Eclectisme dans les petits mondes†

Pierre Fraigniaud[1], Cyril Gavoille[2], Christophe Paul[3].

[1] *CNRS, LRI, Université Paris Sud, 91405 Orsay cedex, France.* pierre@lri.fr
[2] *LaBRI, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence, France.* gavoille@labri.fr
[3] *CNRS, LIRMM, Université de Montpellier, France.* paul@lirmm.fr

Cet article traite des *petits mondes* tels que définis par Kleinberg (2000), c'est-à-dire des graphes obtenus à partir d'une grille à d dimensions en ajoutant des liens choisis aléatoirement suivant une distribution d-harmonique. Ce modèle a pour objet de donner un support formel à la propriété dite des "six pas de séparation" entre individus déduite de l'expérience de Milgram (1967), et confirmée récemment par Dodds, Muhamad et Watts (2003). Kleinberg a en particulier montré que, pour toute constante $d \geq 1$, le *routage glouton* s'exécute en moyenne en $\Theta(\log^2 n)$ étapes dans une grille "augmentée" à d dimensions et n sommets, avec $O(\log n)$ bits de "connaissance topologique" par sommet. Nous montrons que donner $O(\log^2 n)$ bits de connaissance topologique à chaque sommet permet de décroître le nombre moyen d'étapes du routage glouton jusqu'à $O(\log^{1+1/d} n)$ dans une grille augmentée de d dimensions. Nous montrons également que, indépendamment du volume de connaissance topologique donné à chaque sommet, le routage glouton s'exécute en moyenne en $\Omega(\log^{1+1/d} n)$ étapes. En particulier, augmenter la connaissance topologique au delà de l'optimum de $O(\log^2 n)$ bits diminue les performances du routage glouton. Ainsi, nos résultats démontrent que l'efficacité du routage glouton est sensible à la "dimension du monde", au sens que des mondes de plus grandes dimensions permettent d'accélerer le routage glouton. Ce fait n'est pas observable dans le modèle de Kleinberg. Cet article résume la teneur de ces résultats. Les détails, incluant les démonstrations formelles, peuvent être obtenus en consultant le rapport [5].

Keywords: Petits mondes, routage glouton, distribution harmonique

1 Milgram's Experiment

We consider *augmented* graphs as defined in [15], 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 [12], recently reproduced by Dodds, Muhamad, and Watts [4] (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 because, for chains that reached the target‡, the number of intermediate persons between the source and the target ranged from 2 to 10, with a median of 5.

†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.

‡ Many chains did not succeeded in Milgram's experiment. Experiments by Dodds et al. [4] 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.

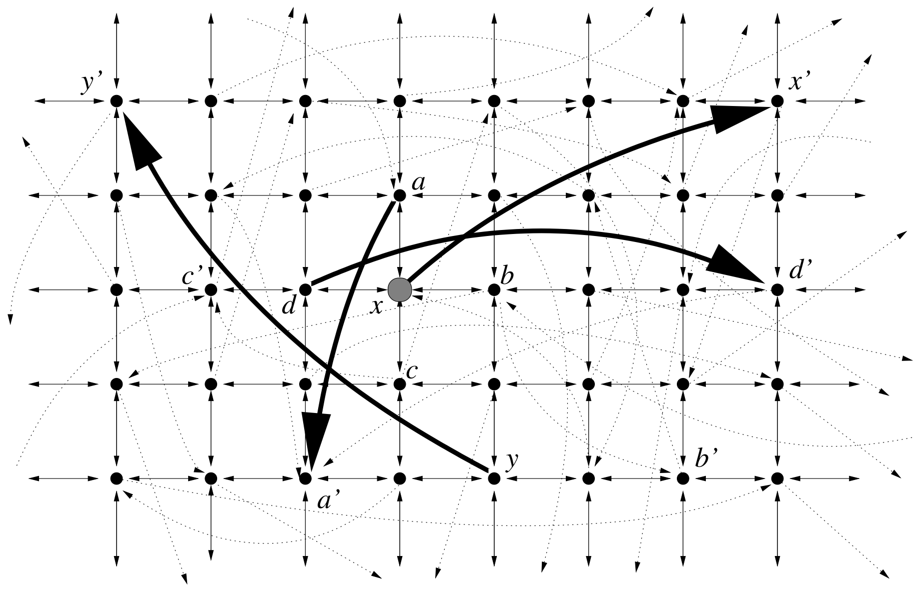

Figure 1: Long-range links in the 2-dimensional mesh. The topological awareness of node x is composed of the four plain long-range links.

2 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. 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 $h(x,y) = 1/(Z_x \cdot \mathrm{dist}(x,y)^d)$ 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[§].

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, 4, 13, 14]. 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.

3 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

[§] Dodds et al. [4] 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.

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[¶], 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 of the greedy routing specification, and not of the links distribution. 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.

4 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. 1). 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 [11] 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

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

	Expected #steps	Amount of awareness (#bits)
Kleinberg's greedy [7]	$O(\frac{1}{c}\log^2 n)$	$c\log n$
Kleinberg's greedy [3]	$\Omega(\frac{1}{c}\log^2 n)$	$c\log n$
Kleinberg's greedy [2]	$\Omega(\frac{1}{c}\log^2 n/\log\log n)$	$c\log n$
NoN-greedy [11]	$O(\frac{1}{c\log c}\log^2 n)$	$c^2\log n$
Decentralized algorithm [10]	$O(\frac{1}{\log^2 c}\log^2 n)$	$c\log n$
Indirect-greedy [This paper]	$O(\frac{1}{c^{1/d}}\log^{1+1/d} n)$	$\log^2 n$

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 at most $O(\log^2 n)$ bits, i.e., $c \leq \log n$ for Kleinberg's greedy routing and Decentralized algorithm, and $c \leq \sqrt{\log n}$ for NoN-greedy. For $c = \log n$, indirect-greedy performs in $O(\log n)$ steps, as Kleinberg's greedy routing. 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. [3] 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.

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 at most $O(\log^2 n)$ bits, our protocol is the fastest.

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). See Figure 2 for a graphic version of these results.

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*. 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.

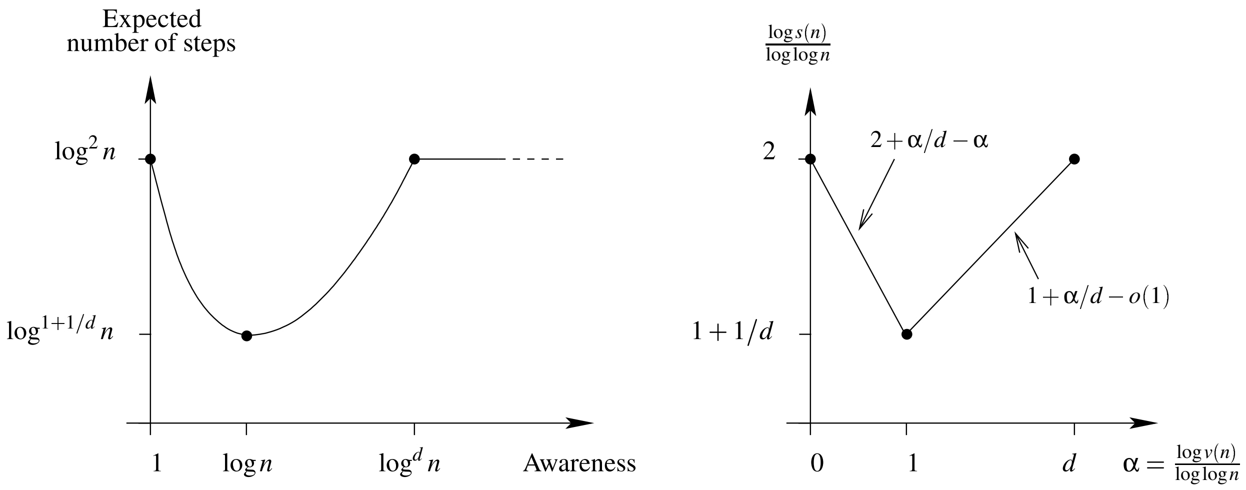

Figure 2: The expected number of steps v.s. the awareness. For $v(n) = (\log n)^\alpha$, the expected number of steps is $s(n) = \Omega((\log n)^{2+\alpha/d-\alpha})$ if $\alpha < 1$, and is $s(n) = \Omega((\log n)^{1+\alpha/d-o(1)})$ if $1 \leq \alpha \leq d$. For $\alpha \geq d$, $s(n) = \Theta(\log^2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

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

[5] P. Fraigniaud, C. Gavoille, and C. Paul. Eclectism Shrinks the World. Tech. Report LRI-1376, Univ. Paris Sud, 2003. (see http://www.lri.fr/˜pierre.)

[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. Tech. Report LIP-2003-55, ENS Lyon, 2003. (see http://www.ens-lyon.fr/LIP/.)

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

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

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

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

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