Universal augmentation schemes for network navigability: overcoming the $\sqrt{n}$-barrier

Pierre Fraigniaud, Cyril Gavoille, Adrian Kosowski, Emmanuelle Lebhar, Zvi Lotker

Universal Augmentation Schemes for Network Navigability: Overcoming the $\sqrt{n}$-Barrier*

Pierre Fraigniaud
CNRS and University of Paris Sud
France

Cyril Gavoille
Bordeaux University
France

Adrian Kosowski
Gdansk University of Technology
Poland

Emmanuelle Lebhar
CNRS and University of Paris 7
France

Zvi Lotker
Ben Gurion University
Israel

Abstract

Augmented graphs were introduced for the purpose of analyzing the "six degrees of separation between individuals" observed experimentally by the sociologist Standley Milgram in the 60's. Formally, an augmented graph is a pair (G, φ) where G is a graph, and φ is a collection of probability distributions $\{\varphi_u, u \in V(G)\}$. Every node $u \in V(G)$ is given an extra link, called a long range link, pointing to some node v, called the long range contact of u. The head v of this link is chosen at random by $\Pr\{u \to v\} = \varphi_u(v)$. In augmented graphs, greedy routing is the oblivious routing process in which every intermediate node chooses among all its neighbors (including its long range contact) the one that is closest to the target according to the distance measured in the underlying graph G, and forwards to it. Roughly, augmented graphs aim at modeling the structure of social networks, while greedy routing aims at modeling the searching procedure applied in Milgram's experiment.

Our objective is to design efficient *universal* augmentation schemes, i.e., augmentation schemes that give to any graph G a collection of probability distributions φ such that greedy routing in (G, φ) is fast. It is known that the uniform scheme φ_{unif} is a universal scheme ensuring that, for any n-node graph G, greedy routing in $(G, \varphi_{\text{unif}})$ performs in $O(\sqrt{n})$ expected number of steps. Our main result is the design of a universal augmentation scheme φ such that greedy routing in (G, φ) performs in $\tilde{O}(n^{1/3})$ expected number of steps for any n-node graph G. We also show that under some more restricted model, the $\sqrt{n}$-barrier cannot be overcome.

Keywords: small world phenomenon, routing, network design.

SUBMISSION SPAA 2007
REGULAR PAPER

*This work was partially done while Zvi Lotker was visiting LRI at University Paris Sud, supported by the COST Action 295 "DYNAMO", and while Adrian Kosowski was visiting LaBRI at University of Bordeaux, also supported by the COST Action 295 "DYNAMO". Pierre Fraigniaud, Cyril Gavoille, and Emmanuelle Lebhar were supported by the ACI project "PairAPair". Pierre Fraigniaud received additional support from the ACI project "FRAGILE", and from the INRIA project "Grand Large".

1 Introduction

Augmented graphs were introduced in [22] and formally defined in [13] for the purpose of understanding the "small world phenomenon". Precisely, augmented graphs give a framework for modeling and analyzing the "six degrees of separation" between individuals observed from Milgram's experiment [18], and stating that short chains of acquaintances between any pair of individuals can be discovered in a distributed manner. The concept of augmented graphs has recently gained interest, and gave rise to an abundant literature (cf., e.g., [1, 2, 3, 7, 9, 10, 11, 13, 15, 16, 17, 21]). We refer to Kleinberg's survey [14] on complex networks for more details on the concept of augmented graphs.

Formally, an augmented graph is a pair (G, φ) where G is an n-node connected graph, and φ is a collection of probability distributions $\{\varphi_u, u \in V(G)\}$. Every node $u \in V(G)$ is given an extra link pointing to some node v, called the *long range contact* of u, chosen at random according to φ_u as follows: $\Pr\{u \to v\} = \varphi_u(v)$. 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 (G, φ) was defined in [13]. It is the oblivious routing protocol where the routing decision taken at the current node u for a message of 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. This process assumes that every node has a 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 $\text{diam}(G, \varphi) = \max_{s,t \in V(G)} \mathbb{E}(\varphi, s, t)$, where $\mathbb{E}(\varphi, s, t)$ is the expected number of steps for traveling from s to t using greedy routing in (G, φ). Let $f : \mathbb{N} \to \mathbb{R}$ be a function. An n-node graph G is f*-navigable* if there exists a collection of probability distributions φ such that $\text{diam}(G, \varphi) \leq f(n)$. Lots of effort have been devoted to characterize the family of graphs that are polylog(n)-navigable (cf. the survey [14]). For instance, it is known [13] 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 [6, 21]. Similarly, all graphs of bounded treewidth, and more generally all graphs excluding a fixed minor are polylog(n)-navigable [1, 10]. All the augmentation schemes proposed in the aforementioned papers are however specifically designed to apply efficiently to each of the considered classes of graphs.

An augmentation scheme is *universal* if it applies to all graphs. The uniform augmentation scheme consists in adding long-range links whose extremities are chosen uniformly at random among all the nodes in the graph. Peleg [19] noticed that any n-node graph is $O(n^{1/2})$-navigable using this scheme. To see why, consider the ball B of radius $\sqrt{n}$ centered at the target. The expected number of nodes visited until the long range contact of the current node belongs to B is $n/|B|$, and thus at most $\sqrt{n}$. Once in B, the distance to the target is at most $\sqrt{n}$. Hence the $O(\sqrt{n})$-navigability of the graph. Up to our knowledge, this $O(n^{1/2})$ upper bound was the best known bound for arbitrary graphs until this paper. On the other hand, it was recently proved [12] that a function f such that every n-node graph is f-navigable satisfies $f(n) = \Omega(n^{1/\sqrt{\log n}})$. A crucial problem in the field of network navigability is to close the gap between these upper and lower bounds for the f-navigability of arbitrary graphs.

1.1 Our results

- We first consider augmentation schemes defined a priori from $n \times n$ matrices $A = (a_{i,j})$, where $a_{i,j}$ is the probability that the node labeled i chooses the node labeled j as its long range contact.
  - As we already mentioned, the uniform matrix $U = (u_{i,j})$ with $u_{i,j} = \frac{1}{n}$ is an augmentation scheme such that, for any n-node graph G, greedy routing in (G, U) performs in $O(\sqrt{n})$ expected number of steps. Such augmentation scheme is called *name-independent* because its efficiency does not depend on the node labeling in $\{1, \ldots, n\}$. We actually prove that the uniform matrix is optimal in the sense that, for any $n \times n$ matrix A, there is a node labeling of the n-node path P from 1 to n such that greedy routing in (P, A) performs in $\Omega(\sqrt{n})$ expected number of steps.
  - To overcome the inefficiency of name-independent schemes, even for graphs as simple as paths, we consider matrix-based augmentation schemes using specific node labelings in $\{1, \ldots, n\}$. We thus consider augmentation schemes φ defined as pairs $(M, \mathcal{L})$ where M is a matrix, and $\mathcal{L}$ is a node-labeling. We describe such a scheme, and analyze its performances in terms of a new parameter, called *pathshape*, achieving tradeoff between pathwidth [20] and pathlength [8]. Precisely, for any n, we describe an $n \times n$ matrix M, and a labeling $\mathcal{L}$ of the nodes of any n-node graph G, such that greedy routing in $(G, (M, \mathcal{L}))$ performs in $O(\min\{\mathrm{ps}(G) \cdot \log^2 n, \sqrt{n}\})$ expected number of steps, where $\mathrm{ps}(G)$ denotes the pathshape of G. This result has many important corollaries. In particular, and in contrast with name-independent schemes, the scheme $(M, \mathcal{L})$ yields polylogarithmic expected number of steps of greedy routing for large classes of graphs such as trees and AT-free graphs, including co-comparability graphs, interval and permutation graphs [5]. These classes were not captured by previous results, since in general they are neither of bounded doubling dimension nor excluding a fixed minor. The matrix M is actually a combination of an ad hoc matrix A with the uniform matrix U.
- In the second part of the paper, we consider augmentation schemes defined a posteriori, that are fully depending on the structure of the graph. We design a universal augmentation scheme that overcomes the $O(\sqrt{n})$ barrier. Precisely, we design a universal scheme φ such that greedy routing in (G, φ) performs in $\tilde{O}(n^{1/3})$ expected number of steps for any n-node graph G, where the big-$\tilde{O}$ notation ignores the polylogarithmic factors.

2 Matrix-Based Augmentation Schemes

This section considers a restricted though rich class of augmentation schemes, those based on matrices.

Definition 1 *An* augmentation matrix *of size n is an $n \times n$ matrix $A = (p_{i,j})_{i,j \in [1,n]}$ such that $0 \leq p_{i,j} \leq 1$ for any $i, j \in [1, n]$, and $\sum_{j=1}^{n} p_{i,j} \leq 1$ for any $i \in [1, n]$.*

Note that an augmentation matrix is not necessarily stochastic, i.e., we may have $\sum_{i=1}^{n} p_{i,j} \neq 1$ for some j. An augmentation matrix of size n can be used to design augmentation schemes of n-node graphs whose nodes are labeled from 1 to n as follows: node i chooses node j as long range contact with probability $p_{i,j}$. This definition yields two remarks:

- One may desire not to have all nodes labeled with distinct labels. In this case, a node labeled i first chooses an index j with probability $p_{i,j}$, and then chooses one of the nodes labeled j uniformly at random among all nodes labeled j.
- Augmentation matrices can also be used independently from any labeling. Such an augmentation scheme is said to be *name-independent*. The performances of a name-independent augmentation scheme based on a matrix A are measured for the worst case labeling of the nodes using distinct labels.

In this section, we first prove that the uniform augmentation scheme is optimal among all name-independent matrix-based augmentation schemes. Then, we analyze matrix-based augmentation schemes in their general context, and prove that there is a way to design a matrix-based augmentation scheme that performs at least as well as the uniform scheme in arbitrary graphs, but offers much better performances than the latter for large classes of graphs.

2.1 Name-Independent Schemes

As we already mentioned in the Introduction, the uniform matrix yields a name-independent augmentation scheme with greedy diameter $O(\sqrt{n})$ for n-node graphs. The following result shows that this is optimal among all matrix-based name-independent augmentation schemes.

Theorem 1 *For any augmentation matrix A of size n, the corresponding name-independent augmentation scheme applied to the n-node path yields greedy diameter $\Omega(\sqrt{n})$.*

Proof. We show that, for any augmentation matrix A of size n, there is a labeling of the n-node path with integers in $\{1,\dots,n\}$ such that the greedy diameter of the labeled path augmented using A is $\Omega(\sqrt{n})$. Let $A=(p_{i,j})_{1\leq i,j\leq n}$ be a $n\times n$ augmentation matrix. We claim that:

$$\exists I \subseteq \{1,\dots,n\}, |I|=\sqrt{n} \text{ and } \sum_{i,j\in I, i\neq j} p_{i,j} < 1.$$

Indeed, assume for the purpose of contradiction that, for any set $I\subseteq\{1,\dots,n\}$ of cardinality $\sqrt{n}$, we have $\sum_{i,j\in I,i\neq j} p_{i,j}\geq 1$. We get $\binom{n}{\sqrt{n}}$ inequalities $\sum_{i,j\in I,i\neq j} p_{i,j}\geq 1$, one for every possible set I, each involving $n-\sqrt{n}$ variables $p_{i,j}$'s, $i\neq j$. By summing all these inequalities, we get:

$$\sum_{I\in\binom{n}{\sqrt{n}}}\ \sum_{i,j\in I,i\neq j} p_{i,j} \geq \binom{n}{\sqrt{n}}.$$

On the other hand, by symmetry, each $p_{i,j}$, $i\neq j$, appears the same number of times in the left hand side of the above inequality. Precisely, every $p_{i,j}$ appears exactly $(n-\sqrt{n})\binom{n}{\sqrt{n}}/n(n-1)$ times. We can group the many occurrences of the variables $p_{i,j}$ in sets of the form

$$\{p_{i,j}, j\in\{1,\dots,n\}\setminus\{i\}\}$$

for fixed i, $1\leq i\leq n$. Since, for any fixed i, $\sum_{j\neq i} p_{i,j}\leq 1$, we get that each of these sets contributes by at most 1 to the sum. Therefore

$$\sum_{I\in\binom{n}{\sqrt{n}}}\ \sum_{i,j\in I,i\neq j} p_{i,j} \leq (n-\sqrt{n})\binom{n}{\sqrt{n}}/(n-1) < \binom{n}{\sqrt{n}},$$

a contradiction. Hence, let us consider a set I, of cardinality $\sqrt{n}$, satisfying $\sum_{i,j\in I, i\neq j} p_{i,j} < 1$.

We assign the labels of the set I to $\sqrt{n}$ consecutive nodes of an n-node path, in an arbitrary order. Let S be this set of nodes, $|S| = |I| = \sqrt{n}$. Let X be the random variable equal to the number of long range links having distinct extremities and both extremities in S. We have $\mathbb{E}(X) = \sum_{i,j\in I, i\neq j} p_{i,j}$, and thus $\mathbb{E}(X) < 1$. Let $0 < \alpha < 1$ be such that $\mathbb{E}(X) \leq 1 - \alpha$, we have $\Pr\{X \geq 1\} \leq 1-\alpha$. Consider two nodes s and t such that s is at distance $|S|/3$ from one extremity of S, t is at distance $|S|/3$ from the other extremity of S, and s and t are at mutual distance $|S|/3$. Let Y be the random variable equal to the number of steps of greedy routing from s to t, we have:

$$\mathbb{E}(Y) \geq \mathbb{E}(Y \mid X < 1) \cdot \Pr\{X < 1\} \geq \frac{|S|}{3}\,\alpha \geq \frac{\sqrt{n}}{3}\,\alpha.$$

Therefore, the greedy diameter of this labeled path is $\Omega(\sqrt{n})$, which completes the proof. □

The previous result shows that no name-independent scheme can yield greedy diameter better than $\Omega(\sqrt{n})$, even for paths. Yet name-independent schemes remain useful. Indeed, in addition to their simplicity, they can be combined with name-dependent schemes that perform well for specific classes of graphs but poorly in general. In particular, the uniform scheme can be combined with a scheme that is efficient for large classes of graphs, in order to preserve the $O(\sqrt{n})$ greedy diameter for general graphs. This is proved in the next section.

2.2 Matrix-Based Augmentation Schemes

In this section, we design a matrix-based augmentation scheme (the matrix is coupled with an appropriate labeling of the nodes) that achieves much better performance than the uniform augmentation scheme for large classes of graphs. Our scheme is based on the new notions of *treeshape* and *pathshape* that establish a tradeoff between the two important notions of treewidth [20] and treelength [8]. These two latter notions have been proved important in many contexts, including algorithm design, routing, and labeling.

Recall that a tree-decomposition of a graph G is a pair (T, X) where T is a tree with node set I of finite size, and $X = \{X_i, i \in I\}$ is a collection of subsets of nodes. T and X must satisfy the three following conditions:

- For any $u \in V(G)$, there exists $i \in I$ for which $u \in X_i$;
- For any $e \in E(G)$, there exists $i \in I$ for which both extremities of e belong to X_i;
- For any $u \in V(G)$, the set $\{i \in I \mid u \in X_i\}$ induces a subtree of T.

The third constraint can be rephrased as: for any triple $(i, j, k) \in I^3$, if j is on the path between i and k in T, then $X_i \cap X_k \subseteq X_j$. The X_is are called *bags*. When the tree T is restricted to be a path, the resulting decomposition is called path-decomposition.

The quality of the tree-decomposition depends on the measure that is applied to the bags X_is. Two measures have been investigated in the past, the width [20] and the length [8]:

$$\mathrm{width}(X_i) = |X_i| - 1, \text{ and } \mathrm{length}(X_i) = \max_{x,y\in X_i} \mathrm{dist}_G(x, y)$$

where dist_G denotes the distance function in the graph G. (Note that $\mathrm{length}(X_i)$ may be much smaller than the diameter of the subgraph induced by X_i; In fact X_i may even not be connected). We introduce a new measure, the *shape*, that will be proved very relevant to augmentation schemes.

Definition 2 *The* shape *of a bag X_i of a tree-decomposition (T, X) of a graph G is defined by*

$$\mathrm{shape}(X_i) = \min\{\mathrm{width}(X_i), \mathrm{length}(X_i)\}$$

The shape of the tree-decomposition is the maximum of the shapes of all its bags. Finally, the treeshape of G (resp., the pathshape of G), denoted by $\mathrm{ts}(G)$ (resp., $\mathrm{ps}(G)$), is the minimum, taken over all tree-decompositions (resp., path-decompositions) of G, of the shape of the decomposition.

We show that path-decompositions with small shape can be used to augment efficiently all graphs using a generic matrix and an appropriate labeling that depends on the path-decomposition.

Theorem 2 *For any $n \geq 1$, there exists an $n \times n$ matrix M and a labeling $\mathcal{L}$ of the nodes of any n-node graph G by integers in $\{1, \ldots, n\}$, such that $(G, (M, \mathcal{L}))$ has greedy diameter*

$$O\Big(\min\{\mathrm{ps}(G) \cdot \log^2 n,\ \sqrt{n}\}\Big).$$

Proof. We start by describing the matrix M, then we describe the node-labeling $\mathcal{L}$, and finally we analyze greedy routing in the graph G augmented by $(M, \mathcal{L})$.

To every integer $x \geq 1$ corresponds a unique integer $k \geq 0$ such that $x = 2^k + \alpha 2^{k+1}$ for some non negative integer α. This integer k is called the *level* of x, denoted by $\mathrm{level}(x)$, and corresponds to the position of the least significant bit of the binary writing of x. An integer x of level k has *ancestors* $y^{(j)}, j \geq 0$, of respective level $k + j$, defined as follows. If $x = 2^k + \sum_{i \geq k+1} x_i 2^i$ with $x_i \in \{0, 1\}$ for all i, then $y^{(j)} = 2^{k+j} + \sum_{i \geq k+j+1} x_i 2^i$. The set of all ancestors of x is denoted by $\mathcal{A}(x)$. (The terminology "ancestor" comes from the fact that this relation applied between consecutive levels induces an infinite binary tree whose leaves are all integers at level 0, i.e., all odd integers).

Let $A = (a_{i,j})_{1 \leq i,j \leq n}$ be the $n \times n$ matrix defined as follows. Assume n satisfies $2^{\nu-1} \leq n < 2^\nu$ for some integer $\nu \geq 1$. Then

$$a_{i,j} = \begin{cases} \frac{1}{1+\log n} & \text{if } j \in \mathcal{A}(i) \cap [1, n]; \\ 0 & \text{otherwise.} \end{cases}$$

A is an augmentation matrix because any index i of level $k \geq 0$ has at most $\nu - k$ ancestors in $[1, n]$, and $\nu - k \leq 1 + \log n$ for every $k \geq 0$. Let $U = (u_{i,j})_{1 \leq i,j \leq n}$ be the uniform matrix, i.e., $u_{i,j} = \frac{1}{n}$ for all i, j. We define $M = (p_{i,j})_{1 \leq i,j \leq n}$ by

$$M = (A + U)/2.$$

That is $p_{i,j} = \frac{1}{2}(a_{i,j} + u_{i,j})$ for all $1 \leq i, j \leq n$. The role of the matrix A, together with the labeling $\mathcal{L}$, is to enable long jumps between bags of a path-decomposition of the considered graph. These jumps are structured according to the hierarchy induced by the different node-levels, and by the ancestor relation. Finding the appropriate long jump requires roughly $O(\mathrm{ps}(G) \cdot \log n)$ expected number of steps, and there are $O(\log n)$ long jumps to be performed. The role of the uniform matrix U is to take care of graphs with large pathshape. It proceeds in parallel with A so as to avoid greedy routing to take more than $O(\sqrt{n})$ expected number of steps in total. The two matrices A and U can be run in parallel while preserving their respective good behavior thanks to the oblivious nature of greedy routing, and to the name-independent nature of the uniform augmentation.

Let G be a connected graph of n nodes, and let (P, X) be a path-decomposition of G, of optimal shape $\mathrm{ps}(G)$. Let b be the number of bags of the decomposition, i.e., the number of nodes of P. W.l.o.g., we can assume $b \leq n$. Indeed, we can restrict ourselves to reduced path-decompositions (i.e., path-decompositions in which no bag is contained in another one) without increasing the shape because if $Y \subseteq Y'$ then $\mathrm{shape}(Y) \leq \mathrm{shape}(Y')$. It is easy to show that the number of bags of a reduced path-decomposition does not exceed $\max\{1, n-1\}$ for an n-node connected graph (cf., e.g., [4]). Label the bags $X_1, \ldots, X_b$ of P consecutively from one extremity of the path to the other. This labeling induces a labeling of the nodes of G as follows.

Let $u \in V(G)$, and let $I_u \subseteq [1, b]$ be the interval of consecutive indices i such that $u \in X_i$ if and only if $i \in I_u$. We set the label $\mathcal{L}(u)$ of node u as the unique index in I_u such that

$$\mathrm{level}(\mathcal{L}(u)) = \max_{i \in I_u} \mathrm{level}(i).$$

The fact that $\mathcal{L}(u)$ is well defined comes from the fact that if $i_1, i_2 \in I_u$ satisfy $\mathrm{level}(i_1) = \mathrm{level}(i_2) = k$, and for any $i \in [i_1, i_2]$, $\mathrm{level}(i) \neq k$, then $i = (i_1 + i_2)/2 \in I_u$, and $\mathrm{level}(i) > k$. All node labels are in $\{1, \ldots, n\}$, but note that several nodes may receive the same label if $b < n$.

We show that the augmented graph $(G, (M, \mathcal{L}))$ has greedy diameter $O(\min\{\mathrm{ps}(G) \log^2 n,\ \sqrt{n}\})$. Let s and t be any two nodes of G. We show that the expected number of steps of greedy routing from s to t in $(G, (M, \mathcal{L}))$ is at most $O(\min\{\mathrm{ps}(G) \log^2 n,\ \sqrt{n}\})$.

If $\sqrt{n} \leq \mathrm{ps}(G) \log^2 n$, the result is clear. Indeed, at any step of greedy routing, the long range contact of the current node has probability at least $\frac{1}{2\sqrt{n}}$ to be at distance at most $\sqrt{n}$ from the target. Hence greedy routing reaches the target in expected time at most $3\sqrt{n}$.

The result when $\sqrt{n} > \mathrm{ps}(G) \log^2 n$ requires some more work. We use the binary hierarchy between the bags induced by the ancestor relation. The target t has label $\mathcal{L}(t)$. For every $i \in \mathcal{A}(\mathcal{L}(t)) \cap [1, b]$, let v_i be the closest node to t in X_i, where ties are broken arbitrarily. Note that $\mathcal{L}(t) \in \mathcal{A}(\mathcal{L}(t))$, and $v_{\mathcal{L}(t)} = t$. These nodes v_is are called *landmarks*. Let u be the current node. Initially, $u = s$. We define the *active* indices at u as the indices of the landmarks that are not further from t than u, i.e., the indices i such that $\mathrm{dist}_G(v_i, t) \leq \mathrm{dist}_G(u, t)$. Clearly, while greedy routing proceeds toward the target t, the number of active indices is non increasing. We compute the expected number of steps of greedy routing for decreasing the number of active indices by at least 1.

For every current node u along the greedy path from s to t, $\mathcal{A}(\mathcal{L}(u)) \cap \mathcal{A}(\mathcal{L}(t)) \cap [1, b] \neq \emptyset$ because the least common ancestor of $\mathcal{L}(u)$ and $\mathcal{L}(t)$ is between $\mathcal{L}(u)$ and $\mathcal{L}(t)$. Moreover, any index in $\mathcal{A}(\mathcal{L}(u)) \cap \mathcal{A}(\mathcal{L}(t)) \cap [1, b]$ is an active index at u. Indeed, by definition of the path-decomposition, a bag X_i, $1 < i < b$, is a separator of G. In particular, it separates $(\cup_{j=1}^{i-1} X_j) \setminus X_i$ from $(\cup_{j=i+1}^{b} X_j) \setminus X_i$. Therefore, $i \in \mathcal{A}(\mathcal{L}(u)) \cap \mathcal{A}(\mathcal{L}(t)) \cap [1, b]$ implies that X_i separates $X_{\mathcal{L}(u)} \setminus X_i$ from $X_{\mathcal{L}(t)} \setminus X_i$. Thus $\mathrm{dist}_G(v_i, t) \leq \mathrm{dist}_G(u, t)$, and hence i is active at u. Let us compute the probability p_0 that the long range contact v of u is at distance at most $\mathrm{ps}(G)$ from v_i for $i \in \mathcal{A}(\mathcal{L}(u)) \cap \mathcal{A}(\mathcal{L}(t)) \cap [1, b]$.

- If $\mathrm{shape}(X_i) < \mathrm{width}(X_i)$, then $\mathrm{shape}(X_i) = \mathrm{length}(X_i)$. Therefore, $p_0 \geq p_{\mathcal{L}(u), i}$ because all nodes in X_i are at mutual distance at most $\mathrm{shape}(X_i)$, that is at most $\mathrm{ps}(G)$.

- If $\mathrm{shape}(X_i) = \mathrm{width}(X_i)$, then, since p_0 is at least the probability that v_i is the long range contact of u, we get that $p_0 \geq p_{\mathcal{L}(u), i}/|X_i|$, and hence $p_0 \geq 1/((1 + \log n)(1 + \mathrm{ps}(G)))$.

Therefore, in both cases, we have

$$p_0 \geq \frac{1}{(1 + \log n)(1 + \mathrm{ps}(G))}.$$

This latter inequality is valid at each intermediate node u along the greedy path from s to t, independently from the fact that this node was reached after traversing a local link or a long link, and, in the latter case, independently from the fact that this long link was induced by the matrix A or the matrix U. Moreover, all trials for reaching a node at distance at most $\mathrm{ps}(G)$ from some v_i where i is an active index at the current node u, are mutually independent. As a consequence, after an expected number of steps at most $(1 + \log n)(1 + \mathrm{ps}(G))$, greedy routing goes from the current node u to its long range contact v at distance at most $\mathrm{ps}(G)$ from some landmark v_i where i is an active index at u. Since greedy routing decreases the distance to the target by at least 1 at each step, $\mathrm{ps}(G) + 1$ additional steps from v leads greedy routing to a node w satisfying $\mathrm{dist}_G(w, t) \leq \mathrm{dist}_G(v, t) - \mathrm{ps}(G) - 1$. Thus, by triangular inequality, $\mathrm{dist}_G(w, t) \leq \mathrm{dist}_G(v_i, t) - 1$. Hence, the index i is no longer active at w.

Therefore, after an expected number of steps at most $(2 + \log n)(1 + \mathrm{ps}(G))$, the number of active indices has decreased by at least 1.

Since there are at most ν active indices at the source s, and since $\nu \leq 1+\log n$, we get that, after an expected number of steps at most $(1+\log n)(2+\log n)(1+\mathrm{ps}(G))$, the number of active indices is at most 1. When only one active index remains, the current node u is in the same bag $X_{\mathcal{L}(t)}$ as the target. Moreover, no shortest path from u to t leaves $X_{\mathcal{L}(t)}$ because otherwise the number of active indices would increase. Hence the distance to the target is at most $\mathrm{shape}(X_{\mathcal{L}(t)}) \leq \mathrm{ps}(G)$. Therefore, after $O(\mathrm{ps}(G) \cdot \log^2 n)$ expected number of steps, greedy routing from s reaches the target t. $\square$

An important corollary of Theorem 2 is that the augmentation scheme $(M, \mathcal{L})$ offers a much better behavior than name-independent schemes for large classes of graphs, namely all those having small pathshape. Note that all classes mentioned in the corollary bellow include paths, for which all name-independent augmentation schemes have $\Omega(\sqrt{n})$ greedy diameter. Note also that the mentioned class of AT-free graphs[1] includes co-comporability graphs, interval graphs, and permutation graphs [5].

Corollary 1 *The universal augmentation scheme of Theorem 2 applied to n-node trees yields greedy diameter $O(\log^3 n)$. Applied to AT-free graphs, it yields greedy diameter $O(\log^2 n)$.*

Proof. Trees have treewidth 1, thus pathwidth at most $O(\log n)$. Hence, they have pathshape at most $O(\log n)$. AT-free graphs have constant pathlength, hence they have pathshape $O(1)$. $\square$

We conclude our analysis of matrix-based augmentation schemes by a discussion about the size of the labels. As we mentioned in the proof of Theorem 2, nodes may not be assigned different labels by the labeling $\mathcal{L}$. A natural question is whether the label set, and hence the matrix size, could be significantly reduced. The following theorem shows that this is impossible if one wants to preserve polylogarithmic greedy diameter for the classes of graphs mentioned in Corollary 1, or even just for paths.

Theorem 3 *Any matrix-based augmentation-labeling scheme using labels of size $\epsilon \log n$ for the n-node path, $0 \leq \epsilon < 1$, yields a greedy diameter $\Omega(n^\beta)$ for any $\beta < \frac{1}{3}(1 - \epsilon)$.*

[1]A graph is AT-free if it does not contain any *Asteroidal Triple*, i.e., a triple of vertices such that for every pair of them there is a path connecting the two vertices that avoids the neighborhood of the remaining vertex.

Proof. Let $0 \leq \epsilon < 1$, and consider an augmentation-labeling scheme using labels of size $\epsilon \log n$ for the n-node path. Let $k \leq n^\epsilon$ be the number of labels used by the labeling. W.l.o.g., these labels are $1, \ldots, k$. The augmentation scheme is described by an augmentation matrix $A = (p_{i,j})_{i,j}$ of size k. The probability that a fixed node u labeled i picks a fixed node v labeled j as long range contact is $p_{i,j}/N_j$ where N_j is the total number of nodes labeled j.

Let $0 < \alpha < \frac{2}{3}(1-\epsilon)$, and let $0 < \beta < \alpha/2$. We divide the n-node path into $n^{1-\beta}$ intervals of length n^β. A label ℓ is said *popular* if at least n^α nodes are labeled ℓ. Among all intervals, at most $n^{\epsilon+\alpha}$ contain a non popular label because there are at most n^ϵ non popular labels, and each of them can appear in at most n^α intervals. Hence there are at least $n^{1-\beta} - n^{\alpha+\epsilon}$ intervals that contain only popular labels. By the settings of α and β, $\alpha + \epsilon < 1 - \beta$, and thus there is at least one interval that contains only popular labels. Let I be such an interval, and let $x \in I$. Let us compute the probability p that x has its long range contact in I. Assuming x is labeled ℓ, we have

$$p = \sum_{i=1}^{k} \frac{C_i}{N_i}\, p_{\ell,i}$$

where C_i is the number of nodes labeled i in the interval I. Since I contains only popular labels, since $\sum_{i=1}^{k} C_i = |I|$, and since $p_{\ell,i} \leq 1$, we get that

$$p \;\leq \sum_{i=1}^{k} \frac{C_i}{n^\alpha} \;= \frac{n^\beta}{n^\alpha}.$$

The expected number of long range links with both extremities in I is $N = p\ |I|$. Therefore, $N \leq n^{2\beta-\alpha} < 1$ because $\beta < \alpha/2$. Let us now consider greedy routing from s to t where these two nodes are at mutual distance $|I|/3$, and each at distance $|I|/3$ from one extremity of the interval. On its way from s to t, greedy routing meets less than 1 expected number of long-range links leading to I, hence intuitively no possible shortcuts. Therefore, the expected number of steps of greedy routing from s to t is at least $\Omega(\mathrm{dist}(s,t))$. More precisely, let X be the random variable defined as the number of shortcuts of greedy routing from s to t, and let Y be the random variable defined as the number of steps of greedy routing from s to t. We have $\mathbb{E}(Y) \geq \mathbb{E}(Y|X<1) \cdot \Pr(X<1)$, and thus

$$\mathbb{E}(Y) \geq \frac{|I|}{3} \cdot \Pr(X<1).$$

From Markov inequality, $\Pr(X \geq 1) \leq N$. For n large enough, we have $N < 1/2$, and thus $\Pr(X<1) \geq 1/2$. Therefore, for n large enough, $\mathbb{E}(Y) \geq |I|/6$. Since $|I| = n^\beta$, we get that the greedy diameter of the considered augmentation is at least $\Omega(n^\beta)$. □

3 An $\tilde{O}(n^{1/3})$-Step Universal Augmentation Scheme

Neither the uniform scheme nor the augmentation scheme of Theorem 2 enables greedy routing to perform better than $\Omega(n^{1/2})$ expected number of steps for all graphs. The existence of a universal augmentation scheme overcoming the $\Omega(n^{1/2})$ barrier was actually open for some time. In this section, we show that there do exist faster schemes.

Theorem 4 *There exists a universal augmentation scheme φ yielding greedy diameter $\tilde{O}(n^{1/3})$ for n-node graphs.*

Proof. We describe the augmentation scheme φ explicitely. Let G be any (connected) graph. For any node $u \in V(G)$, and any integer $r \geq 0$, let $B(u,r) = \{v \in V(G) \mid \mathrm{dist}_G(u,v) \leq r\}$ be the ball of radius r centered at u. G is augmented as follows. First, every node chooses independently an integer $k \in \{1, \ldots, \lceil \log n \rceil\}$ uniformly at random. Then, the long range contact v of a node u that has chosen integer k is selected uniformly at random in $B_k(u) = B(u, 2^k)$. That is, if the *rank* $r(v)$ of a node v is the smallest k such that $v \in B_k(u)$, then

$$\varphi_u(v) = \frac{1}{\lceil \log n \rceil} \sum_{k=r(v)}^{\lceil \log n \rceil} \frac{1}{|B_k(u)|}.$$

We prove that the greedy diameter of (G, φ) is $\tilde{O}(n^{1/3})$. Let $s \in V(t)$ be the source, and $t \in V(G)$ be the target. Let $\mathbf{B}$ be a connected set of $n^{2/3}$ closest nodes to t (ties are broken arbitrarily). We consider five different phases before reaching the target.

Phase 1: Entering $\mathbf{B}$. For any $u \in V(G)$,

$$\Pr(u \to \mathbf{B}) = \sum_{v \in \mathbf{B}} \varphi_u(v) \geq \frac{|\mathbf{B}|}{n \lceil \log n \rceil} = \frac{1}{n^{1/3} \lceil \log n \rceil}.$$

Therefore the expected number of steps of greedy routing for entering $\mathbf{B}$ is at most $\tilde{O}(n^{1/3})$.

Phase 2: Leaving $\mathbf{B}$'s boundary. Since greedy routing decreases the distance to the target by at least 1 at each step, we get that $n^{1/3}$ steps after entering $\mathbf{B}$, the current node u_0 satisfies $B(u_0, n^{1/3}) \subseteq \mathbf{B}$. Thus for $k_0 = \lfloor \frac{1}{3} \log n \rfloor$, $B_{k_0}(u_0) \subseteq \mathbf{B}$.

Phase 3: Increasing the ball size. Starting at u_0, we compute the expected number of steps required to reach a node u_1 such that $t \in B_{k_1}(u_1) \subseteq \mathbf{B}$ for some $k_1 \geq k_0 = \lfloor \frac{1}{3} \log n \rfloor$. For this purpose, assume that the current node u satisfies $B_k(u) \subseteq \mathbf{B}$ for $k \geq k_0$ but $t \notin B_k(u)$. Let P_u be a shortest path from u to t, and let $Q_u = (P_u \cap B_k(u)) \setminus B_{k-1}(u)$. I.e., Q_u is the part of P_u containing all nodes v at distance from u satisfying $2^{k-1} < \mathrm{dist}_G(u,v) \leq 2^k$.

$$\Pr(u \to Q_u) \geq \frac{|Q_u|}{|B_k(u)| \cdot \lceil \log n \rceil} \geq \frac{|Q_u|}{|\mathbf{B}| \cdot \lceil \log n \rceil} = \frac{2^{k-1}}{n^{2/3} \cdot \lceil \log n \rceil}.$$

Therefore, the expected number of steps of greedy routing for reducing the distance by at least 2^{k-1} is at most $n^{2/3} \cdot \lceil \log n \rceil / 2^{k-1}$. Let u' be the current node just after this event occurs. If $B_{k+1}(u') \not\subseteq \mathbf{B}$ and $t \notin B_k(u')$, then one repeats for u' the same arguments as for u. Again, the expected number of steps of greedy routing for reducing the distance by at least 2^{k-1} is at most $n^{2/3} \cdot \lceil \log n \rceil / 2^{k-1}$. Hence, after $2n^{2/3} \cdot \lceil \log n \rceil / 2^{k-1}$ expected number of steps, greedy routing either reaches the target, or reaches a node u'' such that $B_{k+1}(u'') \subseteq \mathbf{B}$. If $t \notin B_{k+1}(u'')$, we repeat for u'' and $k+1$ the same reasoning as for u and k. Eventually, greedy routing reaches the desired node u_1 such that $t \in B_{k_1}(u_1) \subseteq \mathbf{B}$ for some $k_1 \geq k_0$. The expected number of steps from u_0 to u_1 is at most

$$2\, n^{2/3} \lceil \log n \rceil \sum_{k \geq k_0} \frac{1}{2^{k-1}} \leq 4\, n^{2/3} \lceil \log n \rceil \frac{1}{2^{k_0}} \sum_{k \geq 0} \frac{1}{2^k} \leq 8\, \frac{n^{2/3}}{2^{k_0}} \lceil \log n \rceil = \tilde{O}(n^{1/3}).$$

Therefore, the expected number of steps of Phase 3 is at most $\tilde{O}(n^{1/3})$.

Phase 4: Decreasing the ball size. Starting from node u_1, we compute the expected number of steps required to reach a node u_2 such that $t \in B_{k_0}(u_2) \subseteq \mathbf{B}$, for $k_0 = \lfloor \frac{1}{3} \log n \rfloor$. For this purpose, assume that the current node u satisfies $t \in B_k(u) \subseteq \mathbf{B}$ for $k_1 \geq k > k_0$, but $t \notin B_{k-1}(u)$. Again, we consider a shortest path P_u from u to t, and set $Q_u = (P_u \cap B_k(u)) \setminus B_{k-1}(u)$. For the same reason as in Phase 3,

$$\Pr(u \to Q_u) \geq \frac{2^{k-1}}{n^{2/3} \cdot \lceil \log n \rceil}$$

and thus the expected number of steps of greedy routing for reducing the distance by at least 2^{k-1} is at most $n^{2/3} \cdot \lceil \log n \rceil / 2^{k-1}$. When the distance has been reduced by at least 2^{k-1}, the current node u satisfies $t \in B_{k-1}(u) \subseteq \mathbf{B}$. One repeats the same analysis until $t \in B_{k_0}(u) \subseteq \mathbf{B}$. Eventually, greedy routing reaches the desired node u_2 such that $t \in B_{k_0}(u_2) \subseteq \mathbf{B}$. The expected number of steps from u_1 to u_2 is at most

$$n^{2/3} \lceil \log n \rceil \sum_{k \geq k_0} \frac{1}{2^{k-1}} \leq \tilde{O}(n^{1/3}).$$

Therefore, the expected number of steps of Phase 4 is at most $\tilde{O}(n^{1/3})$.

Phase 5: Reaching the target. Since u_2 is at distance at most $2^{k_0} \leq n^{1/3}$ from t, the target is eventually reached after at most $n^{1/3}$ additional steps.

Each of the five phases contributes for $\tilde{O}(n^{1/3})$ to the expected number of steps of greedy routing from s to t in (G, φ). Therefore, the greedy diameter of (G, φ) is $\tilde{O}(n^{1/3})$. □

4 Conclusion

In this paper, we focussed on universal augmentation schemes, i.e., augmentation schemes that can be efficiently applied to all graphs. Indeed, although it can be argued that the underlying graphs representing the acquaintances between individuals (before augmentation) satisfy some specific properties such as bounded doubling dimension or bounded treewidth, these hypotheses are far from being formally established. We believe that understanding the fundamental reasons for the emergence of the small world phenomenon requires a better understanding of what can be achieved in term of augmentation for arbitrary graphs. With this regards, this paper suggests two open problems that should be worth investigating:

- Open problem 1. Closing the gap between the $\Omega(n^{1/\sqrt{\log n}})$ lower bound in [12], and the $\tilde{O}(n^{1/3})$ upper bound in this paper.
- Open problem 2. Is it possible to overcome the $O(\sqrt{n})$ barrier by using matrix-based augmentation schemes? We have proved that achieving this task requires an appropriate node labeling (cf. Theorem 1), and cannot be done using labels significantly smaller than $\Omega(\log n)$ bits if one wants to preserve a polylogarithmic greedy diameter for paths (cf. Theorem 3).

References

[1] I. Abraham and C. Gavoille. Object Location Using Path Separators. 25th ACM Symp. on Principles of Distributed Computing (PODC), pages 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), 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] H. Bodlaender. A partial k-arboretum of graphs with bounded treewidth. Theoretical Computer Science, 209(1-2):1-45, 1998.

[5] A. Brandstädt, V. B. Le, and J. P. Spinrad. *Graph Classes – A survey.* SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, 1999.

[6] 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).

[7] P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Towards small world emergence. In 18th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), pages 225-232, 2006.

[8] Y. Dourisboure. Compact Routing Schemes for Bounded Tree-Length Graphs and for k-Chordal Graphs. 18th International Conference on Distributed Computing (DISC), LNCS 3274, Springer, pages 365-378, 2004.

[9] M. Flammini, L. Moscardelli, A. Navarra, and S. Perennes. Asymptotically optimal solutions for small world graphs. In 19th International Symposium on Distributed Computing (DISC), LNCS 3724, pages 414–428, 2005.

[10] P. Fraigniaud. Greedy routing in tree-decomposed graphs: a new perspective on the small-world phenomenon. In 13th Annual European Symposium on Algorithms (ESA), pages 791–802, 2005.

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

[12] P. Fraigniaud, E. Lebhar, and Z. Lotker. A Doubling Dimension Threshold $\Theta(\log \log n)$ for Augmented Graph Navigability. In 14th European Symposium on Algorithm (ESA), LNCS 4168, Springer, pages 376-386, 2006.

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

[14] J. Kleinberg. Complex networks and decentralized search algorithm. In International Congress of Mathematicians (ICM), 2006.

[15] 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, 2004.

[16] 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), pages 54-63,2004.

[17] C. Martel, and V. Nguyen. Analyzing Kleinberg's (and other) Small-world Models. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), pages 179-188, 2004.

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

[19] D. Peleg. Private communication. Workshop of COST Action 295 "DYNAMO", Les Ménuires, Jan. 2005.

[20] N. Robertson and P. D. Seymour. Graph minors I: Excluding a forest. J. Combin. Theory Ser. B, 35(1):39-61, 1983.

[21] A. Slivkins. Distance estimation and object location via rings of neighbors. In 24th Annual ACM Symposium on Principles of Distributed Computing (PODC), pages 41–50, 2005.

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