

Polylogarithmic Network Navigability Using Compact Metrics with Small Stretch

Pierre Fraigniaud*
CNRS and University Paris Diderot
France

Pierre.Fraigniaud@liafa.jussieu.fr

Cyril Gavoille†
University of Bordeaux
France

gavoille@labri.fr

ABSTRACT

Graph augmentation theory is a general framework for analyzing navigability in social networks. It is known that, for large classes of graphs, there exist augmentations of these graphs such that greedy routing according to the shortest path metric performs in polylogarithmic expected number of steps. However, it is also known that there are classes of graphs for which no augmentations can enable greedy routing according to the shortest path metric to perform better than $\Omega(n^{1/\sqrt{\log n}})$ expected number of steps. In fact, the best known universal bound on the greedy diameter of arbitrary graph is essentially $n^{1/3}$. That is, for any graph, there is an augmentation such that greedy routing according to the shortest path metric performs in $\tilde{O}(n^{1/3})$ expected number of steps. Hence, greedy routing according to the shortest path metric has at least two drawbacks. First, it is in general space-consuming to encode locally the shortest path distances to all the other nodes, and, second, greedy routing according to the shortest path metric performs poorly in some graphs.

We prove that, using semimetrics of small stretch results in a huge positive impact, in both encoding space and efficiency of greedy routing. More precisely, we show that, for any connected n -node graph G and any integer $k \geq 1$, there exist an augmentation φ of G and a semimetric μ on G with stretch $2k - 1$ such that greedy routing according to μ performs in $O(k^2 n^{2/k} \log^2 n)$ expected number of steps. As a corollary, we get that for any connected n -node graph G , there exist an augmentation φ of G and a semimetric μ on G with stretch $O(\log n)$ such that greedy routing according to μ performs in polylogarithmic expected number of steps. This latter semimetric can be encoded locally at every node using only a polylogarithmic number of bits.

*Additional support from the ANR project ALADDIN, from the INRIA project GANG, and from the COST Action 295 DYNAMO.

†Additional support from the ANR project ALADDIN, and from the INRIA project CEPAGE.

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.

SPAA'08, June 14–16, 2008, Munich, Germany.

Copyright 2008 ACM 978-1-59593-973-9/08/06 ...\$5.00.

Categories and Subject Descriptors

C.1.4 [Computer-Communication Networks]: Network Architecture and design—*Network topology*; C.2.2 [Computer-Communication Networks]: Network Protocols—*Routing protocols*; C.2.4 [Computer-Communication Networks]: Distributed systems

General Terms

Algorithms

Keywords

Small world phenomenon, Network Navigability, Social Networks

1. INTRODUCTION

Graph augmentation theory is a general framework for analyzing navigability in social networks. It was introduced in [21] where J. Kleinberg analyzed the *small world phenomenon* observed in Milgram experiment [8, 27]. An augmented graph model is a pair (G, φ) where G is an n -node connected graph with positive edge cost function, and φ is a collection of probability distributions $\{\varphi_u, u \in V(G)\}$. Every node $u \in V(G)$ is given one¹ extra link pointing to some node v , called the *long range contact* of u , chosen at random among all nodes in G according to φ_u as follows:

$$\Pr\{u \rightarrow v\} = \varphi_u(v).$$

The link from a node to its long range contact is called a *long range link*. The links of the underlying graph G are called *local links*. *Greedy routing* in a graph of (G, φ) is the oblivious routing protocol where the routing decision taken at the current node u for a message with destination t consists in

1. selecting a neighbor v of u that is the closest to t according to the distance in G (this choice is performed among all neighbors of u in G and the long range contact of u), and
2. forwarding the message to v .

¹Adding more than one link generally results in speeding up navigability by a factor linear in the number of links, and thus this dimension of the problem is ignored in this paper which focusses on augmentations using one extra link per node. Note that computing the minimum number of links necessary to achieve a given navigability performance is NP-hard [14].

This process assumes that every node has knowledge of the distances in G . On the other hand, every node is unaware of the long range links added to G , except its own long range link. Hence the nodes have no notion of the distances in the augmented graph.

The *greedy diameter* of (G, φ) is defined as the maximum, taken over all source-destination pairs (s, t) , of the expected number of steps for traveling from s to t , using greedy routing in (G, φ) . A graph G is *k-navigable* if there exists a collection φ of probability distributions such that (G, φ) has greedy diameter at most k .

Lots of effort has been devoted to characterize the family of n -node graphs that are polylog(n)-navigable (cf. the surveys [16, 22]). For instance, it is known [21] that for any fixed $d \geq 1$, the d -dimensional meshes are $O(\log^2 n)$ -navigable. More generally, it was proved that all graphs of bounded doubling dimension or bounded growth are polylog(n)-navigable [9, 30]. Similarly, all graphs of bounded treewidth, and more generally all graphs excluding a fixed minor are polylog(n)-navigable [1, 15]. All the augmentation schemes proposed in the aforementioned papers are however specifically designed to apply efficiently to each of the considered classes of graphs. Peleg [28] noticed that any n -node graph is $O(n^{1/2})$ -navigable using the uniform augmentation. This result is improved in [18], where is described a universal augmentation φ such that greedy routing in (G, φ) performs essentially in $n^{1/3}$ expected number of steps for any n -node graph G . On the other hand, it is known [19] that a function f such that every n -node graph is $f(n)$ -navigable satisfies $f(n) = \Omega(n^{1/\sqrt{\log n}})$. This lower bound is mostly due to the strong constraints induced by routing according to the shortest path distance metric that does not allow messages to take short detour for finding short cuts leading closer to the target.

Actually, analyzing the small world phenomenon using shortest path distance metrics also suffers from other problems. In particular, the assumption that all the entities involved in routing know their mutual distances in the base graph G is questionable. Indeed, if most person have some notion of the distances between individuals, these distances are only estimated. This estimation may be accurate enough for geographical distances, but it may certainly become more vague for criteria like professional distances, or distances measured according to sociological parameters like ethical characters, religious beliefs, etc. In fact, for obvious memory space constraints, an individual can hardly store² its exact distance to all other individuals in arbitrary networks.

1.1 Our results

We prove that, by using metrics or semimetrics³ of small *stretch* results in a huge positive impact, in both encoding space and efficiency of greedy routing. More precisely, a semimetric $\mu : V(G) \times V(G) \rightarrow \mathbb{R}^+$ has stretch $\sigma \geq 1$ if and

²It can even be proved that extracting exact distances in any n -node network requires at least $\Omega(n)$ bits of information about the network for some individuals [13, 31].

³A semimetric space generalizes the concept of a metric space by not requiring the condition of satisfying the triangle inequality. We stress that the triangle inequality can be violated in several real world networks, in particular by the interdomain routing latencies (RTT) in the Internet [5, 20].

only if, for any two nodes u and v , we have

$$\text{dist}_G(u, v) \leq \mu(u, v) \leq \sigma \cdot \text{dist}_G(u, v)$$

where $\text{dist}_G(u, v)$ denotes the distance between u and v in G . We show that, for any connected n -node graph G with a positive edge cost function and any integer $k \geq 1$, there exist an augmentation φ of G and a semimetric μ on G with stretch $2k - 1$ such that greedy routing in (G, φ) according to μ performs in $O(k^2 n^{2/k} \log^2 n)$ expected number of steps. A direct consequence of this result is that the best known universal bound $\tilde{O}(n^{1/3})$ for the navigability diameter of arbitrary graphs in [18] is beaten by just using a semimetric of small constant stretch.

As an important corollary, we get that, for any connected n -node graph G , there exist an augmentation φ of G and a semimetric μ on G with stretch $O(\log n)$ such that greedy routing in (G, φ) according to μ performs in polylogarithmic expected number of steps. Therefore, polylogarithmic greedy diameter is achievable for all graphs when using semimetrics of only logarithmic stretch whereas this is impossible to achieve when using metrics (or semimetrics) of stretch 1. Moreover, the semimetric μ can be encoded locally at every node using only polylog(n) $\log \Delta$ bits, where $\Delta = \max_{x \neq y} \text{dist}_G(x, y) / \min_{x \neq y} \text{dist}_G(x, y)$ denotes the *aspect ratio* (or *normalized diameter*) of G .

The improvement provided by our approach, compared to greedy routing according to shortest path distances, is therefore three folded. First, greedy routing is guaranty to perform polylogarithmically in all graphs; Second, the used metric provides good approximations of the real distances in the network (this fits with reality where individuals have only good but not exact estimations of their distances to the other nodes); And, third, this metric can be encoded compactly at every node, as opposed to shortest path distances.

Our proof uses the concept of tree-covers, i.e., spanners obtained as union of trees. It is important to note that a low stretch spanner S of the base graph G is not sufficient to guaranty that greedy routing according to the distance in this spanner be efficient, even if S is polylog(n)-navigable. Nevertheless, we propose a more sophisticated construction of a semimetric μ in G which, combined with an appropriate augmentation φ of G , where both φ and μ depend on a tree-covering of G , insures polylogarithmic performances of greedy routing in (G, φ) according to μ .

1.2 Related work

In addition to the contributions mentioned before, it is worth to mention other important contributions to the analysis of the augmented graph model. Several attempts have been made to increase the performances of greedy routing, by slightly modifying the routing protocol (cf., e.g., [6, 17, 23, 24]). Lower bounds on the performances of greedy routing in rings have been derived in [2, 4, 11, 14]. The distributed setting of the long-range links has been studied in [10]. The structure of augmented graphs has been analyzed in [25, 26]. Finally, models for the emergence of the small world phenomenon have been proposed in [7, 29].

2. NAVIGABILITY DIAMETER

Let G be a connected graph with a positive edge cost function, and φ be an augmenting distribution for G . Greedy routing in $H \in (G, \varphi)$ to a target $t \in V(H)$ according to

a metric, or a semimetric, $\mu : V(G) \times V(G) \rightarrow \mathbb{R}^+$, is the routing process in which every intermediate node $u \in V(H)$

1. selects among all its neighbors in H (i.e., its neighbors in the base graph G plus its long range contact) the node v that minimizes $\mu(v, t)$ (ties are broken arbitrarily), and
2. forwards the message to v .

Note that greedy routing according to μ may not always converge, even if μ is a metric. For instance, let

$$\mu(u, v) = \begin{cases} 1 & \text{if } u \neq v \\ 0 & \text{otherwise.} \end{cases}$$

Then greedy routing does not always converge. On the other hand, if $\mu(u, v) = \text{dist}_G(u, v)$ then greedy routing always converges, simply because the distance to the target (in the base graph G) is reduced by at least $\min_{x \neq y} d_G(x, y) > 0$ at each step.

Actually, greedy routing always converges if $\mu(u, v) = \text{dist}_S(u, v)$ where S is a connected spanning subgraph (spanner) of the base graph G because the neighbor v of the current node u along a shortest path from u to the target t in S satisfies $\mu(v, t) = \text{dist}_S(u, t) - \text{dist}_S(u, v)$, and thus the distance to the target (in the spanner S) can be reduced by at least $\text{dist}_S(u, v) > 0$ at each step.

For any source-target pair $(s, t) \in V(G) \times V(G)$, let $X_{s,t}$ be the random variable that is equal to the number of steps of greedy routing from s to t in (G, φ) according to μ .

DEFINITION 1. *The navigability diameter, $\text{nav}(G, \varphi, \mu)$, of (G, φ, μ) is defined as the maximum, taken over all source-destination pairs (s, t) , of the expected number of steps of greedy routing in (G, φ) from s to t , according to μ . That is,*

$$\text{nav}(G, \varphi, \mu) = \max_{s,t} \mathbb{E}_\varphi(X_{s,t}).$$

Hence the greedy diameter of (G, φ) is the navigability diameter of $(G, \varphi, \text{dist}_G)$. Note that the underlying graph G can be edge-weighted, with $\text{dist}_G(u, v)$ defined as the sum of the edge-weights of a shortest path between u and v . However, the performances of greedy routing are measured in term of number of hops, i.e., ignoring the edge-weights.

One can solve the navigability problem when no constraints are placed on the semimetric. In particular, for any connected n -node graph G , by picking any spanning tree T of G , by augmenting T using the augmentation φ defined for trees in [14], and by setting $\mu(u, v) = \text{dist}_T(u, v)$, one can prove that $\text{nav}(G, \varphi, \mu) = O(\log n)$. The stretch of μ is however $\Omega(n)$, even by selecting the "best" spanning tree T , as witnessed by the n -node cycles. Our objective is, given a graph G , to find an augmentation φ , and a metric or semimetric μ of low stretch such that $\text{nav}(G, \varphi, \mu)$ is small. Unfortunately, using a spanner S of low stretch (i.e., the maximum, over all pairs $\{u, v\}$ of nodes, of the ratio $\text{dist}_G(u, v)/\text{dist}_S(u, v)$ is small) does not imply that greedy routing according to dist_S will be efficient. In fact, the performances of greedy routing according to dist_S may even be independent of the stretch of S . To illustrate this fact, we state the following result.

PROPOSITION 1. *For every $n \geq 1$ and $\epsilon \in (0, 1]$, there exists an edge-weighted graph G of $8n + 4$ vertices, a*

$(1 + \epsilon)$ -spanner S of G , and an augmentation φ such that $\text{nav}(S, \varphi, \text{dist}_S) = \text{polylog}(n)$ but $\text{nav}(G, \varphi, \text{dist}_S) = \Omega(n)$. (For $\epsilon = 1$, the graph G is unweighted).

PROOF. Let $n \geq 1$. We consider the following weighted graph G , as pictured on Figure 1. G has $8n + 4$ nodes, and consists of a cycle $C = (x_0, x_1, \dots, x_{6n+2})$ of $6n + 3$ nodes, plus $2n + 1$ nodes y_0, y_1, \dots, y_{2n} where y_i is connected to the four nodes $x_{3i}, x_{3i+1}, x_{3i+2}, x_{3(i+1)}$, for $i = 0, 1, \dots, 2n$. The edge cost is 1 for all edges of G but for edges between nodes x_{3i} and x_{3i+1} for which the cost is ϵ . Let S be the spanner of G obtained by removing all edges $\{y_i, x_{3i}\}$ for $i = 0, 1, \dots, 2n$. The stretch of S is $1 + \epsilon$. The graph is augmented as follows. For every node $u \in \bigcup_{i=0}^{2n} \{x_{3i}, y_i\}$, we set

$$\Pr(u \rightarrow v) = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{otherwise,} \end{cases}$$

and for every node $u \in \bigcup_{i=0}^{2n} \{x_{3i+1}, x_{3i+2}\}$, we set

$$\Pr(u \rightarrow v) = \begin{cases} 0 & \text{if } v \in \{y_0, y_1, \dots, y_{2n}\} \\ \frac{1}{4H_n - |i-j|} & \text{if } u \in \{x_{3i+1}, x_{3i+2}\} \text{ and} \\ & v \in \{x_{3j+1}, x_{3j+2}\} \text{ with } i \neq j. \end{cases}$$

where $H_n = \sum_{i=1}^n \frac{1}{i}$ is the n th harmonic number. In other words, nodes x_{3i}, y_i are their own long-range contacts, and the long-range contacts of x_{3i+1} and x_{3i+2} are distributed harmonically in the cycle C in a way similar to [21]. Let φ denotes this augmentation.

By the same arguments as in [21], $\text{nav}(S, \varphi, \text{dist}_S) = O(\log^2 n)$.

On the other hand, consider greedy routing in (G, φ) from x_0 to x_{3n} , according to dist_S . For any $i \in \{0, 1, \dots, n-1\}$, the neighbor of x_{3i} that is closest to x_{3n} in S is y_i , and the neighbor of y_i that is closest to x_{3n} in S is $x_{3(i+1)}$, because all these nodes are their own long-range contacts. Therefore, the greedy path from x_0 to x_{3n} is

$$x_0, y_0, x_3, y_1, \dots, x_{3i}, y_i, \dots, x_{3(n-1)}, y_{n-1}, x_{3n}$$

independently from the trials of the long-range links. This path is of length $2n$, and thus $\text{nav}(G, \varphi, \text{dist}_S) \geq 2n$. \square

The proof of Proposition 1 illustrates the (perhaps counterintuitive) fact that adding edges to a graph may result in slowing down the performances of greedy routing considerably. Hence using low-stretch spanners and/or low-stretch metrics are not sufficient by themselves for solving the navigability problem.

3. LOW STRETCH UNIVERSAL GRAPH AUGMENTATION

In this section, we establish our main result about polylogarithmic navigability using semimetrics of logarithmic stretch. Our proof technique is a combination of the use of a certain kind of spanners, the design of specific augmentations for these spanners, and the design of appropriate semimetrics that fit with both the spanners and their augmentations.

THEOREM 1. *For any n -node connected graph G with a positive edge cost function, and any integer $k \geq 1$, there exist an augmenting distribution φ , and a semimetric μ on G with stretch $2k - 1$, such that $\text{nav}(G, \varphi, \mu) = O(k^2 n^{2/k} \log^2 n)$. The semimetric μ can be encoded at every node using*

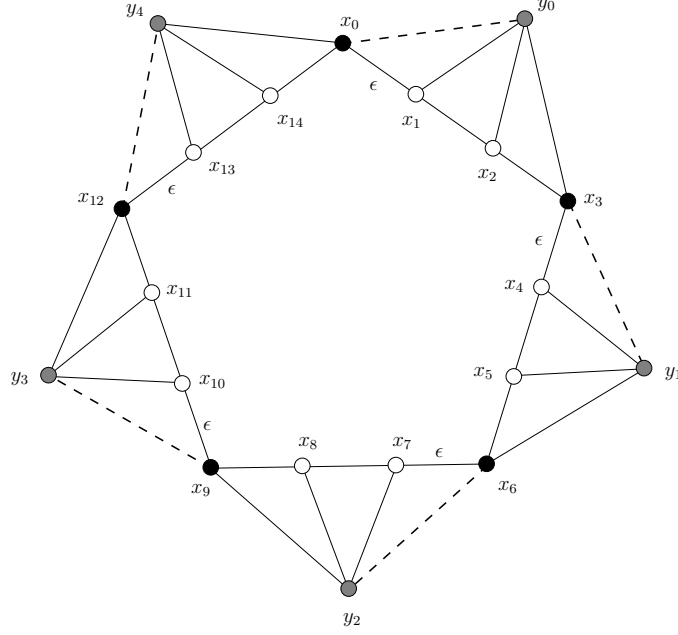


Figure 1: Graph G in the proof of Proposition 1, for $n = 2$. Dotted lines represent the edges not in the spanner S .

$O(kn^{1/k} \log n \log(k\Delta))$ bits, where Δ denotes the normalized diameter of G .

A direct consequence of this result is that the best known universal bound $O(n^{1/3})$ for the navigability diameter of arbitrary graphs in [18] is beaten by just using a semimetric of small constant stretch. By choosing $k = \lceil \log n \rceil$ in Theorem 1, we get the more important following corollary:

COROLLARY 1. *For any n -node connected graph G , there exist an augmenting distribution φ , and a semimetric μ on G with stretch $O(\log n)$, such that $\text{nav}(G, \varphi, \mu) = O(\log^4 n)$. The semimetric μ can be encoded at every node using $O(\log^2 n \log(\Delta \log n))$ bits, where Δ denotes the normalized diameter of G .*

To prove Theorem 1, let us fix a connected n -node graph G . The proof uses the concept of sparse tree-cover [3]. A (σ, δ) -tree-cover of G is a collection \mathcal{T} of trees satisfying the three following conditions:

Cover: Any $T \in \mathcal{T}$ is a subgraph of G , and for any node $u \in V(G)$ there is a tree $T \in \mathcal{T}$ containing u ;

Stretch σ : For any $(u, v) \in V(G) \times V(G)$, there exists $T \in \mathcal{T}$ such that T spans both u and v , and $\text{dist}_T(u, v) \leq \sigma \cdot \text{dist}_G(u, v)$;

Degree δ : For any $u \in V(G)$, $|\{T \in \mathcal{T}, u \in V(T)\}| \leq \delta$.

Theorem 1 directly follows from the lemma below after recalling that any graph has a $(2k - 1, O(kn^{1/k}))$ -tree-cover [31], for any $k \geq 1$.

LEMMA 1. *For any n -node graph G , if G has a (σ, δ) -tree-cover with $\delta \leq n$, then there exist an augmenting distribution*

φ , and a semimetric μ on G with stretch at most σ , such that $\text{nav}(G, \varphi, \mu) \leq (\delta \log n)^2$. The semimetric μ can be encoded at every node using $O(\delta \log n \log(\sigma \Delta))$ bits, where Δ is the normalized diameter of G .

The rest of the section is devoted to the proof of this lemma. Note that, as we underlined in Section 2, the given of a low stretch spanner is not sufficient to guaranty that greedy routing according to the distance in this spanner be efficient. The proof below uses a more sophisticated construction of a semimetric μ which, combined with an appropriate augmentation φ , insures good performances of greedy routing in (G, φ) according to μ .

PROOF. Let \mathcal{T} be a (σ, δ) -tree-cover of G with $\delta \leq n$. To establish the lemma, let $\mu : V(G) \times V(G) \rightarrow \mathbb{R}^+$ be the following map. For any two nodes u and v , let $\mathcal{T}_{u,v}$ denotes the set of trees $T \in \mathcal{T}$ such that both u and v are spanned by T . We then define μ by

$$\mu(u, v) = \min_{T \in \mathcal{T}_{u,v}} \text{dist}_T(u, v).$$

This map μ is well defined because \mathcal{T} satisfies the stretch property, i.e., $\mathcal{T}_{u,v} \neq \emptyset$. It is a semimetric because it inherits this property from the distance metric in trees. It is of stretch at most σ because the stretch of \mathcal{T} is σ .

The augmenting distribution φ for G is set as follow. Fix $u \in V(G)$. We define $\varphi_u \in \varphi$. To select its long-range contact v , node u picks a tree T uniformly at random among all trees in the tree-cover that are spanning u . The long range contact of u is then chosen in T according to the augmentation scheme defined in [15] for graphs of bounded treewidth, applied to trees, i.e., to graphs of treewidth 1. This augmentation is as follows.

Let $|T|$ denote the number of nodes in T . Let c be a centroid of T , i.e., a node such that any trees in $T \setminus \{c\}$

has at most $|T|/2$ nodes. For every node $u \in V(T)$, we set $c^{(u,0)} = c$, and if $u \neq c^{(u,0)}$, define $T^{(u,1)}$ as the tree of $T \setminus \{c^{(u,0)}\}$ containing u . Then, let $c^{(u,1)}$ be a centroid of $T^{(u,1)}$, and, if $u \neq c^{(u,1)}$ let $T^{(u,2)}$ be the tree of $T^{(u,1)} \setminus \{c^{(u,1)}\}$ containing u . And so on. One constructs in this way two finite sequences (see Figure 2):

$$(T^{(u,0)}, T^{(u,1)}, \dots, T^{(u,q^{(u)})}) \quad (1)$$

and

$$(c^{(u,0)}, c^{(u,1)}, \dots, c^{(u,q^{(u)})}) \quad (2)$$

where $T^{(u,0)} = T$, $c^{(u,i)}$ is the centroid of $T^{(u,i)}$ closest to c in T , $T^{(u,i+1)}$ is the tree of $T^{(u,i)} \setminus \{c^{(u,i)}\}$ containing u , and $c^{(u,q^{(u)})} = u$. Note that since $|T| \leq n$, and $|T^{(u,i+1)}| \leq |T^{(u,i)}|/2$, we get that both sequences are of length $q^{(u)} + 1 \leq \log n$.

To set up its long-range link, every node u proceeds as follows:

- it considers the δ_u trees T_1, \dots, T_{δ_u} of the tree-cover it belongs to, where $\delta_u \leq \delta$;
- it picks one of them uniformly at random, say T_j , $1 \leq j \leq \delta_u$; and
- it selects its long range contact as one of the at most $\log n$ centroids $c_j^{(u,k)}$ of T_j , $0 \leq k \leq q_j^{(u)}$, uniformly at random.

In other words, we select the long range contact v of node u according to:

$$\varphi_u(v) = \begin{cases} \frac{1}{\delta_u} \cdot \frac{1}{q_j^{(u)} + 1} & \text{if } v \in \{c_j^{(u,0)}, c_j^{(u,1)}, \dots, c_j^{(u,q_j^{(u)})}\} \\ & \text{for some } 1 \leq j \leq \delta_u \\ 0 & \text{otherwise} \end{cases}$$

Let us now fix a source s and a target t in G , and let us compute the expected number of steps of greedy routing from s to t in (G, φ, μ) . First, observe that greedy routing in (G, φ, μ) does lead from s to t in any augmentation of G . Indeed, if u is the current node, and if $T \in \mathcal{T}_{u,t}$ is the tree for which $\mu(u,t) = \text{dist}_T(u,t)$, then the neighbor v of u along the shortest path from u to t in T satisfies $\text{dist}_T(v,t) < \mu(u,t)$, and thus the next node w on the greedy path from u to t satisfies $\mu(w,t) \leq \mu(v,t) \leq \text{dist}_T(v,t) < \mu(u,t)$. Hence the "distance" measured by the semimetric μ between the current node and the target decreases at each step of greedy routing, and thus greedy routing in (G, φ, μ) always succeeds.

For computing the efficiency of greedy routing in (G, φ, μ) , we consider the potential function Φ defined as follows. Let

$$P = u_0, u_1, \dots, u_\ell$$

with $u_0 = s$ and $u_\ell = t$, be the path followed by greedy routing from s to t . The path P depends on the trial of the long-range links as determined by φ , and the expected number of steps of greedy routing from s to t is equal to $\mathbb{E}_\varphi(|P|)$. For $i = 0, \dots, \ell$, let

$$P_i = u_0, u_1, \dots, u_i$$

be the prefix of P consisting of the $i + 1$ first nodes of P . The potential Φ is a function of the P_i s, more precisely of their extremities u_i s. Φ uses the centroids as beacons, and

counts how many beacons are still closer to the target than the current node is. It is defined by δ_t distinct functions ϕ_j , $1 \leq j \leq \delta_t$, where each ϕ_j is associated to a distinct tree T_j among the δ_t trees in \mathcal{T} spanning the target t . (By the degree property of the (σ, δ) -tree-cover, $\delta_t \leq \delta$). Let

$$(c_j^{(t,0)}, c_j^{(t,1)}, \dots, c_j^{(t,q_j^{(t)})})$$

be the sequence of centroids for $t \in T_j$, as defined by Equations 1 and 2. Each ϕ_j returns a boolean vector of at most $\log n$ coordinates, obtained by the application of at most $\log n$ functions $\phi_j^{(k)}$. More precisely,

$$\phi_j = (\phi_j^{(0)}, \dots, \phi_j^{(q_j^{(t)})})$$

where, for $k = 0, \dots, q_j^{(t)}$, we have

$$\phi_j^{(k)}(P_i) = \begin{cases} 1 & \text{if } \mu(u_i, t) > \mu(c_j^{(t,k)}, t) \\ 0 & \text{otherwise.} \end{cases}$$

In other words, we use the at most $\delta \cdot \log n$ nodes $c_j^{(t,k)}$, $1 \leq j \leq \delta_t$, $0 \leq k \leq q_j^{(t)}$, to represent as many beacons, and we count how many beacons are still closer to the target than the current node is, by defining

$$\Phi(P_i) = \sum_{j=1}^{\delta_t} \|\phi_j(P_i)\|_1, \text{ i.e., } \Phi(P_i) = \sum_{j=1}^{\delta_t} \sum_{k=0}^{q_j^{(t)}} \phi_j^{(k)}(P_i).$$

Since $\phi_j^{(k)}(P_{i+1}) \leq \phi_j^{(k)}(P_i)$, for all i, j, k , we get that

$$\Phi(P_{i+1}) \leq \Phi(P_i)$$

for all $i \geq 0$. Also,

$$\Phi(P_0) \leq \sum_{j=1}^{\delta_t} (1 + q_j^{(t)}) \leq \delta_t \log n \leq \delta \log n. \quad (3)$$

We have $\Phi(P) = \Phi(P_\ell) = 0$, and thus the target t is reached after at most $\Phi(P_0)$ 1-entries successively vanish in the vectors ϕ_j s when traveling along P . Let Z_i be the random variable measuring the number of nodes in P between the vanishing of the i th 1-entry and the vanishing of the $(i+1)$ th 1-entry of Φ . By linearity of the expectation, we have

$$\mathbb{E}_\varphi(|P|) = \sum_{i=0}^{\Phi(P_0)} \mathbb{E}_\varphi(Z_i). \quad (4)$$

The remaining of the proof consists in upper bounding $\mathbb{E}_\varphi(Z_i)$.

The function Φ is based on the centroid decomposition $c_j^{(t,k)}$ for the target node t whereas the routing depends on the long-range links of the current node u_i that are set based on the centroid decomposition $c_j^{(u_i,k)}$ for node u_i . Here is how these two decompositions are related. Let u_i be the last node of P_i , $u_i \neq t$. Let j be the index such that $\mu(u_i, t) = \text{dist}_{T_j}(u_i, t)$. Let k be the largest index such that

$$(c_j^{(u_i,0)}, \dots, c_j^{(u_i,k)}) = (c_j^{(t,0)}, \dots, c_j^{(t,k)})$$

where these sequences are defined for T_j according to Equations 1 and 2. The index k is well defined since $c_j^{(u_i,0)} = c_j^{(t,0)} = c_j$, where c_j is the centroid of T_j . The nodes u_i and t are separated by $c_j^{(u_i,k)} = c_j^{(t,k)}$ in T_j . For instance, on Figure 2, nodes u and v are separated by their

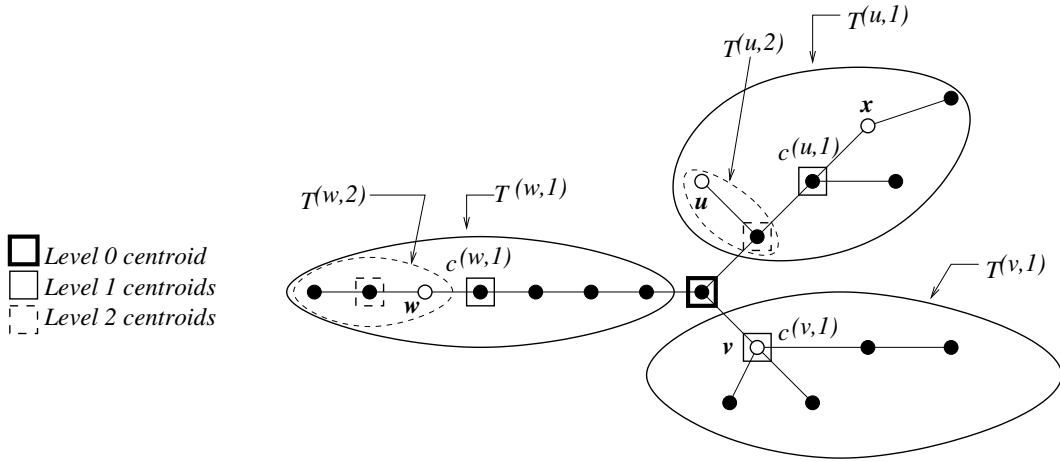


Figure 2: Centroid decomposition.

centroids of level 0: $c^{(u,0)} = c^{(v,0)}$. On the same figure, nodes u and x are separated by their centroids of level 1: $(c^{(u,0)}, c^{(u,1)}) = (c^{(x,0)}, c^{(x,1)})$.

If $u_i = c_j^{(u_i,k)} = c_j^{(t,k)}$, then $\phi_j^{(k)}(P_i) = 0$. On the other hand, $\phi_j^{(k)}(P_{i-1}) = 1$ because, as we observed before, greedy routing according to μ satisfies $\mu(u_{i-1}, t) > \mu(u_i, t)$, and thus $\mu(u_{i-1}, t) > \mu(c_j^{(t,k)}, t)$. Hence at least one 1-entry of Φ vanishes when reaching node u_i . Assume now that $u_i \neq c_j^{(u_i,k)} = c_j^{(t,k)}$. Since

$$\mu(u_i, t) = \text{dist}_{T_j}(u_i, t)$$

and

$$\text{dist}_{T_j}(u_i, t) = \text{dist}_{T_j}(u_i, c_j^{(u_i,k)}) + \text{dist}_{T_j}(c_j^{(u_i,k)}, t),$$

we get that

$$\mu(c_j^{(u_i,k)}, t) \leq \text{dist}_{T_j}(c_j^{(u_i,k)}, t) < \text{dist}_{T_j}(u_i, t) = \mu(u_i, t).$$

As a consequence, if the long-range contact of u_i is $c_j^{(u_i,k)} = c_j^{(t,k)}$, then the next node u_{i+1} along P will satisfy $\mu(u_{i+1}, t) \leq \mu(c_j^{(t,k)}, t)$, and therefore $\phi_j^{(k)}(P_{i+1}) = 0$. Hence at least one 1-entry of Φ vanishes when reaching node u_{i+1} . The event "the long-range contact of u_i is $c_j^{(u_i,k)}$ " occurs with probability at least $1/(\delta \cdot \log n)$ because every node belongs to at most δ trees, and there are at most $\log n$ levels of centroids.

Summarizing the above arguments, at each step along P , either at least one 1-entry of Φ vanishes, or the probability that at least one 1-entry of Φ vanishes at the next step is at least $1/(\delta \cdot \log n)$. As a direct consequence, since all long-range links are chosen independently at each node,

$$\mathbb{E}_\varphi(Z_i) \leq \delta \cdot \log n.$$

By combining this latter inequality with Equations 3 and 4, and using the facts that $\delta_t \leq \delta$ and $1 + q_j^{(t)} \leq \log n$ for all j , we finally get that

$$\mathbb{E}_\varphi(|P|) \leq (\delta \cdot \log n)^2$$

which completes the proof of the Lemma as far as greedy routing is concerned.

We now show how to encode μ compactly at every node. We use the concept of distance-labeling [13]. Roughly, a

distance-labeling scheme for a graph class \mathcal{G} is a pair (λ, α) , where λ is a labeling function applying to the nodes of every graph in \mathcal{G} , and α is a decoding function that, given two labels of two nodes in $G \in \mathcal{G}$, returns the distance between these two nodes in G . That is, every node $u \in V(G)$ is given a label $\lambda(u, G)$, and $\text{dist}_G(u, v) = \alpha(\lambda(u, G), \lambda(v, G))$.

We will provide a compact labeling scheme (L, β) for our semimetric μ . Given this scheme, greedy routing according to μ is executed as follows. The target $t \in V(G)$ is identified by its label $L(t, G)$. Let u be the current node. Node u contacts its neighbors, and collects their labels. Using these labels, u computes, for every neighbor v , the value $\mu(v, t) = \beta(L(v, G), L(t, G))$, and it selects the neighbor with the smallest value for forwarding the message.

It is known (see [12, 13]) that there exists a distance-labeling scheme (λ, α) for trees using $O(\log n_T \log \Delta_T)$ -bit labels, where n_T is the number of nodes of the tree T and Δ_T its normalized diameter of T . The decoding function α is a constant time and constant size algorithm. Let us label the trees of the (σ, δ) -tree-cover \mathcal{T} by pairwise distinct integers from 1 to $|\mathcal{T}| \leq \delta n$. For each tree $T_i \in \mathcal{T}$, $\lambda(u, T_i)$ is the labeling of u in T_i . We define the label $L(u, G)$ by

$$L(u, G) = \left((\ell_1, \lambda(u, T_{\ell_1})), (\ell_2, \lambda(u, T_{\ell_2})), \dots, (\ell_{\delta_u}, \lambda(u, T_{\ell_{\delta_u}})) \right)$$

where T_{ℓ_i} , $i = 1, \dots, \delta_u$, are all trees covering u . Because of the stretch property of μ , (normalized) distances in μ do not exceed $\sigma \Delta$, where Δ is the normalized diameter of G . Thus every label $\lambda(u, T_{\ell_i})$ is on $O(\log n \log(\sigma \Delta))$ bits. Each tree identifier ℓ_i can be represented with $O(\log(\delta n))$ bits, and, by bounding $\delta \leq n$, we get that each label $L(u, G)$ is on at most

$$O(\delta \log n \log(\sigma \Delta) + \log(\delta n)) = O(\delta \log n \log(\sigma \Delta)) \text{ bits.}$$

Given two labels $L(u, G)$ and $L(v, G)$, one can compute $\mu(u, v)$ by selecting all trees that cover both u and v , computing the distance between these two nodes in each of them using α , and taking the minimum of these distances. This completes the proof of Lemma 1, and thus the proof of Theorem 1. \square

4. CONCLUSION

The main result in this paper is the design, for any given connected n -node graph G , of an augmentation φ for G , and of a compact semimetric μ in G with stretch $O(\log n)$, such that $\text{nav}(G, \varphi, \mu)$ is polylogarithmic.

It would be interesting to check whether this is optimal in term of stretch. In particular, is it possible to design, for any given connected n -node graph G , an augmentation φ for G , and a semimetric μ in G with constant stretch, such that $\text{nav}(G, \varphi, \mu)$ is polylogarithmic?

It would also be interesting to check whether it is possible to design a *distance* metric satisfying the same properties as Corollary 1. In particular, is it possible to design, for any given connected n -node graph G , an augmentation φ for G , and a distance metric μ with polylogarithmic stretch that can be stored locally using a polylogarithmic number of bits at each node, such that $\text{nav}(G, \varphi, \mu)$ is polylogarithmic? Again, the use of low stretch spanners does not lead directly to the answer. Indeed, distance labeling requires labels of size $\Omega(\sqrt{n})$ even in graphs of maximum degree 3 (see [13]).

5. REFERENCES

- [1] I. Abraham and C. Gavoille. Object Location Using Path Separators. 25th ACM Symp. on Principles of Distributed Computing (PODC), pp. 188-197, 2006.
- [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), pp. 223-232, 2002.
- [3] B. Awerbuch, D. Peleg. Sparse Partitions. In 31st IEEE Symp. on Foundations of Computer Science (FOCS), pp. 503-513, 1990.
- [4] L. Barrière, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient Routing in Networks with Long Range Contacts. In 15th International Symp. on Distributed Computing (DISC), LNCS 2180, pp. 270-284, Springer, 2001.
- [5] S. Banerjee, T. Griffin, and M. Pias. The interdomain connectivity of PlanetLab nodes. In 5th Workshop on Passive and Active Measurements, LNCS 3015, pp. 73-82, Springer, 2004.
- [6] D. Barbella, G. Kachergis, D. Liben-Nowell, A. Sallstrom, and B. Sowell. Depth of field and cautious-greedy routing in social networks. In 18th International Symp. on Algo. and Computation (ISAAC), LNCS 4835, pp. 574-586. Springer, 2007.
- [7] A. Clauset and C. Moore How Do Networks Become Navigable?
<http://arxiv.org/abs/cond-mat/0309415v2>.
- [8] P. Dodds, R. Muhamad, and D. Watts. An experimental study of search in global social networks. *Science* 301, no5634, pp. 827-829, 2003.
- [9] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Could any graph be turned into a small-world? *Theoretical Computer Science* 355(1): 96-103 (2006).
- [10] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Towards small world emergence. In 18th ACM Symp. on Parallel Algo. and Architectures (SPAA), pp. 225-232, 2006.
- [11] G. Giakkoupis and V. Hadzilacos On the complexity of greedy routing in ring-based peer-to-peer networks In 26th ACM Symp. on Principles of Distributed Computing (PODC), pp. 99-108, 2007.
- [12] C. Gavoille, M. Katz, N. A. Katz, C. Paul, and D. Peleg. Approximate distance labeling schemes. In 9th Europ. Symp. on Algo. (ESA), LNCS 2161, pp. 476-488. Springer, 2001.
- [13] C. Gavoille, D. Peleg, S. Pérennès, and R. Raz. Distance Labeling in Graphs. *J. of Algorithms*, 53(1):85-112, 2004.
- [14] M. Flammini, L. Moscardelli, A. Navarra, and S. Perennes. Asymptotically optimal solutions for small world graphs. In 19th International Symp. on Distributed Computing (DISC), LNCS 3724, Springer, pp. 414-428, 2005.
- [15] P. Fraigniaud. Greedy Routing in Tree-Decomposed Graphs. In 13th Europ. Symp. on Algo. (ESA), LNCS 3669, Springer, pp. 791-802, 2005.
- [16] P. Fraigniaud. Small Worlds as Navigable Augmented Networks: Model, Analysis, and Validation. In 15th Europ. Symp. on Algo. (ESA), LNCS 4698, Springer, pp. 2-11, 2007.
- [17] P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In Proceedings of the 23rd ACM Symp. on Principles of Distributed Computing (PODC), pp. 169-178, 2004.
- [18] P. Fraigniaud, C. Gavoille, A. Kosowski, E. Lebhar, and Z. Lotker. Universal augmentation schemes for network navigability: overcoming the \sqrt{n} -barrier. In 19th ACM Symp. on Parallelism in Algo. and Architectures (SPAA), pp. 1-7, 2007.
- [19] P. Fraigniaud, E. Lebhar, and Z. Lotker. A doubling dimension threshold $\Theta(\log \log n)$ for augmented graph navigability. In 14th Europ. Symp. on Algorithm (ESA), LNCS 4168, Springer, pp. 376-386, 2006.
- [20] P. Fraigniaud, E. Lebhar, and L. Viennot. The inframetric model for the internet. In 27th IEEE Conference on Computer Communications (INFOCOM), 2008.
- [21] J. Kleinberg. The small-world phenomenon: an algorithmic perspective. In 32nd ACM Symp. on Theory of Computing (STOC), pp. 163-170, 2000.
- [22] J. Kleinberg. Complex networks and decentralized search algorithm. Nevanlinna prize presentation at the International Congress of Mathematicians (ICM), Madrid, 2006.
- [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, pp. 894-905, 2004.
- [24] 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), pp. 54-63, 2004.
- [25] C. Martel, and V. Nguyen. Analyzing Kleinberg's (and other) Small-world Models. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), pp. 179-188, 2004.
- [26] C. Martel, and V. Nguyen. Analyzing and characterizing small-world graphs. In 16th ACM-SIAM Symp. on Discrete Algo. (SODA), pp. 311-320, 2005.

- [27] S. Milgram. The Small-World Problem. *Psychology Today*, pp. 60-67, 1967.
- [28] D. Peleg. Private communication. Workshop of COST Action 295 "DYNAMO", Les Ménuires, Jan. 2006.
- [29] O. Sandberg. Neighbor Selection and Hitting Probability in Small-World Graphs. *Annals of Applied Probability* (to appear).
- [30] A. Slivkins. Distance estimation and object location via rings of neighbors. In 24th ACM Symp. on Principles of Distributed Computing (PODC), pp. 41-50, 2005.
- [31] M. Thorup, and U. Zwick. Approximate Distance Oracles. *J. of the ACM*, 52:1-14, 2005.