Eclecticism Shrinks Even Small Worlds*

[Extended Abstract]

Pierre Fraigniaud†
CNRS-LRI
Univ. Paris-Sud & INRIA
France
pierre@lri.fr

Cyril Gavoille
LaBRI
Univ. of Bordeaux & CNRS
France
gavoille@labri.fr

Christophe Paul
CNRS-LIRMM
Univ. of Montpellier
France
paul@lirmm.fr

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. This model aims at giving formal support to the "six degrees of separation" between individuals experienced by Milgram (1967), and verified recently by Dodds, Muhamad, and Watts (2003). In particular, Kleinberg shows that greedy routing performs in $O(\log^2 n)$ expected number of steps in d-dimensional augmented meshes, with $O(\log n)$ bits of topological awareness per node, for any constant $d \geq 1$. We show that giving $O(\log^2 n)$ bits of topological awareness per node decreases the expected number of steps of greedy routing to $O(\log^{1+1/d} n)$ in d-dimensional augmented meshes. We also show that, independently of the amount of topological awareness given to the nodes, 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 performances of greedy routing. Moreover, our model demonstrates that the efficiency of greedy routing is sensible 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 model. 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. Finally, our protocol can obviously be used for the design of DHTs, in the same spirit as Symphony (2003).

Categories and Subject Descriptors

G.2.2 [**Mathematics of Computing**]: Discrete Mathematics—*Graph Theory*

General Terms

Algorithms, Theory

Keywords

Small World Graphs, Routing, Distributed Hash Tables.

1. INTRODUCTION

We consider *augmented* graphs as defined in [17], i.e., the family 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 $\mathrm{dist}_G(x,y)$ 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 ignores that there is an acquaintance between u and v. In particular, x cannot compute the distance $\mathrm{dist}_H(x,y)$ from x to any other node y in H.

Milgram's experiment [14], recently reproduced by Dodds, Muhamad, and Watts [5] (see also [1]), 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 most likely to know 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

*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.
†Additional support from the INRIA project "Grand Large"

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
PODC'04, July 25–28, 2004, St. Johns, Newfoundland, Canada.
Copyright 2004 ACM 1-58113-802-4/04/0007 ...$5.00.

because, for chains that reached the target[1], the number of intermediate persons between the source and the target ranged from 2 to 10, with a median of 5.

1.1 Greedy routing in Augmented Meshes

In his seminal work [7, 8] (see also [9]), Kleinberg gives a formal support to the six degree of separation phenomenon. He considers a d-dimensional mesh augmented with long-range links chosen according to the d-harmonic distribution (see Fig. 2). 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 $h(x, y) = 1/(Z_x \cdot \mathrm{dist}(x, y)^d)$ 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 $Z_x = \sum_{z \neq x} 1/\mathrm{dist}(x, z)^d$. 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 coherent 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[2].

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 computed using the Manhattan distance.

1.2 Substratum of Greedy Routing

We want first to lay stress on two points. First, it was observed (cf., e.g., [6]) that searching for the target in Milgram-like experiments 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 performances. The estimation of the distance to the target is performed thanks to all available criteria. In the 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, and this is the second point that we want to stress, 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 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[3], to locally shuffle the connections. This would however significantly complicate the analysis of the model, without bringing new light on Milgram's experiment. Thus, we chose to stick to Kleinberg's model for analyzing the impact of the number of criteria on the performances of the search. Hence, for the sake of simplicity, we view every dimension of the mesh as a distinct criterion.

In a social context indeed, 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 searching in the social graph. In this context, one would thus expect that the more criteria used the more efficient the search should be. Surprisingly however, Kleinberg's model does not reflect this fact, in the sense that greedy routing has the same performances whether the number of 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 (note that 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 performances, 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 the light of the previous observations, one can conclude that the absence of the dimension parameter from the complexity of greedy routing in augmented meshes is a problem

[1] Many chains did not succeeded in Milgram's experiment. Experiments by Dodds et al. [5] revealed however that this is 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.

[2] 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.

[3] The d-dimensional mesh is the Cartesian product of d paths.

of the greedy routing specification, and not of the links distribution (it was recently shown [13] that the expected diameter of augmented meshes is $O(\log n)$). In this paper, we propose a new greedy routing protocol based on Kleinberg's model. The key feature of our protocol is that its asymptotic complexity depends on the dimension of the mesh.

1.3 Our contributions.

We 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 (see Fig. 2). 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 towards 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* y, i.e., a node y such that its long-range contact is close to t. In the second phase, x applies greedy routing towards y, and forwards the search to some node x'. In x', the same process is applied, a new intermediate destination y' is selected (thanks to x''s awareness $A_{x'}$), and greedy routing is applied towards y'. And so on. Generally, the intermediate destination remains the same at every step of indirect-greedy routing, until the search reaches it. Once the search reaches an intermediate destination y, it is forwarded to y's long-rank contact, which is expected to be not too far from the target t. The same actions are repeated until the search eventually reaches the target.

It is known that giving additional awareness to the nodes speeds up the search. For instance, it is known [12] that if every node is aware of the long-range contacts of its neighbors, then greedy routing performs in $O(\log^2 n/(c \log c))$ expected number of steps with c long-range contacts per node. 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. We conclude that additional topological awareness has positive impact, and that the speed-up factor compared to Kleinberg's greedy routing protocol is $O(\log^{1-1/d} n)$. Moreover, 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[4].

Note that 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. Clearly, obviousness is a desirable property for a routing protocol as it makes simpler the decisions taken at each node. However, it makes its analysis harder as the routes may traverse intermediate nodes that are farther to the destination (according to the Manhattan distance) than the source.

We also show that, surprisingly, the positive impact of additional topological awareness reaches a certain limit. Indeed, 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 degrade the performances 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 performances of indirect-greedy routing, and demonstrate an intrinsic limitation of the greedy routing strategy in augmented graphs. In particular, if every entity has a topological awareness of size n, i.e., is aware of all long-range contacts, then the entities would not perform better than Kleinberg's greedy routing, leading an $\Omega(\log^2 n)$ expected number of steps.

More importantly, our study captures the trade-off that we expected: if entities are living in a d-dimensional world, then giving additional topological awareness of $O(\log^2 n)$ bits to these entities enable 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 to the world's dimension.) In particular, our model demonstrates a significant difference between routing using one criterion, which performs in $O(\log^2 n)$ expected number of steps, and routing using two criteria, which performs in $O(\log^{3/2} n)$ expected number of steps. The relative improvement decreases when the number of dimensions increases, which is coherent with what was observed by Killworth and Bernard [6].

To summarize, given a fixed number of acquaintances $2d + c$ per social entities in an augmented d-dimensional mesh, 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 has more impact on the distance between people than the *number* $2d + c$ of these relations. Our investigation is perhaps a first step towards the formalization of arguments in favor of the sociological evidence stating that eclecticism shrinks the world.

As a final remark, we observe that, in addition to bringing new light to Milgram's experiment, our protocol presents several properties that can obviously be used for the design of DHTs for peer-to-peer systems, in the same spirit as [11] for Kleinberg's routing protocol.

1.4 Organization.

The paper is organized as follows. The next section precisely describes indirect-greedy routing, including the notion of topological awareness. Then, in Section 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 Section 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, Section 5 contains some concluding remarks.

2. TOPLOGICAL AWARENESS AND INDIRECT-GREEDY ROUTING

Our model addresses the following question: what is the

[4] Including NoN-greedy routing [12], defined in the Percolation model of [4].

	Expected #steps	Amount of awareness (#bits)
Kleinberg's greedy [7]	$O(\frac{1}{c}\log^2 n)$	$O(c\log n)$
Kleinberg's greedy [3, 13]	$\Omega(\frac{1}{c}\log^2 n)$	$O(c\log n)$
Kleinberg's 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]	$O(\frac{1}{\log^2 c}\log^2 n)$	$O(c\log n)$
Indirect-greedy [This paper]	$O(\frac{1}{c^{1/d}}\log^{1+1/d} n)$	$O(\log^2 n)$

Table 1: Performances of variants of greedy routing in d-dimensional meshes augmented using d-harmonic distributions, with c long-range contacts per node. For $d \geq 2$, our variant, indirect-greedy, performs faster than all other greedy algorithms, 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. For $c = \sqrt{\log n}$, indirect-greedy performs in $O(\log^{1+1/2d} n)$ 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. [3, 13] is a lower bound on Kleinberg's greedy routing. [2] is a lower bound on Kleinberg's greedy routing in the 1-dimensional directed mesh augmented using *any* distribution. 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.

additional "topological awareness" that could be given to nodes so that greedy 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. Obviously, if entities are given more than one long-range contact, then the performances of greedy routing can be improved, however to a limited extend only. For instance, with c long-range contacts per node, Kleinberg's greedy routing would perform in $O(\frac{1}{c}\log^2 n)$ expected number of steps, which remains $O(\log^2 n)$ 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 social entity. This is motivated by the fact that every individual personally knows a constant number of other individuals only, independently of the size of the world population.

2.1 Topological awareness

Our model is based on the following observation: although every individual personally knows a constant 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. 1). 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 80's. 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

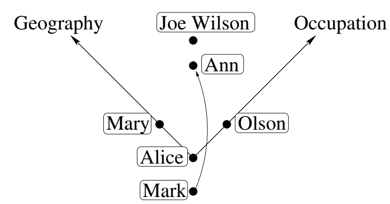

Figure 1: Searching for Joe Wilson.

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 learnt 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 in 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 somebody 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.

In this paper, we define a model that captures the "indirect" routing strategy based on Alice's awareness of the social characteristics of Mark's long-range contact. In this model, we assume 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. This idea is formalized as follows.

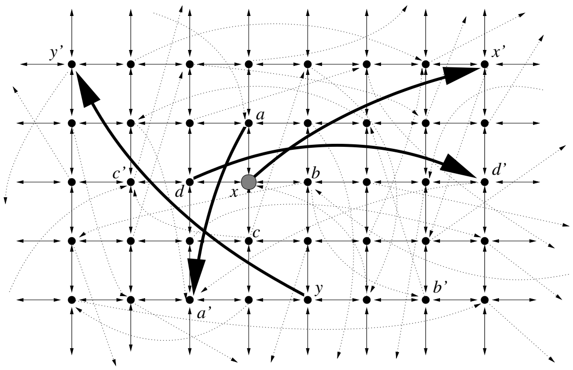

Figure 2: Long-range links in the 2-dimensional mesh. The topological awareness of node x is composed of the four plain long-range links.

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

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 and in Kleinberg's model. For instance, in Fig. 2, 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 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 acquaintance with either y or y', but is just aware of an acquaintance from y to y'. On the other hand, x ignores the long-range contacts of b and c. To be realistic, the number of nodes y that x is aware of should not be too large. Furthermore, these nodes should be preferably located not too far from x. However, it is a reasonable assumption that the topological awareness of every individual grows (though slowly) with the total number of individuals in the world. Indeed, although the total number of individuals has a limited impact on the number of our personal acquaintances (relatives, close friends, etc.), the more individuals, the more stories every individual hears about other individuals, increasing his or her awareness of some inter-individual acquaintances.

This gives rise to the following second question: 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?

2.2 Indirect-greedy routing

To answer the previous question, let us return to our simple example in which Alice is searching for Joe Wilson. According to Kleinberg's greedy routing, Alice 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 to 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. We define *indirect-greedy routing* in which, at every routing step towards a target t, there are two phases. In the first phase, the current node x uses its topological awareness A_x to select an intermediate destination y, i.e., a node y such that its long-range contact is close to t. In the second phase, x applies greedy routing towards y. (Clearly, x makes use of the intermediate destination y only if y is closer to x than t in the mesh. Otherwise, x discards y and simply applies greedy routing towards t.) More formally, we define indirect-greedy routing as follows.

Indirect-greedy routing:

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 $w_1, \ldots, w_{2d}$, and the long-range contact of x is denoted w_0. Finally, let t be the target node, $t \neq x$.

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

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

In the following, the node y selected during Phase 1 is called the *intermediate destination*.

Remark. 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 the search is forwarded during Phase 2, then the intermediate destination for w_i may be different from the intermediate destination 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., $y = \text{tail}(e) = x$. Thus, during Phase 2, the search is forwarded to 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 $e_1, \ldots, e_k$ be the $k \geq 1$ long-range links such that, for every i, $1 \leq i \leq k$, dist(head$(e_i), t)$ is minimum among all long-range links. At every node involved in the search, 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 to travel 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 the latter example, a faster search would be obtained by computing a shortest path from the source to the target in the augmented mesh. However, such a complex computation is assumed to be beyond the computing capabilities of the 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. We underline here that *indirect-greedy routing fully preserves the greediness assumption*.

Remark. As opposed to Kleinberg's greedy routing, the Manhattan distance to the target is not strictly decreasing at each step. Indeed, an intermediate destination can be farther to 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 this phenomenon has, under some weak condition, little impact on the expected performances 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. PERFORMANCES OF INDIRECT GREEDY ROUTING

In this section, we give a necessary and sufficient condition for indirect-greedy routing to converge, i.e., to always route correctly the search 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 $\text{tail}(e) \neq x$, we then define

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

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* $y \in N_x(e)$*.*

Observe 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 certainly 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.

Remark. If all sets $S_x = \{\text{tail}(e) \mid e \in A_x\}$ have the same shape S for all nodes x, in the sense that $S = S_{x_0} = \{\text{tail}(e) \mid e \in A_{x_0}\}$ for some fixed node x_0, and S_x is obtained by translating S_{x_0} along x_0x, then monotonicity is equivalent to the fact that S is x_0-convex, i.e., every shortest path from x_0 to any node in S is included in S. "Be 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. *Indirect-greedy routing converges if and only if the system of awareness is monotone.*

Proof. Assume first that the system of awareness is not monotone, and let us prove that indirect-greedy routing does not always converges in this case, i.e., there is a setting of the long-range contacts for which indirect-greedy does not converge. If the system of awareness is not monotone, then there exists a node x, and $e \in A_x$, such that $e \notin A_y$ where y is the neighbor of x that is the closest to $z = \text{tail}(e)$. We denote by e' the long-range link of x. We construct a setting of the long-range contacts yielding non convergence of the indirect-greedy protocol. First of all, if there are several such y, we choose one that is closest to x. We set $t = \text{head}(e)$ as the target, and we set $\text{dist}(t, \text{head}(f)) > \text{dist}(t, y)$ for every long-range link $f \notin \{e, e'\}$. Now, we consider two cases, depending whether y is the long-range contact of x or not. In both cases, we place t so that x is on a shortest path from y to t.

— If y is not the long-range contact of x, i.e., y is one of the four neighbors of x in the mesh, then we set the long-range link e' of x such that $\text{dist}(t, \text{head}(e')) > \text{dist}(t, y)$. By definition of indirect-greedy routing, x forwards the search to its neighbor y. Next, greedy routing is applied at node y. Since $e \notin A_y$, and since all other long-range links lead away from t, the search is sent back to x, creating an infinite loop between nodes x and y, and thus indirect-greedy routing does not converge.

— If y is the long-range contact of x, i.e., $y = \text{head}(e')$, then, from the setting of all long-range links different from e and e', the search is sent from y to a neighbor y' on the mesh that is on a shortest path from y to t. We place t so that y' is also on a shortest path from y to x. Thus, by the choice of y as the closest node from x satisfying $e \notin A_y$, we have $e \in A_{y'}$. Therefore, the search is sent back from y' to y, creating an infinite loop between y and y', and thus indirect-greedy routing does not converge.

We now assume that the system of awareness is monotone, and we prove that indirect-greedy routing always converges, for any setting of the long-range contacts. 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 as:

$$\phi(s) \;=\; \text{dist}(s, u) + n \cdot \text{dist}(v, t)$$

From s, the search is forwarded 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 $\phi(s') \leq \phi(s) - 1$. 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 $\text{dist}(v', t) \leq \text{dist}(v, t)$. If $\text{dist}(v', t) < \text{dist}(v, t)$ then $\phi(s') = \text{dist}(s', u') + n \cdot \text{dist}(v', t) \leq (n-1) + n \cdot (\text{dist}(v, t) - 1) = \text{dist}(v, t) - 1 < \phi(s)$. If $\text{dist}(v', t) = \text{dist}(v, t)$

then Phase 1 of indirect-greedy routing specifies that since s' chooses u', u' is at least as close to s' as u. Therefore, $\phi(s') = \mathrm{dist}(s', u') + n \cdot \mathrm{dist}(v', t) \le \mathrm{dist}(s', u) + n \cdot \mathrm{dist}(v, t) \le \phi(s) - 1$. Therefore, in all cases, the potential is strictly decreasing after each step of indirect-greedy routing. Thus indirect-greedy routing eventually reaches the target. $\square$

Now, we prove the following:

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 its $O(\log n)$ closest nodes in the mesh, then indirect-greedy routing performs in $O(\log^{1+1/d} n)$ expected number of steps.*

PROOF. Clearly, the system of awareness induced by balls of same radius is monotone (since a ball centered at x is x-convex). 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 $m = \mathrm{dist}(x, t) > c \cdot \log^{1/d} n$ for some constant c large enough. Let us compute the expected number of steps required by indirect-greedy routing for reaching a node x' at Manhattan distance $\le m/2$ from t. Let $B = \{u \mid \mathrm{dist}(u, t) \le m/2\}$. For any node u, let $V(u) = \{v \mid \mathrm{dist}(u, v) \le \log^{1/d} n\}$. Let $\Pr(V(u) \to B)$ be the probability that at least one node in $V(u)$ has its long-range contact in the ball B. We have $\Pr(V(x) \to B) \ge \Pr(V'(x) \to B)$ where $V'(x) = \{u \in V(x) \mid \mathrm{dist}(u, t) \le m\}$. Note that $|V'(x)| \ge \frac{1}{2^d}|V(x)|$ as $t \notin V(x)$, so that $|V'(x)| = \Theta(\log n)$. 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 $\Pr(\mathcal{E}_x) \le \Pr(\mathcal{E}_u)$ for any $u \in V'(x)$, we get $\Pr(V'(x) \to B) \ge 1 - (1 - p)^{|V'(x)|}$. 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 $Z_x = \sum_{i \ge 1} |S_i|/i^d$ 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,

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

Therefore p is at least $\Omega(1/\log n)$. Since $|V'(x)| = \Theta(\log n)$, we get $1 - (1 - p)^{|V'(x)|}$ is at least some constant > 0, and thus $\Pr(V(x) \to B)$ is at least some constant $\beta > 0$.

Let us return to the indirect-greedy routing process, and 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 search is routed from x_0 to x_1 according to Kleinberg's greedy protocol. 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. 4(a)). If such a new node x_2 is discovered, x_1 is discarded, and the new intermediate destination becomes x_2. In this case, x_2 is discovered after performing $O(\log^{1/d} n)$ steps of routing toward x_1 in the worst-case. Indeed, every node is aware of the long-range contacts of its $\log n$ closest neighbors, which correspond to a ball of radius $\Theta(\log^{1/d} n)$. 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, and routing switches to x_3. This phenomenon may occur many times, constructing a sequence $x_1, x_2, x_3, \ldots$ of unreached intermediate destinations (see Fig. 4(a)). The Manhattan distance between every two consecutive unreached intermediate destinations x_i and x_{i+1} satisfies $\mathrm{dist}(x_i, x_{i+1}) \le O(\log^{1/d} n)$, for every $i \ge 0$.

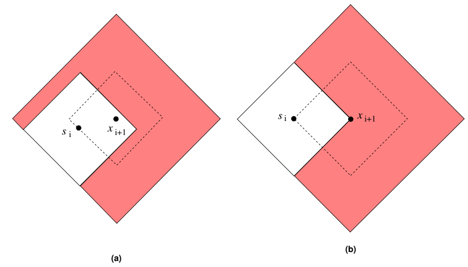

Figure 3: The set C_i is included in the grey area, and in the 2-dimensional mesh $|C_i| \le 3 \log n$.

We show that the expected number of unreached intermediate destinations x_i is a constant. Let s_i be the node where greedy routing switches from x_i to x_{i+1}. 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 to x_{i+1} generated by Kleinberg's greedy routing, where $a_0 = s_i$ and $a_l = x_{i+1}$. 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 x_{i+1} and of radius $\mathrm{dist}(s_i, x_{i+1})$ (see Fig. 3). This inclusion holds even if the path contains long-range links (a_j, a_{j+1}). Hence $|C_i| \le (2^d - 1)\log n$. From this fact, one cannot conclude that $\Pr(C_i \to B) \le (2^d - 1) \cdot \Pr(V(s_i) \to B)$ because the probability of having a long-range contact in B changes with the distance to the target. Nevertheless, since the radius of C_i is only a small fraction of m for c large enough, one can show that, for any $\epsilon > 0$, there is a setting of the constant c such that $\Pr(C_i \to B) \le ((2^d - 1) + \epsilon) \cdot \Pr(V(s_i) \to B)$ for every i such that $\mathrm{dist}(s_i, t) \ge c \cdot \log^{1/d} n$. Therefore, if $\mathrm{dist}(s_i, t) \ge c \cdot \log^{1/d} n$, then the probability that, going from s_i to x_{i+1}, a new intermediate destination is discovered is at most roughly $(2^d - 1)/2^d$. It follows that one can set the constant c large enough so that the expected number of successive intermediate destinations x_i's is a constant. Therefore, after at most $O(\log^{1/d} n)$ expected number of steps, one eventually reaches an intermediate destination y_1 (see Fig. 4(a)).

Starting from y_1, we argue the same as when starting from $y_0 = x$, and thus, after at most $O(\log^{1/d} n)$ expected number steps, one eventually reaches another intermediate destination y_2. And so on, we construct in this way a sequence $y_1, y_2, \ldots$ of intermediate destinations that are reached during indirect-greedy routing (see Fig. 4(b)). Let $\mathcal{E}_i$ be the event "at least one node in $V(y_i)$ has its long-range contact in B". We show that the expected number of reached intermediate destinations y_i before the event $\mathcal{E}_i$ holds is constant. The events $\mathcal{E}_i$ are not pairwise independent. Nevertheless, if $\mathrm{dist}(y_i, y_j) > 2 \cdot \log^{1/d} n$, then $\mathcal{E}_i$ and $\mathcal{E}_j$ are independent. Thus, we consider a subsequence $(y'_i)_{i \ge 0}$ of reached intermediate destinations y_i's such that (1) the events $\mathcal{E}'_i =$"at least one node in $V(y'_i)$ has its long-range contact in B"

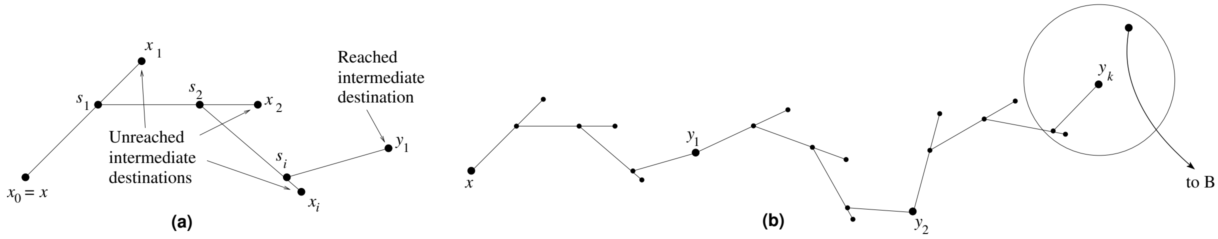

Figure 4: Intermediate destinations before jumping into B.

are pairwise independent, (2) $\mathrm{dist}(y'_i, y'_{i+1}) = \Theta(\log^{1/d} n)$, and (3) $y'_0 = y_0 = x$. (Note that the expected number of reached intermediate destinations between y'_i and y'_{i+1} is constant). Let $p_i = \Pr(\mathcal{E}'_i)$. In particular $p_0 = \beta$. Since $\mathrm{dist}(y'_i, y'_{i+1}) = O(\log^{1/d} n)$, for any positive $\alpha < 1$, one can set the constant c such that $p_{i+1} \geq \alpha p_i$ for any $i \geq 0$. (Recall that c determines how far the current node is from the target.) We are let with a sequence of trials, such that the ith trial succeeds with probability at least $\alpha^i \beta$. The expected number of trials before we get a success is constant. Therefore, starting from x, indirect-greedy routing eventually reaches an intermediate destination y_k, for some $k = O(1)$, such that at least one node in $V(y_k)$ has its long-range contact in B (see Fig. 4(b)). Since going from y_i to y_{i+1} takes $O(\log^{1/d} n)$ expected number of steps, going from x to y_k takes no more than $O(\log^{1/d} n)$ expected number of steps in total.

We are now in the situation in which the current node $x' = y_k$ satisfies that at least one node in $V(x')$ has its long-range contact in B. Indirect-greedy routing applies, that is an intermediate destination x'_1 is selected, and the search goes toward x'_1. The long-range contact of x'_1 is in B because there is a node in $V(x')$ that has its long-range contact in B. As when the search was routed from x to x_1, new long-range links are discovered on the way to x'_1. Thus, a new intermediate destination x'_2 may be selected on the way from x' to x'_1. Again, the long-range contact of x'_2 is in B. In fact, the same analysis as for x can reproduces for x'. One can thus show that, after $O(\log^{1/d} n)$ additional expected number of steps, the search reaches an intermediate destination $z_1 = x'_k$. Similarly to $x'_1, x'_2, \ldots$, node z_1 has its long-range contact in B. Now, the long-range link e going from z_1 to B may not be taken at z_1, because indirect-greedy routing may discover at z_1 a better long-range link, i.e., a long-range link going closer to t than $\mathrm{head}(e)$. However, if such a long-range link f does exist, then $\mathrm{tail}(f)$ is on the frontier of $V(z_1)$. Indeed, otherwise, z_1 would not be a reached intermediate destination because indirect-greedy routing would have switched to $\mathrm{tail}(f)$ before reaching z_1. Since $\mathrm{tail}(f)$ is on the frontier of $V(z_1)$, the probability of existence for f is $O(1/\log^{1/d} n) = o(1)$. Hence this event does not occur too often. Applying the same kind of analysis as before, we consider the sequence of reached intermediate destinations $z_1, z_2, \ldots$, all having their long-range contact in B, and such that $\mathrm{dist}(z_i, z_{i+1}) \leq O(\log^{1/d} n)$. The expected length of such a sequence is constant, and thus indirect-greedy routing eventually reaches an intermediate destination z_ℓ such that the long-range link e of z_ℓ is in B, and all long-range contacts of the nodes in $V(z_\ell)$ are further from t than $\mathrm{head}(e)$. At z_ℓ, indirect-greedy routing applies, and the search is forwarded to $\mathrm{head}(e) \in B$.

Putting everything together, starting from x at Manhattan distance m from t, it takes $O(\log^{1/d} n)$ expected number of steps to reach a node in B. 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 $m \geq c \cdot \log^{1/d} n$ 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 $< c \cdot \log^{1/d} n$ from t.

Hence, it remains to consider the case where the current node x is close to the target t, i.e., $m = \mathrm{dist}(x, t) \leq c \cdot \log^{1/d} n$. 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 $\phi(x) = \mathrm{dist}(x, u) + \mathrm{dist}(v, t) \cdot (1 + \log^{1/d} n)$. From x, the search is forwarded to some node x' on a shortest path from x to u. If the intermediate destination at x' is the same as the one at x, then $\phi(x') \leq \phi(x) - 1$. 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 $\phi(x') = \mathrm{dist}(x', u') + \mathrm{dist}(v', t) \cdot (1 + \log^{1/d} n) \leq \log^{1/d} n + (\mathrm{dist}(v, t) - 1) \cdot (1 + \log^{1/d} n) < \phi(x)$. 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, $\phi(x') = \mathrm{dist}(x', u') + \mathrm{dist}(v', t) \cdot (1 + \log^{1/d} n) \leq \mathrm{dist}(x', u) + \mathrm{dist}(v, t) \cdot (1 + \log^{1/d} n) \leq \phi(x) - 1$. 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 $c \cdot \log^{1/d} n$ 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. □

4. LOWER BOUNDS FOR INDIRECT-GREEDY ROUTING

Theorem 1 shows that, comparatively to Kleinberg's greedy routing, augmenting the awareness up to $O(\log n)$ long-range per node links speeds up indirect-greedy routing. In Theorem 2, we show that the excepted 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 $\log n$ is an optimum for the awareness. If the awareness is smaller than $\log n$ then the expected number of steps is a decreasing function of the awareness. However, after the threshold of $\log n$, the expected number of steps is an increasing function of

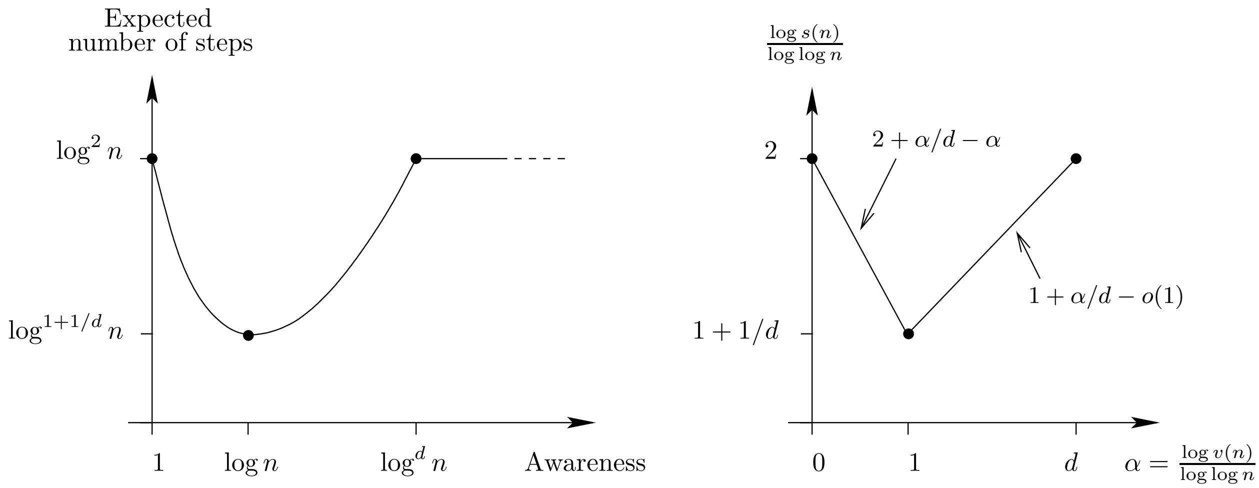

Figure 5: The expected number of steps v.s. the awareness. For $v(n) = (\log n)^\alpha$, the expected number of steps is $s(n) = \Omega((\log n)^{2+\alpha/d-\alpha})$ if $\alpha < 1$ (by Lemma 2), and is $s(n) = \Omega((\log n)^{1+\alpha/d-o(1)})$ if $1 \le \alpha \le d$ (by Lemma 3). For $\alpha \ge d$, $s(n) = \Theta(\log^2 n)$ (by Lemma 3).

the awareness (see Fig. 5).

THEOREM 2. *In the d-dimensional mesh augmented with one long-range link per node chosen according to the d-harmonic distribution, for any $1 \le v(n) \le n$, 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. Moreover, if $d > 1$, then a performance of $O(\log^{1+1/d} n)$ expected number of steps cannot be reached if $v(n) \ne \Theta(\log n)$.*

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

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

PROOF. We assume that the distance $m = \mathrm{dist}(x,t)$ between the current node x and the target t is $\ge c \cdot \log^{1/d} n$ for c large enough. We use the same notations as in the proof of Theorem 1. Let $B = \{u \mid \mathrm{dist}(u,t) \le m/2\}$, and, for any node u, let $V(u) = \{v \mid \mathrm{dist}(u,v) \le v(n)^{1/d}\}$. We have observed that 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 search reaches a sequence $y_1, \ldots, y_k$ of intermediate destinations satisfying that at least one node in $V(y_k)$ has its long-range contact in B, and no node of $V(y_j)$ has its long-range contact in B for $j < k$ (see Fig. 4). Let us compute the expected number of steps required to go from y_j to y_{j+1} using Kleinberg's greedy routing. Let $x_0, x_1, \ldots, x_\ell$ be the sequence of considered intermediate destinations before the search eventually reaches the intermediate destination y_{j+1} starting from y_i. I.e, $x_0 = y_j$ and $x_\ell = y_{j+1}$. Let $r = \mathrm{dist}(x_0, x_1)$ (note $r \le v(n)^{1/d}$ as $x_1 \in V(x_0)$), and let $A = \{u \mid \mathrm{dist}(u, x_1) \le r/2\}$. For every node v such that $\mathrm{dist}(v, x_1) \ge 3r/4$, we have $\Pr(v \to A) \le O(1/\log n)$. Therefore, the probability that a long-range contact is used during the first quarter of the path from x_0 to x_1 is at most $O(r/\log n)$, that is at most $O(v(n)^{1/d}/\log n)$. Thus, with probability $1 - o(1)$, no long-range contacts is used on the path from x_0 to x_1. Since the expected Manhattan distance $\bar{r}$ between x_0 and x_1 is $\Omega(v(n)^{1/d})$, we get that the expected number of steps required to go from x_0 to x_1 using Kleinberg's greedy routing is $\Omega(v(n)^{1/d})$. Actually, the routing does not reach x_1 if a new intermediate destination x_2 is discovered. However, one can easily check that a constant portion of the path from x_0 to x_1 must be traversed before expecting discovering a new intermediate destination. 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, using the same arguments as in the proof of Theorem 1, we prove that the expected number of steps required to go from y_j to y_{j+1} is actually $\Theta(v(n)^{1/d})$ because the sequence $x_0, x_1, \ldots, x_\ell$ is of constant expected length. Since the probability that a long-range contact is used between x_i and x_{i+1} is $o(1)$, 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 of the sequence $y_1, y_2, \ldots, y_k$ is $\Omega(\log n / v(n))$.

To summarize, starting from x at distance m from the target, the search visits an expected number of $\Omega(\log n/v(n))$ intermediate destinations $y_1, \ldots, y_k$, and 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 $\Omega(\log n / v(n)^{1-1/d})$. Now, one can show that, 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 expected Manhattan distance $\Theta(n^{1/d})$ from the target, the expected number of steps to reach a node at distance $< c \cdot \log^{1/d} n$ from the target is $\Omega\left(\frac{\log n}{v(n)^{1-1/d}} \cdot \log\left(\frac{n^{1/d}}{\log^{1/d} n}\right)\right)$, which completes the proof. □

Conversely, if every node is aware of the long-range contacts of its $v(n) \gg \log n$ closest neighbors in the mesh, then 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 3. *If $v(n) = \Omega(\log n)$, then the expected number of steps to reach the target is at least*

$$\Omega\left(\frac{\log n}{\log(v(n)/\log n)} \cdot \min\left\{(\log n) \cdot (\log v(n)),\ v(n)^{1/d}\right\}\right).$$

Proof. We consider first the case $v(n) \ll n$. Assume, in the same spirit as in the proof of Theorem 1, that the distance $m = \text{dist}(x, t)$ between the current node x and the destination t is $\geq c \cdot v(n)^{1/d}$ where c is a constant large enough. Let $B = \{u \mid \text{dist}(u, t) \leq m/2^{r(n)}\}$ where $r(n) = \frac{1}{d}\log(\gamma v(n)/\log n)$ where $\gamma > 0$ is a constant fixed such that $r(n) \geq 1$. From the setting of $r(n)$, one can easily show that $\Pr(V(x) \to B)$ is at least some positive constant. 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

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

Thus, reducing the distance to the target by a factor $2^{r(n)}$ requires $\Omega(\min\{(\log n) \cdot (\log v(n)),\ v(n)^{1/d}\})$ expected number of steps. Therefore, starting from a node at expected distance $\Theta(n^{1/d})$ from the target, the expected number of steps to reach a node at Manhattan distance $< c \cdot v(n)^{1/d}$ from the target is $\Omega\left(\frac{\log n}{r(n)} \cdot \min\left\{(\log n) \cdot (\log v(n)),\ v(n)^{1/d}\right\}\right)$.

If $v(n) = \Theta(n)$, then indirect-greedy routing reduces to Kleinberg's greedy routing since most of the time is spent while routing to an intermediate destination, which is at expected distance $\Omega(n^{1/d})$ from the source. Hence, the expected number of steps to reach the target is $\Omega(\log^2 n)$. □

5. CONCLUSION

In this paper, we proposed a model for the small world phenomenon. This model demonstrates that eclectic relationships are desirable, as far as connectedness to other individuals is concerned. This is coherent 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.

Acknowledgments. The authors are thankful to Maha Abdallah for her careful reading of preliminary versions of this paper, for her comments on its content, and for correcting several grammatical errors.

6. REFERENCES

[1] L. Adamic and E. Adar. How To Search a Social Network. Tech. Report HP Labs, Palo Alto, 2003.

[2] J. Aspnes, Z. Diamadi, and G. Shah. Fault-Tolerant Routing in Peer-to-Peer Systems. In 21st ACM Symp. on Principles of Distributed Computing (PODC), pages 223-232, 2002.

[3] L. Barrière, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient Routing in Networks with Long Range Contacts. In 15th International Symposium on Distributed Computing (DISC), LNCS 2180, pages 270-284, Springer, 2001.

[4] D. Coppersmith, D. Gamarnik, and M. Sviridenko. The Diameter of a Long-Range Percolation Graph. Random Structures and Algorithms 21(1):1-13, 2002.

[5] P. Dodds, R. Muhamad, and D. Watts. An Experimental Study of Search in Global Social Networks. Science 301:827-829, 2003.

[6] P. Killworth and H. Bernard. Reverse Small-World Experiment. Social Networks 1(2):159-192, 1978.

[7] J. Kleinberg. The Small-World Phenomenon: An Algorithmic Perspective. In 32nd ACM Symp. on Theory of Computing (STOC), pages 163-170, 2000.

[8] J. Kleinberg. Navigation in a Small-World. Nature 406:845, 2000.

[9] J. Kleinberg. Small-World Phenomena and the Dynamics of Information. In 15th Neural Information Processing Systems (NIPS), 2001.

[10] E. Lebhar and N. Schabanel. Searching for Optimal paths in long-range contact networks. In 31st International Colloquium on Automata, Languages and Programming (ICALP), 2004.

[11] G. Manku, M. Bawa, and P. Raghavan. Symphony: distributed hashing in a small world. In 4th USENIX Symp. on Internet Technologies and Systems, pages 127-140, 2003.

[12] G. Manku, M. Naor, and U. Wieder. Know Thy Neighbor's Neighbor: The Power of Lookahead in Randomized P2P Networks. In 36th ACM Symp. on Theory of Computing (STOC), 2004.

[13] C. Martel and V. Nguyen. Analyzing Kleinberg's (and other) Small-world Models. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), 2004.

[14] S. Milgram. The Small-World Problem. Psychology Today, 60-67, 1967.

[15] J. Travers and S. Milgram. An Experimental Study of the Small World Problem. Sociometry 32:425, 1969.

[16] D. Watts, P. Dodds, and M. Newman Identity and Search in Social Networks. Science 296:1302-1305, 2002.

[17] D. Watts and S. Strogatz. Collective Dynamics of Small-World Networks. Nature 393:440-442, 1998.