Theoretical Computer Science 410 (2009) 1970–1981

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Universal augmentation schemes for network navigability☆

Pierre Fraigniaud [a,*], Cyril Gavoille [b], Adrian Kosowski [c], Emmanuelle Lebhar [a], Zvi Lotker [d]

[a] CNRS and University Paris Diderot, 75205 Paris, France
[b] University of Bordeaux, 33405 Talence, France
[c] Gdansk University of Technology, 80-952 Gdańsk Wrzeszcz, Poland
[d] Ben Gurion University, Beer-Sheva, Israel

ARTICLE INFO

Article history:
Received 5 October 2007
Received in revised form 8 October 2008
Accepted 8 December 2008
Communicated by D. Peleg

Keywords:
Small world phenomenon
Informative labeling schemes
Routing

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. We define an augmented graph as a pair (G, M) where G is an n-node graph with nodes labeled in $\{1, \ldots, n\}$, and M is an $n \times n$ stochastic matrix. 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 $\Pr\{u \to v\} = M_{u,v}$. In augmented graphs, greedy routing is the oblivious routing process in which every intermediate node chooses from 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. The best augmentation scheme known so far ensures that, for any n-node graph G, greedy routing performs in $O(\sqrt{n})$ expected number of steps.

Our main result is the design of an augmentation scheme that overcomes the $O(\sqrt{n})$ barrier. Precisely, we prove that for any n-node graph G whose nodes are arbitrarily labeled in $\{1, \ldots, n\}$, there exists a stochastic matrix M such that greedy routing in (G, M) performs in $\tilde{O}(n^{1/3})$, where the $\tilde{O}$ notation ignores the polylogarithmic factors.

We prove additional results when the stochastic matrix M is universal to all graphs. In particular, we prove that the $O(\sqrt{n})$ barrier can still be overcame for large graph classes even if the matrix M is universal. This however requires an appropriate labeling of the nodes. If the node labeling is arbitrary, then we prove that the $O(\sqrt{n})$ barrier cannot be overcome with universal matrices.

1. Introduction

Augmented graphs were defined in [18,19] for the purpose of understanding the "small world phenomenon", as a model for greedy decentralized search.Precisely, augmented graphs give one framework for modeling and analyzing the "six degrees of separation" between individuals observed from Milgram's experiment [7,29], 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 the study of navigable small-world networks has given rise to an abundant literature (cf., e.g., [1–3,9,12,13,15,18,23–26,33]). We refer to Kleinberg's survey [21] on complex networks for more details on the concept

☆ A preliminary version of this paper appeared in the proceedings of the 19th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), June 2007, San Diego, CA, USA.

* Corresponding author. Tel.: +33 1 44 27 68 45.
E-mail address: Pierre.Fraigniaud@liafa.jussieu.fr (P. Fraigniaud).

0304-3975/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.12.061

of augmented graphs; in particular, see also [20,24,33] for different models for analyzing the small world phenomenon, and [27] for the application of the concept of augmented graphs to the design of communication networks.

Formally, an augmentation scheme for an n-node graph G is defined by a collection $\varphi = \{\varphi_u, u \in V(G)\}$ of probability distributions, such that every node $u \in V(G)$ is given a certain number of extra links pointing to some nodes, called the *long range contacts* of u. Each long range contact of u is chosen at random according to φ by $\Pr\{u \to v\} = \varphi_u(v)$. The links of the underlying graph G are called *local links*. A link from a node to one of its long range contacts is called a *long range link*. The choices of the long range links are mutually independent.

For the purpose of a detailed analysis of augmented graphs, we slightly refine this definition. We define an augmentation scheme for an n-node graph G as a pair (λ, M) where $\lambda : V(G) \to \{1, \ldots, n\}$ is a one-to-one labeling of the nodes, and $M = (M_{i,j})$ is an $n \times n$ stochastic matrix (i.e., each row consists of nonnegative real numbers summing to 1). Every node $u \in V(G)$ is given one[1] long range contact v, chosen at random according to M as follows: $\Pr\{u \to v\} = M_{\lambda(u),\lambda(v)}$. Clearly, once a node labeling λ has been fixed, and given an augmentation φ, the matrix M is simply defined by $M_{\lambda(u),\lambda(v)} = \varphi_u(v)$. Our definition based on a pair (λ, M) will allow us to distinguish between several cases, including whether or not the matrix M depends on the graph G, and whether or not the labeling λ depends on the matrix M. The triple (G, λ, M) defines a model for random graphs obtained from G by adding long-range links at random according to the labeling λ and the stochastic matrix M.

Greedy routing in augmented graphs was introduced in [18]. It 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(s) of u), and (2) forwarding the message to v. 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(s). Hence the nodes have no notion of the distances in the augmented graph. (Note that the way nodes identify themselves in G, and the way they encode distances in the graph are outside the scope of this paper. In particular, the labeling λ is solely used to augment the graph, not for routing.)

We define the *greedy diameter* of (G, λ, M) as $D_{greedy}(G, \lambda, M) = \max_{s,t \in V(G)} \mathbb{E}(X_{s,t})$, where $X_{s,t}$ is the random variable counting the number of steps for traveling from s to t using greedy routing in (G, λ, M). Note that the expectation here is taken over the choices of the augmented graph, whereas the routing algorithm itself is deterministic. Let $f : \mathbf{N} \to \mathbf{R}$ be a function. An n-node graph G is *f-navigable* if there exists a labeling λ and an $n \times n$ stochastic matrix M such that $D_{greedy}(G, \lambda, M) \le f(n)$.

Lots of effort has been devoted to characterize families of graphs that are polylog(n)-navigable (cf. the survey [21]). For instance, it is known [18] that for any fixed $d \ge 1$, the d-dimensional meshes are $O(\log^2 n)$-navigable. More generally, it was proved that all graphs of bounded ball growth are polylog(n)-navigable [8] (see also [20]), which was even generalized in [33] to all graphs of bounded doubling dimension. Similarly, all graphs of bounded treewidth [13], and more generally all graphs excluding a fixed minor [1] are polylog(n)-navigable. All the augmentation schemes proposed in the aforementioned papers are however specifically designed to apply efficiently to each of the considered classes of 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 [30] 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. This analysis is tight in the sense that Kleinberg [18] proved an $\Omega(\sqrt{n})$ lower bound for uniform augmentation on 2D meshes. To the best of our knowledge, the $O(n^{1/2})$ upper bound was the best known bound for arbitrary graphs until this paper. On the other hand, it was recently proved [16] 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 line of work dealing with small-world network navigability is to close the gap between these upper and lower bounds for the f-navigability of arbitrary graphs. This has been done in [14], but at the price of relaxing the definition of greedy routing: for any connected n-node graph G and any integer $k \ge 1$, there exists X and Y an augmentation of G and a (semi)metric μ on G with stretch $2k - 1$ with respect to the shortest distance metric, 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, there exists X and Y an augmentation and a (semi)metric μ with stretch $O(\log n)$ such that greedy routing according to μ performs in a polylogarithmic expected number of steps).

1.1. Our results

We first consider augmentation schemes that are dependent on the structure of the graph, that is for which the stochastic matrix M depends on the graph. Our main result is the design of a universal augmentation scheme that overcomes the $O(\sqrt{n})$ barrier. Precisely, we prove that for any n-node graph G (and for any labeling λ of the nodes of G) there exists a stochastic matrix M such that

$$D_{greedy}(G, \lambda, M) \le \tilde{O}(n^{1/3}),$$

where the $\tilde{O}$ notation ignores the polylogarithmic factors.

[1] All results stated in this paper also hold if $O(\text{polylog}(n))$ long range contacts are given to each node.

We then consider augmentation schemes defined from stochastic matrices universal to all graphs. In this context, results depends on whether the node labeling λ is also set *a priori*, or as a function of the graph G and the matrix M.

– When the node labeling λ is arbitrary, we prove that the uniform matrix $U = (u_{i,j})$ with $u_{i,j} = \frac{1}{n}$ is optimal. Precisely, we prove that for any $n \times n$ matrix M, there is a node labeling λ of the n-node path P_n such that $D_{greedy}(P_n, \lambda, M) \geq \Omega(\sqrt{n})$. Although this bound demonstrates the limits of augmentation schemes that work for arbitrary labeling, these schemes remain useful. Indeed, in addition to their simplicity, they can be combined with label-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 what is done hereafter.

– When the node labeling λ depends on the matrix M, we prove that there exists a stochastic matrix M such that, for any n-node graph G, there exists a labeling λ for which

$$D_{greedy}(G, \lambda, M) \leq \tilde{O}\left(\min\{\mathrm{pw}(G), \mathrm{pl}(G), \sqrt{n}\}\right), \tag{1}$$

where $\mathrm{pw}(G)$ denotes the pathwidth [31] of G, and $\mathrm{pl}(G)$ denotes the pathlength [11] of G. This result has many important corollaries. In particular, such a scheme yields a 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 graphs and permutation graphs [5]. These classes were not captured by previous results, since in general they are neither of bounded doubling dimension nor exclude a fixed minor. Our result is actually slightly more general than Eq. (1), and is defined in term of the *pathshape* $\mathrm{ps}(G)$ of the graph G, a parameter that we define for achieving better tradeoff between pathwidth and pathlength: $D_{greedy}(G, \lambda, M) \leq \tilde{O}(\min\{\mathrm{ps}(G), \sqrt{n}\})$. This latter bound is essentially in the sense that we also prove that, for any $k \leq \frac{1}{2}\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}$, there exists an n-node graph G of pathshape at most k such that, for any M and any λ, $D_{greedy}(G, \lambda, M) \geq \Omega(k + \log n)$. Finally, we prove that the dimension of the matrices used for augmenting graphs cannot be reduced significantly. Or, in other words, the size of the labels cannot be reduced significantly. This is true if one wants to preserve a polylogarithmic greedy diameter even just for paths. Precisely, we prove that 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 $\beta < \frac{1}{3}(1 - \epsilon)$.

2. An $\tilde{O}(n^{1/3})$-step augmentation scheme

The existence of an augmentation scheme overcoming the $\Omega(n^{1/2})$ barrier was open for some time. In this section, we show that there do exist faster schemes.

Theorem 1. *For any n-node graph G and any labeling λ of the nodes of G in $\{1, \ldots, n\}$, there exists an $n \times n$ stochastic matrix M such that $D_{greedy}(G, \lambda, M) \leq \tilde{O}(n^{1/3})$, where the $\tilde{O}$ notation ignores the polylogarithmic factors.*

Corollary 1. *Any n-node graph G can be augmented with only one long range link per node such that the expected number of steps of greedy routing from any source to any target is $\tilde{O}(n^{1/3})$.*

To capture the intuition of the proof of Theorem 1, let us prove a weaker result, $O(n^{2/5})$. One ingredient in the proof is the choice of the long-range links, that cover several scales of distances. For proving $O(n^{2/5})$, we use just two scales: the long-range contact of a node u is chosen uniformly at random in $V(G)$ with probability $1/2$, and uniformly at random in B_u with probability $1/2$ where B_u is the ball centered at u of radius $n^{2/5}$. Consider now greedy routing from s to t in G. Let $\mathbf{B}$ be the set of the $n^{3/5}$ nodes closest to t in G. The probability of any node to have its long-range contact in $\mathbf{B}$ is $\Omega(n^{-2/5})$, hence the expected number of steps to enter $\mathbf{B}$ is $O(n^{2/5})$. After having entered $\mathbf{B}$, greedy routing reaches a node u such that $B_u \subseteq \mathbf{B}$ after at most $n^{2/5}$ additional steps. The second ingredient in the proof of Theorem 1 is the ability to combine the size of $\mathbf{B}$ with the size of B_u. Let P be a shortest path from u to t in G, and let $Q = P \cap B_u$. Let Q' be the segment of Q consisting in the $\frac{n^{2/5}}{2}$ nodes of Q at distance at least $\frac{n^{2/5}}{2}$ from u. The probability for u to have its long-range link in Q' is $\Omega(|Q'|/|B_u|) \geq \Omega(|Q'|/|\mathbf{B}|) \geq \Omega(n^{-1/5})$. Therefore, a shortcut of length at least $\frac{n^{2/5}}{2}$ is used every $O(n^{1/5})$ steps in expectation. Since the radius of $\mathbf{B}$ is at most its size $n^{3/5}$, using $O(n^{1/5})$ times a shortcut of length at least $\frac{n^{2/5}}{2}$ leads to the target t. This requires $O(n^{2/5})$ steps in expectation. Hence the expected number of steps from s to t is $O(n^{2/5})$ in total. Improving from $O(n^{2/5})$ to $\tilde{O}(n^{1/3})$ requires the use of $\log n$ scales instead of just 2, and a more sophisticated analysis for comparing the size of $\mathbf{B}$ with the size of the $\log n$ balls of different scales centered at the current node.

Proof. We describe the augmentation scheme explicitly. It has a hierarchical structure so that long-range links scale at all distances, and offer the one-over-ball-size nature used in, e.g., [8,20,33]. Let G be any n-node (connected) graph. The node-labeling $\lambda : V(G) \to \{1, \ldots, n\}$ is arbitrary. Hence, for the sake of simplifying the notations, we do not distinguish u and $\lambda(u)$ for the labeling of a node u. 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

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

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

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

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

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

Phase 2: Leaving **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 **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 a distance from u satisfying $2^{k-1} < \mathrm{dist}_G(u, v) \leq 2^k$.

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

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

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

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 a 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 by $\tilde{O}(n^{1/3})$ to the expected number of steps of greedy routing from s to t in (G, λ, M). Therefore, the greedy diameter of (G, λ, M) is $\tilde{O}(n^{1/3})$. □

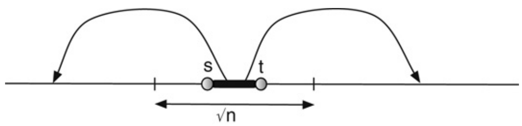

Fig. 1. Intuition of the proof of Theorem 2.

3. Matrix-based augmentation schemes

This section considers a restricted though rich class of augmentation schemes, those for which the stochastic $n \times n$ matrix M is fixed *a priori*, and is universal to all (connected) n-node graphs. We call *matrix-based augmentations schemes* such a class of augmentation schemes. In this context, one may either desire the scheme to be used independently from any node-labeling, or preserve the ability to define an appropriate node-labeling that fits with the matrix M. The former case is called *name-independent*. 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 when the node-labeling can be adapted to the matrix, 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 performance than the latter for large classes of graphs.

3.1. 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 2. *For any $n \times n$ stochastic matrix M, there exists a node-labeling λ of the n-node path P_n such that $D_{greedy}(P_n, \lambda, M) \geq \Omega(\sqrt{n})$.*

Roughly, the proof of Theorem 2 consists in finding a subset $I \subseteq [1, n]$ of size $\sqrt{n}$ such that $\sum_{i,j \in I, i \neq j} M_{i,j} < 1$. By labeling a set of $\sqrt{n}$ consecutive nodes with the labels in I, we get a segment of the path such that all the long-range links of the nodes in this segment point out of the segment, with constant probability (see Fig. 1). Therefore, during its journey from a source s to a target t at distance $\Omega(\sqrt{n})$ in the segment, greedy routing does not find any shortcuts with constant probability. As a consequence, the length of the journey is at least $\Omega(\sqrt{n})$ with constant probability, and thus $\mathbb{E}(X_{s,t}) \geq \Omega(\sqrt{n})$.

Proof. We show that, for any augmentation matrix M of size n, there is a labeling of the n-node path with integers in $[1, n]$ such that the greedy diameter of the labeled path augmented using M is $\Omega(\sqrt{n})$. Let $M = (M_{i,j})_{1 \leq i,j \leq n}$ be an $n \times n$ augmentation matrix. In order to simplify the details of the proof, we make the assumption that $\sqrt{n}$ is an integer. We claim that:

$$\exists I \subseteq [1, n], |I| = \sqrt{n} \quad \text{and} \quad \sum_{i,j \in I, i \neq j} M_{i,j} < 1.$$

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

$$\sum_{I \subset [1,n], |I| = \sqrt{n}} \sum_{i,j \in I, i \neq j} M_{i,j} \geq \binom{n}{\sqrt{n}}.$$

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

$$\{M_{i,j}, j \in [1, n] \setminus \{i\}\}$$

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

$$\sum_{I \subset [1,n], |I| = \sqrt{n}} \sum_{i,j \in I, i \neq j} M_{i,j} \leq (n - \sqrt{n})\binom{n}{\sqrt{n}}/(n-1) < \binom{n}{\sqrt{n}},$$

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

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 $\mathbb{E}(X) = \sum_{i,j\in I, i\neq j} M_{i,j}$, and thus $\mathbb{E}(X) < 1$. From Markov's inequality, we get $\Pr\{X \geq 2\} \leq \mathbb{E}(X)/2 < 1/2$. Partition the set S into three consecutive intervals S_1, S_2 and S_3 of equal cardinality. Consider three pairs of nodes $s_i, t_i \in S_i$, $1 \leq i \leq 3$, such that s_i is at distance $|S_i|/3$ from one extremity of S_i, t_i is at distance $|S_i|/3$ from the other extremity of S_i, and s_i and t_i are at mutual distance $|S_i|/3$, for $1 \leq i \leq 3$.

Let Y_i be the random variable equal to the number of steps of greedy routing from s_i to t_i for $1 \leq i \leq 3$, and let $Y = Y_1 + Y_2 + Y_3$. Note that if the event $\{X = 0\}$ occurs, none of the greedy routes between the three pairs of source and target uses any long range link, and the number of steps is simply their mutual distance. If the event $\{X = 1\}$ occurs, the position of this single long range link in S implies that at least one of the three greedy routes, between s_1 and t_1, s_2 and t_2, or s_3 and t_3 does not use any long range link. We get:

$$\mathbb{E}(Y) \geq \mathbb{E}(Y \mid X < 2) \cdot \Pr\{X < 2\} > \frac{|S|}{9} \cdot \frac{1}{2} \geq \frac{\sqrt{n}}{18}.$$

Therefore, the greedy diameter of this labeled path is at least $\mathbb{E}(Y_1 + Y_2 + Y_3)/3$ which is $\Omega(\sqrt{n})$, which completes the proof. $\square$

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, in certain cases, be combined with label-dependent schemes that perform well for specific classes of graphs but poorly in general. In particular, the uniform scheme can be combined with certain schemes that are efficient for some classes of graphs only, in order to preserve the $O(\sqrt{n})$ greedy diameter for general graphs. This is proven in the next section.

3.2. An $\tilde{O}(\min\{\mathrm{pw}(G), \mathrm{pl}(G), \sqrt{n}\})$*-step matrix-based augmentation scheme*

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 [32] and treelength [11]. These two latter notions have been proved important in many contexts, including algorithm design [6], compact routing [10], and information labeling [17].

Recall that a tree-decomposition [32] 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 following three 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 (i.e., the subgraph of tree T induced by the set $\{i \in I | u \in X_i\}$ is connected).

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 a 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 [31] and the length [11]:

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

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 prove very relevant to augmentation schemes.

Definition 1. 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.

In this paper, we focus on pathshape. By definition, we have $\mathrm{ps}(G) \leq \mathrm{pw}(G)$ and $\mathrm{ps}(G) \leq \mathrm{pl}(G)$ for any graph G, where $\mathrm{pw}(G)$ and $\mathrm{pl}(G)$ denotes the pathwidth and pathlength of G, respectively. More interestingly, there are graphs G for which $\mathrm{ps}(G) \ll \min\{\mathrm{pw}(G), \mathrm{pl}(G)\}$. A trivial example is a long cycle in which one node is replaced by a large clique. A more realistic example is the n-node graph G depicted on Fig. 2, in which a set of trees are pending from the nodes of a dense core graph. (This is in essence the structure of the Internet in which a dense backbone connects smaller domains that are internally hierarchically organized in a tree-like structure [34]). Assume that the core contains m nodes, and that there are k trees pending from the core, with at least three trees of depth at least m'. On the one hand, if the core of G is the complete graph K_m then the treewidth of G is at least $m-1$, and thus $\mathrm{pw}(G) \geq m-1$. (This also holds up to a constant if the core is an m-node bounded-degree expander [22,28]). On the other hand, since at least three paths of length at least m' are pending from the

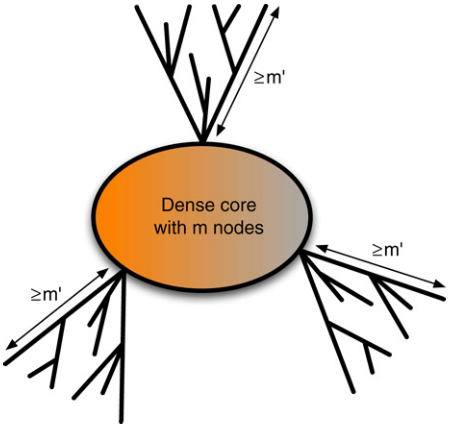

Fig. 2. A graph with small pathshape, but high pathwidth and pathlength.

core, we get by Theorem 6 of [11] that $\mathrm{pl}(G) \geq m'$. We claim that $\mathrm{ps}(G) \leq O(D + k \log n)$ where D is the diameter of the core. To establish the claim, we construct the following path-decomposition. One extremity of the path is a bag containing all nodes in the core. This bag has length D. To this bag are attached the k path-decompositions of width at most $O(\log n)$ of the k trees pending from the core. Every bag but the core-bag in the decomposition is thus the union of at most k bags of width at most $\log n$. Their width is therefore at most $O(k \log n)$. By setting $m \geq \Omega(n)$, $m' \geq \Omega(n)$, $D \leq O(\log n)$, and $k \leq O(1)$, we get that $\mathrm{ps}(G) \leq O(\log n)$ whereas $\min\{\mathrm{pw}(G), \mathrm{pl}(G)\} \geq \Omega(n)$.

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 3. *For any $n \geq 1$, there exists an $n \times n$ stochastic matrix M such that, for any n-node graph G, there exists a node labeling λ satisfying*

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

As a consequence, $D_{greedy}(G, \lambda, M) \leq O\Big(\min\left\{\mathrm{pw}(G) \cdot \log^2 n, \mathrm{pl}(G) \cdot \log^2 n, \sqrt{n}\right\}\Big)$.

Corollary 2. *Any n-node graph G can be augmented with only one long range link per node such that the expected number of steps of greedy routing from any source to any target is* $O\Big(\min\left\{\mathrm{pw}(G) \cdot \log^2 n, \mathrm{pl}(G) \cdot \log^2 n, \sqrt{n}\right\}\Big)$.

Roughly, the matrix M in Theorem 3 is set to simulate the distribution obtained by recursively decomposing the path using balanced separators, like in [13]. The labeling of the nodes is based on a path-decomposition of the graph, with shape $\mathrm{ps}(G)$. Bags are labeled consecutively from 1 to n, and each node takes as its label the label of some specific bag containing it. The analysis of greedy routing is achieved following the same guideline as for the path in [18]. In fact, the $O(\log^2 n)$ factor comes from the same reasons as it appears in the analysis of greedy routing in harmonically augmented paths [18] and in trees augmented using recursive separators [13]. There are however differences between a path and a path-decomposition. In particular, the same bag can be visited several times. Nevertheless, the number of times a bag can be visited while routing greedily from a source to a target is bounded by $\mathrm{ps}(G)$. Indeed, if the bag has width $k \leq \mathrm{ps}(G)$, then it contains k nodes, and each of them can be visited only once. If the bag has length $k \leq \mathrm{ps}(G)$, then its diameter is at most k, and thus a shortest path will not traverse more than k nodes in the bag. There is therefore a slowdown of at most $\mathrm{ps}(G)$ between greedy routing in a path and greedy routing in a path-decomposition of shape $\mathrm{ps}(G)$. The formal proof is given below.

Proof. We start by describing the matrix M, then we describe the node-labeling λ, and finally we analyze greedy routing in the graph G augmented by (G, λ, M).

To every integer $x \geq 1$, we define the *level* of x, denoted by $\mathrm{level}(x)$, by

$$\mathrm{level}(x) = \max\{k \geq 0, x \equiv 0 \bmod 2^k\}.$$

Thus $\mathrm{level}(x)$ corresponds a unique integer $k \geq 0$ such that $x = 2^k + \alpha 2^{k+1}$ for some non-negative integer α. An integer x of level k has *ancestors* $y^{(j)}$, $j \geq 0$, of respective level $k + j$, defined as follows:

$$y^{(j)} = (x \div 2^{k+j+1}) * 2^{k+j+1} + 2^{k+j}$$

where $\div$ denotes the integer division. In other words, 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 $A = (a_{i,j})_{1\le i,j\le n}$ be the $n \times n$ matrix defined as follows. Assume n satisfies $2^{\nu-1} \le n < 2^{\nu}$ for some integer $\nu \ge 1$. 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 $k \ge 0$ has at most $\nu - k$ ancestors in $[1, n]$, and $\nu - k \le 1 + \log n$ for every $k \ge 0$. Let $U = (u_{i,j})_{1\le i,j\le n}$ be the uniform matrix, i.e., $u_{i,j} = \frac{1}{n}$ for all i, j. We define $M = (M_{i,j})_{1\le i,j\le n}$ by

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

That is, $M_{i,j} = \frac{1}{2}(a_{i,j} + u_{i,j})$ for all $1 \le i, j \le n$. The role of the matrix A, together with the labeling λ, 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 guarantee that greedy routing does not 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 $b \le n$. 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 $\mathrm{shape}(Y) \le \mathrm{shape}(Y')$. 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 us define

$$I_u = \{i \in [1, b], u \in X_i\}.$$

By definition of path-decomposition, I_u is an interval of consecutive integers. We set the label of node u as the unique index $j \in I_u$ such that $\mathrm{level}(j) = \max_{i\in I_u} \mathrm{level}(i)$, that is

$$\lambda(u) = \operatorname*{argmax}_{i\in I_u} \mathrm{level}(i).$$

The fact that $\lambda(u)$ is uniquely 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 $i \in [i_1, i_2]$, $\mathrm{level}(i) \ne k$, then $i = (i_1 + i_2)/2 \in I_u$, and $\mathrm{level}(i) > k$. All node labels are in $[1, n]$, but note that several nodes may receive the same label if $b < n$.

We show that the augmented graph (G, λ, M) 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, λ, M) is at most $O(\min\{\mathrm{ps}(G) \log^2 n, \sqrt{n}\})$.

If $\sqrt{n} \le \mathrm{ps}(G) \log^2 n$, 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 a 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 $\lambda(t)$. For every $i \in \mathcal{A}(\lambda(t)) \cap [1, b]$, let v_i be the closest node (according to dist_G) to t in X_i, where ties are broken arbitrarily. Note that $\lambda(t) \in \mathcal{A}(\lambda(t))$, and $v_{\lambda(t)} = t$. 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 $\mathrm{dist}_G(v_i, t) \le \mathrm{dist}_G(u, t)$. 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, $\mathcal{A}(\lambda(u)) \cap \mathcal{A}(\lambda(t)) \cap [1, b] \ne \emptyset$ because the least common ancestor of $\lambda(u)$ and $\lambda(t)$ is between $\lambda(u)$ and $\lambda(t)$. Moreover, any index in $\mathcal{A}(\lambda(u)) \cap \mathcal{A}(\lambda(t)) \cap [1, b]$ 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, $i \in \mathcal{A}(\lambda(u)) \cap \mathcal{A}(\lambda(t)) \cap [1, b]$ implies that X_i separates $X_{\lambda(u)} \setminus X_i$ from $X_{\lambda(t)} \setminus X_i$. (The case $\lambda(u) = \lambda(t)$ is considered separately in the final stage of the proof.) Thus $\mathrm{dist}_G(v_i, t) \le \mathrm{dist}_G(u, t)$, 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 $i \in \mathcal{A}(\lambda(u)) \cap \mathcal{A}(\lambda(t)) \cap [1, b]$.

- If $\mathrm{shape}(X_i) < \mathrm{width}(X_i)$, then $\mathrm{shape}(X_i) = \mathrm{length}(X_i)$. Therefore, $p_0 \ge M_{\lambda(u),i}$ 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 $p_0 \ge M_{\lambda(u),i}/|X_i|$, and hence $p_0 \ge 1/((1 + \log n)(1 + \mathrm{ps}(G)))$.

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 lead greedy routing to a node w satisfying $\mathrm{dist}_G(w,t) \leq \mathrm{dist}_G(v,t) - \mathrm{ps}(G) - 1$. Thus, by triangle 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 $X_{\lambda(t)}$ as the target. Moreover, no shortest path from u to t leaves $X_{\lambda(t)}$ because otherwise the number of active indices would be more than 1. Hence the distance to the target is at most shape $(X_{\lambda(t)}) \leq \mathrm{ps}(G)$. 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 3 is that the augmentation scheme (M,λ) offers much better behavior than name-independent schemes for large classes of graphs. Note that all classes mentioned in the corollary below 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[2] includes co-comparability graphs, interval graphs, and permutation graphs [5].

Corollary 3. *The augmentation scheme of Theorem 3 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 $O(\log n)$. It can be easily checked that AT-free graphs have constant pathlength, hence they have pathshape $O(1)$. $\square$

The upper bound in Theorem 3 is essentially tight, in the sense that we have the following:

Theorem 4. *For any $n \geq 1$ and any $k \leq \frac{1}{2}\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}$, there exists an n-node graph of pathshape at most k such that, for any $n \times n$ stochastic matrix M, and any node labeling λ, $D_{greedy}(G,\lambda,M) \geq \Omega(k+\log n)$.*

Proof. To establish the proof, we make a preliminary statement. Let G_1 and G_2 be graphs of n_1 and n_2 nodes, respectively. Let H be the graph of $n = n_1 + n_2$ nodes obtained from G_1 and G_2 by linking them via an edge connecting some node x_1 of G_1 with some node x_2 of G_2 (see Fig. 3(a)). For $i = 1, 2$, let y_i be a node at some distance d_i from x_i in H. Let $B_i = \{z \in V(H), \mathrm{dist}_H(y_i, z) < d_i/3\}$. Let s_i, t_i be two nodes of B_i. We have $\mathrm{dist}_H(s_i,t_i) < 2d_i/3$, and $\mathrm{dist}_H(t_i,x_i) \geq 2d_i/3$. Therefore, for any augmentation of H, greedy routing from s_i to t_i traverses only nodes of G_i. As a consequence we get that the greedy diameter of H is at least the maximum of the greedy diameters of B_1 and B_2. More formally, let M_i be a $n_i \times n_i$ stochastic matrix, and λ_i be a node labeling in $\{1,\ldots,n_i\}$ such that the expected number of steps $D_{greedy}(B_i,\lambda_i,M_i)$ of greedy routing in (G_i,λ_i,M_i) between nodes of B_i is minimized. Then for any $n \times n$ stochastic matrix M, and any node labeling λ, we have

$$D_{greedy}(H,\lambda,M) \geq \max_{i=1,2} D_{greedy}(B_i,\lambda_i,M_i). \tag{2}$$

We will use this inequality to establish the theorem, by using two specific graphs G_1 and G_2. One of them is a path. The description of the other graph follows.

We consider the graph $G(p,d)$, where p and d are integers, defined as follows. Nodes are d-tuples $(x_1,\ldots,x_d)$, $x_i \in \mathbb{Z}_p$, and node $(x_1,\ldots,x_d)$ is connected to all nodes $(x_1+\delta_1,\ldots,x_d+\delta_d)$, where $\delta_i \in \{-1,0,+1\}$ and additions are taken modulo p. The graph $G(p,d)$ can be seen as a d-dimensional grid of side p with all "diagonal edges". This graph has p^d nodes, and we can check that its diameter is $\lfloor p/2 \rfloor$. The following lower bound on the greedy diameter of $G(p,d)$ is a rephrasing of a result in [16]: For all p, d such that $p \geq 2^d \geq 1$,

$$D_{greedy}(G(p,d),\lambda,M) \geq \lfloor p/4 \rfloor \tag{3}$$

for any $p^d \times p^d$ stochastic matrix M, and any node labeling λ. The proof of Eq. (3) in [16] uses the fact that, for any augmentation, the greedy path from a source $s = (s_1,\ldots,s_d)$ to a target $t = (t_1,\ldots,t_d) = (s_1+\delta_1\ \tau,\ldots,x_d+\delta_d\ \tau)$,

[2] 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.

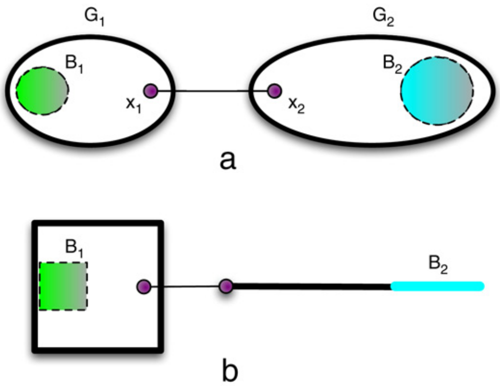

Fig. 3. Construction of H as in the proof of Theorem 4.

with $\delta_i \in \{-1, +1\}$ and $\tau > 0$, traverses only nodes on the same "diagonal", i.e., of the form $(s_1 + \delta_1 \ell, \ldots, x_d + \delta_d \ell)$ for $0 \le \ell < 2\tau$. Then a counting argument based on the fact that there are 2^d different diagonals enables us to establish that the expected number of steps before using a long-range link along the adequate diagonal is at least 2^d. The same reasoning applies if one restricts the analysis to source-target pairs inside a ball B of radius $p/6$. Inside such a ball, the bound of Eq. (3) becomes: for any augmentation (λ, M) of $G(p, d)$ the expected number of steps $D_{greedy}(B, \lambda, M)$ of greedy routing between nodes of B satisfies

$$D_{greedy}(B, \lambda, M) \ge \Omega(p). \tag{4}$$

We are now ready to prove the theorem. Let $d(k) = \lfloor \log k \rfloor + 1$. For every integer k such that $1 \le k \le \frac{1}{2}\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}$, we consider the graph H obtained from $G_1 = G(2k, d(k))$ by adding a path G_2 with $\ell = n - (2k)^{d(k)}$ nodes that is connected to G_1 by an edge linking one of the extremities of G_2 to an arbitrary node of G_1 (see Fig. 3(b)).

Graph H has n nodes. Let us show that the number of nodes ℓ of G_2 satisfies $\ell \ge n/2$, or equivalently, that the number of nodes in G_1 is at most $n/2$. Since $k \le \frac{1}{2}\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}$, we have:

$$d(k) \le \left\lfloor \log\left(\frac{1}{2}\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}\right)\right\rfloor + 1 < \sqrt{\log n}$$

and thus the number of nodes of G_1 is:

$$(2k)^{d(k)} \le \left(\left(\frac{n}{2}\right)^{1/\sqrt{\log n}}\right)^{\sqrt{\log n}} = \frac{n}{2}.$$

Let us show that $\mathrm{ps}(H) \le k$. We construct for H a path-decomposition composed of the bags $X_0', X_0, X_1, \ldots, X_{\ell-1}$, where X_0' is the set of nodes of G_1, X_0 is the edge between G_1 and G_2, and X_i, for $i \in \{1, \ldots, \ell - 1\}$, is the endpoint of the ith edge of G_2, starting from the node linked to G_1. This forms a path-decomposition of shape k, since $\mathrm{width}(X_i) = 1$ for $i \ge 0$, and $\mathrm{length}(X_0') = k$ the diameter of G_1. So, $\mathrm{ps}(H) \le k$.

It remains to prove that $D_{greedy}(H, \lambda, M) \ge \Omega(k + \log n)$ for any matrix M and labeling λ. We use Eq. (2) with one ball B_1 in G_1 at distance at least $k/3$ from the edge connecting G_1 and G_2, and another ball B_2 in G_2 at distance at least $2\ell/3$ from that edge.

Let (λ_1, M_1) be an augmentation of G_1 that minimizes the expected time of greedy routing in B_1. By Eq. (4), $D_{greedy}(B_1, \lambda_1, M_1) \ge \Omega(k)$.

Let (λ_2, M_2) be an augmentation of G_2 that minimizes the expected time of greedy routing in B_2. We have $D_{greedy}(B_2, \lambda_2, M_2) \ge \Omega(\log \ell)$ because nodes of B_2 have out-degree at most three, which implies that, for every realization of the augmenting distribution, there is a constant fraction of the pairs that are at distance $\Omega(\log \ell)$. Hence $D_{greedy}(B_2, \lambda_2, M_2) \ge \Omega(\log n)$ since $\ell \ge n/2$.

It follows from Eq. (2) that $D_{greedy}(H, \lambda, M) = \Omega(k + \log n)$, completing the proof. □

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 3, nodes may not be assigned different labels by the labeling λ. 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 a polylogarithmic greedy diameter for the classes of graphs mentioned in Corollary 3, or even just for paths.

Theorem 5. *Any matrix-based augmentation-labeling scheme using labels of size $\epsilon \log n$ for the n-node path, $0 \le \epsilon < 1$, yields a greedy diameter $\Omega(n^{\beta})$ for any $\beta < \frac{1}{3}(1 - \epsilon)$.*

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 $M = (M_{i,j})_{i,j}$ of size k. The probability that a fixed node u labeled i picks a fixed node v labeled j as its long range contact is $M_{i,j}/N_j$ 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 to be *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

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

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 $M_{\ell,i} \leq 1$, 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 there are 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 $\Omega(n^{\beta})$. □

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 terms of augmentation for arbitrary graphs. With this regard, 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 [16], 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 augmentation schemes where the $n \times n$ matrix is defined *a priori*, for all n-node graphs? We have proved that achieving this task requires an appropriate node labeling (cf. Theorem 2), 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 5).

Acknowledgments

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".

References

[1] I. Abraham, C. Gavoille, Object location using path separators, in: 25th ACM Symp. on Principles of Distributed Computing, PODC, 2006, pp. 188–197.
[2] J. Aspnes, Z. Diamadi, G. Shah, Fault-tolerant routing in peer-to-peer systems, in: 21st ACM Symp. on Principles of Distributed Computing, PODC, 2002, pp. 223–232.

[3] L. Barrière, P. Fraigniaud, E. Kranakis, D. Krizanc, Efficient routing in networks with long range contacts, in: 15th International Symposium on Distributed Computing, DISC, in: LNCS, vol. 2180, Springer, 2001, pp. 270–284.
[4] H. Bodlaender, A partial k-arboretum of graphs with bounded treewidth, Theoretical Computer Science 209 (1–2) (1998) 1–45.
[5] A. Brandstädt, V.B. Le, J.P. Spinrad, Graph classes — A survey, SIAM Monographs on Discrete Mathematicsand Applications, Philadelphia, 1999.
[6] A. Dawar, M. Grohe, S. Kreutzer, N. Schweikardt, Approximation schemes for first-order definable optimisation problems, in: 21st Annual IEEE Symposium on Logic in Computer Science, LICS, IEEE Computer Society Press, 2006, pp. 411–420.
[7] P. Dodds, R. Muhamad, D. Watts, An experimental study of search in global social networks, Science 301 (5634) (2003) 827–829.
[8] P. Duchon, N. Hanusse, E. Lebhar, N. Schabanel, Could any graph be turned into a small-world? Theoretical Computer Science 355 (1) (2006) 96–103.
[9] P. Duchon, N. Hanusse, E. Lebhar, N. Schabanel, Towards small world emergence, in: 18th Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA, 2006, pp. 225–232.
[10] Y. Dourisboure, Compact routing schemes for bounded tree-length graphs and for k-chordal graphs, Journal of Graph Algorithms and Applications 9 (2) (2005) 277–297.
[11] Y. Dourisboure, C. Gavoille, Tree-decompositions with bags of small diameter, Discrete Mathematics 307 (16) (2007) 2008–2029.
[12] M. Flammini, L. Moscardelli, A. Navarra, S. Perennes, Asymptotically optimal solutions for small world graphs, in: 19th International Symposium on Distributed Computing, DISC, in: LNCS, vol. 3724, 2005, pp. 414–428.
[13] P. Fraigniaud, Greedy routing in tree-decomposed graphs: A new perspective on the small-world phenomenon, in: 13th Annual European Symposium on Algorithms, ESA, 2005, pp. 791–802.
[14] P. Fraigniaud, C. Gavoille, Polylogarithmic network navigability using compact metrics with small stretch, in: 20th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA, 2008, pp. 62–69.
[15] P. Fraigniaud, C. Gavoille, C. Paul, Eclecticism shrinks even small worlds, in: Proceedings of the 23rd ACM Symposium on Principles of Distributed Computing, PODC, 2004, pp. 169–178.
[16] P. Fraigniaud, E. Lebhar, Z. Lotker, A doubling dimension threshold $\Theta(\log\log n)$ for augmented graph navigability, in: 14th European Symposium on Algorithm, ESA, in: LNCS, vol. 4168, Springer, 2006, pp. 376–386.
[17] C. Gavoille, M. Katz, N.A. Katz, C. Paul, D. Peleg, Approximate distance labeling schemes, in: 9th Annual European Symposium on Algorithms, ESA, in: LNCS, vol. 2161, Springer, 2001, pp. 476–488.
[18] J. Kleinberg, The Small-world phenomenon: An algorithmic perspective, in: 32nd ACM Symp. on Theory of Computing, STOC, 2000, pp. 163–170.
[19] J. Kleinberg, Navigation in a small-world, Nature 406 (2000) 845.
[20] J. Kleinberg, Small-world phenomena and the dynamics of information, in: 15th Neural Information Processing Systems, NIPS, 2001.
[21] J. Kleinberg, Complex networks and decentralized search algorithm, in: International Congress of Mathematicians, ICM, 2006.
[22] J. Kleinberg, R. Rubinfeld, Short paths in expander graphs, in: 37th IEEE Symp. on Foundations of Computer Science, FOCS, 2007, pp. 86–95.
[23] E. Lebhar, N. Schabanel, Searching for optimal paths in long-range contact networks, in: 31st International Colloquium on Automata, Languages and Programming, ICALP, 2004, pp. 894–905.
[24] G. Manku, M. Naor, 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, pp. 54–63.
[25] C. Martel, V. Nguyen, Analyzing Kleinberg's (and other) small-world models, in: 23rd ACM Symp. on Principles of Distributed Computing, PODC, 2004, 179–188.
[26] C. Martel, V. Nguyen, Analyzing and characterizing small-world graphs, in: 16th ACM-SIAM Symp. on Discrete Algorithms, SODA, 2005, pp. 311–320.
[27] C. Martel, V. Nguyen, Designing networks for low weight, small routing diameter and low congestion, in: 25th Conference of the IEEE Communications Society, INFOCOM, 2006.
[28] D. Marx, Can you beat treewidth, in: 48th IEEE Symp. on Foundations of Computer Science, FOCS, 2007, pp. 169–179.
[29] S. Milgram, The small-world problem, Psychology Today (1967) 60–67.
[30] D. Peleg, Private communication, in: Workshop of COST Action 295 ''DYNAMO'', Les Ménuires, Jan. 2006.
[31] N. Robertson, P.D. Seymour, Graph minors. I. Excluding a forest, Journal of Combinatorial Theory Series B 35 (1) (1983) 39–61.
[32] N. Robertson, P.D. Seymour, Graph minors. III. Planar tree-width, Journal of Combinatorial Theory Series B 36 (1) (1984) 49–64.
[33] A. Slivkins, Distance estimation and object location via rings of neighbors, in: 24th Annual ACM Symposium on Principles of Distributed Computing, PODC, 2005, pp. 41–50.
[34] The CAIDA Web Site. http://www.caida.org/.