of d. (This bound is tight as it was shown in [3] that greedy routing performs in at least $\Omega(\log^2 n)$ expected number of steps, independently of d). Kleinberg has also shown that augmenting the d dimensional mesh with the r-harmonic distribution, $r \neq d$, results in poor performance, i.e., $\Omega(n^{\alpha_r})$ expected number of steps for some positive constant α_r. Furthermore, it is shown in [2] that, in the 1-dimensional mesh augmented according to *any* probabilistic distribution, greedy routing performs in $\Omega(\log^2 n/\log\log n)$ expected number of steps, and this lower bound is conjectured to hold in higher dimensions.

In light of the previous lower bounds combined with the fact that the expected diameter of augmented meshes is $O(\log n)$ (cf. [13]), one can conclude that the absence of the dimension parameter from the complexity of greedy routing in augmented meshes is a problem of the greedy routing specification, and not of the links distribution. We thus propose a new greedy protocol, called *indirect-greedy routing*, based on additional *topological awareness* given to the nodes, meaning that every node x is aware of the existence of a list A_x of long-range links. (Hence note that by additional topological awareness we do *not* mean adding more long-range contacts to nodes). Kleinberg's model can actually be seen as a special case of our model in which the awareness of every node is reduced to its own long-range contact, i.e., to $O(\log n)$ bits. At every step of indirect-greedy routing toward a target t, there are two phases. In the first phase, the current node x uses its awareness A_x to select an *intermediate destination* $\widehat{x}$, i.e., a node such that its long-range contact is close to t. In the second phase, x applies greedy routing toward $\widehat{x}$, and forwards to some neighbor y. In y, the same process is applied, a new intermediate destination $\widehat{y}$ is selected (thanks to y's awareness A_y), and greedy routing is applied toward $\widehat{y}$. And so on. The intermediate destination may or may not remain the same at every step of indirect-greedy routing. Once the routing reaches a node x for which $x = \widehat{x}$, greedy routing applies, and forwards to the neighbor of x that is closest to the target t. The same actions are repeated at every node until the routing eventually reaches the target.

1.1 Our results

We show that if every node is given a topological awareness of size $O(\log^2 n)$ bits or, more specifically, if every node is aware of the long-range contacts of its $O(\log n)$ closest nodes in the d-dimensional mesh, then indirect-greedy routing performs in $O(\log^{1+1/d} n)$ expected number of steps. Comparing the indirect-greedy protocol with other greedy protocols of the literature (cf. Table 1) demonstrates that, for an awareness of $\Theta(\log^2 n)$ bits, our protocol is the fastest. Indeed, this table displays the performances of variants of greedy routing in d-dimensional meshes augmented using d-harmonic distributions, with c long-range contacts per node.[1] For $d \geq 2$, indirect-greedy routing performs faster than any other greedy algorithm, for any value of c such that the amount of awareness is $\Theta(\log^2 n)$ bits, i.e., $c = \log n$ for Kleinberg's greedy routing and Decentralized algorithm, and $c = \sqrt{\log n}$ for NoN-greedy routing [12], defined in the percolation model of [4]. For $c = \sqrt{\log n}$, indirect-greedy performs in $O(\log^{1+1/2d} n)$ expected number of steps, that is faster than $O(\log^{3/2} n/\log\log n)$ steps for NoN-greedy. For $c = \log n$, indirect-greedy performs in $O(\log n)$ steps, as Kleinberg's greedy routing. The Decentralized algorithm [10] visits $O(\log^2 n/\log^2 c)$ nodes, and distributively discovers routes of expected length $O(\log n(\log\log n)^2/\log^2 c)$ links using headers of size $O(\log^2 n)$ bits.

Table 1 Performance of variants of greedy routing in d-dimensional meshes augmented using d-harmonic distributions, with c long-range contacts per node

Routing algorithm	Expected #steps	Amount of awareness (#bits)
Greedy [7]	$O(\frac{1}{c}\log^2 n)$	$O(c\log n)$
Greedy [3, 13]	$\Omega(\frac{1}{c}\log^2 n)$	$O(c\log n)$
Greedy [2]	$\Omega(\frac{1}{c}\log^2 n/\log\log n)$	$O(c\log n)$
NoN-greedy [12]	$O(\frac{1}{c\log c}\log^2 n)$	$O(c^2\log n)$
Decentralized algorithm [10]	$O(\frac{1}{\log^2 c}\log^2 n)$	$O(c\log n)$
Non oblivious [13]	$O(\frac{1}{c^{1/d}}\log^{1+1/d} n)$	$O(\log^2 n)$
Indirect-greedy [This paper]	$O(\frac{1}{c^{1/d}}\log^{1+1/d} n)$	$O(\log^2 n)$

The algorithm in [13] has the same performance as indirect-greedy. It is however not oblivious. In contrast, our protocol is totally oblivious, i.e., there is no header modification along the path from the source to the target, and the routing decision depends only on the target, and on information stored locally at each node. Obliviousness is a desirable property for a routing protocol because the decisions are taken locally at each node independently of the past, hence insuring better fault-tolerance. (This is of course true up to a reasonable tradeoff between performance and simplicity/fault-tolerance). Our interest in obliviousness is actually motivated by Milgram's experiment in which the intermediate persons performed in an oblivious manner.

Surprisingly, the positive impact of additional topological awareness reaches a certain limit, as far as indirect-greedy routing is concerned. Indeed, if the number c of long-range contacts of each node is constant, then indirect-greedy routing performs in $\Omega(\log^{1+1/d} n)$ expected number of steps, independently of the topological awareness given to the nodes, that is independently of the lists A_x, and of their sizes. Above a certain limit, augmenting the topological awareness of the nodes not only becomes useless, but also degrades the performance of indirect-greedy routing. Precisely, this limit is $\Theta(\log^2 n)$ bits of topological awareness per node (i.e., the awareness of $\Theta(\log n)$ long-range links).

These results prove that there is no trade-off between the amount of topological awareness given to the nodes and the performance of indirect-greedy routing, and demonstrate an intrinsic limitation of this strategy in augmented graphs.

[1] The coefficient $1/c^{1/d}$ in front of the performance of indirect-greedy routing comes from the fact that if every node has c long-range contacts, then to get an awareness of $O(\log n)$ long-range links, every node just needs to be aware of the long-range contacts of all nodes at distance $O((\frac{\log n}{c})^{1/d})$ from it. (The same holds for [13]).

In particular, if every node has a topological awareness of size n, i.e., is aware of all long-range contacts, then indirect-greedy routing would not perform better than Kleinberg's greedy routing, leading to an $\Omega(\log^2 n)$ expected number of steps.

More importantly, our study captures the trade-off that we expected: if social entities are living in a d-dimensional world, then giving additional topological awareness of $O(\log^2 n)$ bits to these entities enables indirect-greedy routing to perform in $O(\log^{1+1/d} n)$ expected number of steps. (Again, this is in contrast with Kleinberg's greedy routing which performs in $\Theta(\log^2 n)$ number of steps, independently of the world's dimension.) In particular, our model demonstrates a significant difference between routing using one criterion (i.e., in the 1-dimensional mesh), which performs in $O(\log^2 n)$ expected number of steps, and routing using two criteria (i.e., in the 2-dimensional mesh), which performs in $O(\log^{3/2} n)$ expected number of steps. (Note that in both cases, every node has only one long-range contact). The relative improvement decreases when the number of dimensions increases, which is consistent with what was observed by Killworth and Bernard [6].

To summarize, given a fixed number of "acquaintances" $2d + c$ per node in an augmented d-dimensional mesh with c long-range contacts per node, greedy routing performs in $O(\frac{1}{c}\log^2 n)$ expected number of steps, whereas indirect-greedy routing performs in $O(\frac{1}{c^{1/d}}\log^{1+1/d} n)$ expected number of steps. These results lead to the conclusion that the *variety* d of our relationships seems to have more impact on the distance between people than the *number* $2d + c$ of these relations, as far as Milgram's experiment is concerned. Our investigation is perhaps a first step toward the formalization of arguments in favor of the sociological evidence stating that eclecticism shrinks the world.

1.2 Organization

The paper is organized as follows. The next section precisely describes indirect-greedy routing, including the notion of topological awareness. Then, in Sect. 3, we give a necessary and sufficient condition for indirect-greedy routing to converge, and we compute an upper bound on the expected number of steps of indirect-greedy routing when nodes are aware of the long-range contacts of their $O(\log n)$ closest neighbors in the mesh. In Sect. 4, we compute a tight lower bound on the expected number of steps of indirect-greedy routing, independently of the amount of awareness given to the nodes. Finally, in Sect. 5, we give further motivations to our model, by revisiting it in the context of Milgram's experiment. In particular:

- we will expand on the surprising fact that giving more awareness does not necessarily improve performances, at least as far as Milgram's experiment is concerned, and
- we will motivate our interpretation of the dimensions of the mesh in terms of criteria based on which routing is performed.

The reader unaware of the details of Milgram's experiments and of Kleiberg's results can consult Sects. 5.1 and 5.2 respectively.

2 Topological awareness and indirect-greedy routing

We address the following question: is there some additional "topological awareness" that could be given to nodes so that greedy-like routing performs in less than $\Theta(\log^2 n)$ expected number of steps in the augmented d-dimensional mesh, at least for $d > 1$? By additional topological awareness we do *not* mean adding long-range contacts to nodes (in the remainder, there is only one long-range contact per node). Obviously, if nodes are given more than one long-range contact, then the performance of greedy routing can be improved, however to a limited extent only. For instance, with c long-range contacts per node, Kleinberg's greedy routing would perform in $\Omega(\frac{1}{c}\log^2 n)$ expected number of steps [3], which remains $\Omega(\log^2 n)$ for $c = O(1)$. We propose a model in which the $\log^2 n$ barrier can be overcome, with a constant number c (say, $c = 1$) of long-range contacts per node.

Kleinberg's "traditional" greedy routing fails to discover short routes for at least two reasons. First, the path toward the target may never pass exactly by nodes possessing long-range links leading close to the target, and, second, the path toward the target does not consider long-range links for which a small detour is necessary. To address these two problems, indirect-greedy routing considers more long-range links (thanks to the "awareness" of each node), and allows detours going away from the target when this enables to find a long-link leading close to the target.

2.1 Topological awareness

In our model, we assume that, in addition to the underlying graph, and to its long-range contact in the augmented graph, every node is aware (say, assuming that the nodes model social entities, thanks to some rumors) of some list of "acquaintances" between pairs of other nodes. This idea is formalized as follows.

Definition 1 The *topological awareness* of a node x is a list A_x of long-range links in the augmented graph.

In Kleinberg's model $A_x = \{e_x\}$ where e_x is the long-range link of x. We consider the case in which $A_x = \{e_1, e_2, \ldots, e_k\}$ with $e_x \in A_x$ and where, for every i, e_i is a long-range link not necessarily incident to x. Note that the degree of x remains unchanged compared to Kleinberg's model, i.e., the number of long-range contacts of every node x is the same in our model than in Kleinberg's model. For instance, in Fig. 1, node x has four neighbors in the 2-dimensional mesh: a, b, c, and d. It also has one long-range contact x'. The topological awareness of x is $A_x = \{(x, x'), (a, a'), (d, d'), (y, y')\}$. This means that node x is

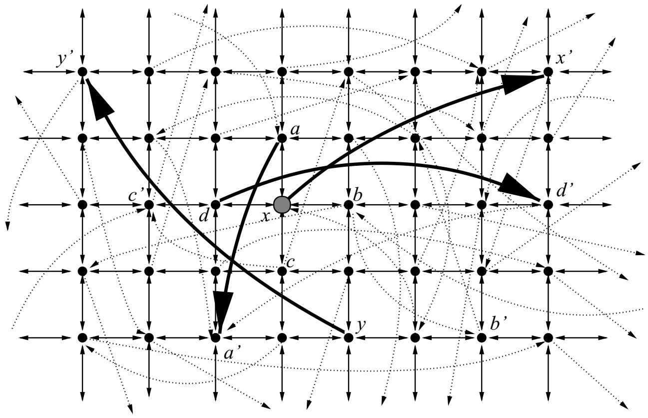

Fig. 1 Long-range links in the 2-dimensional mesh. The topological awareness of node x is composed of the four plain long-range links

aware that there is a long-range link from a to a', from d to d', and from y to y'. Note that x does not have any long-range link to either y or y', but is just aware that there is a long-range link from y to y'. On the other hand, x does not know the long-range contacts of b and c.

This gives rise to the following: how to benefit from the additional topological awareness given to the nodes to perform simple (i.e., greedy) routing in the augmented d-dimensional mesh? To answer this question, we define indirect-greedy routing.

2.2 Indirect-greedy routing

To define indirect-greedy routing, let us introduce some notation. For a directed edge $e = (u, v)$, we denote $u = \text{tail}(e)$, and $v = \text{head}(e)$. The $2d$ neighbors of the current node x in the d-dimensional mesh are denoted by $w_1, \dots, w_{2d}$, and the long-range contact of x is denoted w_0. Finally, let t be the target node, $t \neq x$. The function $\text{dist}(u, v)$ is the Manhattan distance between nodes u and v in the mesh.

Phase 1. Among all edges in $\{(x, w_1), \dots, (x, w_{2d})\} \cup A_x$, x selects an edge e such that $\text{head}(e)$ is closest to the target t in the mesh (according to the Manhattan distance). If there are several such edges e, x selects one such that $\text{tail}(e)$ is the closest to x in the mesh. Possible remaining ties are broken arbitrarily. If $\text{tail}(e) = x$ or if $\text{dist}(x, \text{tail}(e)) \geq \text{dist}(x, t)$, then set $\widehat{x} = t$, otherwise set $\widehat{x} = \text{tail}(e)$.

Phase 2. Node x selects, among its $2d + 1$ neighbors $w_0, w_1, \dots, w_{2d}$, the one that is the closest to $\widehat{x}$, and forwards to that neighbor.

In the following, the node $\widehat{x}$ selected during Phase 1 is called the *intermediate destination* for x. Note that we set $\widehat{x} = t$ if $\text{dist}(x, \text{tail}(e)) \geq \text{dist}(x, t)$. We could replace this latter condition by $\text{dist}(x, \text{tail}(e)) + \text{dist}(\text{head}(e), t) \geq \text{dist}(x, t)$ but this would not improve the performance of indirect-greedy routing. In fact, the condition $\text{dist}(x, \text{tail}(e)) \geq \text{dist}(x, t)$ is somewhat more consistent with the fact that routing from x to $\text{tail}(e)$ is performed by traditional greedy routing, whereas routing from $\text{head}(e)$ to t is performed by indirect-greedy routing.

Remark 1 Indirect-greedy routing is totally *oblivious*, i.e., there is no header modification along the path from the source to the target, and the routing decision depends only on the target, and on information stored locally at each node. That is, in contrast with non-oblivious protocols (see, e.g., [10, 13]), the computation of the intermediate destination is performed at every node involved in the routing process. In particular, if x is the current node, and if w_i is the neighbor of x to which it forwarded during Phase 2, then the intermediate destination $\widehat{w_i}$ for w_i may be different from the intermediate destination $\widehat{x}$ for x.

Let us take two extreme examples to illustrate the behavior of indirect-greedy routing:

(a) If the topological awareness of every node is reduced to its own long-range contact, then the edge e selected during Phase 1 is necessarily incident to the current node x, i.e., $\text{tail}(e) = x$ and thus $\widehat{x} = x$. Thus, during Phase 2, x forwards to $\text{head}(e)$. Therefore, indirect-greedy routing reduces to greedy routing in this case.

(b) If the topological awareness of every node is the whole graph, i.e., if every node is aware of all long-range contacts (a very unrealistic hypothesis), then let $e_1, \dots, e_k$ be the $k \geq 1$ long-range links such that, for every i, $1 \leq i \leq k$, $\text{dist}(\text{head}(e_i), t)$ is minimum among all long-range links. At every node involved in routing, the intermediate destination is $y_i = \text{tail}(e_i)$ for some i. (The intermediate destination may change if the current node is at equal distance from two intermediate destinations.) For a source s, let $m = \min_{1 \leq i \leq k} \text{dist}(s, y_i)$. Most of the process actually consists in traveling distance m in the mesh, from s to one of the y_i's, using Kleinberg's greedy routing. Hence, indirect-greedy routing also reduces to greedy routing in this case. Obviously, in this example, a faster routing would be obtained by computing a shortest path from the source to the target in the augmented mesh, but this would be a quite unrealistic model as far as social networks are concerned (see Sect. 5).

Remark 2 As opposed to Kleinberg's greedy routing, the Manhattan distance to the target is not strictly decreasing at each step of indirect-greedy routing. Indeed, an intermediate destination can be farther from the target than the current node, and thus going to this intermediate destination may result in increasing the Manhattan distance to the target. We will see in the next section that, under a weak condition, this phenomenon has little impact on the expected performance of indirect-greedy routing because it is counter balanced by the fact that the intermediate destination has a long-range contact leading close to the target.

3 Performance of indirect greedy routing

In this section, we give a sufficient condition for indirect-greedy routing to converge, i.e., to always route correctly for any setting of the long-range links. We later prove that if every node is aware of the long-range contacts of its $O(\log n)$ closest nodes in the d-dimensional mesh, then indirect-greedy routing performs in $O(\log^{1+1/d} n)$ expected number of steps.

Let A_x be the topological awareness given to every node x. The set $\{A_x \mid x \in V\}$ is called the *system of awareness* of the augmented mesh $H = (V, E)$. Now, for every node x, let us denote by N_x the set of x's neighbors in H (thus including x's long-range contact). For every link e with $\mathrm{tail}(e) \neq x$, we then define

$$N_x(e) = \{y \in N_x \mid \mathrm{dist}(y, \mathrm{tail}(e)) \\ \leq \mathrm{dist}(z, \mathrm{tail}(e)) \text{ for every } z \in N_x\}.$$

$N_x(e)$ is the set of neighbors of x closest to $\mathrm{tail}(e)$, i.e., those nodes to which x forwards when applying Kleinberg's greedy routing toward $\mathrm{tail}(e)$. Our condition for convergence of indirect-greedy routing is based on the following definition.

Definition 2 A system of awareness $\{A_x \mid x \in V\}$ is *monotone* if, for every x, and for every $e \in A_x \setminus \{e_x\}$ where e_x is the long-range link of x, we have $e \in A_y$ for every $y \in N_x(e)$.

Remark 3 If all sets $S_x = \{\mathrm{tail}(e) \mid e \in A_x\}$ have the same shape S for all nodes x, in the sense that $S = S_{x_0} = \{\mathrm{tail}(e) \mid e \in A_{x_0}\}$ for some fixed node x_0, and S_x is obtained by translating S_{x_0} along the vector $x_0 \to x$, then monotonicity is equivalent to the fact that every shortest path in the mesh from x_0 to any node in S is included in S. "Being monotone" is more general than "having the same shape" because it does not require the structure of the topological awareness to be the same for all nodes.

Lemma 1 *If the system of awareness is monotone then indirect-greedy routing converges.*

Proof Let s be the current node, and let t be the target. Let u be the current intermediate destination, and let v be the long-range contact of u. We define the *potential* of s with respect to destination t as:

$$\phi_t(s) = \mathrm{dist}(s, u) + n \cdot \mathrm{dist}(v, t)$$

From s, the route goes to some node s' on a shortest path from s to u. If the intermediate destination at s' is the same as the one at s, then $\phi_t(s') \leq \phi_t(s) - 1$. If the intermediate destination changes, then let u' be the new intermediate destination, and let v' be its long-range contact. Since the system of awareness is monotone, we have $(u, v) \in A_{s'}$. Therefore, $\mathrm{dist}(v', t) \leq \mathrm{dist}(v, t)$.

– If $\mathrm{dist}(v', t) < \mathrm{dist}(v, t)$ then

$$\phi_t(s') = \mathrm{dist}(s', u') + n \cdot \mathrm{dist}(v', t) \leq (n - 1) \\ + n \cdot (\mathrm{dist}(v, t) - 1) = \mathrm{dist}(v, t) - 1 < \phi_t(s).$$

– If $\mathrm{dist}(v', t) = \mathrm{dist}(v, t)$ then Phase 1 of indirect-greedy routing specifies that since s' chooses u', u' is at least as close to s' as u. Therefore,

$$\phi_t(s') = \mathrm{dist}(s', u') + n \cdot \mathrm{dist}(v', t) \leq \mathrm{dist}(s', u) \\ + n \cdot \mathrm{dist}(v, t) \leq \phi_t(s) - 1.$$

Therefore, in all cases, the potential is strictly decreasing after each step of indirect-greedy routing. Thus indirect-greedy routing eventually reaches the target. □

Let d be any fixed positive integer (the dimension of the mesh).

Theorem 1 *In the d-dimensional mesh augmented with one long-range link per node chosen according to the d-harmonic distribution, if every node is aware of the long-range contacts of all nodes at Manhattan distance $\leq \log^{1/d} n$ in the mesh, then indirect-greedy routing performs in $O(\log^{1+1/d} n)$ expected number of steps.*

The remainder of this section is dedicated to the proof of Theorem 1. Notice that the system of awareness induced by balls of the same radius is monotone (cf. Remark 3). Therefore, thanks to Lemma 1, indirect-greedy routing converges. We compute the expected number of steps to reach any target from any source. Let x be the current node, and t be the target node. First, we consider the case where x is far from the target t in the mesh, that is $\mathrm{dist}(x, t) > \lambda \cdot \log^{1/d} n$ for a sufficiently large constant λ that will be determined later.

Remark 4 The general argument of the proof consists in computing the expected number of steps for reducing the distance to the target by a factor at least 2, and to reapply iteratively this argument every time the distance to the target has been reduced by a factor at least 2. It is crucial to note that the decision taken by the algorithm at the current node is independent from the history of the algorithm to reach this node. Moreover, the harmonic distribution is such that finding a long-range link halving the distance m to the target is independent from m. Therefore, the expected number of steps to decrease the distance to the target by a factor at least 2, conditioned to the fact that the current node is x, is in fact independent from x. This is why we can sum up the conditional expectations to get the total expected number of steps for reaching the target. On the other hand, the number of fresh long-range links in the awareness of the current node x depends on how x was reached. For instance, if x is reached via a link of the underlying mesh, then there are less fresh long-range links in the awareness of x than if x would have been reached via a long link. This type of dependency is taken into account in our analysis of indirect-greedy routing.

Lemma 2 *Starting at a node x at Manhattan distance $m > \lambda \log^{1/d} n$ from the target, $\lambda > 1$, indirect-greedy routing reaches a node at Manhattan distance $\leq \lambda \log^{1/d} n$ from the target in at most $O(\log^{1+1/d} n)$ expected number of steps.*

Proof Let $m = \mathrm{dist}(x, t) > \lambda_1 \log^{1/d} n$ for some $\lambda_1 > 1$, and let us compute the expected number of steps required by indirect-greedy routing for reaching a node x' at Manhattan distance $\leq m/2$ from t. Let

$B = \{u \mid \mathrm{dist}(u, t) \leq m/2\}$.

For any node u, let

$V(u) = \{v \mid \mathrm{dist}(u, v) \leq \log^{1/d} n\}$.

$V(u)$ corresponds to the set of all possible tails of the long-range links known by u. We define the subset $V'(u)$ of $V(u)$ as follows:

$V'(u) = \{v \in V(u) \mid \mathrm{dist}(v, t) \leq m\}$.

For two node sets X and Y, let $\Pr(X \to Y)$ be the probability that at least one node in X has its long-range contact in Y. *Claim* 1 $\Pr(V(x) \to B)$ is asymptotically at least some constant $\beta > 0$ (depending only on the dimension d of the mesh).

Proof We have $\Pr(V(x) \to B) \geq \Pr(V'(x) \to B)$. Note that $|V'(x)| \geq \frac{1}{2^d}|V(x)|$ as $t \notin V(x)$ since $\lambda_1 > 1$. Therefore, $|V'(x)| = \Theta(\log n)$. Let $a > 0$ and $n_1 > 0$ be such that $|V'(x)| \geq a \log n$ for any $n \geq n_1$. For any node u, let $\mathcal{E}_u$ be the event "u has its long-range contact in B." We have $\Pr(V'(x) \to B) = 1 - \Pi_{u \in V'(x)}(1 - \Pr(\mathcal{E}_u))$. Let $p = \Pr(\mathcal{E}_x)$. Since $\Pr(\mathcal{E}_x) \leq \Pr(\mathcal{E}_u)$ for any $u \in V'(x)$, we get $\Pr(V'(x) \to B) \geq 1 - (1-p)^{|V'(x)|}$. Now, we have

$$p = \sum_{u \in B} h(x, u) = \frac{1}{Z_x} \sum_{u \in B} 1/\mathrm{dist}(x, u)^d$$

where $Z_x = \sum_{w \neq x} 1/\mathrm{dist}(x, w)^d$.

On one hand $Z_x = \sum_{i \geq 1} |S_i|/i^d$ where S_i is the set of nodes at Manhattan distance exactly i from x. We have $|S_i| = O(i^{d-1})$ for any i. Thus $Z_x = O(\log n)$.

On the other hand,

$$\sum_{u \in B} 1/\mathrm{dist}(x, u)^d \geq |B|/(3m/2)^d \geq \Omega(m^d)/(3m/2)^d \geq \Omega(1).$$

Therefore p is at least $\Omega(1/\log n)$. Let $b > 0$ and $n_2 \geq n_1$ be such that $p \geq b/\log n$ for any $n \geq n_2$. We have

$$(1-p)^{|V'(x)|} \leq (1 - b/\log n)^{a \log n}$$

for any $n \geq n_2$. Since $(1 - b/z)^{az} \approx e^{-ab}$ for large z, we get that $1 - (1-p)^{|V'(x)|} \geq f(n)$ where $f(n) \approx 1 - e^{-ab}$ for large n. Let $0 < \beta < 1 - e^{-ab}$. There exists $n_3 \geq n_2$ such that $\Pr(V(x) \to B) \geq \beta$ for any $n \geq n_3$. □

Let $x_1 \in V(x)$ be the intermediate destination selected by $x = x_0$ during phase 1 of indirect-greedy routing. In phase 2, the route goes from x_0 to x_1 according to Kleinberg's greedy routing. However, on the way to x_1, new long-range links are discovered, and possibly a new node x_2 whose long-range contact is a node closer to t than the long-range contact of x_1 is discovered (see Fig. 2a). If such a new node x_2 is discovered (on Fig. 2a, x_2 is discovered at node

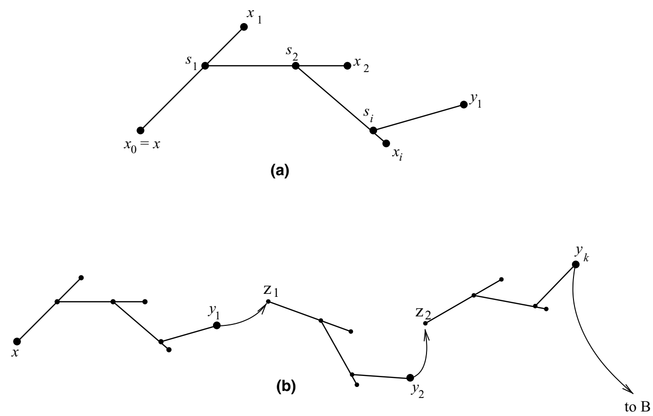

Fig. 2 Intermediate destinations before jumping into B

s_1), x_1 is discarded, and the new intermediate destination becomes x_2. In this case, x_2 is discovered after performing at most $\log^{1/d} n$ steps of routing toward x_1 in the worst-case. Indeed, every node is aware of the long-range contacts in a ball of radius $\log^{1/d} n$. Again, on the way to x_2, possibly a new node x_3 whose long-range contact leads to a node closer to t than the long-range contact of x_2 is discovered at, say s_2, and routing switches to x_3. This phenomenon may occur many times, constructing a sequence $x_0, x_1, x_2, \ldots$ of intermediate destinations, with $x_0 = x$ (see Fig. 2a). More formally, we define the following:

Definition 3 An intermediate destination v is *good* if (1) the path constructed by indirect-greedy routing reaches v, (2) the intermediate destination $\widehat{v}$ for v satisfies $\widehat{v} = t$, and (3) the long-link of v is used by indirect-greedy routing at v. The intermediate destination v is *bad* otherwise.

In the sequence $(x_i)_{i \geq 0}$ of intermediate destinations defined above, since x_{i+1} is the intermediate destination for s_i, the Manhattan distance between every two consecutive intermediate destinations x_i and x_{i+1} satisfies

$$\mathrm{dist}(x_i, x_{i+1}) \leq 2 \log^{1/d} n \quad \text{for every } i \geq 0. \tag{1}$$

Claim 2 There exists a constant $\lambda_2 > 0$ such that starting from x at distance $> \lambda_2 \cdot \log^{1/d} n$ from the target, the expected number of bad intermediate destinations x_i's is at most a constant γ depending only on the dimension of the mesh, or routing reaches a node at distance $\leq \lambda_2 \log^{1/d} n$ from the target.

Proof Let s_i be the node where indirect-greedy routing switches from x_i to x_{i+1}, with possibly $s_i = x_i$ if x_i is bad. Let C_i be the set of all tails of the new long-range links discovered while going from s_i to x_{i+1}, and let $a_0, a_1, a_2, \ldots, a_l$ be the path from s_i toward x_{i+1} generated by Kleinberg's greedy routing, where $a_0 = s_i$ and $a_l = s_{i+1}$. This path is the one generated by Phase 2 of indirect-greedy routing. By definition, we have $C_i = (\cup_{j=1}^{l} V(a_j)) \setminus V(s_i)$. The path $a_0, a_1, a_2, \ldots, a_l$ is included in the ball centered at

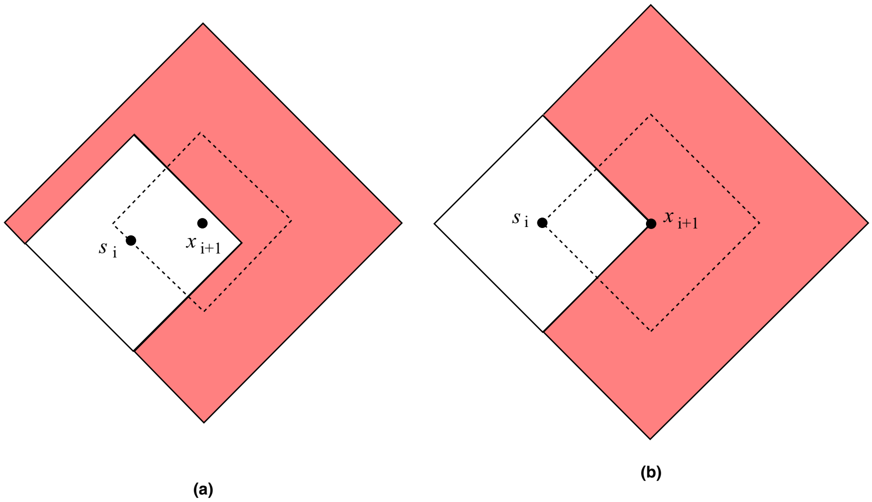

Fig. 3 The set C_i is included in the Grey area, and in the 2-dimensional mesh $|C_i| \leq 3 \log n$

x_{i+1} and of radius $\text{dist}(s_i, x_{i+1})$ because Kleinberg's greedy routing always decreases the distance to the target (here the target is x_{i+1}). This inclusion holds even if the path contains long-range links (a_j, a_{j+1}). Hence $|C_i| \leq (2^d - 1) \log n$ (see Fig. 3). Let $\mathcal{E}$ be the event "there is a long-link e such that $\text{tail}(e) \in C_i$ and $\text{head}(e)$ is closer to t than the long-range contact of x_{i+1}." One cannot directly state that

$$\Pr(\mathcal{E}) \leq |C_i| \ / \ (|V(s_i)| + |C_i|)$$

because the probability of having a long-range contact close to t changes with the distance to the target. Nevertheless, since C_i is included in the ball centered at x_{i+1} and of radius $\log^{1/d} n$, the maximum distance between two nodes in C_i is only a small fraction of m if $m = \Omega(\log^{1/d} n)$, and thus this probability does not change much along the path $a_0, a_1, a_2, \ldots, a_l$. Therefore, for any $\epsilon > 0$, there exists $\lambda_\epsilon > 0$ such that

$$\Pr(\mathcal{E}) \leq (1 + \epsilon) \cdot |C_i| \ / \ (|V(s_i)| + |C_i|)$$

for every i such that $\text{dist}(s_i, t) > \lambda_\epsilon \cdot \log^{1/d} n$. Therefore, since $|V(s_i)| = \log n$ and $|C_i| \leq (2^d - 1) \log n$, we get that for any $\epsilon > 0$, there exists $\lambda_\epsilon > 0$ such that, if $\text{dist}(s_i, t) > \lambda_\epsilon \cdot \log^{1/d} n$, then the probability that, while going from s_i to x_{i+1}, a better intermediate destination is discovered is at most

$$p_\epsilon \leq (1 + \epsilon) \cdot \frac{1}{1 + \frac{1}{2^d - 1}}.$$

Let $\epsilon < 1/(2^d - 1)$ so that $p_\epsilon < 1$. The expected number of successes of trials which succeed each with probability at most p_ϵ is constant γ (i.e., depending only on d and ϵ but not on n). Therefore, by setting $\lambda_2 = \max\{\lambda_1, \ \lambda_\epsilon\}$, we get that starting from x at distance $\lambda_2 \cdot \log^{1/d} n$ from the target, the expected number of bad intermediate destinations x_i's is at most γ (or routing reaches a node a node at distance $\leq \lambda_2 \log^{1/d} n$ from the target). □

From Eq. (1) and Claim 2, after at most $2\gamma \log^{1/d} n$ expected number of steps, one eventually reaches a good intermediate destination y_1 (see Fig. 2a). Since y_1 is good, the long-link is used, leading to some node z_1 (see Fig. 2b). If $z_1 \in B$ then we are done. Otherwise, starting from z_1, indirect-greedy routing eventually reaches another good intermediate destination y_2. Since y_2 is good, the long-link is used, leading to some node z_2. And so on. We construct in this way the sequence $z_1, z_2, \ldots$ of the long-range contacts of the good intermediate destinations $y_1, y_2, \ldots$ that are reached during indirect-greedy routing (see Fig. 2b). Let $\mathcal{E}_i$ be the event "at least one node in $V(z_i)$ has its long-range contact in B."

Claim 3 There exists a constant $\lambda_3 > 0$ such that starting from x at distance $> \lambda_3 \cdot \log^{1/d} n$ from the target, the expected number of good intermediate destinations y_i that are visited before the event $\mathcal{E}_i$ holds is constant (i.e., depending only on the dimension of the mesh), or routing reaches a node a node at distance $\leq \lambda_3 \log^{1/d} n$ from the target.

Proof We observe the two following points.

(1) By construction, the long-range contact z_1 of y_1 is not farther to t than any other node that was visited by indirect-greedy routing before. Actually, z_1 is not farther to t than any end-point of long-range links in the awareness of nodes visited before.
(2) The expected distance between y_1 and z_1 is at least $\log^{1/d} n$. Indeed, the expected distance between y_1 and z_1 is

$$\sum_{i \geq 1} i \Pr(\text{dist}(y_1, z_1) = i)$$

$$\geq \sum_{i \geq \mathrm{dist}(y_1,t)/2} i \Pr(\mathrm{dist}(y_1, z_1) = i)$$
$$\geq \frac{\mathrm{dist}(y_1, t)}{2} \Pr(\mathrm{dist}(y_1, z_1) \geq \mathrm{dist}(y_1, t)/2)$$
$$\geq \frac{\mathrm{dist}(y_1, t)}{2} \Pr(\mathrm{dist}(z_1, t) \leq \mathrm{dist}(y_1, t)/2).$$

By the same analysis as Claim 3, $\Pr(\mathrm{dist}(z_1, t) \leq \mathrm{dist}(y_1, t)/2) \geq \beta$. Thus the expected distance between y_1 and z_1 is at least β $\mathrm{dist}(y_1, t)/2$. Setting $\lambda_3 = 4/\beta$, and assuming that $\mathrm{dist}(y_1, t) > \lambda_3 \log^{1/d} n$, the expected distance between y_1 and z_1 is such that the two balls centred at y_1 and z_1, and of radius $\log^{1/d} n$ do not intersect.

Combining these two observations, we get that the long-links whose tails are in $V'(z_1)$ have never been considered so far by indirect-greedy routing. Therefore, Claim 1 can be applied to z_1 and we get that $\Pr(V'(z_1) \to B) \geq \beta$. Thus we can repeat the same analysis for z_1 as we did for x, yielding that after at most $\gamma \log^{1/d} n$ expected number steps indirect-greedy routing reaches y_2, and from there z_2. By repeating the same analysis at every z_i, we get that

$$\Pr(V'(z_i) \to B) \geq \beta$$

for every $i \geq 1$. Therefore the number of good intermediate destinations visited before $\mathcal{E}_i$ holds is $1/\beta$. □

Set $\lambda = \max\{\lambda_1, \lambda_2, \lambda_3\}$. From Claim 3, starting from x at Manhattan distance $m > \lambda \log^{1/d} n$ from t, it takes at most $\frac{1}{\beta} \log^{1/d} n$ expected number of steps to reach a node in B, or a node at distance $\leq \lambda \log^{1/d} n$ from the target. In other words, decreasing the Manhattan distance by a factor of 2 takes at most $O(\log^{1/d} n)$ expected number of steps. Therefore, from any source at Manhattan distance $m > \lambda \cdot \log^{1/d} n$ from t, it takes $O((\log m) \cdot (\log^{1/d} n)) = O(\log^{1+1/d} n)$ expected number of steps to reach a node at Manhattan distance $\leq \lambda \cdot \log^{1/d} n$ from t. This completes the proof of Lemma 2. □

It remains to consider the case where the current node x is close to the target t, i.e., $m = \mathrm{dist}(x, t) \leq \lambda \cdot \log^{1/d} n$ for some constant λ.

Lemma 3 *Starting at a node x at distance $\leq \lambda \log^{1/d} n$ from the target, indirect-greedy routing reaches the target in at most $O(\log^{1+1/d} n)$ number of steps on expectation.*

Proof Let u be the current intermediate destination (i.e., the one selected by x), and let v be the long-range contact of u. We proceed similarly as in the proof of Lemma 1, and define the *potential* of x as

$$\phi_t(x) = \mathrm{dist}(x, u) + \mathrm{dist}(v, t) \cdot (1 + \log^{1/d} n).$$

From x, the route goes to some node x' on a path from x to u. If the intermediate destination at x' is the same as the one at x, then $\phi_t(x') \leq \phi_t(x) - 1$. If the intermediate destination changes, then let u' be the new intermediate destination, and let v' be its long-range contact. Since balls form a monotone system of awareness, we have $(u, v) \in A_{x'}$. Therefore $\mathrm{dist}(v', t) \leq \mathrm{dist}(v, t)$.

If $\mathrm{dist}(v', t) < \mathrm{dist}(v, t)$ then

$$\phi_t(x') = \mathrm{dist}(x', u') + \mathrm{dist}(v', t) \cdot (1 + \log^{1/d} n) \leq \log^{1/d} n + (\mathrm{dist}(v, t) - 1) \cdot (1 + \log^{1/d} n) < \phi_t(x).$$

If $\mathrm{dist}(v', t) = \mathrm{dist}(v, t)$ then Phase 1 of indirect-greedy routing specifies that since x' chooses u', $\mathrm{dist}(x', u') \leq \mathrm{dist}(x', u)$. Therefore,

$$\phi_t(x') = \mathrm{dist}(x', u') + \mathrm{dist}(v', t) \cdot (1 + \log^{1/d} n) \leq \mathrm{dist}(x', u) + \mathrm{dist}(v, t) \cdot (1 + \log^{1/d} n) \leq \phi_t(x) - 1.$$

Therefore, in all cases, the potential is strictly decreasing after each step of indirect-greedy routing. The potential of a node x at distance m from t is at most $\log^{1/d} n + m \cdot (1 + \log^{1/d} n)$. Thus, a node at distance at most $\lambda \cdot \log^{1/d} n$ from t has potential $\leq O(\log^{2/d} n) \leq O(\log^{1+1/d} n)$. Therefore, the target is reached after at most $O(\log^{1+1/d} n)$ steps, which completes the proof of Lemma 3. □

Theorem 1 directly follows from Lemmas 2 and 3.

4 Lower bounds for indirect-greedy routing

Theorem 1 shows that indirect-greedy routing with topological awareness of the $O(\log n)$ closest neighbors in the mesh routes faster than greedy routing. Hereafter, in Theorem 2, we show that the expected number of steps of indirect-greedy routing is $\Omega(\log^{1+1/d} n)$ for any amount of awareness. More interestingly, Theorem 2 demonstrates that $\log n$ links is an optimum for the awareness. If the amount $v(n)$ of awareness is smaller than $\log n$ links then the expected number of steps is a decreasing function of the awareness (see Fig. 4). However, after the threshold of $v(n) = \log n$, the expected number of steps is an increasing function of the amount of awareness (see Fig. 4).

Theorem 2 *In the d-dimensional mesh augmented with one long-range link per node chosen according to the d-harmonic distribution, for any $1 \leq v(n) \leq n$, if every node is aware of the long-range contacts of its $v(n)$ closest nodes in the mesh, then indirect-greedy routing performs in $\Omega(\log^{1+1/d} n)$ expected number of steps. More precisely, if $d > 1$, and $v(n) = \log^\alpha n$ for some $\alpha \geq 0$, then a performance of $O(\log^{1+1/d} n)$ expected number of steps cannot be reached if $\alpha \neq 1$.*

To prove Theorem 2, we first prove the following:

Lemma 4 *Reaching a node at distance m using (Kleinberg's) greedy routing requires:*

- *at least m expected number of steps if $m = \log^\alpha n$ for some $\alpha < 1$;*
- *at least $\log m \log n$ expected number of steps if $m = \log^\alpha n$ for some $\alpha > 1$:*

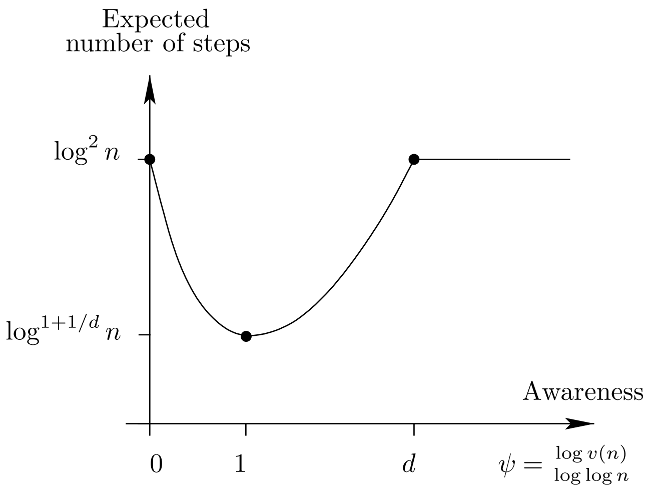

Fig. 4 The expected number of steps vs. the awareness $v(n) = (\log n)^\alpha$. The expected number of steps is $\Omega((\log n)^{2+\alpha/d-\alpha})$ if $\alpha < 1$ (by Lemma 5), and $\Omega((\log n)^{1+\alpha/d-o(1)})$ if $1 \le \alpha < d$ (by Lemma 6). For $\alpha > d$, the expected number of steps is $\Theta(\log^2 n)$ (by Lemma 6)

Proof Let s be the source node, t be the target node, and $m = \text{dist}(s,t)$. Assume first that $m = \log^\alpha n$ for some $\alpha < 1$. Let $r \le m$, and let B be the ball of radius $2r$ centered at x. By definition of the d-harmonic distribution, we have

$$\Pr(x \to B) = \sum_{b\in B} \Pr(x \to b) \approx \frac{1}{\log n} \sum_{b\in B} \frac{1}{\text{dist}(x,b)^d} \approx \frac{1}{\log n} \sum_{i=1}^{2r} \frac{|S_i|}{i^d}$$

where S_i is the set of nodes at distance exactly i from x. Thus

$$\Pr(x \to B) \approx \frac{1}{\log n} \sum_{i=1}^{2r} \frac{i^{d-1}}{i^d} \approx \frac{\log r}{\log n} \le \frac{\log m}{\log n}.$$

Therefore, while going from s to t using Kleinberg's greedy routing, the expected number of discovered long range links that connect to nodes closer to t than the current node is $O(\frac{m \log m}{\log n})$. Since $m = \log^\alpha n$ with $\alpha < 1$, this number goes to zero as n goes to infinity, and thus no long range link is used between s and t, resulting in m expected routing steps.

Assume now that $m = \log^\alpha n$ for some $\alpha > 1$. Then, for $i \ge 0$, let

$$B_i = \{u \mid \text{dist}(u,t) \le m/2^i\}.$$

For any node $x \in B_0 \setminus B_1$, and any $i \ge 1$, we have (ignoring the constant depending on d only):

$$\Pr(x \to B_i) = \sum_{b\in B_i} \Pr(x \to b) \approx \frac{1}{\log n} \sum_{b\in B_i} \frac{1}{\text{dist}(x,b)^d} \le \frac{1}{\log n} \cdot \frac{|B_i|}{m^d(1-\frac{1}{2^i})^d}$$

The latter inequality follows from the fact the d-harmonic distribution decreases with the distance. Since $|B_i| \approx (m/2^i)^d$, we get that, up to a constant,

$$\Pr(x \to B_i) \le \frac{1}{\log n} \cdot \frac{(m/2^i)^d}{m^d(1-\frac{1}{2^i})^d} = \frac{1}{(2^i-1)^d \log n}.$$

Therefore, while traveling in $B_0 \backslash B_1$, the probability to visit a node whose long-range contact is in B_1 is only $O(1/\log n)$. Thus while traveling in $B_0 \backslash B_1$, the expected number of steps before visiting a node whose long-range contact is in B_i is $\Omega(\log n)$ for any $i \ge 1$. Since $m \gg \log n$, such a node will eventually be visited. However, since $\Pr(x \to B_i)$ decreases exponentially with i, the expected value of the index i such that greedy routing reaches a node in B_i while entering B_1 for the first time is a constant. As a consequence, an expected number of $\Omega(\log n)$ steps are required to decrease the distance to the target by at most a constant expected factor. Therefore, starting from a node at distance m from the target, at least $\Omega(\log m \log n)$ expected number of steps are required. □

To prove Theorem 2, we consider separately the cases $v(n) \ll \log n$, and $v(n) \gg \log n$. Intuitively, if every node is aware of the long-range contacts of its $v(n) \ll \log n$ closest neighbors, then reaching an intermediate destination is fast, but a large number of intermediate destinations must be visited before expecting reaching a node whose long range-contact leads close to the target. In fact, we show the following:

Lemma 5 *If $v(n) = \log^\alpha n$, for some $0 \le \alpha < 1$, then the expected number of steps to reach the target is at least*

$$\Omega(((\log n)/v(n))^{1-1/d} \cdot \log^{1+1/d} n).$$

Proof Let $m = \text{dist}(x,t)$ be the distance between the current node x and the target t. We use the same notations as in the proof of Theorem 1. Let $B = \{u \mid \text{dist}(u,t) \le m/2\}$, and, for any node u, let $V(u) = \{v \mid \text{dist}(u,v) \le v(n)^{1/d}\}$. From the definition of the d-harmonic distribution, an expected number of $\Omega(\log n)$ long-range contacts must be considered before finding one that leads to a node in B. Hence, we compute the expected number of steps required to learn about $\Omega(\log n)$ long-range contacts. Starting from x, the route visits a sequence $y_1, y_2, \ldots$ of good intermediate destinations (see Fig. 2).

Claim 4 The expected number of steps required to go from y_j to y_{j+1} is $\Theta(v(n)^{1/d})$.

Proof Let $x_0, x_1, \ldots, x_\ell$ be the sequence of bad intermediate destinations that are considered while traveling to y_{j+1} starting from y_j, until the route eventually reaches the good intermediate destination y_{j+1}. I.e, $x_0 = y_j$ and $x_\ell = y_{j+1}$. Let $r = \text{dist}(x_0, x_1)$ (note that $r \le v(n)^{1/d}$ since $x_1 \in V(x_0)$).

Since the expected Manhattan distance $\bar{r}$ between x_0 and x_1 is $\Omega(v(n)^{1/d})$, and $v(n) = \log^\alpha n$, $\alpha < 1$, it follows from

Lemma 4 that the expected number of steps required to go from x_0 to x_1 is $\Omega(v(n)^{1/d})$. Actually, the routing does not reach x_1 if a new intermediate destination x_2 is discovered. However, a constant portion of the path from x_0 to x_1 must be traversed before expecting to discover a new intermediate destination. Indeed, to discover the same order of magnitude of new long-range links as $v(n)$, one must go at expected distance $\Omega(v(n)^{1/d})$ from x_0. Therefore, the portion of the path from x_0 to x_1 that is traversed before possibly switching toward x_2 requires $\Omega(v(n)^{1/d})$ expected number of steps. Hence, the expected number of steps required to go from y_j to y_{j+1} is $\Omega(v(n)^{1/d})$.

On the other hand, by Claim 2, the expected number of steps required to go from y_j to y_{j+1} is actually $O(v(n)^{1/d})$ because the sequence $x_0, x_1, \dots, x_\ell$ is of constant expected length. □

From Claim 4, $\text{dist}(y_j, y_{j+1}) \leq O(v(n)^{1/d})$. As a consequence, the expected number of long-range contacts discovered while going from y_j to y_{j+1} is $O(v(n))$. Therefore, learning about an expected number of $\Omega(\log n)$ long-range contacts implies that the expected length k of the sequence $y_1, y_2, \dots, y_k$ is $\Omega(\log n/v(n))$. Hence, starting from x at distance m from the target, the route visits an expected number of $\Omega(\log n/v(n))$ good intermediate destinations $y_1, \dots, y_k$, and, by Claim 4, the expected number of steps required to go from y_j to y_{j+1} is $\Omega(v(n)^{1/d})$. Therefore, the expected number of steps required to reach B, and thus to reduce the distance to the target by a factor at least 2, is $\Omega(\log n/v(n)^{1-1/d})$. By the same arguments as in the second part of the proof of Lemma 4, after this amount of steps from a node at distance m from the target t, the distance from t is reduced by an expected constant factor. Therefore, starting from a node at Manhattan distance $\Theta(n^{1/d})$ from the target, the expected number of steps to reach the target is $\Omega(\frac{\log n}{v(n)^{1-1/d}} \cdot \log n^{1/d})$, which completes the proof of Lemma 5. □

Conversely, if every node is aware of the long-range contacts of its $v(n) \gg \log n$ closest neighbors in the mesh, then, intuitively, it is easy to find a long-range link that leads close to the target. However, traveling from the current node to the intermediate destination that is the tail of this long-range link requires a large number of steps. More precisely, we show the following:

Lemma 6 *If $v(n) = \log^\alpha n$ for $\alpha \geq 1$, then the expected number of steps to reach the target is at least*

$$\Omega\left(\frac{\log n}{\log(v(n)/\log n)} \cdot v(n)^{1/d}\right) \quad \text{if } \alpha < d$$

and

$$\Omega\left(\frac{\log n}{\log(v(n)/\log n)} \cdot \log n \log v(n)\right) \quad \text{if } \alpha > d.$$

Proof Assume that the distance $m = \text{dist}(x, t)$ between the current node x and the destination t is $> c \cdot v(n)^{1/d}$ where c is a constant large enough. Let $B = \{u \mid \text{dist}(u, t) \leq m/2^{r(n)}\}$ where $r(n) = \frac{1}{d}\log(2^d v(n)/\log n)$. We have $r(n) \geq 1$. From the setting of $r(n)$, we get:

Claim 5 $\Pr(V(x) \to B)$ is asymptotically at least some positive constant.

Proof We have $\Pr(V(x) \to B) = 1 - \Pi_{y \in V(x)}(1 - \Pr(y \to B))$. Now, ignoring the constants, we have

$$\begin{aligned}
\Pr(y \to B) &= \sum_{b \in B} \Pr(y \to b) \approx \frac{1}{\log n} \sum_{b \in B} \cdot \frac{1}{\text{dist}^d(y, b)} \\
&\geq \frac{1}{\log n} \cdot \frac{|B|}{(m + v(n)^{1/d} + m/2^{r(n)})^d} \\
&\approx \frac{1}{\log n} \cdot \frac{(m/2^{r(n)})^d}{(m + v(n)^{1/d} + m/2^{r(n)})^d} \\
&\approx \frac{1}{\log n} \cdot \frac{1/2^{d \cdot r(n)}}{(1 + \frac{v(n)^{1/d}}{m} + \frac{1}{2^{r(n)}})^d} \\
&\approx \frac{1}{v(n)} \cdot \frac{1}{(1 + (v(n)/m)^{1/d})^d} \\
&\geq \frac{1}{v(n)} \cdot \frac{1}{(1 + 1/c)^d}.
\end{aligned}$$

Therefore $\Pr(V(x) \to B)$ is lower bounded by a function of n that is $\approx 1 - (1 - \frac{1}{v(n)})^{v(n)}$. The latter is asymptotically constant, and the claim follows. □

From Claim 5, with constant probability, the current node x finds a long-range link leading to B in its awareness, i.e., a long-range link decreasing the distance to the target by a factor $2^{r(n)}$. The expected Manhattan distance between x to a node in $V(x)$ whose long-range contact is in B is $\Omega(v(n)^{1/d})$. To travel such a distance using Kleinberg's greedy routing, the expected number of steps is, from Lemma 4, $\Omega(v(n)^{1/d})$ if $\alpha < d$, and $\Omega(\log n \log v(n))$ if $\alpha > d$. Thus, reducing the distance to the target by a factor $2^{r(n)}$ requires $\Omega(v(n)^{1/d})$ expected number of steps if $\alpha < d$, and $\Omega(\log n \log v(n))$ expected number of steps if $\alpha > d$. Therefore, starting from a node at distance $\Theta(n^{1/d})$ from the target, the expected number of steps to reach the target is $\Omega(\frac{\log n}{r(n)} \cdot v(n)^{1/d})$ if $\alpha < d$, and $\Omega(\frac{\log n}{r(n)} \cdot \log n \log v(n))$ if $\alpha > d$. □

5 Social networks perspectives

The aim of this section is to give further motivations to our model, by revisiting it in the context of Milgram's experiment, and in light of Kleinberg's results.

5.1 Milgram's experiment

Augmented graphs as defined in [17] have been introduced as a model for the "small world phenomenon." They consist in families of graphs $H = (G, \mathcal{D})$ obtained from a graph

G by adding links chosen at random according to a probabilistic distribution $\mathcal{D}$. The graph G models an *awareness* common to all the *social entities* represented by the nodes of H. In other words, nodes of H are aware of the topology G. In particular, any node x can compute the distance $\text{dist}_G(x, y)$ from x to any other node y in G. The links in G model acquaintances between social entities that can be easily deduced from characteristics of the social entities (geographical positions, hobbies, professional activities, etc.). The added links, called long-range links, model acquaintances that cannot be deduced globally because they correspond to random events which created acquaintances between entities that have generally little in common. If (u, v) is an edge of G, then any node x is aware that u and v have some acquaintance. However, if (u, v) is a long-range link non-incident to x, then x does not know that there is an acquaintance between u and v. In particular, x cannot compute the distance $\text{dist}_H(x, y)$ from x to any other node y in H.

Milgram's experiment reports that there are short chains of acquaintances between individuals, and that these chains can be discovered in a greedy manner. Roughly speaking, given an arbitrary source person s (e.g., living in Wichita, KA), and an arbitrary target person t (e.g., living in Cambridge, MA), a letter can be transmitted from s to t via a chain of individuals related on a personal basis. The transmission rule is that the letter held by an intermediate person x is passed to the next person y who, as judged by x, is closer to the target among all persons x knows on a first-basis. Milgram's experiment conclusion is often summarized as the "six degrees of separation" phenomenon because, for chains that reached the target,[2] the number of intermediate persons between the source and the target ranged from 2 to 10, with a median of 5.

5.2 Greedy routing in augmented meshes

In his seminal work [7, 8] (see also [9]), Kleinberg gives a formal support to the six degrees of separation phenomenon. He considers a d-dimensional mesh augmented with long-range links chosen according to the d-harmonic distribution (see Fig. 1). More precisely, the underlying graph G is the d-dimensional mesh $n^{1/d} \times \cdots \times n^{1/d}$, and the augmented graph H is obtained by adding exactly one out-going link to every node x. If there is a long-range link from x to y, then y is called the *long-range contact* of x. The probability that x chooses y as long-range contact is $h(x, y) = 1/(Z_x \cdot \text{dist}(x, y)^d)$ where dist() is the Manhattan distance in the mesh (i.e., the distance in the L_1 metric), and the normalizing coefficient Z_x satisfies $Z_x = \sum_{z \neq x} 1/\text{dist}(x, z)^d$. In Kleinberg's model, long-range links are directed, i.e., a long-range link from x to y does not imply a long-range link from y to x. This is consistent with what can be observed in the human society. In particular, human relationships are not always symmetric. More importantly, although directed long-range links produce nodes with high in-degree, these "hubs" remain with only an out-degree of one. Hence the impact of hubs is kept limited in the model.[3]

A salient property of Kleinberg's model is that it is a "small world," i.e., a graph in which not only the expected distance between nodes is small, but also greedy routing is able to discover short routes between any pair of nodes.

Greedy routing is a metaphor of the way social entities proceed to search for resources or information in the graph representing their acquaintances [1, 5, 15, 16]. These entities are given very limited computational power. This restriction is motivated by the fact that social entities (e.g., humans) have bounded storage capability, and are usually unable to perform complex computations involving more than a small number of objects. Typically, computing shortest paths in a graph with more than few vertices is assumed to be a too complex task to be performed by social entities. Greedy routing performs as follows: at the current node x, a search for a target node t is forwarded to the neighboring node y of x, including its long-range contact, which is the closest to t in the mesh. In other words, a social entity optimizes locally the discovery of the target by choosing, among all its acquaintances, the one that is likely to be the closest to the target. The distance to the target is however estimated using the Manhattan distance.

5.3 Criteria vs. dimensions

It was observed (cf., e.g. [6]) that searching for the target in Milgram's experiment is performed based on at least two criteria (e.g., geography and occupation), and that performing the search based on one criterion only (e.g., geography) results in poorer performance. The estimation of the distance to the target is performed thanks to all available criteria. In Kleinberg's model, the estimation of the distance to the target is performed based on the coordinates of the nodes in the mesh. That is, the mesh is *not* aiming at modeling geography only, but at capturing all possible criteria used for the search. In other words, the mesh includes all criteria per se, and the long-range links model random events capturing the fact that our acquaintances are not necessarily living close to us, do not necessarily practice the same religion (if they do), do not necessarily occupy the same social position, etc.

On the other hand, there is no one-to-one correspondence between the dimensions of the mesh and the criteria used for the search. In particular, moving along one axis preserves all the coordinates of the mesh, which is not perfectly true in real life. Nevertheless, most of the time, our acquaintances have characteristics very similar to ours. (The

[2] Many chains did not succeeded in Milgram's experiment. Experiments by Dodds et al. [5] revealed however that this is perhaps not due to the inability of reaching the target, but rather due to the fact that individuals do not necessarily benefit from their connectedness: they often stop retransmission simply because they believe that there is no short chain to the target, although such a chain does exist.

[3] Dodds et al. [5] observed that, in contrast with what is often believed, the presence of hubs appears to have a limited relevance to social search. Thus it is desirable that a model keeps the role of hubs limited.

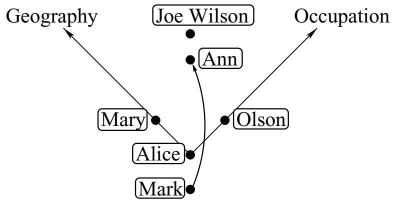

Fig. 5 Searching for Joe Wilson

rare cases of acquaintances with characteristics very different from ours are modeled by long-range contacts.) A model aiming at capturing the slight variations of the characteristics of our acquaintances could be obtained by introducing some randomness in the Cartesian product operation, to locally shuffle the connections. This would however significantly complicate the analysis of the model, without bringing new light on Milgram's experiment. Thus, in this paper we have chosen to stick to Kleinberg's model for analyzing the impact of the number of criteria on the performance of the search. Hence, for the sake of simplicity, we have viewed every dimension of the mesh as a distinct criterion.

5.4 Substratum of topological awareness

Our model was based on the following observation: although every individual personally knows a small number of other individuals only, he or she is often aware of a large number of personal acquaintances between individuals that he or she does not personally know. Let us take a simple example to illustrate this observation (see Fig. 5).

Consider Milgram's experiment in which the goal is to send a letter to Joe Wilson, who is located at Revelstoke, Alberta, Canada. In addition to Wilson's location, we are also given the facts that Wilson is a designer, and that he won a downhill ski Canadian championship in the 1980s. The letter is currently held by Alice, a librarian in San Francisco. Alice has a friend, Mary, living in Seattle, an uncle, Olson, living in Bergen where he is training the Norwegian cross country ski team, and finally a former schoolfriend, Mark, who is a pianist in the Vienna symphony orchestra.

Based on her acquaintances, Alice may forward the letter either to Mary or to Olson. In the former case, there is a geographical improvement. In the latter case, there is also an improvement because a cross country ski trainer is somewhat close (in terms of occupation) to a downhill ski champion. On the other hand, Alice would certainly not forward the letter to Mark because Mark is geographically farther from Joe Wilson than Ann, and Mark's vitae has little to do with Wilson's vitae.

Now, assume that in Alice's recent phone conversation with Mark, she learned that Mark moved to a new house, entirely designed by his new girlfriend, Ann, an architect who graduated from Vancouver. Based on this "topological awareness," it makes sense for Alice to forward the letter to Mark, because he may then forward it to his girlfriend Ann. Once the letter will be in Ann's hands, the improvement will be significant because an architect who graduated from Vancouver is reasonably close to a designer living in Alberta. Note that there is no personal acquaintance between Alice and Ann (she hardly remembers her name). However, Alice is aware that there is an acquaintance between Mark and some architect from Vancouver. This acquaintance is a long-range link because an acquaintance between a member of the Vienna symphony orchestra and a Canadian architect can be hardly guessed. The fact that Alice is aware of Mark's long-range contact significantly improves the search for Joe Wilson. This phenomenon cannot be captured by Kleinberg's model because, in his model, a social entity is not aware of any long-range links not incident to it.

5.5 Substratum of indirect-greedy routing

Our model captures the "indirect" routing strategy based on Alice's awareness of the social characteristics of Mark's long-range contact. In this model, we assumed that, in addition to the underlying graph G, and to its long-range contact in the augmented graph H, every social entity is aware of some list of acquaintances between pairs of other entities.

According to Kleinberg's greedy routing, when Alice is searching for Joe Wilson, she chooses, among all her personal acquaintances, the one who is most likely to know Wilson. As we mentioned before, this strategy results in having Alice choosing either Olson or Mary, but not Mark, although Mark is more likely to be closer to Wilson than both Olson and Mary. Being aware of Mark's long-range contact Ann, Alice may then decide to use Mark as an "intermediate destination." Mark is farther from the target Joe Wilson than Alice. However, from Mark, the search may be forwarded close to Wilson, thanks to the long-range link Mark-to-Ann.

Obviously, a faster search would be obtained by computing short cuts from the source to the target in the augmented mesh using the local awareness of every node. However, such a complex computation is assumed to be beyond the computing capabilities of social entities. For instance, although most humans would be able to go through a reasonably large directory to select one key (say, the smallest), most humans would be unable to sort a directory based on the keys contained into it.

The convergence of indirect-greedy routing requires the system of awareness to be monotone. It is reasonable to assume that monotonicity is a property that a system of awareness usually satisfies. Indeed, if a social entity x is aware of the acquaintance that some node u has with v, then a node y that is closer to u than x is probably also aware if this acquaintance. For instance, if you become aware that Bob, the companion of the sister Sophie of your friend Tom, meets some unrelated guy Charles in a plane, then certainly Tom is aware of that, and this is even more certainly the case of Sophie. One may argue the other way though, by saying that if you become aware of some relation between two of your friends, your neighbor in the street may not know that, even if he lives closer to your friends than you do. Nevertheless, our definition of convergence is very restrictive, and even if the system of awareness is not properly monotone, indirect-greedy routing will converge for most setting of the long range contacts, and non convergence may occur for only few pairs source-target.

As a final remark concerning or model, note that we assumed that every social entitiy personnaly knows a constant number of other entities (its $2d$ neighbors in the mesh plus it c long-range contacts). In contrast we have assumed that every social entity is aware of $\log n$ long-range links. This is of course debatable, but it is reasonable to assume that the number of people we know *personally* is less impacted by the world population than the number of rumors we hear about other people.

5.6 What did we learn out of indirect-greedy routing?

We have defined our model having in mind the way social entities may plausibly have routed the letters in Milgram's experiment, i.e., (1) by using intermediate destinations, and (2) in an oblivious manner. The latter is imposed by the way the experiment was performed. The former is our conjecture. By interpreting the dimensions of the mesh as many criteria on which greedy routing is based, our model demonstrates that eclectic relationships are desirable, as far as connectedness to other individuals is concerned. This is consistent with what can be observed in every-day life. In particular, searching using two criteria is significantly faster than searching using only one criterion. For instance, Killworth and Bernard [6] have observed that, in a search for an individual, at least two criteria (occupation and geography) were used by the participants. Determining whether individuals involved in Milgram's experiment used intermediate destinations (consciously or unconsciously) to route the letter to the target would allow us to validate our model.

Acknowledgements We are thankful to Maha Abdallah and Vassos Hadzilacos for their careful reading of preliminary versions of this paper, for their comments on its content, and for correcting several technical and grammatical errors. We also deeply appreciated the suggestions of George Giakkoupis for simplifying the proof of Theorem 1. Last but not least we are also thankful to the referees whose comments were extremely valuable. *Financial supports*: The three authors are supported by the Actions Spécifiques CNRS "Dynamo" and "Algorithmique des grands graphes," and by the project "PairAPair" of the ACI Masses de Données. Pierre Fraigniaud received additional supports from the INRIA project "Grand Large."

References

1. Adamic, L., Adar, E.: How To Search a Social Network. Technical Report, HP Labs, Palo Alto (2003)
2. Aspnes, J., Diamadi, Z., Shah, G.: Fault-tolerant routing in peer-to-peer systems. In: 21st ACM Symposium on Principles of Distributed Computing (PODC), pp. 223–232 (2002)
3. Barrière, L., Fraigniaud, P., Kranakis, E., Krizanc, D.: Efficient routing in networks with long range contacts. In: 15th International Symposium on Distributed Computing (DISC). LNCS 2180, pp. 270–284. Springer-Verlag, Berlin Heidelberg, New York (2001)
4. Coppersmith, D., Gamarnik, D., Sviridenko, M.: The diameter of a long-range percolation graph. Random Struct. Algorithms **21**(1), 1–13 (2002)
5. Dodds, P., Muhamad, R., Watts, D.: An experimental study of search in global social networks. Science **301**, 827–829 (2003)
6. Killworth, P., Bernard, H.: Reverse small-world experiment. Soc. Networks **1**(2), 159–192 (1978)
7. Kleinberg, J.: The small-world phenomenon: an algorithmic perspective. In: 32nd ACM Symposium on Theory of Computing (STOC), pp. 163–170 (2000)
8. Kleinberg, J.: Navigation in a small-world. Nature **406**, 845 (2000)
9. Kleinberg, J.: Small-world phenomena and the dynamics of information. In: 15th Neural Information Processing Systems (NIPS) (2001)
10. Lebhar, E., Schabanel, N.: Searching for optimal paths in long-range contact networks. In: 31st International Colloquium on Automata, Languages and Programming (ICALP). LNCS 3142, pp. 894–905 (2004)
11. Manku, G., Bawa, M., Raghavan, P.: Symphony: distributed hashing in a small world. In: 4th USENIX Symposium on Internet Technologies and Systems, pp. 127–140 (2003)
12. Manku, G., Naor, M., Wieder, U.: Know thy neighbor's neighbor: the power of lookahead in randomized P2P networks. In: 36th ACM Symposium on Theory of Computing (STOC) (2004)
13. Martel, C., Nguyen, V.: Analyzing Kleinberg's (and other) small-world models. In: 23rd ACM Symposium on Principles of Distributed Computing (PODC), pp. 179–188 (2004)
14. Milgram, S.: The small-world problem. Psychol. Today, pp. 60–67 (1967)
15. Travers, J., Milgram, S.: An experimental study of the small world problem. Sociometry **32**, 425 (1969)
16. Watts, D., Dodds, P., Newman, M.: Identity and search in social networks. Science **296**, 1302–1305 (2002)
17. Watts, D., Strogatz, S.: Collective dynamics of small-world networks. Nature **393**, 440–442 (1998)