Recovering the Long Range Links in Augmented Graphs

Pierre Fraigniaud, Emmanuelle Lebhar, Zvi Lotker

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Détecter les longs liens dans les graphes augmentés

Pierre Fraigniaud — Emmanuelle Lebhar — Zvi Lotker

N° 6197

Mai 2007

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6197--FR+ENG

Détecter les longs liens dans les graphes augmentés

Pierre Fraigniaud* , Emmanuelle Lebhar* , Zvi Lotker† ‡

Thème COM — Systèmes communicants
Projet Gang

Rapport de recherche n° 6197 — Mai 2007 — 27 pages

Résumé : Le modèle des *graphes augmentés*, introduit par Kleinberg (STOC 2000), est un modèle attrayant pour analyser la navigabilité dans les réseaux sociaux. De façon informelle, ce modèle est défini par une paire (H, φ), où H est un graphe dont les distances entre les noeuds sont supposées simples à calculer, ou au moins à estimer. Ce graphe est "augmenté" par des liens, appelés *longs liens*, qui sont sélectionnés selon la distribution de probabilité φ. Le modèle des graphes augmentés permet d'analyser le *routage glouton* dans les graphes augmentés $G \in (H, \varphi)$. Dans le routage glouton, chaque noeud intermédiaire détenant un message pour une cible t sélectionne, parmi ses voisins dans G, celui qui est le plus proche de t dans H, et lui transmet le message.

Cet article s'intéresse au problème de vérifier si un graphe donné G est un graphe augmenté. Il répond partiellement à la question posée par Kleinberg dans son Problème 9 (Int. Congress of Math. 2006). Plus précisément, étant donné $G \in (H, \varphi)$, nous cherchons à extraire le graphe de base H et les longs liens R à partir de G. Nous montrons que si H a un coefficient de *clustering* élevé et une dimension doublante bornée, alors un algorithme simple permet de partitionner les arêtes de G en deux ensembles H' et R' tels que $E(H) \subseteq H'$ et que les arêtes dans $H' \setminus E(H)$ soient peu distordues, c.-à-d. que la carte de H n'est pas trop perturbée par les longs liens non détectés qui sont restés dans H'. La perturbation est en fait si faible que l'on peut démontrer que les performances du routage glouton dans G en utilisant les distances dans H' sont proches de celles du routage glouton dans (H, φ) (en espérance). Bien que ce dernier résultat puisse sembler intuitif au premier abord puisque $H' \supseteq E(H)$, il ne l'est pas, et nous montrons que le routage glouton avec une carte plus précise que H peut voir ses performances sensiblement endommagées. Enfin, nous montrons qu'en l'absence d'hypothèse sur le coefficient de clustering, toute tentative d'extraction des longs liens de façon structurelle ratera la détection d'au moins $\Omega(n^{5\varepsilon}/\log n)$ longs liens de

* CNRS, INRIA and University Paris 7, France

† Ben Gurion University, Israel

‡ This work was partially done while the third author was visiting University Paris 7 at LIAFA. Additional support by University Paris 7.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

distorsion au moins $\Omega(n^{1/5-\varepsilon})$ pour tout $0 < \varepsilon < 1/5$, et donc la carte de H ne pourra pas être reconstruite avec une bonne précision.

En résumé, nous résolvons le Problème 9 de Kleinberg dans le sens que nous montrons que la reconstruction des graphes augmentés est possible si et seulement si le graphe de base présente une fort coefficient de clustering.

Mots-clés : petit monde, dimension doublante, croissance bornée

Recovering the Long Range Links in Augmented Graphs

Abstract: The *augmented graph* model, as introduced by Kleinberg (STOC 2000), is an appealing model for analyzing navigability in social networks. Informally, this model is defined by a pair (H, φ), where H is a graph in which inter-node distances are supposed to be easy to compute or at least easy to estimate. This graph is "augmented" by links, called *long range* links, which are selected according to the probability distribution φ. The augmented graph model enables the analysis of *greedy routing* in augmented graphs $G \in (H, \varphi)$. In greedy routing, each intermediate node handling a message for a target t selects among all its neighbors in G the one that is the closest to t in H and forwards the message to it.

This paper addresses the problem of checking whether a given graph G is an augmented graph. It answers part of the questions raised by Kleinberg in his Problem 9 (Int. Congress of Math. 2006). More precisely, given $G \in (H, \varphi)$, we aim at extracting the base graph H and the long range links R out of G. We prove that if H has a high clustering coefficient and bounded doubling dimension, then a simple algorithm enables to partition the edges of G into two sets H' and R' such that $E(H) \subseteq H'$ and the edges in $H' \setminus E(H)$ are of small stretch, i.e., the map H is not perturbed too greatly by undetected long range links remaining in H'. The perturbation is actually so small that we can prove that the expected performances of greedy routing in G using the distances in H' are close to the expected performances of greedy routing in (H, φ). Although this latter result may appear intuitively straightforward, since $H' \supseteq E(H)$, it is not, as we also show that routing with a map more precise than H may actually damage greedy routing significantly. Finally, we show that in absence of a hypothesis regarding the high clustering coefficient, any structural attempt to extract the long range links will miss the detection of at least $\Omega(n^{5\varepsilon}/\log n)$ long range links of stretch at least $\Omega(n^{1/5-\varepsilon})$ for any $0 < \varepsilon < 1/5$, and thus the map H cannot be recovered with good accuracy.

To sum up, we solve Kleinberg's Problem 9 in the sense that we show that reconstructing augmented graphs is achievable if and only if the base graph has a high clustering coefficient.

Key-words: small world, doubling dimension, bounded growth

1 Introduction

Numerous papers that appeared during the last decade tend to demonstrate that several types of interaction networks share common statistical properties, encompassed under the broad terminology of *small worlds* [36, 37, 38]. These networks include the Internet (at the router level as well as at the AS level) and the World Wide Web. Actually, networks defined in frameworks as various as biology (metabolic and protein networks), sociology (movies actors collaboration network), and linguistic (pairs of words in english texts that appear at most one word apart) also share these statistical properties [22]. Specifically, a network is said small world [41] if it has low density (i.e., the total number of edges is small, typically linear in the number of nodes), the average distance between nodes is small (typically polylogarithmic as a function of the number of nodes), and the so-called *clustering* coefficient, measuring the local edge density, is high (i.e., significantly higher than the clustering coefficient of Erdös-Rényi random graphs $\mathcal{G}_{n,p}$). Other properties often shared by the aforementioned networks include *scale free* properties [7] (i.e., fat tailed shapes in the distributions of parameters such as node degree), limited growth of the ball sizes [2, 23], or low doubling dimension [40].

Lots remain to be done to understand why the properties listed above appear so frequently, and to design and analyze models capturing these properties. Nevertheless, there is now a common agreement on their presence in interaction networks. The reason for this agreement is that, although the statistical validity of some measurements is still under discussion [5], many tools (including the controversial Internet Traceroute) have been designed to check whether a network satisfies the aforementioned properties.

This paper addresses the problem of checking an other important property shared by social networks : the *navigability* property.

It was indeed empirically observed that social networks not only possess small average inter-node distance, but also that short routes between any pair of nodes can be found by simple decentralized processes [10, 35]. One of the first papers aiming at designing a model capturing this property is due to Kleinberg [25], where the notion of *augmented graphs* is introduced. Informally, an augmented graph is aiming at modeling two kinds of knowledge of distances available to the nodes : a global knowledge given by a base graph, and a local knowledge given by one extra random link added to each node. The idea is to mimic the available knowledge in social networks, where individuals share some global distance comparison tool (e.g. geographical of professional), but also have private connections (e.g. friendship) unknown from the other individual. We define an *augmented graph model* as a pair (H,φ) where H is a graph, called *base graph*, and φ is a probability distribution, referred to as an *augmenting distribution* for H. This augmenting distribution is defined as a collection of probability distributions $\{\varphi_u, u \in V(H)\}$. Every node $u \in V(H)$ is given one extra link[1], called *long range link*, pointing to some node, called the *long range contact* of u. The destination v of such a link is chosen at random according to $\Pr\{u \to v\} = \varphi_u(v)$. (If $v = u$ or v is a neighbor of u, then no link is added). In this paper, a graph $G \in (H,\varphi)$

[1] By adding $k_u \geq 1$ long range links to node u, for every $u \in V(H)$, instead of just one, with $\Pr(k_u = k) \sim 1/k^\alpha$ for some $\alpha > 1$, the model can also capture the scale-free property. For the sake of simplicity however, we will just assume $k_u = 1$ for every $u \in V(H)$.

will often be denoted by $H + R$ where H is the base graph and R is the set of long rang links resulting from the trial of φ yielding G.

A important feature of this model is that it enables to define simple but efficient decentralized routing protocols modeling the search procedure applied by social entities in Milgram's [35] and Dodd's et al [10] experiments. In particular, *greedy routing* in (H, φ) is the oblivious routing process in which every intermediate node along a route from a source $s \in V(H)$ to a target $t \in V(H)$ chooses among all its neighbors (including its long range contact) the one that is the closest to t *according to the distance measured in* H, and forwards to it. For this process to apply, the only "knowledge" that is supposed to be available at every node is its distances to the other nodes in the base graph H. This assumption is motivated by the fact that, if the base graph offers some nice properties (e.g., embeddable in a low dimensional metric with small distorsion) then the distance function dist_H is expected to be easy to compute, or at least to approximate, locally.

Lots of efforts have been done to better understand the augmented graph model (see, e.g., [1, 6, 8, 12, 16, 13, 14, 17, 30, 31, 32, 33], and the survey [27]). Most of these works tackle the following problem : given a family of graph $\mathcal{H}$, find a family of augmenting distributions $\{\varphi_H, H \in \mathcal{H}\}$ such that, for any $H \in \mathcal{H}$, greedy routing in (H, φ_H) performs efficiently, typically in $O(\text{polylog}(n))$ expected number of steps, where $n = |V(H)|$. Kleinberg first showed that greedy routing performs in $O(\log^2 n)$ expected number of steps on any regular mesh augmented with an appropriate harmonic distribution [25]. Among the works that followed Kleinberg's seminal results, an informative result due to Duchon et al [11] states that any graph of bounded growth can be augmented so that greedy routing performs in $O(\text{polylog}(n))$ expected number of steps. Slivkins [40] extended this result to graphs of bounded doubling dimension, and even doubling dimension at most $O(\log \log n)$. This bound on the doubling dimension is tight since [18] proved that, for any function $d(n) \in \omega(\text{polylog}(n))$, there is a family of graphs of doubling dimension $d(n)$ for which any augmentation yields greedy routing performing in $\omega(\text{polylog}(n))$ expected number of steps[2].

Despite these progresses in analyzing the augmented graph model for small worlds, the key question of its validity is still under discussion. In [27], Kleinberg raised the question of how to check that a given network is an augmented graph (Problem 9). This is a critical issue since, if long range links are the keystone of the small world phenomenon, they should be present in social networks, and their detection should be greatly informative. This paper answers part of this detection problem.

This paper addresses the following reconstruction problem : given an n-node graph $G = H + R \in (H, \varphi)$, for some unknown graph H and unknown distribution φ, extract a good approximation H' of H such that greedy routing in G using distances in H' performs approximately as well as when using the distances in the "true" base graph H. More precisely, the expected number of steps of greedy routing in H' has to be the one in H up to a polylogarithmic factor. To measure the quality of the approximation H' of H, we define the *stretch* of a long range link between u and v as $\text{dist}_H(u, v)$. Then, the extracted base graph

[2]The notation $d(n) \in \omega(f(n))$ for some functions f and d means that $d(n)/f(n)$ tends to infinity when n goes to the infinity.

H' is considered to be of good quality if it contains H and does not contain too many long range links of large stretch. Indeed, we want to approximate H by H' as close as possible not only for the purpose of efficient routing using the metric of H', but also because the augmented graph model assumes that distances in H are easy to compute or approximate. Therefore, the map of distances of H' is wished to be close to the one of H.

In addition to its fundamental interest, the reconstruction problem may find important applications in network routing. In particular, if the base graph H offers enough regularity enabling navigability using routing tables of small size, then critical issues of storage and quick access to routing information (such as the ones currently faced for Internet) can be addressed, see [28, 34].

Our results

Motivated by the aforementioned properties of interaction networks, we focus on graphs with bounded doubling dimension, hence including the case of graphs with bounded growth. The results in [11, 25, 40] use very close augmenting distributions where $\varphi_u(v)$ is inversely proportional to the size of the ball of radius $\text{dist}(u, v)$ centered at u. We call such kind of augmenting distributions *density based* distributions. They are the ones enabling an efficient augmentation of graphs with bounded growth, and, up to modifying the underlying metric by weighting nodes, of graphs with bounded doubling dimension.

First, we present a simple algorithm called EXTRACT that, given an n-node graph $G = H+R \in (H, \varphi)$, where H has high clustering coefficient, and φ is a density-based augmenting distribution, computes a partition (H', R') of $E(G)$. This partition satisfies $E(H) \subseteq H'$ and, if X is the random variable counting the number of links in $R' \setminus R$ of stretch at least $\log^{\beta+1} n$, then $\Pr\{X > \log^{2\beta+1}(n)\} \leq 1/n$, for any $\beta \geq \Omega(\log \Delta / \log \log n)$, where Δ is the maximum degree of H. That is, Algorithm EXTRACT is able to almost perfectly reconstruct the map H of G, up to long range links of polylogarithmic stretch. It is worth mentioning that Algorithm EXTRACT runs in time close to linear in $|E(G)|$, and thus is applicable to large graphs with few edges, which is typically the case of small world networks.

Our main positive result (Theorem 1) is that if in addition H has bounded growth, then greedy routing in G using the distances in H' performs in $O(\text{polylog}(n))$ expected number of steps between any pair. This result is crucial in the sense that Algorithm EXTRACT is able to approximate the base graph H and the set R of long range links accurately enough so that greedy routing performs efficiently. In fact, we prove that the expected slow down of greedy routing in G using the distances in H' compared to greedy routing in (H, φ) is only $O(\log^{2+\beta(4+\alpha)} n)$. Although this latter result may appear intuitively straightforward since $H' \supseteq E(H)$, we prove that routing with a map more precise than H may actually damage greedy routing performances significantly.

We also show how these results can be generalized to the case of graphs with bounded doubling dimension.

Our assumption regarding the clustering coefficient may seem artificial since this parameter does not show up in the analysis of greedy routing. Nevertheless, and surprisingly enough, our main negative result (Theorem 2) proves that the clustering coefficient plays a crucial role for extracting the long range links of an augmented graph. Indeed, we prove that any structural attempt to extract the long range links in some augmented graph with low clustering coefficient would fail. In fact, this is true even in the case of cycles augmented using the harmonic distribution, that is even in the case of basic graphs at the kernel of the theory of augmented graphs [25]. We prove that any (deterministic or randomized) algorithm extracting the long range links based on the structure of the harmonically augmented cycle fails to detect at least $\Omega(n^{5\varepsilon}/\log n)$, of the long range links of length $\Omega(n^{1/5-\varepsilon})$ for any $0 < \varepsilon < 1/5$. As a consequence, the stretch of the undetected long range links cannot be polylogarithmically bounded.

To sum up, we solve Kleinberg's Problem 9 in the sense that we show that reconstructing augmented graphs is achievable if and only if the base graph has a high clustering coefficient.

Related works

Our approach is very much related to the ones in [4, 9]. In these two papers the authors introduce an hybrid model that resembles the augmented graph model, defined by a *local* graph H and a *global* graph D, over the same set of vertices. In both papers, D is a random power law graph. In [9], the local graph H has a local connectivity characterized by a certain number of edge-disjoint paths of bounded length connecting the two extremities of any edge. In [4], the local connectivity is characterized by an amount of flow that can be pushed from one extremity of an edge to the other extremity, along routes of bounded length. In both cases, in addition to presenting a set of informative results about their models, the authors give an algorithm that can recover the local graph almost perfectly. As far as the recovering of the local graph is concerned, our approach goes a bit further than [4, 9] by proving that one can recover the local graph accurately enough so that the performances of greedy routing remain unchanged (up to polylogarithmic factors).

Greedy routing in scale-free graphs has been considered in the literature. In [24], the authors perform simulations on a model of power law networks to compare a random walk search strategy with a search strategy guided by high degree nodes. They observe that the latter search strategy performs better than the former. Nevertheless, the search strategy performs in a polynomial number of steps. In [3], a mere mean-field analysis, somewhat confirmed by simulations, tends to show that the expected number of steps required to find an object in a random power law network with n nodes and power law exponent α, scales sub-linearly as $n^{3(1-2/\alpha)}$ for $2 < \alpha < 3$. The power of search can however be improved by publishing over many nodes. For instance, [39] proposes that every node publishes its data at every node along a random walk of length L. The search strategy then proceeds along a random walk of same length, and every node traversed by the walk starts partially flooding the network (the search is sent through every edge with probability less than the percolation threshold of the network). It is then shown that this search efficiently locates the data by setting $L \sim n^{1-2/\alpha}$ for $2 < \alpha < 3$. More recently, the simulations performed in [15] on real traces extracted from a P2P file-sharing system tend to show that the search strategy guided by high degree nodes would probably perform well in practice, i.e., in a logarithmic number of steps in average.

2 Extracting the long range links

In this section, we first focus on the task of extracting the long range links from an augmented graph $G = H + R \in (H, \varphi)$ without knowing H. The efficiency of our extraction algorithm in terms of greedy routing performances will be analyzed in the next section. As will be shown in Section 4, extracting the long range links from an augmented graph is difficult to achieve in absence of a priori assumptions on the base graph H and on the augmenting distribution φ. Before presenting the main result of the section we thus present the assumptions made on H and φ.

The clustering coefficient of a graph H is aiming at measuring the probability that two distinct neighboring nodes u, v of a node w are neighbors. Several similar formal definitions of the clustering coefficient appear in the literature. In this paper, we use the following definition. For any node u of a graph H, let $N_H(u)$ denotes the open neighborhood of u, i.e., the set of all neighbors of u in G. Let $N_H[u]$ denote the closed neighborhood of node u, i.e., $N_H[u] = N_H(u) \cup \{u\}$.

Definition 1 *An n-node graph H has clustering $c \in [0, 1]$ if and only if, for any edge $\{u, v\} \in E(H)$,*

$$\frac{|N_H[u] \cap N_H[v]|}{n} \geq c.$$

For instance, according to Definition 1, a random graph $G \in \mathcal{G}_{n,p}$ with $p \simeq \frac{\log n}{n}$ has expected clustering $1/n^3$ up to polylogarithmic factors. In our results, motivated by the fact that interaction networks have a clustering coefficient much larger than uniform random graphs, we consider graphs in (H, φ) for which the clustering coefficient of H is slightly more that $1/n$.

We also focus on a specific kind of augmenting distributions : those that are known to be efficient ways to augment graphs of bounded growth (or bounded doubling dimension) [11, 25, 40]. For any node u of a graph H, and any $r > 0$, let $B_H(u, r)$ denote the ball centered at u of radius r in H, i.e., $B_H(u, r) = \{v \in V(G) \mid \mathrm{dist}_H(u, v) \leq r\}$.

Definition 2 *An augmenting distribution φ of a graph H is* density-based *if and only if $\varphi_u(u) = 0$, and for every two distinct nodes u and v of H,*

$$\varphi_u(v) = \frac{1}{Z_u} \frac{1}{|B_H(u, \mathrm{dist}_H(u, v))|}$$

where $Z_u = \sum_{w \neq u} 1/|B_H(u, \mathrm{dist}_H(u, w))|$ is the normalizing coefficient.

Density-based distributions are motivated by their kernel place in the theory of augmented graphs, as well as by experimental studies in social networks. Indeed, density-based distributions applied to graphs of bounded growth roughly give a probability $1/k$ for a node u to have its long range contact at distance k, which distributes the long range links equivalently over all scales of distances, and thus yields efficient greedy routing. In addition, Liben-Nowell et al. [29] showed that in some social networks, two-third of the friendships are actually geographically distributed this way : the probability of befriending a particular person is inversely proportional to the number of closer people.

According to the previous discussion, the following definition specifies the family of augmented graph models that will be considered in this paper.

Definition 3 *For any $n, \beta \geq 1$, $\mathcal{M}(n,\beta)$ is the family of n-node augmented graph models (H,φ) such that (1) H has clustering $c \geq \Omega(\frac{\log n}{n\ \log\log n})$, (2) H has maximum degree $\Delta \leq O(\log^{\beta} n)$, and (3) φ is density-based.*

We describe below a simple algorithm, called EXTRACT, that, given an n-node graph G and a real $c \in [0,1]$, computes a partition (H', R') of the edges of G. This simple algorithm will be proved quite efficient for reconstructing a good approximation of the base graph H and a good approximation the long range links of a graph $G \in (H,\varphi)$ when H has high clustering and φ is density-based.

Algorithm EXTRACT **:**
Input : a graph G, $c \in [0,1]$;
  $R' \leftarrow \emptyset$;
  For every $\{u,v\} \in E(G)$ **do**
    If $|N_G[u] \cap N_G[v]| < c\,n$ **then** $R' \leftarrow R' \cup \{u,v\}$;
  $H' \leftarrow E(G) \setminus R'$;
Output : (H', R').

Note that the time complexity of Algorithm EXTRACT is $O(\sum_{u \in V(G)} (\deg_G(u))^2)$, i.e., close to $|E(G)|$ for graphs of constant average degree. More accurate outputs could be obtained by iterating the algorithm using the test $|N_{H'}[u] \cap N_{H'}[v]| < c\,n$ until H' stabilizes. However, this would significantly increase the time complexity of the algorithm without significantly improving the quality of the computed decomposition (H', R'). The main quantifiable gain of iterating Algorithm EXTRACT would only be that H' would be of clustering c, and would be maximal for this property. The result hereafter summarizes the main features of Algorithm EXTRACT.

Lemma 1 *Let $(H,\varphi) \in \mathcal{M}(n,\beta)$ for $n, \beta \geq 1$ and $G \in (H,\varphi)$. Let c be the clustering coefficient of H. Assume $G = H + R$. Then Algorithm* EXTRACT *with input (G,c) returns a partition (H', R') of $E(G)$ such that $E(H) \subseteq H'$, and :*

$$\Pr(X > \log^{2\beta+1} n) \leq \mathcal{O}\left(\frac{1}{n}\right),$$

where X is the random variable counting the number of links in $R \setminus R'$ of stretch at least $\log^{\beta+1} n$.

Proof. Since H has clustering c, for any edge $\{u,v\}$ in $E(H)$, $|N_H[u] \cup N_H[v]| \geq c\,n$, and therefore $\{u,v\}$ is not included in R' in Algorithm EXTRACT. Hence, $E(H) \subseteq H'$. For the purpose of upper bounding $\Pr(X > \log^{2\beta+1} n)$, we first lower bound Z_u, for any $u \in G$. We have the following claim (the proofs of all the claims can be found in appendix).

Claim 1 *For any $u \in G$, $Z_u \geq Z_{\min} =_{\text{def}} \frac{1}{4}$, for n large enough.*

Let $\mathcal{S} \subseteq R$ be the set of long range links that are of stretch at least $\log^{\beta+1} n$. We say that an edge $\{u,v\} \in R$ *survives* if and only if it belongs to H'. For each edge $e \in \mathcal{S}$, let X_e be the random variable equal to one if e survives and 0 otherwise.

Let $e = \{u,v\} \in \mathcal{S}$. For e to be surviving in H', it requires that u and v have at least $c \cdot n$ neighbors in common in G. If w is a common neighbor of u and v in G, then, since $\text{dist}_H(u,v) \geq \log n > 2$, at least one of the two edges $\{w,u\}$ or $\{w,v\}$ has to belong to R. Note that u and w can only have one common neighbor w such that both of these edges are in R because we add exactly one long range link to every node. Thus, there must be at least $c \cdot n - 1$ common neighbors w for which only one of the edges $\{w,u\}$ or $\{w,v\}$ is in R. Again, since there is only one long range link per node, there can be at most two common neighbors w for which the edge $\{w,u\}$ or $\{w,v\}$ is the long range link of u or v. Therefore, there must be at most $c \cdot n - 2$ common neighbors having a long range link to u or v. The following claim upper bound the probability of this event.

Claim 2 $\Pr\{X_e = 1\} \leq 1/n$

To compute the probability that at most $\log^{2\beta+1} n$ edges of $\mathcal{S}$ survive in total, we use virtual random variables that dominate the variables X_e, $e \in R$, in order to bypass the dependencies between the X_e. Let us associate to each $e \in \mathcal{S}$ a random variable Y_e equals to 1 with probability $1/n$ and 0 otherwise. By definition, Y_e dominates X_e for each $e \in \mathcal{S}$ and the Y_e are independently and identically distributed. Note that, the fact that some long range link e survives affects the survival at most Δ^2 other long range links of R, namely, all the potential long range links between $N_H(u)$ and $N_H(v)$. Therefore the probability that k links of $\mathcal{S}$ survive is at most the probability that k/Δ^2 of the variables Y_e are equal to one. In particular we have : $\Pr\{\sum_{e\in\mathcal{S}} X_e > \log^{2\beta+1} n\} \leq \Pr\{\sum_{e\in\mathcal{S}} Y_e > \log^{2\beta+1} n/\Delta^2\}$. Using Chernoff' inequality, we have the following claim.

Claim 3 $\Pr\{\sum_{e\in\mathcal{S}} Y_e > \log^{2\beta+1} n/\Delta^2\} \leq 1/n$.

We conclude that $\Pr\{\sum_{e\in\mathcal{S}} X_e > \log^{2\beta+1} n\} \leq \frac{1}{n}$. □

3 Navigability

In the previous section, we have shown that we can almost recover the base graph H of an augmented graph $G \in (H,\varphi)$: very few long range links of large stretch remain undetected with high probability. In this section, we prove that our approximation H' of H

is good enough to preserve the efficiency of greedy routing. Indeed, although it may appear counterintuitive, being aware of more links does not necessarily speed up greedy routing. In other words, using a map $H' \supseteq H$ may not yield better performances than using the map H, and actually it may even significantly damage the performances. This phenomenon occurs because the augmenting distribution φ is generally chosen to fit well with H, and this fit can be destroyed by the presence of a few more links in the map. This is illustrated by the following property (its proof can be found in the appendix) .

Property 1 *There exists an n-node augmented graph model (H, φ) and a long range link e such that, for $\Omega(n)$ source-destination pairs, the expected number of steps of greedy routing in (H, φ) is $O(\log^2 n)$, while greedy routing using distances in $H \cup \{e\}$ takes $\omega(\text{polylog}(n))$ expected number of steps.*

Property 1 illustrates that being aware of some of the long range links may slow down greedy routing dramatically, at least for some source-destination pairs. Nevertheless, we show that algorithm EXTRACT is accurate enough for the undetected long range links not to cause too much damage. Precisely, we show that for bounded growth graphs as well as for graphs of bounded doubling dimension, greedy routing using distances in H' can slow down greedy routing in (H, φ) only by a polylogarithmic factor.

Definition 4 *A graph G has (q_0, α)-*expansion *if and only if, for any node $u \in V(G)$, and for any $r > 0$, we have : $|B_G(u, r)| \geq q_0 \Rightarrow |B_G(u, 2r)| \leq 2^\alpha\, |B_G(u, r)|$. In the bulk of this paper, we will set $q_0 = O(1)$, and refer to α as the* expanding dimension *of G, and to 2^α as the* growth rate *of G.*

Definition 4 is inspired from Karger and Ruhl [23]. The only difference with Definition 1 in [23] is that we exponentiate the growth rate. Note that, according to Definition 4, a graph has bounded growth if and only if its expanding dimension is $O(1)$.

Theorem 1 *Let $(H, \varphi) \in \mathcal{M}(n, \beta)$ for $n, \beta \geq 1$ be such that H has (q_0, α)-expansion, with $2 < q_0 = O(1)$ and $\alpha = O(1)$. Let $G \in (H, \varphi)$. Algorithm* EXTRACT *outputs (H', R') such that (a) $E(H) \subseteq H'$, (b) with high probability H' contains at most $\log^{2\beta+1} n$ links of stretch more than $\log^{\beta+1} n$, and (c) for any source s and target t, the expected number of steps of greedy routing in G using the metric of H' is at most $O(\log^{4+4\beta+(\beta+1)\alpha} n)$.*

Proof. The fact that $E(H) \subseteq H'$ and that with high probability H' contains at most $\log^{2\beta+1} n$ links of stretch more than $\log^{\beta+1} n$ is a direct consequence of Lemma 1. Let $H'' = H' \setminus \mathcal{S}$. Note that the maximum stretch in H'' is $\log^{\beta+1} n$. For any $x \in V(H)$, let $L(x)$ denote the long range contact of x. Let Z_u be the normalizing constant of the augmenting distribution at node u. We have the following claim (the proofs of all claims can be found in appendix).

Claim 4 *For any $u \in V(G)$, $Z_u \leq 2^\alpha \log n =_{\text{def}} Z_{\max}$.*

Let us analyze greedy routing in G from $s \in V(G)$ to $t \in V(G)$ using the distances in H'. Let $\mathcal{S} = \{\{u_1, v_1\}, \ldots, \{u_k, v_k\}\}$ be the set of the surviving long range links (i.e. in $R \cap H'$) that have stretch more than $\log^{\beta+1} n$, v_i being the long range contact of u_i for all $1 \le i \le k$. For the homogeneity of the notations, let $u_0 = v_0 = t$.

Let τ be the current step of greedy routing from s to t, and x the current node. We define the *concerned index* at step τ as the unique index j defined by :

$$j = \min_{i \in \{1,\ldots,k\}} \{i \,|\mathrm{dist}_{H'}(x,t) = \mathrm{dist}_{H''}(x,u_i) + 1 + \mathrm{dist}_{H'}(v_i,t)\}.$$

In other words, $\{u_j, v_j\}$ is the first surviving long range link encountered along the shortest path from x to t in H'. If there is no such index, set $j = 0$.

Claim 5 *Let x be the current node of greedy routing, and j be the concerned index at the current step. If $x \in B_{H''}(u_j, r)$, but $\mathrm{dist}_{H''}(x, u_j) > r/2$, and for some $r > 0$, then :*

$$\Pr\{L(x) \in B_{H''}(u_j, r/2)\} \ge \frac{1}{2^{4\alpha} \log^{1+\alpha(\beta+1)} n},$$

and if $L(x) \in B_{H''}(u_j, r/2)$ then greedy routing routes inside $B_{H''}(u_j, r/2)$ at the next step.

Claim 6 *Let x and x' be two nodes on the greedy route reached at respective steps τ and τ', $\tau < \tau'$. Assume that the concerned index at steps τ and τ' is the same, denoted by j, $j \le k = |\mathcal{S}|$. If $x \in B_{H''}(u_j, r)$ for some $r > 0$, then $x' \in B_{H''}(u_j, r)$.*

For any $0 \le i \le \log n$, $0 \le j \le k$, and $\tau > 0$, let $\mathcal{E}_j^i(\tau)$ be the event : "greedy routing from s to t already entered $B_{H''}(u_j, 2^i)$ during the first τ steps". Note that, for any $0 \le j \le k$ and any $\tau > 0$, $\mathcal{E}_j^0(\tau) \subseteq \ldots \subseteq \mathcal{E}_j^{\log n}(\tau)$. We describe the current state of greedy routing at step τ by the event $\mathcal{E}_0^{i_0}(\tau) \cap \ldots \cap \mathcal{E}_k^{i_k}(\tau)$ where for every $0 \le j \le k$, $i_j = \min\{i \,|\, \mathcal{E}_j^i(\tau) \text{ occurs}\}$.

Note that greedy routing has reached t at step τ if and only if $\mathcal{E}_0^0(\tau)$ has occured. Clearly, at step 0 (in s), the event $\mathcal{E}_0^{\log n}(0) \cap \mathcal{E}_2^{\log n}(0) \ldots \cap \mathcal{E}_k^{\log n}(0)$ occurs.

Claim 7 *Assume that the state of greedy routing at step τ is $\mathcal{E}_0^{i_0}(\tau) \cap \ldots \cap \mathcal{E}_k^{i_k}(\tau)$, for some $i_0, \ldots i_k \in \{0, \ldots, \log n\}$. Then, after at most $(k+1) \cdot 2^{4\alpha} \log^{1+\alpha(\beta+1)} n$ steps on expectation, there exists an index $0 \le \ell \le k$ such that the state of greedy routing is $\mathcal{E}_0^{j_0}(\tau') \cap \ldots \cap \mathcal{E}_\ell^{j_\ell}(\tau') \ldots \cap \mathcal{E}_k^{j_k}(\tau')$, with $j_\ell < i_\ell$, $\tau' > \tau$.*

Let X be the random variable counting the number of steps of greedy routing from s to t. As noticed before, $\mathbb{E}(X)$ is at most the expected number of steps τ to go from state $\mathcal{E}_0^{\log n}(0) \cap \mathcal{E}_1^{\log n}(0) \ldots \cap \mathcal{E}_k^{\log n}(0)$ to state $\mathcal{E}_0^0(\tau) \cap \mathcal{E}_1^{i_1}(\tau) \ldots \cap \mathcal{E}_k^{i_k}(\tau)$, for some $i_1, \ldots i_k \in \{0, \ldots, \log n\}$. From Claim 7, we get : $\mathbb{E}(X) \le (k+1) \log n \times ((k+1) \cdot 2^{4\alpha} \log^{1+\alpha(\beta+1)} n)$. And, from Lemma 1, $\Pr\{k > \log^{2\beta+1} n\} \le 1/n$. Therefore, we have :

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X \,|\, k \le \log^{2\beta+1} n) \cdot \Pr\{k \le \log^{2\beta+1} n\} + \mathbb{E}(X \,|\, k > \log^{2\beta+1} n) \cdot \Pr\{k > \log^{2\beta+1} n\} \\ &\le 2^{4\alpha} \log^{2+\alpha(\beta+1)\alpha+2(2\beta+1)} n + n \cdot (1/n) = O(\log^{4+4\beta+(\beta+1)\alpha} n). \end{aligned}$$

□

Remark. Graphs of bounded expanding dimension and graphs of bounded doubling dimension are very closely related. Indeed, it can be shown that, assigning a specific weight function to a graph of bounded doubling dimension (the doubling measure of its metric), it can be made bounded growth by considering the ball sizes with nodes multiplicity corresponding to their weight [20]. Moreover, this weight function can be computed in polynomial time [21]. This allows us to extend Theorem 1 to graphs of bounded doubling dimension, up to a constant factor change in the exponent of greedy routing performances[3].

4 Impossibility results

In this section, we show that the high clustering coefficient hypothesis used in all our previous results is crucial. We prove that, without this hypothesis, any "structural algorithm" aiming at detecting the long range links of a graph in (H, φ), where H is of bounded growth and φ is density-based, fails to detect a polynomial number of long range links with polynomial stretch. By "structural algorithm", we mean any algorithm *à la* EXTRACT, i.e., that considers only the statistical structure of the input graph G. More precisely an algorithm $\mathcal{A}$ for recovering the base graph H from $G \in (H, \varphi)$ is *structural* if $\mathcal{A}$ decides whether or not an edge $e \in E(G)$ is a long range link according to the value of $\Pr(G \mid e \in E(H))$. Algorithm EXTRACT is an extreme case in this class of algorithms applied to graphs with high clustering coefficient. Indeed, if $e = \{u, v\} \in E(H)$, one must have $\frac{1}{n}|N_G(u) \cap N_G(v)| \geq c$. Hence, if $\frac{1}{n}|N_G(u) \cap N_G(v)| < c$, then $\Pr(G \mid e \in E(H)) = 0$, and therefore it would identify e as a long range link. Algorithm EXTRACT is only failing in the detection of few long range links with large stretch (Lemma 1) because, for a link $e = \{u, v\}$ with large stretch, $\Pr(\frac{1}{n} \cdot |N_G(u) \cap N_G(v)| \geq c)$ is small. We show that in absence of the high clustering coefficient, the number of long range links with large stretch that are not detected can be much higher, for any structural algorithm.

This impossibility result even holds in the case of a $(2n+1)$-node cycle C_{2n+1} augmented using the harmonic distribution $h_u^{(n)}(v) = 1/(2H_n \cdot \text{dist}(u, v))$, where $H_n = \sum_{i=1}^{n} \frac{1}{i}$ is the nth harmonic number. Note that $h^{(n)}$ is density-based, but C_{2n+1} has a clustering coefficient equal to zero. It was proved in [25] that greedy routing in $(C_{2n+1}, h^{(n)})$ performs in $O(\log^2 n)$ expected number of steps between any pair.

Theorem 2 *For any $0 < \varepsilon < 1/5$, any structural algorithm for recovering the base graph C_{2n+1} in $G \in (C_{2n+1}, h^{(n)})$ fails in the detection of an expected number $\Omega(n^{5\varepsilon}/\log n)$ of long range links of stretch $\Omega(n^{1/5-\varepsilon})$.*

Proof. Let $C_{2n+1} = \{x_1, \ldots, x_{2n+1}\}$ with nodes numbered clockwise. We divide C_{2n+1} into intervals of length L and we consider a specific configuration of long range links on each of these intervals. More precisely, let $I = \{x_1, \ldots, x_L\}$ be an interval of length

[3] The interested reader can find the details in Section C of the Appendix.

L in $G \in (C_{2n+1}, h^{(n)})$, and let $\mathcal{E}_I$ be the event="G contains the six long range links $(x_1, x_4), (x_4, x_1)$, $(x_2, x_{L-1}), (x_{L-1}, x_2)$, $(x_3, x_L), (x_L, x_3)$" (see Figure 1 page 20). Let us write $e_1 =_{\text{def}} \{x_1, x_4\}, e_2 =_{\text{def}} \{x_2, x_{L-1}\}$ and $e_3 =_{\text{def}} \{x_3, x_L\}$. Let $\mathcal{A}$ be some structural algorithm for detecting the long range links. Note that conditionally to $\mathcal{E}_I$, $\mathcal{A}$ can make a specific mistake, called *swap mistake*, by returning the six long range links as local edges of C_{2n+1}, and returning the local C_{2n+1} edges $\{x_1, x_2\}$, $\{x_3, x_4\}$, and $\{x_{L-1}, x_L\}$ as six long range links created by $h^{(n)}$. Conditionally to $\mathcal{E}_I$, counting the number of times $\mathcal{A}$ does the swap mistake on I is a lower bound on the number of possible mistakes it does on I. Note that the swap mistake induces a modification of the distances perceived in C_{2n+1} of at most 2. For instance, the distance in C_{2n+1} from y to z in Figure 1 is k, but it would appear as being $k+2$ if $\mathcal{A}$ does the swap mistake because the local edge $\{x_{L-1}x_L\}$ in the ring is replaced by the path $(e_2, \{x_2, x_3\}, e_3)$ of length 3. The key of the proof is to show that, when L is large, this modification is too tiny to be detectable on expectation by any structural algorithm. We have the following claim (the proofs of all the claims can be found in the Appendix)

Claim 8 *Let e be a link in $G \in (C_{2n+1}, h^{(n)})$, and let $\mathcal{B}$ be a structural algorithm which systematically decides that*

$$\begin{aligned}
&e \in E(C_{2n+1}) \quad \textit{if} \quad \Pr\{G \,|\, e \in E(C_{2n+1})\} > \Pr\{G \,|\, e \notin E(C_{2n+1})\},\\
&e \notin E(C_{2n+1}) \quad \textit{if} \quad \Pr\{G \,|\, e \in E(C_{2n+1})\} < \Pr\{G \,|\, e \notin E(C_{2n+1})\},\\
&\textit{uniform random choice otherwise.}
\end{aligned}$$

Then, the expected number of mistakes of $\mathcal{B}$ on e is at most the expected number of mistakes of any structural algorithm on e.

From Claim 8, since we compute a lower bound on the expected number of mistakes of $\mathcal{A}$, we can assume that $\mathcal{A}$ always decides $e \in E(C_{2n+1})$ if $\Pr\{G \,|\, e \in E(C_{2n+1})\} > \Pr\{G \,|\, e \notin E(C_{2n+1})\}$ and $e \in E(C_{2n+1})$ otherwise. In case of equality, $\mathcal{A}$ choses uniformly at random between the two possibilities.

Again in the sake of computing a lower bound on the number of mistakes, we can also assume that $\mathcal{A}$ has extra information on H and knows that $H = C_{2n+1}$ (in particular, it knows that the degree in H is 2), that $h^{(n)}$ is the harmonic distribution, and also knows all local links in I except $\{x_1, x_2\}, \{x_3, x_4\}, \{x_{L-1}, x_L\}$. Hence, if $\mathcal{A}$ decides $e_1 \notin H$, then $\{e_2, e_3\} \notin H$ from degree considerations (indeed, there is exactly one outgoing long range link at each node). Therefore

$$\Pr\{G \,|\, e_1 \notin C_{2n+1}\} = \Pr\{G \,|\, \{e_1, e_2, e_3\} \cap E(C_{2n+1}) = \emptyset\}.$$

Let Ω_I be the probability space describing the set of the $L-3$ other long range links outgoing from I. A configuration $C \in \Omega_I$ can be written as $C = \big\{\{(\ell_5^o, \sigma_5^o), \mathcal{L}_5\}, \ldots, \{(\ell_{L-2}^o, \sigma_{L-2}^o), \mathcal{L}_{L-2}\}\big\}$ where for each $5 \le i \le L-2$:

– $\ell_i^o \in \{0, \ldots n\}$ is the length of the long range link of x_i,

- $\sigma_i^o \in \{-1,+1\}$ is the *direction* of the long range link : it is equal to $+1$ if the link goes clockwise, and -1 otherwise,
- and $\mathcal{L}_i = \{(\ell_i^1,\sigma_i^1),\ldots,(\ell_i^{p_i},\sigma_i^{p_i})\}$ is the list of the lengths and directions of the $p_i \geq 0$ incoming long range links arriving at x_i.

Note that, for any long range link with its both extremities between x_5 and x_{L-2}, its probability of existence is unchanged whether the edges e_1, e_2, e_3 belong to C_{2n+1} or not. On the contrary, any long range link of a node x_i, for $5 \leq i \leq L-2$, that has length $\ell_i^o > L-i$ and direction $\sigma_i^o = +1$ has probability $1/(2H_n(\ell_i^o+2))$ to exist if e_1, e_2 and e_3 are in C_{2n+1}, and probability $1/(2H_n\ell_i^o)$ if e_1, e_2 and e_3 are not in C_{2n+1}. Therefore, the probability of existence of the long range link of x_i is greater when e_1, e_2 and e_3 are not in C_{2n+1}, which is the event $\mathcal{E}_I$. Symmetrically, if the direction is $\sigma_i^o = -1$, and $\ell_i^o > i$, the probability of existence of x_i's long range link is $1/(2H_n\ell_i)$ if e_1, e_2 and e_3 are in C_{2n+1}, and $1/(2H_n(\ell_i^o+2))$ otherwise : the probability of existence is lower conditionally to $\mathcal{E}_I$ than to $\neg\mathcal{E}_I$. Informally, we deduce from these observations that, for any two configurations $C, \tilde{C} \in \Omega_I$ that are "symmetric" with respect to the middle of $\{x_5,\ldots,x_{L-2}\}$, $\mathcal{A}$ has to make a swap mistake on one of them. The idea of the proof is therefore to group such "symmetric" configurations in pairs in order to lower bound the expected number of swap mistakes by 1/2 on each pair.

More formally, we say that two configurations $C = \big\{\{(\ell_i^o,\sigma_i^o),\mathcal{L}_i\}, 5 \leq i \leq L-2\big\}$ and $\tilde{C} = \big\{\{(\tilde{\ell}_i^o,\tilde{\sigma}_i^o),\mathcal{L}'_i\}, 5 \leq i \leq L-2\big\}$ are *symmetric* if and only if :

1. the long range contacts of outgoing long range links outgoing from C or $\tilde{C}$ are not in $\{x_2, x_3, x_L, x_{L+1}\}$,
2. none of the origins of the long range contacts ingoing in C or $\tilde{C}$ is x_{L+1},
3. for all $0 \leq j \leq L-7$, $(\ell_{5+j}^o,\sigma_{5+j}^o) = (\tilde{\ell}_{L-2-j}^o, -\tilde{\sigma}_{L-2+j}^o)$, and $(\ell_{5+j}^m,\sigma_{5+j}^m) = (\tilde{\ell}_{L-2-j}^m, -\tilde{\sigma}_{L-2+j}^m)$ for any $(\ell_{5+j}^m,\sigma_{5+j}^m) \in \mathcal{L}_{5+j}$.

Note that, because of conditions 1 and 2, not all the configurations in Ω_I can be *symmetrized*. Let Ω_I^s be the set of configuration of Ω_I that can be symmetrized. Let X_I be the random variable counting the number of swap mistakes done by $\mathcal{A}$ on I. We have :

$$\mathbb{E}(X_I \,|\, \mathcal{E}_I) = \sum_{C\in\Omega_I} \mathbb{E}(X_I \,|\, \mathcal{E}_I \text{ and } C)\cdot \Pr\{C\} \geq \sum_{C\in\Omega_I^s} \frac{1}{2}\cdot \Pr\{C\},$$

since $\mathcal{A}$ makes at least one swap mistake for two symmetric configurations C and $\tilde{C}$ in Ω_I^s on expectation. It remains to evaluate $\sum_{C\in\Omega_I^s}\Pr\{C\}$. A configuration $C = \big\{\{(\ell_i^o,\sigma_i^o),\mathcal{L}_i\}, 5 \leq i \leq L-2\big\}$ can be symmetrized if and only if : 1) for all $5 \leq i \leq L-2$, $\ell_i^o \notin \{i-2, i-3, L-i, L-i+1\}$, and 2) the long range contact of x_{L+1} is not in $\{x_5,\ldots,x_{L-2}\}$. We get the following claim.

Claim 9 $\sum_{C\in\Omega_I^s}\Pr\{C\} \geq e^{-(2\log L)/H_n}$.

Thus, $\mathbb{E}(X_I \,|\, \mathcal{E}_I) \geq \frac{1}{2}e^{-2(\log L)/H_n} \geq \frac{1}{2e}$, since $L \leq n$.

Let X be the random variable counting the total number of swap mistakes of $\mathcal{A}$ on G. Let $I_1, I_2, \ldots, I_{\lfloor (2n+1)/L \rfloor}$ be the largest set of adjacent and disjoint intervals of length L on C_{2n+1}. We have :

$$\mathbb{E}(X) \geq \sum_{i=1}^{\lfloor (2n+1)/L \rfloor} \mathbb{E}(X_{I_i}) \geq \sum_{i=1}^{\lfloor (2n+1)/L \rfloor} \mathbb{E}(X_{I_i} \,|\, \mathcal{E}_{I_i}) \cdot \Pr \mathcal{E}_{I_i} \geq \sum_{i=1}^{\lfloor (2n+1)/L \rfloor} \frac{1}{2e} \cdot \frac{1}{(2H_n)^6 \cdot 3^2 \cdot (L-2)^4}$$

because $1/\big((2H_n)^6 \cdot 3^2 \cdot (L-2)^4\big)$ is the probability of existence of the six long range links described in $\mathcal{E}_{I_i}$. Finally :

$$\mathbb{E}(X) \geq \Omega\Big(\frac{n}{L^5 \log^6 n}\Big).$$

Specifically, taking $L = n^{\frac{1}{5}-\varepsilon}$ for some $0 < \varepsilon < 1/5$, we get $\mathbb{E}(X) \geq \Omega(n^{5\varepsilon}/\log n)$, which means that $\mathcal{A}$ fails in detecting at least $\Omega(n^{5\varepsilon}/\log n)$ links of length $\Theta(n^{\frac{1}{5}-\varepsilon})$ on expectation, since one swap mistake of $\mathcal{A}$ on some interval means that the long range edges e_2 and e_3 of this interval have not been detected as long range links. $\square$

Acknowledgments : The authors are thankful to Dmitri Krioukov for having raised to them the question of how to extract the based graph of an augmented graph, and for having pointed to them several relevant references. They are also thankful to Augustin Chaintreau and Laurent Viennot for fruitful discussions.

Références

[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] I. Abraham, D. Malkhi, and O. Dobzinski. LAND : Stretch $(1+\epsilon)$ locality aware networks for DHTs. In ACM-SIAM Symposium on Discrete Algorithms (SODA), 2004.

[3] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman. Search in power law networks. Physical Review E, vol. 64, pages 46135-46143, 2001.

[4] R. Andersen, F. Chung, and L. Lu. Modeling the small-world phenomenon with local network flow. Internet Mathematics 2, no. 3, pages 359–385, 2006.

[5] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of Traceroute sampling, or : power-law degree distributions in regular graphs. In 37th ACM Symposium on Theory of Computing (STOC), 2005.

[6] 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.

[7] A. Barabási and R. Albert. Emergence of scaling in random networks. Science 286, pages 509-512, 1999.

[8] 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, 2001.

[9] F. Chung and L. Lu. The small world phenomenon in hybrid power law graphs. Lect. Notes Phys. 650, pages 89-104, 2004.

[10] P. Dodds, R. Muhamad, and D. Watts. An experimental study of search in global social networks. Science 301, no5634, pages 827-829, 2003.

[11] 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).

[12] 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.

[13] 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.

[14] 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.

[15] P. Fraigniaud, P. Gauron, and M. Latapy. Combining the use of clustering and scale-free nature of user exchanges into a simple and efficient P2P System. In 11th European conference on Parallelism (EuroPar), LNCS 3648, pages 1163-1172, 2005.

[16] P. Fraigniaud, C. Gavoille, A. Kosowski, E. Lebhar, and Z. Lotker. Universal augmentation schemes for network navigability : overcoming the $\sqrt{n}$-barrier. To appear in proc. of the 19th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), 2007.

[17] 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.

[18] 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, pages 376-386, 2006.

[19] A. Gupta, R. Krauthgamer, and J. Lee. Bounded geometries, fractals, and low- distortion embeddings. Proceedings of the 44th Annual IEEE Symposium on Foun- dations of Computer Science (FOCS), 534 ?543, 2003.

[20] J. Heinonen. Lectures on analysis on metric spaces. Springer-Verlag, 2001.

[21] S. Har-Peled and M. Mendel. Fast Construction of Nets in Low Dimensional Metrics, and Their Applications. SICOMP 35 (5), pages,1148–1184, 2006.

[22] A. Iamnitchi, M. Ripeanu, and I. Foster. Small-world file-sharing communities. In 23rd Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), pages 952-963, 2004.

[23] D. Karger and M. Ruhl. Finding nearest neighbors in growth-restricted metrics. In 34th ACM Symp. on the Theory of Computing (STOC), pages 63–66, 2002.

[24] B. Kim, C. Yoon, S. Han, and H. Jeong. Path finding strategies in scale-free networks. Physical Review E, vol. 65, pages 027103-1–027103-4, 2002.

[25] J. Kleinberg. The small-world phenomenon : an algorithmic perspective. In 32nd ACM Symp. on Theory of Computing (STOC), pages 163-170, 2000.

[26] J. Kleinberg. Small-World Phenomena and the Dynamics of Information. Advances in Neural Information Processing Systems (NIPS) 14, 2001.

[27] J. Kleinberg. Complex networks and decentralized search algorithm. Nevanlinna prize presentation at the International Congress of Mathematicians (ICM), Madrid, 2006.

[28] D. Krioukov, K. Fall, and X. Yang. Compact routing on Internet-like graphs. In 23rd Conference of the IEEE Communications Society (INFOCOM), 2004.

[29] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins Geographic routing in social networks Proc. of the Natl. Academy of Sciences of the USA, vol. 102/3, pages 11623–11628.

[30] 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.

[31] 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.

[32] 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.

[33] C. Martel, and V. Nguyen. Analyzing and characterizing small-world graphs. In 16th ACM-SIAM Symp. on Discrete Algorithms (SODA), pages 311-320, 2005.

[34] C. Martel, and V. Nguyen. Designing networks for low weight, small routing diameter and low congestion. In 25th Conference of the IEEE Communications Society (INFOCOM), 2006.

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

[36] M. Newman. The Structure and Function of Complex Networks. SIAM Review 45, pages 167-256, 2003.

[37] M. Newman, A. Barabasi, and D. Watts (Eds). The Structure and Dynamics of Complex Networks. Princeton University Press, Princeton, 2006.

[38] R. Pastor-Satorras and A. Vespignani. Evolution and Structure of the Internet : A Statistical Physics Approach. Cambridge University Press, 2004.

[39] N. Sarshar, P. Boykin, and V. Roychowdhury. Percolation search in power law networks : making unstructured peer-to-peer networks scalable. In 4^{th} International Conference on Peer-to-Peer Computing, pages 2-9, 2004.

[40] 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.

[41] D. Watts and S. Strogatz. Collective dynamics of small-world networks. Nature 393, pages 440-443, 1998.

– Figure –

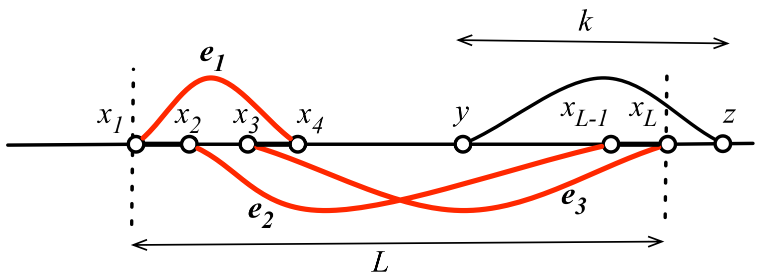

FIG. 1 – Configuration of the links e_1, e_2 and e_3 of $\mathcal{E}_I$ in the proof of Theorem 2.

– Appendix –

A Proof of Property 1

Let H be the $2n$-node graph consisting in a path P of n nodes $u_1, \ldots, u_n$ connected to a d-dimensional ℓ_∞-mesh M of n nodes. Precisely, M is the n-node graph consisting of k^d nodes labeled $(x_1, \ldots, x_d)$, $x_i \in \mathbf{Z}_k$ for $1 \leq i \leq d$, where $k = n^{1/d}$. Node $(x_1, \ldots, x_d)$ of M is connected to all nodes $(x_1 + a_1, \ldots, x_d + a_d)$ where $a_i \in \{-1, 0, 1\}$ for $1 \leq i \leq d$, and all operations are taken modulo k. Note that, by construction of M, the distance between any two nodes $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ is $\max_{1 \leq i \leq d} \min\{|y_i - x_i|,\ k - |y_i - x_i|\}$. Hence, the diameter of M is $\lfloor n^{1/d}/2 \rfloor$. Assume that P is augmented using the harmonic augmenting distribution h, and M is augmented using some augmenting distribution ψ. It is proved in [18] that, for any augmenting distribution ψ for M, there is a pair $s_0, t_0 \in V(M)$, with $2^{d-1} - 1 \leq \text{dist}(s_0, t_0) \leq 2^d$ such that the expected number of steps of greedy routing from s_0 to t_0 is at least $\Omega(2^d)$ whenever $d < \sqrt{\log n}$. Let $d = \sqrt{\log n}/2$. To construct H, we connect the extremity u_n of P to the node t_0 of M (see Figure 2). In P, we use a slight modification $\hbar$ of the harmonic distribution h : $\hbar$ is exactly h except at node u_1 where $\hbar_{u_1}(s_0) = 1$ (i.e. for any trial of $\hbar$, the long range contact of u_1 is s_0). Consider the augmented graph model $(H, \hbar \cup \psi)$, and set $e = \{u_1, s_0\}$. In $(H, \hbar \cup \psi)$, greedy routing within P takes

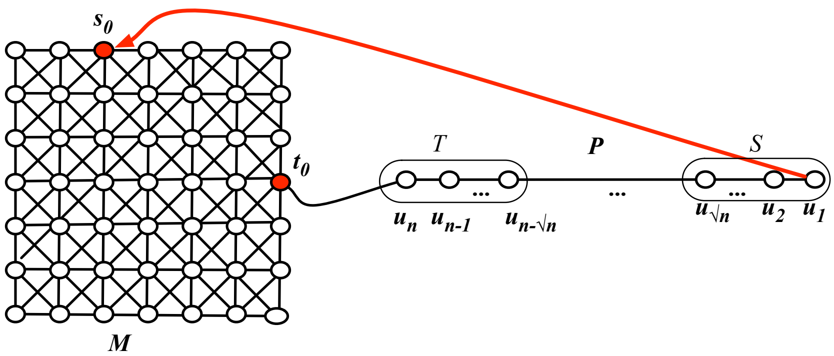

FIG. 2 – Graph H in the proof of Property 1.

$O(\log^2 n)$ expected number of steps [25]. Let $H' = H \cup \{e\}$. We consider greedy routing using distances in H' between the two following sets :

$$S = \{u_2, \ldots, u_{\sqrt{n}}\} \text{ and } T = \{u_{n-\sqrt{n}}, \ldots, u_n\}.$$

Hence, for any $s \in S$ and $t \in T$, the shortest path from s to t in H' goes through e. Indeed, their shortest path in H is of length at least $n - 2\sqrt{n}$, while in H' it is of length at most $2\sqrt{n} + \text{dist}(s_0, t_0) + 2 \leq 2\sqrt{n} + 2^{\sqrt{\log n}/2} + 2$ using e, which is less than $n - 2\sqrt{n}$.

Let $\mathcal{B} = B_H(u_{n-\sqrt{n}}, 2\sqrt{n} + n^{1/d})$. For any node $x \in S$, the probability that the long range contact of x is in $\mathcal{B}$ is at most $O\left(\frac{1}{\sqrt{n}\cdot\log n}\right)$. Therefore, the expected number of steps required to find such a link in S is at least $\Omega(\sqrt{n} \cdot \log n)$ which is larger than $|S|$. As a consequence, with constant probability, greedy routing from a node $s \in S$ to a node $t \in T$, using the distances in H', routes to u_1 and, from there to s_0. This implies that greedy routing from s to t will take at least as many steps as greedy routing from s_0 to t_0 within (M, ψ), that is $\Omega(2^{\sqrt{\log n}})$ expected number of steps, which is $\omega(\mathrm{polylog}(n))$. □

B Proofs of the claims

Proof of Claim 1. Let D be the diameter of H and Δ be the maximum degree of H.

$$\begin{aligned} Z_u &= \sum_{v \neq u} \frac{1}{|B_H(u, \mathrm{dist}(u,v))|} = \sum_{r=1}^{D} \frac{|B_H(u,r)| - |B_H(u,r-1)|}{|B_H(u,r)|} = \sum_{r=1}^{D} \left(1 - \frac{|B_H(u,r-1)|}{|B_H(u,r)|}\right) \\ &\geq \sum_{r=1}^{D} \left(1 - \frac{|B_H(u,r-1)|}{|B_H(u,r-1)|+1}\right) = \sum_{r=1}^{D} \frac{1}{|B_H(u,r-1)|+1}, \end{aligned}$$

since $|B_H(u,r-1)| \leq |B_H(u,r)| - 1$ for any $1 \leq r \leq D$. Moreover, since H has maximum degree Δ, $|B_H(u,r)| \leq \Delta^r$ for any $1 \leq r \leq D$. Therefore :

$$Z_u \geq \sum_{r=1}^{D} \frac{1}{\Delta^{r-1}+1} \geq \frac{1}{2} \cdot \frac{1-(1/\Delta)^D}{1-(1/\Delta)}.$$

Since $2 \leq \Delta \leq n$, there exists $n_0 \geq 1$ such that for any $n \geq n_0$, $\frac{1-(1/\Delta)^D}{1-(1/\Delta)} \geq \frac{1}{2}(1+\frac{1}{\Delta}) \geq \frac{1}{2}$. Finally, for any $n \geq n_0$ and for all $u \in G$

$$Z_u \geq \frac{1}{4}.$$

□

Proof of Claim 2. Let $w \in V(H)$ and assume that $\{w, v\} \in H$. Since $\mathrm{dist}_H(u,w) \geq \mathrm{dist}_H(u,v) - 1 \geq \log^{\beta+1} n - 1$, and $d_H(v,w) = 1$, we get that $B_H(w, \mathrm{dist}(w,u))$ contains at least $|N_H(v)| + \log^{\beta+1} n - 2$ nodes. Therefore, the probability that u is the long range contact of w is at most :

$$\frac{1}{Z_{\min}} \cdot \frac{1}{|N_H(v)| + \mathrm{dist}_H(u,w) - 2} \leq \frac{4}{|N_H(v)| + \log^{\beta+1} n - 2}.$$

The probability that u and v have at least $c \cdot n$ neighbors in common in G is at most the probability that there are k_1 of the nodes $w \in N_H(v)$ such that the long range contact of w is u and k_2 nodes $w \in N_H(u)$ such that the long range contact w is v, with $k_1 + k_2 \geq c \cdot n - 2$.

Using the previous upper bound on the probability of each of these events, we get that $\Pr\{X_e = 1\}$, i.e. the probability for e to survive, is at most :

$$\sum_{k_1,k_2\geq 0,\, k_1+k_2\geq c\cdot n-2} \binom{|N_H(v)|}{k_1}\binom{|N_H(u)|}{k_2} \prod_{j=1}^{k_1} \frac{4}{|N_H(v)| + \log^{\beta+1} n - 2} \prod_{i=1}^{k_2} \frac{4}{|N_H(u)| + \log^{\beta+1} n - 2}$$
$$\leq \frac{1}{[(\log^{\beta+1} n - 2)/(4N)]^{c\cdot n-2}} \sum_{k_1,k_2\geq 0,\, k_1+k_2\geq c\cdot n-2} \binom{|N_H(v)|}{k_1}\binom{|N_H(u)|}{k_2} \frac{1}{|N_H(v)|^{k_1}} \frac{1}{|N_H(u)|^{k_2}},$$

where $N = \max\{|N_H(u)|, |N_H(v)|\}$. Since the maximum degree is $\Delta = O(\log^\beta n)$ and $N \leq \Delta$, we have $((\log n^{\beta+1} - 2)/(4N))^{-1} \leq 8/\log n$. Moreover, for any $a \geq \in \mathbb{N}$, since $a!/(a-b)! \leq a^b$, we have $\binom{a}{b}\frac{1}{a^b} \leq \frac{1}{b!}$. Finally, we get that $\Pr\{X_e = 1\}$ is at most :

$$\frac{1}{(\log n/8)^{c\cdot n-2}} \sum_{k_1,k_2\geq 0,\, k_1+k_2\geq c\cdot n-2} \frac{1}{k_1!k_2!} \leq \frac{1}{(\log n/8)^{c\cdot n-2}} \sum_{i\geq c\cdot n-2} \frac{2^i}{i!} \leq \frac{\mathcal{O}(1)}{(\log n/8)^{c\cdot n-2}} \leq \frac{1}{n},$$

for n large enough, since $c > \Omega(\frac{\log n}{n \log\log n})$.

□

Proof of Claim 3. The variables Y_e, $e \in \mathcal{S}$, are i.i.d., and $\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\} = |\mathcal{S}|/n \leq 1$. From Chernoff's inequality, we get, for any $\delta > 0$:

$$\Pr\{\sum_{e\in\mathcal{S}} Y_e > (1+\delta)\cdot\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}\} < \Big(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\Big)^{\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}}$$
$$= n^{-\big((1+\delta)\log(1+\delta) - \delta\log e\big)\frac{\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}}{\log n}}$$

Let $(1+\delta) = \frac{\log^{2\beta+1} n}{\Delta^2} \cdot \frac{1}{\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}}$. Note that $(1+\delta) \geq \log n/\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\} \geq \log n$, then $\delta\log e \leq (\delta+1)\log(\delta+1)/2$ for $n \geq n_1$ for some $n_1 > 0$. Therefore :

$$\Pr\{\sum_{e\in\mathcal{S}} Y_e > (1+\delta)\cdot\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}\} < n^{-\frac{1}{2}\log(1+\delta)} \leq n^{-\frac{1}{2}\log\log n} \leq \frac{1}{n}.$$

Finally :

$$\Pr\{\sum_{e\in\mathcal{S}} Y_e > \log^{2\beta+1} n/\Delta^2\} = \Pr\{\sum_{e\in\mathcal{S}} Y_e > (1+\delta)\cdot\mathbb{E}\{\sum_{e\in\mathcal{S}} Y_e\}\} \leq \frac{1}{n}.$$

□

Proof of Claim 4.

$$
\begin{aligned}
Z_u &= \sum_{v\neq u} \frac{1}{|B_H(u,\mathrm{dist}_H(u,v))|} = \sum_{r=1}^{D} \frac{|B_H(u,r)| - |B_H(u,r-1)|}{|B_H(u,r)|} \\
&= \sum_{i=1}^{\log(D+1)} \sum_{k=2^{i-1}}^{2^i-1} \frac{|B_H(u,k)| - |B_H(u,k-1)|}{|B_H(u,k)|} \\
&\leq \sum_{i=1}^{\log(D+1)} \frac{1}{|B_H(u,2^{i-1})|} \sum_{k=2^{i-1}}^{2^i-1} (|B_H(u,k)| - |B_H(u,k-1)|) \\
&\leq \sum_{i=1}^{\log(D+1)} \frac{|B_H(u,2^i-1)|}{|B_H(u,2^{i-1})|} - 1 \leq \sum_{i=1}^{\log(D+1)} (2^\alpha - 1) \leq 2^\alpha \log n.
\end{aligned}
$$

□

Proof of Claim 5. We have $B_H(u_j, r/2) \subseteq B_{H''}(u_j, r/2)$. Therefore, the probability for $L(x)$ to lie in $B_{H''}(u_j, r/2)$ is at least the probability to lie in $B_H(u_j, r/2)$. Moreover, the largest distance in H from x to a node in $B_H(u_j, r/2)$ is at most $\log^{\beta+1} n \cdot (3r/2)$, because the stretch is at most $\log^{\beta+1} n$ in H''. It follows :

$$\Pr\{L(x) \in B_{H''}(u_j, r/2)\} \geq \frac{1}{Z_{\max}} \cdot \frac{|B_H(u_j, r/2)|}{|B_H(x, \log^{\beta+1} n \cdot (3r/2))|}.$$

On the other hand $B_H(x, \log^{\beta+1} n \cdot (3r/2)) \subseteq B_H(u_j, \log^{\beta+1} n \cdot (5r/2))$. And from the expanding dimension α of H we get :

$$|B_H(u_j, \log^{\beta+1} n \cdot (5r/2))| \leq 2^{\alpha(\log 5 + (\beta+1)\log\log n)} |B_H(u_j, r/2)|.$$

Finally :

$$\Pr\{L(x) \in B_{H''}(u_j, r/2)\} \geq \frac{1}{Z_{\max}} \cdot \frac{1}{2^{3\alpha} \log^{\alpha(\beta+1)} n} = \frac{1}{2^{4\alpha} \log^{1+\alpha(\beta+1)} n}.$$

Assume that the event "$L(x) \in B_{H''}(u_j, r/2)$" occurs. Suppose for the purpose of contradiction that the next step of greedy routing is the node y with $y \notin B_{H''}(u_j, r/2)$. From greedy routing strategy, it must be that $\mathrm{dist}_{H'}(y,t) \leq \mathrm{dist}_{H'}(L(x),t)$. Since x has only one long range link, y has to be a neighbor of x in H. Therefore, $\mathrm{dist}_H(x,y) = 1 = \mathrm{dist}_{H''}(x,y)$. Moreover, since j is the concerned index for x, y has to be on the shortest path in H'' from x to u_j (otherwise y would be further from t than x in H'). We have :

$$
\begin{aligned}
\mathrm{dist}_{H'}(y,t) \geq \mathrm{dist}_{H'}(x,t) - 1 &= \mathrm{dist}_{H''}(x,y) + \mathrm{dist}_{H''}(y,u_j) + \mathrm{dist}_{H'}(v_j,t) \\
&> r/2 + 1 + \mathrm{dist}_{H'}(v_j,t).
\end{aligned}
$$

On the other hand, we have $\mathrm{dist}_{H'}(L(x),t) \leq r/2 + 1 + \mathrm{dist}_{H'}(v_j,t)$, and therefore we obtain that $\mathrm{dist}_{H'}(L(x),t) < \mathrm{dist}_{H'}(y,t)$. This is in contradiction with greedy routing strategy and concludes the proof of the claim. □

Proof of Claim 6. Since $\tau' > \tau$, greedy routing strategy enforces that $\text{dist}_{H'}(x', t) < \text{dist}_{H'}(x, t)$. On the other hand, by definition of the concerned index we have :

$$\begin{aligned} & \text{dist}_{H'}(x,t) = \text{dist}_{H''}(x,u_j) + 1 + \text{dist}_{H'}(v_j,t) \leq r + 1 + \text{dist}_{H'}(v_j,t) \\ \text{and} \quad & \text{dist}_{H'}(x',t) = \text{dist}_{H''}(x',u_j) + 1 + \text{dist}_{H'}(v_j,t), \end{aligned}$$

therefore $\text{dist}_{H''}(x', u_j) \leq r$ and thus $x' \in B_{H''}(u_j, R)$. □

Proof of Claim 7. At any step τ, we have for any $0 \leq j \leq k$:

$$\Pr\{\mathcal{E}_j^{i-1}(\tau+1) \,|\, \mathcal{E}_j^i(\tau) \text{ and } j \text{ is the concerned index at step } \tau\} \geq 1/(2^{4\alpha} \log^{1+\alpha(\beta+1)} n). \tag{1}$$

Indeed, from Claim 6, if $\mathcal{E}_j^i(\tau)$ and if j is the concerned index at the current step, then the current node x satisfies $\text{dist}_{H''}(x, u_j) \leq 2^i$ and we can apply Claim 5 which provides the above inequality.

For any fixed j, from Equation 1, if there exist $i > 0$ and $\tau > 0$ such that $\mathcal{E}_j^i(\tau)$ occurs, then greedy routing does not perform more than $2^{4\alpha} \log^{1+\alpha(\beta+1)} n$ steps on expectation before $\mathcal{E}_j^{i-1}(\tau')$ occurs for some $\tau' > \tau$. Besides, since every step $\tau > 0$ has a concerned index, after at most $(k+1) \cdot 2^{4\alpha} \log^{1+\alpha(\beta+1)} n$ steps on expectation, there must exist one index j for which $\mathcal{E}_j^{i-1}(\tau')$ occurs for some $\tau' > 0$. This conclude the proof of the claim. □

Proof of Claim 8. Let $\mathcal{B}'$ be some structural algorithm, and let $\alpha \in [0,1]$ such that $\mathcal{B}'$ decides $e \in E(C_{2n+1})$ with probability α and $e \notin E(C_{2n+1})$ with probability $1-\alpha$. $\mathcal{B}'$ makes a mistake on e if $e \in E(C_{2n+1})$ while it decides $e \notin E(C_{2n+1})$ or vice versa. On expectation, such a mistake occurs

$$\alpha \cdot \Pr\{G \,|\, e \notin E(C_{2n+1})\} + (1-\alpha) \cdot \Pr\{G \,|\, e \in E(C_{2n+1})\} \text{ times.}$$

If $\Pr\{G \,|\, e \in E(C_{2n+1})\} > \Pr\{G \,|\, e \notin E(C_{2n+1})\}$, this number is strictly greater than $\Pr\{G \,|\, e \notin E(C_{2n+1})\}$. But in this case, $\mathcal{B}$ makes $\Pr\{G \,|\, e \notin E(C_{2n+1})\}$ mistakes on e on expectation. Similarly, if $\Pr\{G \,|\, e \in E(C_{2n+1})\} < \Pr\{G \,|\, e \notin E(C_{2n+1})\}$, $\mathcal{B}$ makes strictly less mistakes on expectation. Finally, if $\Pr\{G \,|\, e \in E(C_{2n+1})\} = \Pr\{G \,|\, e \notin E(C_{2n+1})\}$, the expected number of mistakes of $\mathcal{B}'$ is one while the one of $\mathcal{B}$ is $1/2$. We conclude that the expected number of mistakes of $\mathcal{B}'$ is larger than the one of $\mathcal{B}$ on e. □

Proof of Claim 9.

$$\begin{aligned} \sum_{C \in \Omega_I^s} \Pr\{C\} &\geq \Big(1 - \frac{H_{L-3}}{2H_n}\Big) \prod_{5 \leq i \leq L-2} \Big(1 - \frac{1}{2H_n}\big(\frac{1}{i-2} + \frac{1}{i-3} + \frac{1}{L-i} + \frac{1}{L-i+1}\big)\Big) \\ &\geq \Big(1 - \frac{\log(L-3)}{\log n}\Big) \prod_{5 \leq i \leq L-2} \Big(1 - \frac{1}{H_n}\big(\frac{1}{i-2} + \frac{1}{L-i}\big)\Big) \end{aligned}$$

Then, for n large enough,

$$\ln\Big(\sum_{C\in\Omega_I^s}\Pr\{C\}\Big) \geq -\frac{1}{2}\frac{\log(L-3)}{\log n} - \frac{1}{2}\frac{1}{H_n}\sum_{5\leq i\leq L-2}\big(\frac{1}{i-2}+\frac{1}{L-i}\big)$$
$$\geq -\frac{2\log L}{H_n}.$$

□

C Graphs of bounded doubling dimension

In this section, we briefly sketch how the results for graphs of bounded expanding dimension given in Theorem 1 can be extended to graphs of bounded doubling dimension.

Definition 5 *A graph G is (q_0,α)-doubling if and only if, for any node $u \in V(G)$, and for any $r > 0$, we have : $|B_G(u,r)| \geq q_0 \;\Rightarrow\; \exists W \subseteq V(G),\; |W| \leq 2^\alpha,\; B_G(u,2r) \subseteq \bigcup_{w\in W} B_G(w,r)$. In the bulk of this paper, we will set $q_0 = O(1)$, and refer to α as the doubling dimension of G.*

It is easy to check that if G has (q_0,α)-expansion then G is $(q_0,4\alpha)$-doubling (see e.g. [19]). The reverse is not true in general, except by providing to the nodes appropriate positive weights. More precisely, let us define the expansion of a node-weighted graph as in Definition 4 where the cardinality of a ball is replaced by the sum of the weights of its nodes. The following lemma is folklore (see e.g., [20, 21]).

Lemma 2 *If G has (q_0,α)-expansion then G is $(q_0,4\alpha)$-doubling. If G is (q_0,α')-doubling, then there exists a function $\mu : V(G) \to \mathbf{R}^+$ such that the node-weighted graph (G,μ) has $(q_0,13\alpha')$-expansion. Moreover, the weights $\{\mu(u), u \in V(G)\}$ can be computed in polynomial time. We say that μ is a* doubling measure *for G.*

Using the weights introduced in Lemma 2, one can extend Definition 2 : an augmenting distribution φ of a node-weighted graph $\langle H,\mathbf{w}\rangle$ is $\mathbf{w}$-density-based if and only if $\varphi_u(u) = 0$, and for every two distinct nodes u and v of H,

$$\varphi_u(v) = \frac{1}{W_u}\;\frac{1}{\sum_{x\in B_H(u,\mathrm{dist}_H(u,v))}\mathbf{w}(x)}$$

where $W_u = \sum_{w\neq u}(1/\sum_{x\in B_H(u,\mathrm{dist}_H(u,w))}\mathbf{w}(x))$ is the normalizing coefficient. Using these concepts, Theorem 1 can easily be extended to the following.

Theorem 3 *Let $G = (H,\varphi)$ be such that (a) H has clustering $c \geq \Omega(\frac{\log n}{n\,\log\log n})$, (b) H is (q_0,α)-doubling $2 < q_0 \leq O(1)$, (c) h has maximum degree $\Delta = O(\log^\beta n)$ for some $\beta \geq 1$,* *and (d)* φ *is* $\mathbf{w}$*-density based, where* $\mathbf{w}$ *is a doubling measure for* H*. Algorithm* EXTRACT *outputs a partition* (H', R') *of* $E(G)$ *such that* $E(H) \subseteq H'$ *and for any source-target pair* $(s,t) \in V(G) \times V(G)$*, the expected number of steps of greedy routing in* G *using the distance metric of* H' *is at most* $O(\log^{4+4\beta+13\alpha(\beta+1)} n)$.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399