A New Perspective on the Small-World Phenomenon: Greedy Routing in Tree-Decomposed Graphs

Pierre Fraigniaud*
CNRS
LRI, University of Paris Sud
91405 Orsay, France
pierre@lri.fr

Abstract

Milgram's experiment (1967) demonstrated that there are short chains of acquaintances between individuals, and that these chains can be discovered in a greedy manner. Kleinberg (2000) gave formal support to this so-called "small world phenomenon" by using meshes augmented with long-range links chosen randomly according to harmonic distributions. In this paper, we propose a new perspective on the small world phenomenon by considering arbitrary graphs augmented according to distributions guided by tree-decompositions of the graphs. We show that, for any n-node graph G of treewidth $\leq k$, there exists a tree-decomposition-based distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(k \log^2 n)$ expected number of steps. We argue that augmenting a graph with long-range links chosen according to a tree-decomposition-based distribution is plausible in the context of social networks. However, social networks can have unbounded treewidth. Nevertheless, we note that these networks have few long chordless cycles because of their high clustering coefficient. We prove that if G has chordality $\leq k$, then the tree-decomposition-based distribution $\mathcal{D}$ insures that greedy routing in $(G, \mathcal{D})$ performs in $O((k + \log n) \log n)$ expected number of steps. In particular, for any n-node graph G of chordality $O(\log n)$ (e.g., chordal graphs), greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(\log^2 n)$ expected number of steps. This latter result stresses the fact that our model may well explain why greedy routing is so efficient in social networks, such as observed in Milgram's experiment.

Keywords: Milgram's Experiment, Social Networks, Complex Networks.

Technical Report LRI-1397 – February 2005
`Laboratoire de Recherche en Informatique`
`UMR CNRS 8623 - Université Paris Sud`
`Bâtiment 490 - 91405 Orsay cedex - France`

*The author received additional supports from the project "PairAPair" of the ACI Masses de Données, from the project "Fragile" of the ACI Sécurité Informatique, and from the project "Grand Large" of INRIA.

1 Introduction

In his seminal work [20], Kleinberg gave a formal support to the "six degrees of separation" phenomenon, defined after the Milgram's experiment [28], recently reproduced by Dodds, Muhamad, and Watts [11]. This experiment demonstrated that there are short chains of acquaintances between individuals, and that these chains can be discovered in a greedy manner. More precisely, 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 target is identified by its name, its professional occupation, and by the US state of its home town. 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-name 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.

Expanding on [31], Kleinberg modeled Milgram's experiment as follows (cf. [20, 21]). Let $\mathcal{M}$ be the set of all 2-dimensional square meshes (i.e., the $n \times n$ grids, for $n \geq 1$). For $M \in \mathcal{M}$, every node x of M is given an additional directed link pointing to some node y. The head y of the added link (x, y), called the *long-range contact* of x, is chosen according to the 2-harmonic distribution $\mathcal{H}$, i.e., the probability that x chooses y, $y \neq x$, as long-range contact, is $\texttt{Prob}_x(y) = 1/(H_x \cdot \mathrm{dist}^2(x, y))$ where $\mathrm{dist}(x, y)$ denotes the Manhattan distance between x and y in M, and $H_x = \sum_{y \neq x} 1/\mathrm{dist}^2(x, y)$ is a normalizing coefficient. The resulting graph is called an *augmented* mesh, and the set of graphs M augmented by $\mathcal{H}$ is denoted by $(M, \mathcal{H})$. Then, Kleinberg defined *greedy routing* in any graph of $(M, \mathcal{H})$ as the following process: Given a target node t, and a current node x, x selects among all its neighbors (including its long-range contact) the one that is closest to t in the mesh M (i.e., according to the Manhattan distance), and forwards to this neighbor. Kleinberg proved that greedy routing in the n-node mesh augmented with long-range links set according to the 2-harmonic distribution performs in $O(\log^2 n)$ expected number of steps.

In Kleinberg's model, the choice of the 2-harmonic distribution for the 2-dimensional meshes is crucial. Indeed, Kleinberg also proved that greedy routing in 2-dimensional meshes augmented with the k-harmonic distribution $\texttt{Prob}_x(y) = 1/(H_x^{(k)} \cdot \mathrm{dist}^k(x, y))$ where $H_x^{(k)} = \sum_{y \neq x} 1/\mathrm{dist}^k(x, y)$, performs poorly if $k \neq 2$, i.e., in $\Omega(n^\alpha)$ expected number of steps, for some $\alpha > 0$ that depends on k. Therefore, finding the right distribution for 2-dimensional meshes was far from being obvious, and there is no distribution $\mathcal{D}$ for which greedy routing in $(M, \mathcal{D})$ is known to perform in $O(\mathrm{polylog}(n))$ expected number of steps, but (distributions structurally equivalent to) the 2-harmonic distribution. More generally, the design of an appropriate distribution $\mathcal{D}$ for an arbitrary given graph G seems to be uneasy. Formally, we rise the following question:

Problem 1 *For any n-node graph $G = (V, E)$, is there a distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(\mathrm{polylog}(n))$ expected number of steps?*

By greedy routing in the augmented graph $(G, \mathcal{D})$, it is meant the following process:

Definition 1 *(Greedy Routing.) For any target node $t \in V$, the current node $x \in V$ selects among all its neighbors (including its long-range contact chosen according to $\mathcal{D}$) the neighbor y that is closest to t in the underlying graph G, and forwards to y.*

Note that it is not sufficient to place a graph with small diameter on top of the underlying graph G for greedy routing to perform in a small number of steps. Indeed, greedy routing optimizes the choice of the current node's neighbor according to a distance measured in G, and not in the graph including the long-range links.

Beside its own theoretical interest as a natural generalization of Kleinberg's work on the mesh, solving Problem 1 would have a significant impact on our understanding of routing in social networks, as illustrated by Milgram's experiment. Indeed, although Kleinberg's model is a powerful tool for analyzing greedy routing strategies, there is no evidence that the network formed by social acquaintances looks like an augmented mesh. There are however some evidences that the social entities share a common knowledge about their relative distances, based on their geographical positions, on their professional occupations, on their hobbies, or on any criteria available to the entities. This common knowledge could be modeled by a graph G. Then, the random events of life create connections between individuals who have, a priori, very little in common. This could be modeled by random links added on top of the underlying graph G. Therefore there is some support to the hypothesis that a social network can reasonably be modeled by an augmented graph $(G, \mathcal{D})$. Still, this gives rise to several questions: what is the graph G? What is the distribution $\mathcal{D}$? Why the long-range links are structured according to some specific distribution rather than to another? By considering Problem 1, this paper is an attempt to solve these questions.

Last but not least, solving Problem 1 by the affirmative would have some impact on routing in fully decentralized P2P systems. Indeed, given a set of peers connected according to some topology G, then, in the same spirit as in [24] for the ring, it would be possible to enhance G with additional links so that a simple greedy search procedure would proceed in a polylogarithmic expected number of steps, with no flooding nor any use of sophisticated routing protocols.

1.1 Our Results

First, we address Problem 1 in tree-decomposed graphs. Informally, the *treewidth* of a graph measures how far the graph is from a tree. Graphs of bounded treewidth form a large class of graphs, including trees, outer-planar graphs, series-parallel graphs, etc. In addition to their connection to the graph-minor theory (cf, e.g., [30]), they have a wide range of applications in graph searching [29] and routing [15, 16]. They also play a central role in complexity and logic. In particular, it is known that several NP-hard problems can be solved in polynomial time if instances are restricted to graphs of bounded treewidth [2, 5]. Actually, on graphs of treewidth at most k, where k is fixed, every decision or optimization problem expressible in monadic second-order logic has a linear algorithm [10]. We show that, for any n-node graph G of treewidth $\mathtt{tw}(G)$, there exists a *tree-decomposition-based* distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G, \mathcal{D})$ performs in

$$O(\mathtt{tw}(G) \log^2 n) \tag{1}$$

expected number of steps. In particular, for graphs of bounded treewidth, there exists a tree-decomposition-based distribution such that greedy routing in the augmented graph performs in $O(\log^2 n)$ expected number of steps. This latter bound is close to optimal as it is known that, in the n-node directed ring, no distribution enables greedy routing to perform better than $\Omega(\log^2 n/ \log\log n)$ expected number of steps [3]. We also give a constructive variant of our result. More precisely, given any n-node graph G, we show how to construct (in polynomial time) a distribution $\mathcal{D}$ such that

Underlying graph	Distribution	Expected #steps	References
d-dimensional meshes	d-harmonic	$O(\log^2 n)$	[20]
d-dimensional meshes	k-harmonic, $k \neq d$	$\Omega(n^\alpha), \alpha > 0$	[20]
ring	1-harmonic	$\Omega(\log^2 n)$	[4]
directed ring	any	$\Omega(\log^2 n / \log\log n)$	[3]
d-dimensional meshes, $d > 1$	d-harmonic	$\Omega(\log^2 n)$	[17]
moderate growth graphs	1/ball-size	$O(\text{polylog}(n))$	[13]
graphs of treewidth $\leq k$	tree-decomposition-based	$O(k \log^2 n)$	[this paper]
graphs of chordality $\leq \gamma$	tree-decomposition-based	$O((\gamma + \log n) \log n)$	[this paper]

Table 1: Performances of (pure) greedy routing (with 1 long-range contact per node)

greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(\texttt{tw}(G) \cdot \log \texttt{tw}(G) \cdot \log^2 n)$ expected number of steps.

Next, we focus our attention on social networks, and argue that our tree-decomposition-based distribution is plausible in this context, i.e., a social network is well modeled by a graph G augmented with long-range links chosen according to a tree-decomposition-based distribution $\mathcal{D}$. In particular, as opposed to hierarchical models which define the hierarchical structure *a priori* (cf., e.g., [22]), the hierarchy of our model is inherited from the natural structure of the social network.

However, social networks do not necessarily have bounded treewidth. Thus the bound of Equation 1 does not directly explain why greedy routing performs so well in the Milgram's experiment. Nevertheless, social networks possess specific topological properties which strongly impact the performances of greedy routing. In particular social networks generally have high *clustering coefficient*, a parameter measuring the probability that two nodes having a common neighbor be neighbors too. A high clustering coefficient implies that the network contains a lot of triangles. More generally, a high clustering coefficient implies that a cycle of more than three nodes is likely to have a chord. This property motivated us to investigate greedy routing in graphs of bounded *chordality* (the chordality is the length of the longest chordless cycle). We prove that if G has chordality γ, then the tree-decomposition-based distribution $\mathcal{D}$ insures that greedy routing in $(G, \mathcal{D})$ performs in

$$O((\gamma + \log n) \log n) \tag{2}$$

expected number of steps. In particular, for any n-node graph G of chordality $O(\log n)$ (e.g., chordal graphs), greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(\log^2 n)$ expected number of steps, where $\mathcal{D}$ is the tree-decomposition-based distribution. It is important to note that, as opposed to Equation 1, the performances of greedy routing in graphs of bounded chordality are independent from the treewidth of these graphs, although the treewidth of n-node chordal graphs can take any value between 1 and $n - 1$.

Since high clustering coefficient implies that, in average, long cycles have chords, *these results stress the fact that our model may well explain why greedy routing is so efficient in social networks*, such as observed in Milgram's experiment.

All known complexity results (including ours) relative to the performances of greedy routing in graphs augmented with one long-range contact per node are summarized in Table 1. This table does not list results related to variants of greedy routing, such as the ones mentioned in Section 1.2.

1.2 Related Works

Several authors expanded on [20, 21]. In [4], it is shown that the $O(\log^2 n)$ upper bound of [20] is tight in the ring augmented with the 1-harmonic distribution, i.e., greedy routing performs in $\Omega(\log^2 n)$ expected number of steps. More generally, [3] shows that in the directed ring augmented with *any* distribution, greedy routing performs in $\Omega(\log^2 n/\log\log n)$ expected number of steps. In [23], a decentralized routing algorithm for augmented meshes is described. The routing visits $O(\log^2 n)$ nodes, and distributively discovers routes of expected length $O(\log n(\log\log n)^2)$ links using headers of size $O(\log^2 n)$ bits. Neighbor-of-neighbor greedy routing defined in [9, 25] performs in $O(\frac{1}{c\log c}\log^2 n)$ expected number of steps, with c long-range contacts per node. The non-oblivious routing protocol described in [26] performs in $O(\log^{1+1/d} n)$ expected number of steps in the d-dimensional mesh. The oblivious Indirect-greedy routing protocol described in [17] performs in $O(\log^{1+1/d} n)$ expected number of steps in the d-dimensional mesh. [17] also shows that the $O(\log^2 n)$ upper bound of [20] is tight in the d-dimensional mesh augmented with the d-harmonic distribution, for any $d \geq 1$. [13] generalizes Kleinberg's result to the family of "moderate growth graphs", namely the graphs such that, roughly speaking, the size of the ball of radius r centered at any node x is equal to $r^{d_x(r)}$ where d_x is a function that is $\mathcal{C}^1$ and whose derivative is in $O(1/(r\log r))$. Finally, [27] recently proposed several constructions of small worlds, based on adding links with probability proportional to the inverse distance. These constructions generalize both [20] and [22] (reference [22] will be discussed in more detail in Section 4).

2 Definitions and Notations

2.1 Performances of Greedy Routing

Let $G = (V, E)$ be a connected graph with n nodes, and let $\mathcal{D} = \{\mathtt{Prob}_x,\ x \in V\}$, where, for any $x \in V$, $\mathtt{Prob}_x$ is a probability distribution on $V \setminus \{x\}$. An augmentation of G according to $\mathcal{D}$ is a graph obtained from G by adding at every node $x \in V$ one directed edge (x, y) where y is chosen with probability $\mathtt{Prob}_x(y)$. For every ordered pair $(s, t) \in V \times V$, let $\mathtt{X}_{s,t}$ be the random variable specifying the number of steps required by greedy routing to go from s to t in the augmented graph. Let $\mathtt{EX}_{s,t}$ be the expected value of $\mathtt{X}_{s,t}$. Kleinberg proved that, if G is a 2-dimensional square mesh, and $\mathcal{D}$ is the 2-harmonic distribution, then $\mathtt{EX}_{s,t} = O(\log^2 n)$. Greedy routing insures that, at every step, one gets closer in G to the target, i.e.:

Fact 1 *If x is the current node, and greedy routing forwards to y, then $dist_G(y, t) < dist_G(x, t)$ where $dist_G()$ is the distance function in the underlying graph G.*

As a consequence, greedy routing has no loop, and it requires at most $\mathrm{dist}_G(s, t)$ steps to go from s to t. In particular, in graphs with polylogarithmic diameter, there is no need to add long-range contacts for greedy routing to perform in polylogarithmic number of steps.

2.2 Treewidth

A *tree-decomposition* of graph G is a pair (T, X) where T is a tree, and $X = \{X_v, v \in V(T)\}$ is a collection of subsets of $V(G)$ satisfying the following three conditions:

- `C1`: $V(G) = \cup_{v \in V(T)} X_v$;
- `C2`: For any edge e of G, there is a set X_v such that both end-points of e are in X_v;
- `C3`: For any triple u, v, w of nodes in $V(T)$, if v is on the path from u to w in T, then $X_u \cap X_w \subseteq X_v$.

Condition `C3` can be rephrased as: for any node x of G, $\{v \in V(T) \mid x \in X_v\}$ is a subtree of T. The sets X_vs are called *bags*. The *width*, $\omega(T, X)$, of a tree-decomposition (T, X) is defined as $\max_{v \in V(T)} |X_v| - 1$, i.e., the width of (T, X) is roughly the maximum size of its bags. The treewidth $\mathtt{tw}(G)$ is defined as $\min \omega(T, X)$ where the minimum is taken over all tree-decompositions (T, X) of G. For instance, trees have treewidth 1, cycles have treewidth 2, and n-node cliques have treewidth $n - 1$.

Internal bags of a (non-artificial) tree-decomposition are separators of the graph. In fact, a tree-decomposition can be obtained by recursively separating the graph. (This is essentially the way treewidth is $O(\log n)$-approximated in [6, 7].) We will intensively use this fact throughout all the paper. Let (T, X) be a tree-decomposition of a graph G. Let x and y be two nodes of G, and let b be a bag of T containing neither x nor y, i.e., $b \cap \{x, y\} = \emptyset$. Removing b from T results in a forest of $k \geq 1$ trees $T_1, \ldots, T_k$. Since $b \cap \{x, y\} = \emptyset$, `C1` and `C3` imply that there is a unique i (resp., j) in $\{1, \ldots, k\}$ such that x (resp., y) belongs to some bag(s) of T_i (resp., T_j). Assume that $i \neq j$, then the following is folklore:

Fact 2 *The bag b is an (x, y)-separator in G (i.e., all paths from x to y in G go through some node(s) in b).*

3 Tree-Decomposition-Based Long-Range Contacts Distribution

This section is dedicated to the definition of the tree-decomposition-based distribution of the long-range contacts, and to the proof of the following result:

Theorem 1 *For any connected n-node graph G of treewidth $\leq k$, there is a distribution $\mathcal{D}$ such that, for any source-destination pair (s, t), $\mathtt{EX}_{s,t} = O(k \log^2 n)$.*

Corollary 1 *For any connected n-node graph G of bounded treewidth, there is a distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(\log^2 n)$ expected number of steps.*

Corollary 1 is close to optimal since [3] shows that greedy routing performs in $\Omega(\log^2 n / \log \log n)$ expected number of steps in the directed ring augmented with *any* distribution. In both Theorem 1 and Corollary 1, the distribution $\mathcal{D}$ is a tree-decomposition-based distribution, as defined in the proof bellow.

Proof of Theorem 1. For any $k \geq 2$, let $\mathcal{G}_k$ be the class of connected graphs of treewidth $< k$. Let $G \in \mathcal{G}_k$ be a graph of n nodes, and let T be a tree-decomposition of G, of width $< k$. We can choose T with at most n bags (cf., e.g., Theorem 4.8 and Proposition 4.16 in [19]). In order to describe the distribution $\mathcal{D}$, we describe the $\mathcal{D}_x$s, i.e., we describe the setting of the long-range contact of every node x in G. Recall that a *centroid* of an r-node tree is a node whose removal from the tree results in a forest with at most $r/2$ nodes in each subtree. A tree has either one or two centroids, and if a tree has two centroids, then they are neighbors.

Let c be a centroid of T. For every node $x \in V$, let us denote by $\widehat{x}$ the bag containing x that is closest to c in T. Note that, by C3 of the treewidth definition, $\widehat{x}$ is uniquely defined. We set $c_x^{(0)} = c$, and define $T_x^{(1)}$ as the subtree of $T \setminus \{c_x^{(0)}\}$ containing $\widehat{x}$. Then, let $c_x^{(1)}$ be a centroid of $T_x^{(1)}$, and let $T_x^{(2)}$ be the subtree of $T_x^{(1)} \setminus \{c_x^{(1)}\}$ containing $\widehat{x}$. And so on. One constructs in this way two sequences

$$(T_x^{(0)}, T_x^{(1)}, \ldots, T_x^{(q_x)}) \text{ and } (c_x^{(0)}, c_x^{(1)}, \ldots, c_x^{(q_x)})$$

where (cf. Fig. 1):

1. $T_x^{(0)} = T$;
2. $c_x^{(i)}$ is the centroid of $T_x^{(i)}$ closest to c in T;
3. $T_x^{(i+1)}$ is the subtree of $T_x^{(i)} \setminus \{c_x^{(i)}\}$ containing $\widehat{x}$;
4. $c_x^{(q_x)} = \widehat{x}$.

Note that since $|T| \leq n$, and $|T_x^{(i+1)}| \leq |T_x^{(i)}|/2$, we get that both sequences are of length $q_x + 1 \leq \log n$. The result hereafter directly follows from the definition of these two sequences.

Lemma 1 *For any two nodes u and v, and for any index i, if $\widehat{v} \in T_u^{(i)}$, then $c_v^{(j)} = c_u^{(j)}$ for $j = 0, \ldots, i$, and $c_v^{(j)} \in T_u^{(i)}$ for $j = i, \ldots, q_v$.*

Tree-decomposition-based distribution $\mathcal{D}$. Node x picks its long-range contact as follows:

- First x selects an index $i \in \{0, \ldots, q_x\}$ with $\mathtt{Prob}_x(i) = 1/(q_x + 1)$;
- Next, x selects a node y chosen uniformly at random in the bag $c_x^{(i)}$.

Node y is the long-range contact of x (cf. Fig. 1).

We show that with this setting of the long-range contacts, for any source node s, and any target node t, $\mathtt{EX}_{s,t} = O(k \log^2 n)$. Let G^+ be an instance of the graph G augmented with the long-range contacts set as above. Note that G^+ is directed since the edge from a node to its long-range contact is directed (edges of the underlying graph G remain undirected). Let $t \in V(G)$, and let $i \in \{1, \ldots, q_t\}$. Let

$$U_i = \{v \in V(G) \mid \widehat{v} \in T_t^{(i)}\}.$$

Lemma 2 *The node-set $\cup_{j=0}^{i-1} c_t^{(j)} \subseteq V(G)$ separates U_i and $V(G) \setminus U_i$ in G^+, i.e., any path in G^+ from a node in U_i to a node in $V(G) \setminus U_i$ goes through a node in $\cup_{j=0}^{i-1} c_t^{(j)}$.*

Proof. Let P be a path from a node in U_i to a node in $V(G) \setminus U_i$. Let $e = (v, w)$ be an edge of P from $v \in U_i$ to $w \in V(G) \setminus U_i$. If $v \in c_t^{(j)}$ for some $j \in \{0, \dots, i-1\}$, then we are done. Thus assume $v \notin \cup_{j=0}^{i-1} c_t^{(j)}$. Since $w \notin U_i$, we have $\widehat{w} \notin T_t^{(i)}$. We consider separately the case where e is an edge of G, from the case where e is a long-range link.

If $e \in E(G)$, then let b be a bag containing both v and w (this bag exists from C2). On the one hand, by C3, w belongs to all bags on the path in T from b to $\widehat{w}$. On the other hand, we have b further from c than $\widehat{v}$, i.e., $\widehat{v}$ is on the path from b to c in T. Now, by construction of the sequence $\{c_t^{(j)}, 0 \leq j \leq q_t\}$, the neighborhood of $T_t^{(i)}$ in T (i.e., the set of bags not in $T_t^{(i)}$ but adjacent to some bag in $T_t^{(i)}$) is included in $\cup_{j=0}^{i-1} c_t^{(j)}$. Hence $b \in T_t^{(i)}$ since otherwise v would belong to some bag of the neighborhood of $T_t^{(i)}$ (by C3), which would imply $v \in \cup_{j=0}^{i-1} c_t^{(j)}$. Since $\widehat{w}$ is closer to c than b, but $\widehat{w} \notin T_t^{(i)}$, there is some $c_t^{(j)}$, $j \in \{0, \dots, i-1\}$ on the path from b to $\widehat{w}$ in T. Therefore, $w \in \cup_{j=0}^{i-1} c_t^{(j)}$, proving Lemma 2.

If $e \notin E(G)$, then w is the long-range contact of v. By the setting of the long-range contacts, $w \in \cup_{j=0}^{q_v} c_v^{(j)}$. By Lemma 1, all bags $c_v^{(j)}$ for $j \geq i$ are nodes of $T_t^{(i)}$. If node w belongs to some bag b of $T_t^{(i)}$, then, as in the case $e \in E(G)$, combining C3 with the fact that $\widehat{w} \notin T_t^{(i)}$ yields $w \in \cup_{j=0}^{i-1} c_t^{(j)}$, and we are done. Thus assume that w does not belong to any bag of $T_t^{(i)}$. Therefore $w \in \cup_{j=0}^{i-1} c_v^{(j)}$. From Lemma 1, since $\widehat{v} \in T_t^{(i)}$, $c_v^{(j)} = c_t^{(j)}$ for $j = 0, \dots, i$. Therefore $w \in \cup_{j=0}^{i-1} c_t^{(j)}$, which completes the proof of Lemma 2. ◇

Let $(T_t^{(0)}, T_t^{(1)}, \dots, T_t^{(q_t)})$ and $(c_t^{(0)}, c_t^{(1)}, \dots, c_t^{(q_t)})$ be the sequences of subtrees and centroids corresponding to the target t. Let x be the current node. (Initially, x is the source node s.) Let i be the largest index such that $\widehat{x} \in T_t^{(i)}$. Let $x_0, x_1, \dots, x_r$ be the sequence of nodes visited by greedy routing from $x = x_0$ until either it reaches a node x_r with $\widehat{x_r} \notin T_t^{(i)}$, or it reaches t.

Lemma 3 *Let $y \in \cup_{j=0}^{i} c_t^{(j)}$. For every $\ell = 0, \dots, r-1$, the probability that y is the long-range contact of x_ℓ is at least $1/(k \log n)$. The probability that the long-range contact of x is in $c_t^{(j)}$ is at least $1/\log n$ for every $j = 0, \dots, i$.*

Proof. We have $\widehat{x_\ell} \in T_t^{(i)}$ for every $\ell < r$. Thus, from Lemma 1, for any $\ell < r$, $c_{x_\ell}^{(j)} = c_t^{(j)}$ for $j = 0, \dots, i$. Therefore, every node x_ℓ, $\ell < r$, has its long-range contact in a specific bag $c_t^{(j)}$, $0 \leq j \leq i$, with probability $1/(1 + q_{x_\ell})$. A node $y \in c_t^{(j)}$ for some $j \leq i$ is the long-range contact of x_ℓ with probability at least $1/(|c_t^{(j)}|(1 + q_{x_\ell}))$. Since $|c_t^{(j)}| \leq k$, and $1 + q_{x_\ell} \leq \log n$, Lemma 3 follows. ◇

Lemma 4 *The path from s to t constructed by greedy routing does not visit any bag $c_t^{(0)}, \dots, c_t^{(q_t)}$ more than k times.*

Proof. Since T has width $< k$, no bag contains more than k nodes. Thus, from Fact 1, no bag can be visited by greedy routing more than k times on the way from s to t. This is true in particular for bags $c_t^{(0)}, \dots, c_t^{(q_t)}$. ◇

Finally, we will make use of the following simple result. Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables in $\{0,1,\dots,N\}$ with

$$\begin{array}{ll}\texttt{Prob}(\{X_i = j\}) = p/N & \text{if } j \in \{1,\dots,N\};\\ \texttt{Prob}(\{X_i = 0\}) = 1-p; & \end{array}$$

for some $0<p<1$. We consider the following iterative process. Let $S_0=\{b_1,\dots,b_N\}$ be a set of N non negative integers. After the ith trial, if $X_i>0$, then all integers $b_j \geq b_{X_i}$ in the current set are removed, i.e., $S_i = S_{i-1} \setminus \{b_j \mid b_j \geq b_{X_i}\}$. Let $\texttt{Y}$ be the random variable specifying the number of trials i until S_i becomes empty.

Lemma 5 $\texttt{EY} \leq N/p$.

Proof. The set becomes empty after the first trial i such that $X_i = j$, where $b_j = \min_\ell b_\ell$. This occurs with probability at least p/N. $\diamond$

Let P be the path followed by greedy routing from s to t in G^+. We decompose P into a sequence of subpaths $P_0P_1P_2\dots P_{q_t}$ where the first node of P_0 is s, the last node of P_{q_t} is t, and, for every $i=0,1,\dots,q_t$, $P_i \subseteq T_t^{(i)}$ and is minimal for that property. More explicitely, let $P=x_0,\dots,x_r$ with $x_0=s$ and $x_r=t$. For $i=0,\dots,q_t$, let a_i be the smallest index such that, for every $j\geq a_i$, $\widehat{x_j} \in T_t^{(i)}$. In particular, $a_0=0$ since $\widehat{s} \in T = T_t^{(0)}$ and every node of G belongs to some bag of T. Similarly, $a_{q_t} \leq r$ since greedy routing eventually reaches $t \in c_t^{(q_t)} \in T_t^{(q_t)}$. We define P_i as the path in G which starts at x_{a_i}, and ends at $x_{a_{i+1}-1}$, but P_{q_t} which ends at t. (If $a_{i+1}=a_i$, then P_i is the empty path.) We have:

$$|P| \;=\; \sum_{i=0}^{q_t} |P_i|. \tag{3}$$

Let $i \in \{0,1,\dots,q_t\}$, and consider P_i. By definition, while traveling along P_i, greedy routing never goes out of $T_t^{(i)}$. Thus, from Lemma 2, it does not visit nodes x such that $\widehat{x} \in \cup_{j=0}^{i-1} c_t^{(j)}$. P_i may however go in and out of $T_t^{(i+1)}$. From Lemma 2, the only way P_i goes in and out of $T_t^{(i+1)}$ is through $c_t^{(i)}$. From Lemma 3, for each node x of P_i, the long-range contact y of x is in $c_t^{(i)}$ with probability at least $1/\log n$. Assume success, i.e., $y \in c_t^{(i)}$. From Fact 1, no node z with $\text{dist}_G(z,t) > \text{dist}_G(y,t)$ will be ever visited by greedy routing after x. In particular, no node $z \in c_t^{(i)}$ with $\text{dist}_G(z,t) > \text{dist}_G(y,t)$ will be ever visited by greedy routing after x. In the same spirit as for Lemma 5, we just say that those nodes y and $z \in c_t^{(i)}$ are "removed". We are in the situation of Lemma 5 with $p \geq 1/\log n$, and $N = |c_t^{(i)}| \leq k$. Thus, after an expected number of at most $O(k\log n)$ trials, all nodes in $c_t^{(i)}$ are removed. Therefore, from Lemma 4, after this expected amount of trials, no nodes of $c_t^{(i)}$ will be ever visited by greedy routing. Hence, once in P_i, the path P enters P_{i+1} after at most $O(k\log n)$ expected number of steps. In other words, the expected length of P_i is $O(k\log n)$. Therefore, from Eq. 3, the expected length of the path P is at most $O(q_t k \log n) \leq O(k\log^2 n)$ which completes the proof of Theorem 1. $\blacksquare$

Theorem 1 is an existential result. Nevertheless, a combination of this theorem with known results from the literature allows us to explicitly construct a long-range contact distribution for any graph G. This is however to the price of a $\log \mathtt{tw}(G)$ factor in the performances of greedy routing. More precisely, a tree-decomposition T of any graph G, with width $\leq O(\mathtt{tw}(G) \log \mathtt{tw}(G))$, can be computed in polynomial time (see, e.g., [7]). Once this is done, since the distribution $\mathcal{D}$ in Theorem 1 can obviously be computed in polynomial time, we get:

Corollary 2 *There is an polynomial time algorithm that, for any connected n-node graph G of treewidth $\leq k$, computes a distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G, \mathcal{D})$ performs in $O(k \log k \log^2 n)$ expected number of steps.*

4 The Case of Social Networks

4.1 Substratum of Tree-Decomposition

As already mentioned in the Introduction, social entities share a common knowledge about their relative distances, based on their geographical positions, on their professional occupations, on their hobbies, or on any criteria available to the entities. This common knowledge could be modeled by a graph G. The random events of life create connections between individuals who have, a priori, very little in common. This could be modeled by random links added on top of the underlying graph G. This is essentially the main acknowledged justification to the hypothesis that a social network can reasonably be modeled by an augmented graph $(G, \mathcal{D})$ (cf., e.g., [17, 21, 23]). Hereafter, we argue that, in addition, $\mathcal{D}$ can reasonably be assumed to be a tree-decomposition-based distribution.

For each criterion, the individuals can be grouped in large families. For instance: Africans, Americans, Europeans, etc., or artists, scientists, farmers, etc. This can be done recursively. For instance, Europeans can be grouped according to their countries of leaving, while scientists can be grouped according to their scientific domains. And so on. This clustered and hierarchical structure of the social networks was already pointed out by several authors (cf., e.g., [1, 8, 12, 14, 31]). The model in [22] was the first model specified to capture this hierarchy (see also [27]). However, this model assumes that the hierarchy is induced by a specific structured graph, defined a priori. (More precisely, in the model, nodes are leaves of a complete b-ary tree, and the downer is the lowest common ancestor of two nodes, the more likely these two nodes are to be connected by a long-range link.) Moreover the model in [22] reflects one type of hierarchy only (e.g., arts/music/opera) whereas social entities belong to several *interleaved* hierarchies such as those based on the place of living, the professional activity, the recreative activity, etc. In contrast, a tree-decomposition of the "natural" acquaintances (i.e., the acquaintances described by the graph G) determines a hierarchy that is inherited from these acquaintances, and not specified a priori. This hierarchy is the expression of *all* the underlying interleaved hierarchies. (This "general" hierarchy could be interpreted a posteriori in the same way one interprets the result of a Principal Components Analysis, but this is beyond the scope of this paper.) The long-range contacts enable jumping across this hierarchy, as illustrated on Fig. 1, and one may jump upwards as well as downwards across the hierarchy. In addition, the hierarchy is viewed differently from each node. In particular, nodes that are placed far apart in the tree-decomposition have very different views of the hierarchy.

4.2 Clustered Networks

Social networks are known to have high clustering coefficient [31]. That is, two social entities having both an acquaintance with a third one are likely to know each other. A high clustering coefficient limit the number of long chordless cycles in the acquaintance graph, simply because the probability that a cycle of length ℓ be chordless is at most $(1-p)^\ell$ where p is the probability that two nodes sharing a common neighbor be connected. This motivated us to study greedy routing using tree-decomposition-based long-range contact distributions in graphs of bounded *chordality*. Formally, the chordality of a graph G is the maximum length of a chordless cycle in G. In particular, a graph of chordality 3 is a *chordal* graph.

Theorem 2 *For any connected graph G of n nodes and chordality γ, there is a tree-decomposition-based distribution $\mathcal{D}$ enabling greedy routing in $(G,\mathcal{D})$ to perform in $O((\gamma+\log n)\log n)$ expected number of steps, for any source-destination pair.*

The proof uses the same arguments as in the proof of Theorem 1, combined with the following two additional facts.

1. For any n-node connected graph of chordality γ, there is a tree-decomposition with at most n bags such that two nodes in the same bag are at distance at most $\gamma/2$ (cf., e.g., [18]).
2. Let $x_0 = s, x_1, x_2, \ldots, x_r = t$ be the path followed by greedy routing from s to t; If $\mathrm{dist}_G(x_i, x_j) \leq d$ then $|i-j| \leq d$.

Note that the result of Theorem 2 is independent from the treewidth of the graph. It has an important consequence:

Corollary 3 *For any n-node graph G of chordality $O(\log n)$ (in particular for any chordal graph), there is a tree-decomposition-based distribution $\mathcal{D}$ such that greedy routing in the augmented graph $(G,\mathcal{D})$ performs in $O(\log^2 n)$ expected number of steps.*

5 Concluding Remark

Assuming that social networks are well modeled by graphs of small chordality augmented with long-range links set according to a tree-decomposition-based distribution, Corollary 3 may well explain why greedy routing is so efficient in social networks, such as observed in Milgram's experiment.

Acknowledgments: The author is thankful to Cyril Gavoille, Emmanuelle Lebhar, and Nicolas Nisse, for their help and comments.

References

[1] M. Aldenderfer and R. Blashfield. Cluster Analysis. Quantitative Applications in the Social Sciences, Vol. 44, SAGE Publications, London, 1984.

[2] S. Arnborg, B. Courcelle, A. Proskurowski, and D. Seese. An algebraic theory of graph reduction. Journal of the ACM 40(5), pages 1134–1164, 1993.

[3] 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.

[4] 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.

[5] H. Bodlaender. Treewidth: algorithmic techniques and results. In 22nd Symp. on Mathematical foundations of computer science (MFCS), LNCS 1295, pages 19–36, Springer, 1997.

[6] H. Bodlaender, J. Gilbert, H. Hafsteinsson, and T. Kloks. Approximating Treewidth, Pathwidth, Frontsize and Shortest Elimination Trees. Journal of Algorithms 18, pages 238–255, 1995.

[7] V. Bouchitté, D. Kratsch, H. Müller, and I. Todinca. On treewidth approximations. Discrete Applied Mathematics 136, pages 183–196, 2004.

[8] A. Clauset, M. Newman, and C. Moore. Finding community structure in very large networks. Phys. Rev. E 70, 066111 (2004).

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

[10] B. Courcelle, J. Makowsky, and U. Rotics. Linear-time solvable optimization problems on graphs of bounded cliquewidth. In 24th Workshop on Graph-Theoretic Concepts in Computer Science (WG), LNCS 1517, pages 1–16, Springer, 1998.

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

[12] L. Donetti and M. Muñoz. Detecting Network Communities: A New Systematic and Efficient Algorithm. J. of Statistical Mechanics: Theory and Experiment, P10012, 2004.

[13] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Could any graph be turned into a small world? Technical Report LIP RR-2004-61, ENS-Lyon, Dec 2004.

[14] B. Everitt, S. Landau and M. Leese. Cluster Analysis. Arnold, London, 4th edition, 2001.

[15] P. Fraigniaud and C. Gavoille. Routing in trees. In 29th International Colloquium on Automata, Languages and Programming (ICALP), LNCS 2076, pages 757–772, Springer, 2001.

[16] P. Fraigniaud and C. Gavoille. End-to-end routing. In 17th Symposium on Distributed Computing (DISC), LNCS 2848, pages 211–223, Springer, 2003.

[17] P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), pages 169–178, 2004.

[18] C. Gavoille, M. Katz, N. Katz, C. Paul, and D. Peleg. Approximate Distance Labeling Schemes. In 9th European Symposium on Algorithms (ESA), LNCS 2161, pages 476–487, Springer, 2001.

[19] M. Golumbic. Algorithmic Graph Theory and Perfect Graphs. Academic Press, New-York, 1980.

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

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

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

[23] E. Lebhar and N. Schabanel. Searching for optimal paths in long-range contact networks. In 31st International Colloquium on Automata, Languages and Programming (ICALP), LNCS 3142, pages 894–905, Springer, 2004.

[24] 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.

[25] 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.

[26] C. Martel and V. Nguyen. Analyzing Kleinberg's (and other) Small-world Models. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), pages 178–187, 2004.

[27] C. Martel and V. Nguyen. Analyzing and Charaterizing Small-World Graphs. In 16th ACM Symp. on Discrete Algorithms (SODA), 2005.

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

[29] N. Megiddo, S. Hakimi, M. Garey, D. Johnson and C. Papadimitriou. The complexity of searching a graph. Journal of the ACM 35(1), pages 18–44, 1988.

[30] N. Robertson and P. D. Seymour. Graph minors II, Algorithmic Aspects of Tree-Width. Journal of Algorithms 7, pages 309–322, 1986.

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

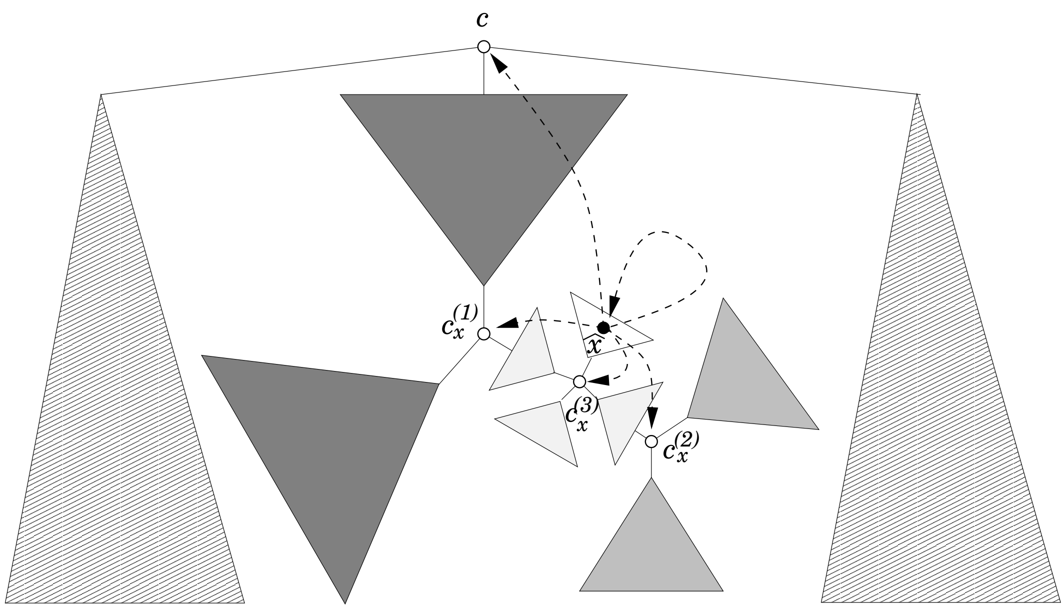

Figure 1: Tree-decomposition-based long-range contacts distribution

This figure displays all possible locations of the long-range contacts of some node $x \in V(G)$. The tree-decomposition T has $c = c_x^{(0)}$ as centroid, and $\widehat{x} \in V(T)$ is displayed as the black node. Removing $c_x^{(0)}$ from T results in three subtrees. Node $\widehat{x}$ is in the middle subtree $T_x^{(1)}$, whose centroid is $c_x^{(1)}$. Removing $c_x^{(1)}$ from $T_x^{(1)}$ results in three subtrees. Node $\widehat{x}$ is in the bottom-right subtree $T_x^{(2)}$, whose centroid is $c_x^{(2)}$. Removing $c_x^{(2)}$ from $T_x^{(2)}$ results again in three subtrees. Node $\widehat{x}$ is in the upper-left subtree $T_x^{(3)}$, whose centroid is $c_x^{(3)}$. Removing $c_x^{(3)}$ from $T_x^{(3)}$ results in four subtrees. Node $\widehat{x}$ is in the upper-right subtree $T_x^{(4)}$, whose centroid is $c_x^{(4)}$. The process stops here by assuming that $\widehat{x} = c_x^{(4)}$. Therefore $q_x = 4$. Node $x \in V(G)$ chooses its long-range contact $y \in V(G)$ as one of the nodes in one of the bags $c_x^{(0)}, c_x^{(1)}, c_x^{(2)}, c_x^{(3)}, c_x^{(4)}$.