HAL open science

Algorithmes de routage et modèles aléatoires pour les graphes petits mondes

Emmanuelle Lebhar

To cite this version:

Emmanuelle Lebhar. Algorithmes de routage et modèles aléatoires pour les graphes petits mondes. Réseaux et télécommunications [cs.NI]. Ecole normale supérieure de lyon - ENS LYON, 2005. Français. NNT : . tel-00011646

HAL Id: tel-00011646

https://theses.hal.science/tel-00011646v1

Submitted on 19 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : 341
N° attribué par la bibliothèque : 05ENSL0 341
École doctorale : Mathématiques et Informatique fondamentale

Thèse

présentée pour l'obtention du titre de

Docteur de l'École Normale Supérieure de Lyon

Spécialité : Informatique

par **Emmanuelle LEBHAR**

Algorithmes de routage et modèles aléatoires pour les graphes petits mondes

Soutenue publiquement le 6 décembre 2005 devant la commission d'examen composée de

Directeurs de thèse :	Michel Morvan et Nicolas Schabanel
Rapporteurs :	Pierre Fraigniaud et Jean-Claude König
Examinateurs :	Jose Rolim et Laurent Viennot

École Normale Supérieure de Lyon
Laboratoire de l'Informatique du Parallélisme
46 allée d'Italie - 69007 Lyon

Mis en page avec la classe Phdlasl

Remerciements

Au début de ma thèse, l'étude des grands réseaux d'interaction était un sujet en plein essor et riche de nombreuses problématiques de recherche. J'ai alors eu la chance de rencontrer Nicolas Schabanel qui désirait ouvrir ses recherches à ces problèmes. Sa grande culture algorithmique et sa vision originale de ces problèmes m'ont permis de me lancer dans l'étude des aspects algorithmiques de petits mondes. Ce fut un plaisir de travailler avec Nicolas Schabanel ; son optimisme et ses idées originales m'ont permis de maintenir le cap sur des problèmes nouveaux et difficiles, et d'ouvrir mes recherches à de nouveaux domaines avec engouement. Je tiens à le remercier d'avoir accepté de co-encadrer ma thèse, ainsi que pour sa disponibilité, sa sympathie et son soutien.

J'ai rencontré Michel Morvan après mon stage de DEA, il m'a alors fait découvrir les problématiques des grands réseaux d'interaction avec beaucoup d'enthousiasme. Je le remercie de m'avoir proposé de travailler avec lui sur ce sujet de recherche passionnant. Sa grande connaissance de divers aspect de l'informatique, sa culture pluridisciplinaire ainsi que son dynamisme m'ont permis d'avoir de riches discussions avec lui sur les problèmes que j'abordais, m'apportant toujours un regard nouveau sur mes recherches, je l'en remercie.

Je remercie Pierre Fraigniaud et Jean-Claude König d'avoir accepté d'être les rapporteurs de ma thèse. La pertinence et la précision de leurs commentaires m'ont ouvert de nouvelles perspectives de recherche.

Je remercie Michel Habib et Christophe Paul qui ont co-encadré mon stage de DEA au LIRMM, et m'ont réellement fait découvrir la recherche et ses attraits en me proposant un sujet intéressant, sur lequel j'ai pris plaisir à travailler et écrire en leur compagnie dans une ambiance sympathique.

Je tiens à remercier les membres de l'équipe MC2 du LIP dans laquelle j'ai pu effectuer ma thèse de façon très agréable. Je remercie en particulier Marianne Delorme et Jacques Mazoyer avec qui, depuis le début de mes études à l'ENS Lyon, j'ai pu avoir des discussions enrichissantes et sympathiques, tant sur la science que sur le monde qui nous entoure.

Je remercie également mes camarades de thèse et de bureaux avec qui j'ai

pu partager les difficultés ou les plaisirs d'une thèse dans la bonne humeur et l'humour, en particulier Nazim Fatès, Emmanuel Jeandel, Éric Boix, Guillaume Theyssier, Frédéric Chavanon, Gilles Radenne et Sylvain Périfel.

Lors de ma thèse, j'ai également eu la chance d'avoir des échanges scientifiques enthousiasmants et motivants avec Philippe Duchon, Christoph Dürr, Pierre Fraigniaud, Cyril Gavoille, Nicolas Hanusse, Matthieu Latapy, Zvi Lotker, Cristopher Moore et Ivan Rapaport, je les en remercie.

Pour mener à bien ma thèse, j'ai bénéficié du soutien d'une équipe de choc que je souhaite remercier grandement : merci à Fred&Mum, Emilie, Magali, Fanfan ainsi qu'à tous les amis cachés dans ces points de suspension... Et un grand merci à Clément !

Enfin, je remercie la machine à café et les Beatles sans qui rien de tout cela n'aurait été possible.

Dear Sir or Madam will you read my book,
It took me years to write will you take a look,
[...] It's a thousand pages give or take a few,
I'll be writing more in a week or two,
I can make it longer if you like the style [...].
—Paperback writer, The Beatles—

Table des matières

Introduction **5**
Le contexte des grands réseaux d'interactions 5
L'effet petit monde et la navigabilité 8
Problématiques des petits mondes navigables 10
Modèles pour les réseaux d'interactions 11
Organisation du document 14

1 Préliminaires : modèle de Kleinberg et principales définitions **17**
1.1 Formalisme des graphes 17
1.2 Routage décentralisé dans un graphe augmenté 18
1.3 Le modèle de petit monde de J. Kleinberg (2000) 20
1.4 Classe des réseaux de Kleinberg 21

I Aspect dynamique de l'effet petit monde **25**

2 Algorithmes de routage efficaces et graphes petits mondes **27**
Introduction 27
2.1 L'algorithme glouton de Kleinberg 27
2.2 Améliorer l'efficacité du routage grâce à une exploration restreinte 28
2.2.1 Compromis entre le recoupement et la profondeur d'exploration 30
2.2.2 *Lien valide* et *zone de sécurité* 31
2.3 Notre algorithme de routage 32
2.4 Analyse de l'algorithme en dimension 1 35
2.4.1 Analyse d'un processus de branchement générique 38
2.4.2 Analyse de l'algorithme 41
2.5 Généralisation à toute dimension 50
2.5.1 Lemmes géométriques préliminaires 51
2.5.2 Analyse de l'algorithme 54
2.6 Remarques générales sur les performances et applications 60

3 Charge du réseau **63**
Introduction 63
3.1 Charge et espérance du nombre de noeuds visités 65
3.1.1 Espérance du nombre de noeuds visités 66
3.1.2 Une première estimation de la charge 66
3.2 Charge de deux algorithmes décentralisés 69
3.2.1 Charge induite par l'algorithme glouton 71
3.2.1.1 Distance critique source-cible 71
3.2.1.2 Probabilité de visite d'un noeud 72
3.2.1.3 Charge d'un lien de longueur $\ell \geqslant 1$. 76
3.2.2 Analyse de l'algorithme d'exploration 79
3.2.2.1 Charge d'un lien longue-distance pour une source et une cible données 80
3.2.2.2 Charge d'un lien longue-distance de longueur ℓ 84
3.3 Autres paramètres de comparaison 86
3.3.1 Taux moyen de liens locaux utilisés dans un routage . . . 87
3.3.2 Comparaison des longueurs de chemin 91
3.4 Conclusion et perspectives 93

II Aspect structurel de l'effet petit monde **97**

4 Problématique de la *petit-mondisation* **99**
Introduction 99
4.1 Vers une généralisation du modèle de Kleinberg 100
4.1.1 Les généralisations pour les réseaux sociaux 100
4.1.2 Caractéristiques et limites structurelles du modèle 101
4.1.3 Extension aux grilles infinies 105
4.2 Métriques de graphes et petits mondes 108
4.2.1 Graphes *petit-mondisables* 109

5 Processus de *petit-mondisation* **113**
Introduction 113
5.1 Graphes à croissance modérée 113
5.2 *Petit-mondisation* des graphes à croissance modérée 115
5.3 Une caractérisation non exhaustive des graphes *petit-mondisables* 122
5.4 D'autres distributions sont-elles possibles ? 123
5.5 Un modèle qui peut engendrer (presque) toutes les distributions de degrés 126
5.6 Une autre approche : la *petit-mondisation* par plongement 127
5.7 Discussion 132

6 *Petit-mondisation* distribuée **135**
Introduction 135
6.1 Application aux réseaux informatiques décentralisés 136
6.2 Modèle étudié 137
6.3 *Petit-mondisation* par échantillonnage multi-niveaux 137
6.3.1 Le processus de *petit-mondisation* distribué 138
6.3.2 Analyse de l'algorithme de routage glouton dans le graphe augmenté 141
6.4 Implémentations décentralisées et performances 142
6.4.1 L'étape d'exploration de LEADER 143
6.4.2 L'étape des requêtes de contact longue-distance 145
6.4.3 Performances 150
6.5 Conclusion et perspectives 151

Conclusion générale **153**

Index **161**

Bibliographie **163**

Bibliographie personnelle **169**

Introduction

Cette thèse est consacrée à l'étude des aspects algorithmiques du routage dans les grands réseaux d'interactions présentant la propriété petit monde. Dans cette introduction, nous présentons le contexte de cette étude ainsi que les problématiques que nous avons choisi de développer.

Le contexte des grands réseaux d'interactions

On appelle *réseau d'interactions* tout ensemble d'entités interagissant de façon individuelle. Les grands réseaux d'interactions recouvrent ainsi des réseaux aussi divers que le réseau des routeurs d'Internet, le réseau des contacts sociaux entre individus, ou le réseau des réactions chimiques entre protéines dans le métabolisme d'un être vivant. On parlera également de *réseaux réels*. Alors que les interactions locales sont généralement bien connues (la communication entre deux routeurs, la réaction entre deux protéines), le résultat *global* de l'ensemble des interactions est encore mal compris. La compréhension de ces propriétés globales touche pourtant à des problématiques essentielles : la dynamique des interactions dans un réseau social ou un réseau informatique est par exemple liée à la problématique de la propagation des virus (informatiques ou biologiques), celle d'un réseau de distribution d'électricité, au problème de la robustesse d'un grand réseau. L'augmentation récente des capacités de traitement et de collecte d'un grand nombre de données statistiques sur ces réseaux a permis l'essor des études de ces objets. En particulier, on a observé expérimentalement que ces réseaux, *a priori* éloignés, partageaient des propriétés macroscopiques communes. Une de ces propriétés est l'*effet petit monde* qui est l'objet principal de cette thèse.

L'étude conjointe des propriétés macroscopiques de différents types de réseaux contient plusieurs problématiques. Pour donner un sens à une propriété observée sur plusieurs ensembles de données statistiques, il faut s'assurer de sa pertinence. En particulier, il faut garantir que la méthode de collecte des données n'est pas biaisée et que la propriété est spécifique, c'est-à-dire qu'on ne l'observe pas sur tout ensemble de données arbitraire, ou aléatoire uniforme. Une fois

qu'une propriété pertinente est isolée se pose la question de son apparition dans le réseau : est-ce dû à un processus dynamique de construction du réseau ? Comment peut-on la reproduire ? Révèle-telle un principe général sous-jacent ? Les réponses à ces questions peuvent avoir des applications intéressantes si la propriété présente un avantage pratique (on peut par exemple penser aux mécanismes de diffusion de rumeurs dans un réseau social pour une entreprise de marketing). Pour résumer, les études sur les grands réseaux d'interactions ont trois objectifs principaux :

- isoler les propriétés communes pertinentes,
- les reproduire,
- et les expliquer.

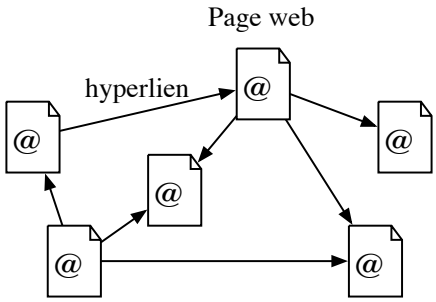


FIG. 1 : Exemple de grand réseau d'interactions : le réseau des pages web qui contient plusieurs millions de noeuds [BKM+00].

Nous en venons à présent aux trois principales propriétés qui ont été observées expérimentalement : le petit diamètre, la distribution des degrés suivant une loi de puissance et la forte densité locale ou *clustering*.

1. Petit diamètre. Le diamètre d'un réseau est formellement le plus long des plus courts chemins entre deux entités, ou *noeuds*, du réseau, via ses connexions. Dans l'étude des grands réseaux d'interactions, il est souvent fait référence à un petit diamètre pour décrire l'observation d'une petite distance moyenne, le diamètre étant généralement trop coûteux à obtenir (puisqu'il s'agit d'un pire cas). Dans de nombreux réseaux d'interactions, la distance moyenne observée est, de façon surprenante, de l'ordre du logarithme du nombre total de noeuds, alors que le nombre de connexions, ou *liens*, reste très inférieur au carré du nombre de noeuds. Newman [New01] donne l'exemple d'un réseau de co-auteurs d'articles de biologie de 1 520 251 noeuds et 11 803 064 liens, dans lequel la longueur moyenne d'un chemin est 4,9. Un autre exemple est l'étude du réseau Internet de 10 597 noeuds et 31 992 liens effectuée par Faloutsos *et al.* [FFF99], où la longueur moyenne d'un chemin est de 3,3.

Si les chemins sont très courts par rapport à la densité de liens des réseaux, la pertinence de cette propriété peut toutefois être remise en cause lorsque l'on remarque qu'un réseau aléatoire uniforme, où chaque paire de noeuds est reliée indépendamment avec une probabilité fixée, présente également un diamètre logarithmique en son nombre de noeuds lorsque le nombre d'arête est très faible (de l'ordre de $n \log n$ pour n noeuds).

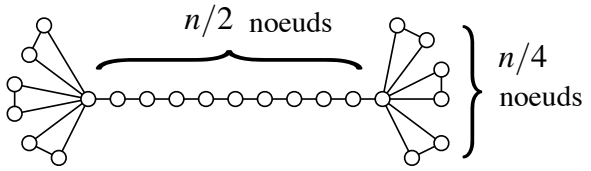
2. Distribution des degrés suivant une loi de puissance. On dit qu'un réseau présente une *distribution des degrés* suivant une loi de puissance si le nombre de noeuds de degré k est proportionnel à $1/k^\alpha$, pour une constante $\alpha > 0$, sur un intervalle de plusieurs ordres de grandeur (par exemple entre $k = 10$ et $k = 10^6$). En 1999, Faloutsos *et al.* [FFF99] ont observé que le réseau Internet présentait cette propriété. Par la suite, elle a également été observée dans des réseaux de pages web, et des réseaux de distribution d'électricité [New03]. Cette découverte a été cruciale pour les travaux sur la propagation des virus dans les réseaux réels. Avant cette découverte, le modèle usuel pour ce type d'étude était un réseau aléatoire uniforme, sur lequel on observe un effet de seuil pour la transmission d'un virus, c'est-à-dire qu'en dessous d'une fraction d'individus infectés, le virus cesse de se répandre. Mais une étude similaire menée sur un modèle présentant une distribution de degrés en loi de puissance a donné des résultats différents [PSV00], en particulier l'effet de seuil disparaît. Une telle observation a donc remis en cause les mécanismes mis en place pour freiner la propagation des virus et les modèles utilisés jusqu'alors.

Récemment, les travaux d'Achlioptas *et al.* [ACKM05] ont toutefois mis en doute la pertinence de la loi de puissance comme propriété caractéristique, en mettant en évidence un risque de biais lié à la méthode de collecte des données. Ils démontrent qu'en parcourant un réseau aléatoire uniforme selon un processus similaire à celui utilisé pour les parcours expérimentaux du réseau des routeurs Internet (le parcours par *traceroute* à partir d'un faible nombre de sources), on obtient des données statistiques présentant une distribution des degrés suivant une loi de puissance. Il est pourtant connu que le réseau aléatoire uniforme présente une distribution des degrés suivant une loi de Poisson [ER59]. Cette étude souligne le risque de baser la construction de modèles sur une propriété qui pourrait n'être qu'un biais de la mesure expérimentale.

3. Forte densité locale ou *clustering*. On parle de forte densité locale d'un réseau lorsque les voisins d'un même noeud sont très reliés entre eux. Dans un réseau social, par exemple, cela signifie que les amis d'un même individu ont une grande probabilité d'être amis entre eux. Pour quantifier cette propriété, Watts et Strogatz ont introduit en 1998 la notion formelle de *coefficient de clustering* [WS98]. Il s'agit de la moyenne, sur tous les noeuds $\mathbf{u}$, du ratio du nombre de voisins de $\mathbf{u}$

qui sont reliés entre eux sur le nombre total de liens qui pourraient potentiellement exister entre ces voisins. Ils observent que ce coefficient s'élève à 0,2 dans un réseau de collaborations entre acteurs de cinéma comprenant 449 913 noeuds et 25 516 582 liens, alors qu'en construisant un réseau aléatoire uniforme ayant le même nombre de noeuds et de liens, on obtient un coefficient de l'ordre de 10^{-4} seulement.

Toutefois, cette propriété semble encore mal définie par le coefficient de *clustering*. Il est en effet possible de construire des réseaux ayant un fort coefficient de *clustering* alors que, majoritairement, les noeuds n'ont pas de voisins reliés entre eux. Un exemple est illustré sur la figure ci-dessous, où le coefficient de *clustering* vaut environ $1/2$ alors que $n/2$ noeuds sont disposés en chaîne et n'ont aucun de leurs voisins reliés entre eux.



Ces trois propriétés contiennent des éléments essentiels aux problématiques de diffusion dans les réseaux, toutefois leur pertinence n'est pas assurée. Dans cette thèse, nous étudions l'effet petit monde, décrit dans la section suivante, qui inclut les propriétés de petit diamètre et forte densité locale, et possède une propriété pertinente supplémentaire : le routage efficace, ou *navigabilité*. On pourra se reporter à [AB02] et [New03] pour une vue d'ensemble des études sur les grands réseaux d'interactions jusqu'en 2002 et 2003 respectivement.

L'effet petit monde et la navigabilité

L'effet petit monde tient son nom de l'expression populaire « le monde est petit »désignant la surprise de constater que deux connaissances d'un même individu, *a priori* sans rapport, se connaissent entre elles. L'étude des relations sociales en tant que réseau d'interactions date des années 30, en particulier la création par le psychologue Jacob-Levy Moreno de la sociométrie. Celle-ci a pour but la mesure objective des relations sociales au sein d'un groupe. C'est dans une période ultérieure de développement de la science des réseaux sociaux que les psychologues se sont intéressés à l'effet petit monde [TM69, dSPK78, KB78]. En particulier, Stanley Milgram a effectué une expérience sociologique en 1967 [Mil67] en demandant à 300 habitants du Nebraska et de Boston de faire parvenir une lettre à un habitant de Boston dont ils ne connaissaient que le lieu d'habitation et la profession, en ne la retransmettant qu'à une personne qu'ils connaissaient

personnellement (et en itérant le processus jusqu'à atteindre la personne cible). Relativement peu de lettres sont arrivées à destination (environ un quart), mais le résultat surprenant fut que la longueur moyenne d'une chaîne de porteurs du message de son origine à sa destination était très faible (5,2) en regard du nombre d'individus et de leur éloignement géographique et social. Cette expérience a par ailleurs été reproduite récemment par Dodds *et al.* [DMW03] sur 60 000 individus, échangeant cette fois des *e-mails*, et a aboutit à des chaînes de messages de longueur moyenne 4,1 entre individus de continents différents. On pourrait penser qu'il n'est pas vraiment surprenant que les chaînes soient si courtes, puisqu'en supposant que chaque individu ait seulement 10 connaissances, chacun pourrait *a priori* atteindre 10^6 individus à distance 6. Mais les réseaux sociaux présentent un fort coefficient de *clustering*, il est donc probable que les connaissances immédiates d'un individu n'aient qu'une seule connaissance étrangère à ce voisinage, et que l'on atteigne seulement 11 individus à distance 2 par exemple, c'est pourquoi cette propriété est remarquable.

La notion de petit monde n'a pas aujourd'hui de définition formelle ; elle est définie, dans certains articles ([WS98]), comme la combinaison d'un fort coefficient de *clustering* et d'un petit diamètre. L'objet de cette thèse est d'approfondir l'explication dynamique et algorithmique de ce phénomène proposée par Kleinberg en 2000 [Kle00] qui y introduisit la notion de *navigabilité*. Il s'agit de caractériser non seulement la longueur des chemins, mais aussi la façon dont ils sont découverts. Dans l'expérience de Milgram, les individus n'utilisent que leurs contacts locaux pour renvoyer la lettre ; il s'agit donc d'un routage *décentralisé*, en ce sens que l'on n'utilise qu'une vue locale du réseau pour transmettre le message. C'était aussi le cas de la navigation à travers le réseau des pages web il y a quelques années[1] , qui se faisait d'une page à l'autre sans connaître la carte globale du réseau [AJB99, Kai99]. Par ailleurs, la découverte des chemins de façon décentralisée est une nécessité pour les réseaux d'interactions réels qui comportent un très grand nombre de sommets et où une recherche classique des plus courts chemins n'est pas possible, car très coûteuse en temps.

La propriété de navigabilité de l'effet petit monde est une caractérisation plus spécifique que la propriété de petit diamètre utilisée jusqu'alors, pour le définir. En particulier, même si un réseau aléatoire uniforme présente un diamètre logarithmique en le nombre de noeuds, Kleinberg montre qu'il n'existe aucun algorithme décentralisé qui puisse découvrir ces chemins courts (*i.e.* polylogarithmiques) [Kle00]. On appellera par la suite *petit monde navigable* tout réseau dont le diamètre est polylogarithmique en le nombre de noeuds, et dont les chemins polylogarithmiques peuvent être calculés par un algorithme décentralisé. On parlera

[1] Aujourd'hui, l'utilisation massive de moteurs de recherche comme *Google* pour accéder aux pages web nous amène néanmoins à nuancer ce fait.

simplement de *petit monde* lorsqu'il n'y a pas d'ambiguïté.

Problématiques des petits mondes navigables

En étudiant l'effet petit monde sous l'angle algorithmique de la navigabilité, nous abordons le problème de la compréhension d'un phénomène réel qui contient des problématiques fondamentales pour l'informatique.

Compréhension de l'effet petit monde. La compréhension du phénomène petit monde observé dans les réseaux sociaux contient deux questions essentielles :

- quel est le type d'algorithme de routage décentralisé utilisé pour transmettre un message dans un réseau social ?
- existe-t-il une structure sous-jacente particulière qui permet la construction d'un réseau petit monde navigable ?

Ces questions sont intimement liées puisque l'algorithme de routage utilise la structure du réseau pour naviguer. Dans cette thèse, nous allons décrire ces deux aspects du problème comme l'aspect dynamique et l'aspect structurel de l'effet petit monde :

- d'une part, un réseau petit monde navigable doit avoir une structure qui présente un faible diamètre et qui permet de stocker de manière locale les informations requises pour évaluer les positions relatives des noeuds (aspect structurel) ;
- d'autre part, il doit exister un processus de routage décentralisé qui peut en tirer parti (aspect dynamique).

Nous traiterons l'aspect dynamique dans la première partie de ce document et l'aspect structurel dans la seconde.

Problématiques informatiques des petits mondes. Ces dernières années, le réseau Internet a connu un changement d'échelle rendant irréalistes des algorithmes de routage centralisés. Le développement des algorithmes décentralisés qui n'utilisent qu'une vue locale du réseau a permis de passer à grande échelle et de créer de nouveaux grands réseaux informatiques totalement décentralisés, comme les réseaux pair-à-pair. Un protocole pair-à-pair est un protocole d'échange de fichiers entre ordinateurs jouant tous le même rôle, c'est-à-dire qu'il n'y a pas, en général, de serveur central (on peut citer les réseaux CAN [RFH+01] et Chord [SMK+01]). Outre la propriété de passage à l'échelle, un tel réseau présente également une bonne résistance aux pannes et aux attaques ciblées, contrairement à un réseau centralisé, puisque la mise hors service d'un des ordinateurs du réseau a peu d'influence sur son fonctionnement global. Par ailleurs, les fichiers échangés sur les réseaux pair-à-pair sont souvent de grande taille (fi-

chiers vidéo), il est donc nécessaire de construire des algorithmes dédiés qui sont décentralisés *et* calculent des chemins courts. La spécificité du routage dans les grands réseaux informatiques décentralisés et dans les petits mondes est donc similaire ; nous verrons dans cette thèse que cette symétrie permet un va-et-vient très riche entre les problèmes et applications de l'un et l'autre domaine. On distingue deux problématiques principales de l'effet petit monde pour l'informatique :

- construire de nouveaux algorithmes décentralisés efficaces, dédiés aux grands réseaux ;
- construire de nouvelles architectures de grands réseaux de petit diamètre, dont les chemins courts peuvent être calculés de façon décentralisée.

Modèles pour les réseaux d'interactions

Dans la première partie de cette thèse, nous commençons par baser notre étude de l'aspect dynamique de l'effet petit monde sur le seul modèle de petit monde navigable existant au début de cette thèse : le modèle de Kleinberg. Dans la deuxième partie, nous nous employons avec succès à définir un modèle plus général. De nombreux modèles ont été introduits pour représenter les réseaux d'interactions, ou certaines de leur propriétés statistiques, et ont inspiré le modèle de Kleinberg. Nous rappelons les principaux dans cette section.

Un réseau se représente de façon naturelle par un *graphe*. Nous utiliserons dans toute la suite l'un ou l'autre vocabulaire de façon indifférenciée : un *noeud* correspond à un *sommet* et un *lien* à une *arête*. On trouvera la définition formelle d'un graphe dans le préambule.

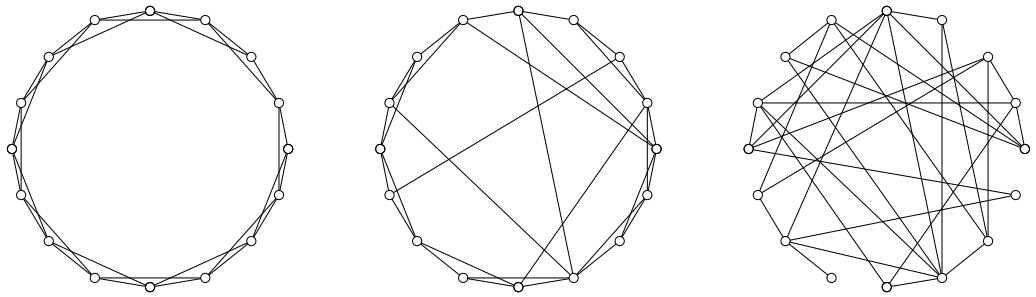
Graphe aléatoire uniforme d'Erdös et Rényi. Le premier graphe étudié comme un modèle possible pour les réseaux réels fut le graphe aléatoire $G(n, p)$ d'Erdös et Rényi [ER59]. Il s'agit d'un graphe aléatoire uniforme sur n sommets où il existe une arête entre deux sommets avec une probabilité p constante. Cet objet mathématique a été très étudié, on pourra se référer à [Bol85] pour une vue d'ensemble des résultats. Nous nous intéressons seulement ici à son rôle historique en tant que modèle pour les réseaux d'interactions, car il était considéré jusqu'à récemment comme le seul modèle, par défaut, pour les réseaux réels. Or les récentes observations expérimentales (qui étaient impossibles ou très partielles auparavant) ont mis en lumière d'importantes différences. Ainsi, la distribution des degrés suit une loi de Poisson (exponentielle), alors qu'il s'agit d'une loi de puissance pour la grande majorité des réseaux réels. Il s'agit là d'une différence importante puisqu'elle caractérise l'hétérogénéité du réseau. Les noeuds jouant tous le même rôle dans un graphe d'Erdös-Rényi, le graphe ne

possède pas de propriété discriminante, et les degrés sont naturellement répartis de façon égale autour de la moyenne. Pour la même raison, ce graphe présente un faible coefficient de *clustering*, puisque les voisins d'un noeud n'ont aucune raison d'être davantage reliés entre eux que deux noeuds pris au hasard. Enfin, même si ce graphe présente un diamètre polylogarithmique en le nombre de sommets, on peut montrer que tout algorithme décentralisé y calcule des chemins de longueur au moins polynomiale [Kle00] ; ce n'est donc pas un petit monde navigable. Ces différences ont montré l'importance de trouver un modèle plus fidèle.

Modèle d'Albert et Barabasi pour la distribution des degrés. Suite à la découverte de distributions de degrés suivant une loi de puissance dans de nombreux réseaux réels, la construction de nouveaux modèles a été dirigée vers la reproduction de cette propriété. En 1999, Albert et Barabasi [BA99] ont popularisé un modèle dynamique permettant d'obtenir une distribution des degrés suivant une loi de puissance. Ce modèle consiste à construire un réseau noeud par noeud, en reliant chaque nouveau noeud préférentiellement aux sommets existants de plus hauts degrés. Il est connu sous le nom de l'*attachement préférentiel*. Des processus similaires avaient été introduits dès les années 20 par des mathématiciens [EP23, Yul24, Zip49], puis étudiés en sociologie [Sim55, dSP76], mais ce fut la première étude de ce processus en tant que modèle d'un phénomène physique. L'intérêt de ce modèle est sa construction dynamique, puisque dans de nombreux réseaux réels, des noeuds et des liens sont fréquemment ajoutés et enlevés au cours du temps (on peut penser au réseau des pages web par exemple).

Nous présentons maintenant les deux principaux modèles qui ont été développés spécifiquement pour reproduire l'effet petit monde.

Modèle de petit monde non navigable de Watts et Strogatz. En introduisant la notion formelle de coefficient de *clustering*, Watts et Strogatz [WS98] ont proposé un modèle qui présente à la fois un petit diamètre et un fort coefficient de *clustering*. Une variante du modèle a également été développée et analysée par Newman et Watts [NW99]. Le modèle est construit de la façon suivante : à partir d'un anneau régulier de n sommets et k arêtes par sommet, distribuées régulièrement par rapport à leur origine, on redirige indépendamment chaque extrémité d'arête avec une probabilité p constante, donnée en paramètre, vers un sommet de l'anneau choisi de manière aléatoire uniforme. La figure 2 illustre ce modèle pour $p = 0$, $p = 1$ et une valeur intermédiaire $0 < p < 1$ qui donne lieu à l'apparition des deux propriétés de petit diamètre et fort *clustering* simultanément. Le nombre k d'arêtes de départ n'influe pas sur le modèle.



p=0
Grand diamètre
et fort *clustering*.

0<p<1
Petit diamètre
et fort *clustering*.

p=1
Petit diamètre
et faible *clustering*.

FIG. 2 : Modèle de Watts et Strogatz

En cherchant à reproduire à la fois le petit diamètre présent dans les graphes aléatoires uniformes et le fort coefficient de *clustering* présent dans les grilles régulières, Watts et Strogatz ont eu l'idée d'introduire une part d'aléatoire sur un réseau régulier et ils ont obtenu les deux propriétés simultanées pour un assez grand intervalle de valeurs du paramètre p. Toutefois, ce modèle ne présente pas la propriété dynamique de navigabilité. Précisément, Kleinberg [Kle00] a montré que tout algorithme de routage décentralisé calcule, entre deux sommets de ce graphe, un chemin de longueur au moins polynomiale en n. Cela signifie que, même lorsque le paramètre p donne naissance à des chemins polylogarithmiques en n entre toute paire de sommets, il n'existe pas d'algorithme qui puisse les découvrir en n'ayant qu'une vue locale du graphe.

Modèle de petit monde navigable de Kleinberg. Le premier modèle de petit monde présentant la propriété dynamique de navigabilité a été introduit par Kleinberg en 2000 [Kle00]. Il s'agit d'une grille régulière de dimension 2, augmentée d'un arc aléatoire par sommet $\mathbf{u}$, dont la destination est $\mathbf{v}$ avec probabilité proportionnelle à $1/|\mathbf{u}-\mathbf{v}|^s$, pour chaque $\mathbf{v}$, où $|\mathbf{u}-\mathbf{v}|$ est la distance ℓ_1 entre les deux noeuds dans la grille régulière, et $s > 0$ un paramètre[2]. La principale différence avec le modèle de Watts et Strogatz est la distribution non uniforme des liens aléatoires, elle est en effet fortement liée aux positions des noeuds sur la grille sous-jacente. Kleinberg montre que lorsque $s = 2$, un algorithme de routage glouton, très simple et décentralisé, calcule un chemin de longueur polylogarithmique en la taille du graphe ($O(\log^2 n)$ pour n noeuds) entre toute paire de sommets. Nous étudierons l'optimalité de ce routage dans la première partie.

[2]On trouvera une définition détaillée de ce modèle dans le préambule, page 22

Ce modèle met en valeur deux types d'informations contenues dans un réseau : une information globale, mais partielle, partagée par tous (représentée par la grille) ; et une information locale qui permet la navigation (représentée par les liens aléatoires). Un algorithme est alors dit décentralisé s'il calcule des chemins en connaissant les positions de tous les noeuds sur la grille, mais en n'accédant à la position de la destination d'un lien aléatoire qu'en visitant son origine. Kleinberg propose une analogie avec un réseau social, où la grille serait l'ensemble des positions géographiques des individus et les liens aléatoires les connaissances amicales d'un individu dont lui seul connaît les positions. Comme les liens supplémentaires sont distribués selon une loi reliée à la structure sous-jacente de l'information globale, ils *stockent* cette information dans un certain sens, puisque les décisions de routage, prises en fonction des liens aléatoires, sont corrélées à la position des noeuds dans la structure sous-jacente. Il s'agit d'une caractéristique structurelle de l'effet petit monde pour ce modèle que nous étudierons dans la deuxième partie.

Un résultat récent de Clauset et Moore [CM03] vient par ailleurs renforcer la pertinence du modèle de Kleinberg et de sa distribution. Clauset et Moore ont simulé un processus de reconnexion dynamique des liens dans l'anneau que l'on peut comparer, par exemple, au comportement d'un internaute qui créerait un nouveau marque-page vers une page web lorsque la recherche d'une page a pris un temps supérieur à un certain seuil. Ils ont alors observé expérimentalement que la distribution des liens convergeait vers une loi harmonique (la distribution des arcs aléatoires du modèle de Kleinberg pour l'anneau) tandis que l'espérance de longueur des chemins calculés par l'algorithme de routage glouton convergeait vers $O(\log^2 n)$, pour n noeuds.

Du point de vue informatique, le modèle de Kleinberg présente des caractéristiques algorithmiques intéressantes que ne présentaient pas les modèles précédents (modèles d'Erdös et Rényi, et de Watts et Strogatz). Ajoutons que si le modèle de Kleinberg reproduit bien la propriété de petit monde navigable, une critique fréquente est qu'il ne reproduit pas la distribution des degrés observée sur les réseaux réels, puisque chaque sommet a un degré sortant constant.

Organisation du document

Cette thèse se structure en trois parties : un premier chapitre de préliminaires qui donne les définitions et propriétés des principaux objets utilisés dans la suite du document, une première partie consacrée à l'aspect dynamique de l'effet petit monde et une seconde partie consacrée à son aspect structurel.

La première partie est divisée en deux chapitres. Dans le chapitre 2, nous étudions l'efficacité des algorithmes de routage existant dans le cadre du modèle

de Kleinberg original et nous présentons un nouvel algorithme de routage décentralisé qui reste le meilleur, à notre connaissance, en terme de longueur de chemin. Nous étudions également la pertinence des différents algorithmes sous l'angle de la compréhension de l'effet petit monde dans un réseau social. Après avoir étudié les améliorations du routage décentralisé en terme de longueur de chemin, nous analysons, dans le chapitre suivant, leurs conséquences sur la charge du réseau. Cette étude accentue l'importance du compromis entre la charge et la longueur du chemin dans un tel réseau et nous permet d'obtenir un moyen de discriminer deux types d'algorithmes via l'observation de la charge qu'ils induisent sur les liens les plus longs.

Dans la deuxième partie, nous tentons d'extraire du modèle étudié jusqu'ici les caractéristiques intrinsèques qui permettent l'émergence de la navigabilité, en particulier les propriétés de la structure sous-jacente. Dans le chapitre 5, nous proposons des modèles de petits mondes navigables qui généralisent celui de Kleinberg. Pour cela, nous présentons deux processus d'augmentation d'une grande classe de graphes en graphes petits mondes par l'ajout d'un arc aléatoire par noeud : l'un basé la croissance des boules du graphes, et l'autre sur l'existence d'un plongement de faible distorsion du graphe dans un espace euclidien de faible dimension. Dans le chapitre 6, nous étendons notre étude en exhibant un processus entièrement distribué qui permet d'augmenter un graphe en un petit monde, en ajoutant un seul lien par noeud. Ces résultats mettent en valeur l'importance de la création de raccourcis dans le réseau pour l'émergence de l'effet petit monde, tandis que la métrique sous-jacente semble jouer un rôle moins déterminant.

Les résultats présentés dans cette thèse sont pour la plupart issus de collaborations avec d'autres chercheurs : Philippe Duchon, Nicolas Hanusse, Cristopher Moore et Nicolas Schabanel. Ces résultats ont fait l'objet de publications dans des conférences et journaux internationaux : [1, 2] pour le chapitre 2 et [3, 4] pour le chapitre 5. Le chapitre 6 a fait l'objet d'un article soumis pour publication [7].

Pour la cohérence de ce document, j'ai choisi de présenter ici mon travail de thèse portant sur les aspects algorithmique des petits mondes. J'ai par ailleurs travaillé durant ma thèse sur les problèmes de graphes sandwichs, dont la problématique est disjointe de celle de ce document. Il s'agit d'un problème central en algorithmique des graphes, généralisant les problèmes de complétion et de reconnaissance. Étant donnés un graphe G et un surgraphe G' de G, on cherche à compléter G de façon à ce qu'il satisfasse une propriété de graphe donnée en n'utilisant seulement les arêtes présentes dans le surgraphe G'. On appelle un graphe correspondant éventuel un graphe sandwich pour la propriété. En collaboration avec Michel Habib et Christophe Paul, nous avons montré que le problème de l'énumération des modules sandwichs pouvait être exponentiel [6], contrairement à ce qu'affirmait la littérature. D'autre part, nous avons étendu l'étude

des problèmes sandwichs aux ordres partiels. En collaboration avec Michel Habib, David Kelly et Christophe Paul, nous avons donné des algorithmes polynomiaux pour certaines classes d'ordres partiels, et démontré la NP-complétude du problème sandwich pour d'autres classes [5]. Cela a mis en évidence qu'il n'existe pas de relation de complexité directe entre les problèmes sandwich non orientés et les problèmes d'ordres partiel sandwich correspondants.

Chapitre 1

Préliminaires : modèle de Kleinberg et principales définitions

Ce chapitre donne les définitions formelles des objets et notions étudiés dans la suite du document. Il présente les principaux résultats pré-existants sur le modèle de petit monde de Kleinberg.

1.1 Formalisme des graphes

Nous rappelons ici la définition usuelle d'un graphe, naturellement utilisé pour représenter un réseau.

Définition 1.1 (Graphe non orienté)
Un graphe non orienté *est un couple* (V, E) *où* V *est un ensemble d'objets, appelés* sommets *du graphe, et* E *est un sous-ensemble de* $\mathcal{P}_2(V)$*, l'ensemble des parties à deux éléments de* V*, dont les éléments sont appelés* arêtes *du graphe. On note* $\{\mathbf{u}, \mathbf{v}\}$ *l'arête entre deux sommets* $\mathbf{u}, \mathbf{v} \in V$ *lorsque* $\{\mathbf{u}, \mathbf{v}\} \in E$*.*

Dans de nombreux réseaux, les connexions entre les noeuds ne sont pas symétriques mais *orientées*. On peut citer l'exemple du graphe des pages web où une page populaire (celle du groupe The Beatles) est pointée par d'autres pages web (celles d'amateurs des Beatles) sans avoir de liens réciproques vers ces pages. Dans notre étude du routage dans les grands réseaux d'interactions, un chemin de routage entre deux noeuds n'est valide qu'à la condition que chaque connexion d'un noeud vers un autre soit orientée dans le bon sens (on ne peut pas accéder à la page web d'un amateur des Beatles à partir de la page web des Beatles, à moins d'y être arrivé via la page web de cet amateur).

Définition 1.2 (Graphe orienté)
Un graphe orienté *est un couple* (V, E) *où* V *est l'ensemble des sommets du graphe et* E *est un sous-ensemble de* V^2 *dont les éléments sont appelés* arcs *du graphe. On note* $(\mathbf{u}, \mathbf{v})$ *l'arc du sommet* $\mathbf{u}$ *vers le sommet* $\mathbf{v}$ *lorsque* $(\mathbf{u}, \mathbf{v}) \in E$.

On remarquera que tout graphe non orienté peut être transformé en un graphe orienté en remplaçant chaque arête $\{\mathbf{u}, \mathbf{v}\}$ par deux arcs symétriques $(\mathbf{u}, \mathbf{v})$ et $(\mathbf{v}, \mathbf{u})$. Dans toute la suite, on ne considérera que des graphes orientés.

Les définitions suivantes nous seront utiles pour formaliser le routage dans un réseau.

Définition 1.3 (Voisin ou contact)
Étant donné un sommet $\mathbf{u}$ *d'un graphe orienté* $G = (V, E)$*, tout sommet* $\mathbf{v} \in V$ *tel que* $(\mathbf{u}, \mathbf{v}) \in E$ *est appelé* voisin *ou* contact *de* $\mathbf{u}$.

Définition 1.4 (Chemin)
Étant donné un graphe orienté $G = (V, E)$ *et deux sommets* $\mathbf{u}, \mathbf{v} \in V$*, on appelle* chemin *de* $\mathbf{u}$ *vers* $\mathbf{v}$ *de longueur* ℓ *toute suite* $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}, \mathbf{x}_\ell$ *où* $\mathbf{x}_0 = \mathbf{u}$*,* $\mathbf{x}_\ell = \mathbf{v}$ *et pour tout* $0 \leqslant i \leqslant \ell - 1$*,* $(\mathbf{x}_i, \mathbf{x}_{i+1}) \in E$.

Définition 1.5 (Distance)
Étant donné un graphe orienté $G = (V, E)$ *et deux sommets* $\mathbf{u}, \mathbf{v} \in V$*, on appelle* distance *de* $\mathbf{u}$ *à* $\mathbf{v}$ *la longueur du plus court chemin de* $\mathbf{u}$ *vers* $\mathbf{v}$.

Définition 1.6 (Diamètre)
Le diamètre *d'un graphe* G *est la plus grande des distances entre les paires de sommets de* G.

Enfin, nous ne considérerons que des graphes connexes dans toute la suite, c'est-à-dire pour lesquels il existe un chemin entre tout couple de sommet. En effet, l'étude d'algorithmes de routage dans un graphe non connexe peut se faire indépendamment sur chaque composante connexe, c'est-à-dire chaque sous-ensemble de sommets appartenant à la même classe d'équivalence de la relation « est relié par un chemin à ».

1.2 Routage décentralisé dans un graphe augmenté

Dans cette section, nous définissons formellement la notion de routage décentralisé. Contrairement au routage centralisé, le routage décentralisé utilise seulement la connaissance locale de chaque noeud visité pour construire un chemin. Au sens fort, cette connaissance locale est réduite à l'ensemble des contacts du noeud. On peut cependant étendre cette connaissance dans le cas où le réseau

possède deux ensembles de liens, en s'inspirant du modèle de Kleinberg. Le premier ensemble de liens est supposé connu de tous les noeuds tandis que pour l'autre ensemble, chaque lien n'est connu que de son origine. On parle alors de *graphe augmenté*. Nous donnons la définition d'un algorithme décentralisé dans le cadre de ce type de graphe.

Définition 1.7 (Graphe augmenté)
Un graphe augmenté $G = (V, E, E')$ *est un graphe obtenu à partir d'un graphe* $H = (V, E)$*, en ajoutant un ensemble d'arêtes supplémentaires* E' *sur* V*.*
La distance sous-jacente *de* $\mathbf{u}$ *à* $\mathbf{v}$ *dans* G *est la distance de* $\mathbf{u}$ *à* $\mathbf{v}$ *dans* H*.*

Définition 1.8 (Algorithme de routage)
Un algorithme de routage $\mathcal{A}$ *dans un graphe* (V, E) *reçoit en entrée la donnée d'une* source $\mathbf{s} \in V$ *et d'une* cible *(ou* destination*)* $\mathbf{t} \in V$ *et renvoie en sortie un chemin* $\mathbf{P}$ *de* $\mathbf{s}$ *à* $\mathbf{t}$*.* $\mathbf{P}$ *est appelé* chemin de routage *de* $\mathbf{s}$ *à* $\mathbf{t}$*.*

L'efficacité d'un algorithme de routage décentralisé se mesure par deux paramètres principaux : la longueur des chemins calculés et la *latence* qui est le nombre de noeuds visités ou interrogés. Durant le calcul d'un chemin, l'algorithme peut en effet interroger des noeuds, c'est-à-dire acquérir les informations qu'ils contiennent sans que ces noeuds appartiennent au chemin renvoyé par l'algorithme (il peut par exemple s'agir d'une phase d'exploration du réseau).

Définition 1.9 (Latence)
La latence *d'un algorithme de routage sur un graphe donné est le plus grand nombre de noeuds interrogés pour calculer un chemin entre deux sommets.*

Dans le cadre du routage décentralisé, le graphe $H = (V, E)$ sous-jacent à un graphe augmenté $G = (V, E, E')$ représente la structure globale sur les noeuds qui est connue de tous, comme la position géographique des individus d'un réseau social. C'est relativement à cette structure que l'on repère la position des liens inconnus E' qui peuvent représenter les connaissances amicales du réseau social.

Dans toute la suite, on mesurera la distance entre les noeuds d'un graphe augmenté dans la structure sous-jacente. Le terme de *distance* sera donc toujours utilisé pour la distance sous-jacente s'il s'agit d'un graphe augmenté, et ne référera jamais à la distance dans le graphe augmenté qui est *a priori* inconnue.

Définition 1.10 (Algorithme de routage décentralisé)
Un algorithme de routage dans un graphe augmenté (V, E, E') *est* décentralisé *si et seulement s'il a pour seule connaissance le graphe sous-jacent* (V, E)*, la position de la cible sur* (V, E) *et peut seulement interroger les noeuds connus ou les voisins de noeuds connus.*

On peut illustrer cette définition avec l'expérience de Milgram. Supposons que le graphe sous-jacent représente la connaissance globale des positions géographiques des individus. Supposons par exemple que Germaine, de Paris, connaît Paul, de Liverpool, qu'elle a rencontré lors d'un concert. C'est une information *locale*, inconnue des autres individus. Durant la transmission d'une lettre dans le réseau social, on ne peut accéder à cette information qu'au moment où Germaine reçoit la lettre. Si la lettre est destinée à un Anglais, Germaine va sans doute choisir de la retransmettre à Paul, en utilisant l'information *globale* des positions entre individus selon laquelle Paul est *a priori* plus proche du destinataire car il habite Liverpool.

1.3 Le modèle de petit monde de J. Kleinberg (2000)

En 2000, Kleinberg [Kle00] propose le premier modèle de petit monde présentant la propriété de navigabilité, c'est-à-dire le premier modèle de graphe dont le diamètre est polylogarithmique en le nombre de noeuds *et* dont des chemins polylogarithmiques peuvent être découverts par un algorithme décentralisé entre tout couple de sommets.

Il s'agit d'une grille carrée de dimension 2 et de côté n que l'on augmente en ajoutant à tout sommet $\mathbf{u}$ un nombre $k > 0$ constant d'arcs aléatoires tirés indépendamment. La destination du j-ème arc de $\mathbf{u}$, pour $1 \leqslant j \leqslant k$, est $\mathbf{v}$ avec une probabilité proportionnelle à $1/|\mathbf{u} - \mathbf{v}|^s$, où $|\mathbf{u} - \mathbf{v}|$ est la distance ℓ_1 entre $\mathbf{u}$ et $\mathbf{v}$ dans la grille[1], et $s > 0$ est une constante. On remarquera que la grille est constituée de liens symétriques (ou arêtes) tandis que les liens (ou arcs) ajoutés aléatoirement sont orientés.

La figure 1.1 illustre un exemple de réseau de Kleinberg de dimension 2 ayant un lien aléatoire sortant par noeud (en gras). Les liens aléatoires ne sont pas tous représentés pour la lisibilité.

Kleinberg démontre que seul le cas où $s = 2$ permet d'obtenir la propriété de navigabilité. Il exhibe, pour ce cas, un algorithme de routage décentralisé qui calcule des chemins dont l'espérance de la longueur est $O(\log^2 n/k)$ entre tout couple de noeuds (on dit que la longueur est $O(\log^2 n/k)$ *en espérance*). Il s'agit de l'algorithme glouton $\mathcal{G}$ (algorithme 1) qui transmet, en chaque noeud, le message au voisin le plus proche de la cible du routage.

Ce résultat a été généralisé par la suite à toute grille de dimension $d > 0$ par Barrière *et al.* [BFKK01] qui démontrent que l'on obtient la propriété de navigabilité pour le seul exposant $s = d$. Ils démontrent également que l'espérance de la

[1]Dans une grille d-dimensionnelle, la distance ℓ_1 entre deux noeuds $\mathbf{x} = (x_1, \ldots, x_d)$ et $\mathbf{y} = (y_1, \ldots, y_d)$ est définie par : $|\mathbf{x} - \mathbf{y}| = \sum_{i=1}^{d} |x_i - y_i|$, elle est également appelée *distance de Manhattan*.

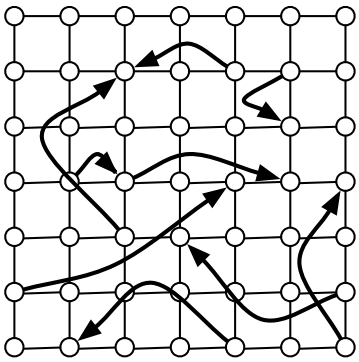


FIG. 1.1 : Exemple de réseau de Kleinberg $\mathcal{K}^2_{1,n}$.

Algorithme 1 Algorithme de routage glouton

Entrée : la source s et la cible $\mathbf{t}$.

Initialisation : $\mathbf{x} \leftarrow \mathrm{s}$. *($\mathbf{x}$ est le porteur courant du message)*

tant que $\mathbf{x} \neq \mathbf{t}$ **faire**

$\mathbf{x} \leftarrow \mathbf{y}$ où $\mathbf{y}$ est le contact local ou longue-distance de $\mathbf{x}$ qui minimise $|\mathbf{y}-\mathbf{t}|$ (choix aléatoire uniforme en cas d'égalité).

fin tant que

longueur du chemin construit par l'algorithme glouton $\mathcal{G}$ est $\Theta(\log^2 n/k)$.

Lorsque $s \neq d$, ils montrent que tout algorithme de routage décentralisé calcule, entre deux noeuds, un chemin de longueur au moins polynomiale en espérance en le nombre de sommets. Ce résultat démontre en particulier qu'aucune distribution uniforme ($s = 0$) n'est navigable sur le tore d-dimensionnel, c'est le cas par exemple des modèles d'Erdös-Rényi [ER59] et de Watts et Strogatz [WS98].

1.4 Classe des réseaux de Kleinberg

Dans cette section, nous définissons de façon précise la classe de graphes aléatoires des réseaux de Kleinberg navigables qui correspondent au modèle de Kleinberg où l'exposant s est pris égal à la dimension d.

Définition 1.11 (Tore régulier)
Un tore régulier *de dimension $d \geqslant 1$ et de côté $n \geqslant 1$ est un graphe dont l'ensemble des sommets est $\{-n, \ldots, 0, \ldots, n\}^d$ et tel qu'il existe une arête entre*

deux sommets $(x_1, \ldots, x_d)$ *et* $(y_1, \ldots, y_d)$ *si et seulement s'il existe une coordonnée* $1 \leqslant i \leqslant d$ *telle que* $x_i = y_i \pm 1 \mod 2n+1$, *et* $x_j = y_j$ *pour tout* $j \neq i$.

Définition 1.12 (Réseau de Kleinberg [Kle00])
Étant donnés $k, n, d \geqslant 1$, *un* réseau de Kleinberg *de dimension* d, *noté* $\mathcal{K}^d_{k,n}$, *est un tore régulier de dimension* d *et de côté* n *augmenté de* k *liens (ou arcs) longue-distance sortants par noeud, aléatoires et indépendants, tels que pour tout noeud* $\mathbf{u}$, *la probabilité que la destination de son* j*-ème lien longue-distance soit* $\mathbf{v}$ *est proportionnelle à* $1/|\mathbf{u} - \mathbf{v}|^d$.

On appelle *lien local* toute arête du tore sous-jacent. Sa destination est un *contact local* de son origine. L'ensemble des liens locaux représente l'information globale connue par tous les noeuds du réseau. Ces liens sont symétriques.

On appelle *lien longue-distance* tout arc aléatoire du modèle. Sa destination est un *contact longue-distance* de son origine. Un lien longue-distance représente une information partielle, sans carte globale. Ces liens sont orientés.

On note $L_j(\mathbf{u})$ le j-ème contact longue-distance d'un sommet $\mathbf{u}$.

Dans toute la suite, $\ln$ représente le logarithme népérien, $\log$ le logarithme en base 2 et H_n la somme harmonique jusqu'à l'indice n : $H_n = \sum_{i=1}^{n} 1/i$.

Distribution exacte des liens longue-distance. Pour déterminer de façon exacte la distribution des liens longue-distance dans un réseau de Kleinberg, nous devons en calculer le facteur de normalisation, défini ci-après.

Dans un réseau $\mathcal{K}^d_{n,k}$, étant donné un entier j, $1 \leqslant j \leqslant k$ et deux noeuds $\mathbf{u}$ et $\mathbf{v}$, on a :

$$\Pr\{L_j(\mathbf{u}) = \mathbf{v}\} = \frac{1}{Z_{n,d}(\mathbf{u})} \frac{1}{|\mathbf{u} - \mathbf{v}|^d},$$

où $Z_{n,d}(\mathbf{u})$ est le facteur de normalisation de la loi de probabilité pour $\mathbf{u}$, défini de la façon suivante :

$$Z_{n,d}(\mathbf{u}) = \sum_{\mathbf{v} \in \mathcal{K}^d_{k,n}, \mathbf{u} \neq \mathbf{v}} \frac{1}{|\mathbf{u} - \mathbf{v}|^d}.$$

Le tore régulier d-dimensionnel est *sommet-transitif*, cela signifie qu'il existe une bijection de l'ensemble des sommets qui préserve l'ensemble des arêtes. Grossièrement, chaque sommet a la même « vue »du graphe que tous les autres. Par conséquent, le facteur de normalisation $Z_{n,d}(\mathbf{u})$ ne dépend pas de $\mathbf{u}$. On le notera donc simplement $Z_{n,d}$.

En dimension 1, le tore sous-jacent est un anneau régulier et le nombre de noeuds à distance i d'un noeud donné $\mathbf{u}$ vaut 2 pour $1 \leqslant i \leqslant n$. On peut alors

réécrire $Z_{n,1}$:

$$Z_{n,1} = \sum_{i=1}^{n} \frac{2}{i} = 2H_n.$$

Comme la somme harmonique vérifie l'encadrement

$$\ln(n+1) \leqslant H_n \leqslant 1 + \ln n,$$

on obtient dans $\mathcal{K}^1_{k,n}$:

$$\frac{c}{|\mathbf{u}-\mathbf{v}|\log n} \leqslant \Pr\{L_j(\mathbf{u}) = \mathbf{v}\} \leqslant \frac{c'}{|\mathbf{u}-\mathbf{v}|\log n},$$

où $c, c' > 0$ sont deux constantes que l'on peut déduire des remarques précédentes.

En dimension supérieure, $d > 1$, nous devons évaluer le nombre de points à distance $1 \leqslant i \leqslant n$ d'un noeud $\mathbf{u}$ dans un tore d-dimensionnel. Il s'agit du cardinal de la sphère de rayon i centrée sur $\mathbf{u}$. On note $\mathcal{S}(\mathbf{u}, r)$ et $\mathcal{B}(\mathbf{u}, r)$, respectivement la sphère ℓ_1 et la boule ℓ_1 centrées sur $\mathbf{u}$ et de rayon r, de cardinalités respectives $s_{\mathbf{u}}(r)$ et $b_{\mathbf{u}}(r)$ (on omettra éventuellement l'indice du noeud centre, par sommet-transitivité).

Clairement, pour $r \leqslant n$, $s(r) = \Theta(r^{d-1})$ et $b(r) = \Theta(r^d)$. Plus précisément, pour $r \leqslant n$,

$$s(r) = \frac{2^d}{(d-1)!} r^{d-1} + \xi(r)$$

$$\text{et} \quad b(r) = \frac{2^d}{d!} r^d + \eta(r),$$

où $\xi(r)$ et $\eta(r)$ sont deux polynômes strictement positifs de degrés respectifs $d-2$ et $d-1$. Ces expressions sont aussi des majorants de $s(r)$ et $b(r)$ pour $r > n$. On obtient alors les minorations et majorations suivantes sur le facteur de normalisation :

$$Z_{n,d} \geqslant \sum_{i=1}^{n} \frac{s(i)}{i^d} \geqslant \frac{2^d}{(d-1)!} H_n,$$

$$\text{et} \;\; Z_{n,d} \leqslant \sum_{i=1}^{nd} \left(\frac{2^d}{(d-1)!} \frac{i^{d-1}}{i^d} + O\left(\frac{1}{i^2}\right) \right) \leqslant \frac{2^d}{(d-1)!} H_n + O(1).$$

Théorème 1.1 ([Kle00, BFKK01])
Étant donné une source et une cible à distance mutuelle δ dans un réseau de Kleinberg de dimension d, de côté n et avec k arcs aléatoires sortants par noeud, $k, n, d \geqslant 1$, l'algorithme de routage glouton calcule un chemin de longueur $\Theta(\log n \log \delta / k)$ en espérance .

Première partie

Aspect dynamique de l'effet petit monde

Chapitre 2

Algorithmes de routage efficaces et graphes petits mondes

Introduction

Dans ce chapitre, nous étudions la navigabilité à travers les algorithmes de routage décentralisés qui peuvent être utilisés sur les graphes petits mondes. Nous nous intéressons à deux questions principales sur cet aspect dynamique de l'effet petit monde : la conception d'algorithmes efficaces pour les réseaux petits mondes et la pertinence des processus de routage pour un réseau réel. En particulier, nous construisons un algorithme de routage décentralisé qui reste, à notre connaissance, le plus efficace en terme de longueur de chemin sur un réseau de Kleinberg.

2.1 L'algorithme glouton de Kleinberg

L'algorithme glouton présenté par Kleinberg [Kle00] présente l'avantage d'être à la fois très simple et très efficace, puisqu'il calcule des chemins de longueur $\Theta\left((\log^2 n)/k\right)$ en espérance entre toute paire de noeuds d'un réseau de Kleinberg de côté n. On peut alors se poser la question de l'optimalité des chemins calculés, c'est-à-dire déterminer si le diamètre de ce réseau est $\Theta\left((\log^2 n)/k\right)$. Si ce diamètre est inférieur, cela soulève une autre question : existe-t-il un algorithme décentralisé qui peut trouver ces chemins plus courts ? La propriété de navigabilité se caractérise en effet par la capacité de trouver les chemins courts par un algorithme décentralisé. Ces observations nous conduisent à la notion de *diamètre apparent* d'un graphe qui est l'espérance de la longueur du plus long des plus courts chemins qui peuvent être calculés par un algorithme décentralisé. Par exemple, le modèle de Watts et Stogatz [WS98] présente un diamètre polylogarithmique, mais son diamètre apparent est polynomial. Une

question est donc de déterminer si $\Theta\left((\log^2 n)/k\right)$ est le diamètre apparent du modèle de Kleinberg.

Kleinberg compare cet algorithme glouton à l'expérience de Milgram [Mil67] de transmission de lettres, ce qui soulève la question de la pertinence de cet algorithme pour les réseaux sociaux. En effet, l'algorithme de routage glouton renvoie le message au voisin le plus proche de la cible, il semble pourtant plus réaliste qu'un individu sonde un voisinage à une distance supérieure à 1 avant de renvoyer la lettre. L'expérience de Milgram a été récemment reproduite sous forme de transmission d'*e-mails* par Dodds et al. [DMW03] qui ont recueilli des statistiques sur la façon dont les *e-mails* ont été transmis. Cette expérience a révélé que les individus avaient tendance à choisir de transmettre le message à une de leur connaissance qu'ils pensaient être plus à-même de le retransmettre. Il ne s'agit donc pas dans ce cas de transmettre le message à la connaissance considérée comme la « plus proche »de la destination, mais plutôt à celle qui a une connaissance plus proche.

Dans ce chapitre, nous présentons un nouvel algorithme de routage décentralisé dédié aux graphes petits mondes. Nous verrons qu'il apporte des réponses aux problématiques soulevées ci-dessus, par son efficacité en longueur de chemin et son schéma de routage. En particulier, il effectue une recherche locale restreinte, pertinente dans le cadre des réseaux sociaux. Nous présentons et analysons cet algorithme dans un réseau de Kleinberg $\mathcal{K}^d_{k,n}$, $k, n, d \geqslant 1$.

2.2 Améliorer l'efficacité du routage grâce à une exploration restreinte

Pour rechercher des améliorations de l'efficacité du routage glouton, nous commençons par revenir sur son analyse. L'analyse globale du routage glouton $\mathcal{G}$ entre une source et une cible à distance δ se fait par dichotomie : on découpe l'exécution de l'algorithme en $\log \delta$ phases correspondant aux divisions successives de la distance courante par deux. La i-ème phase, $1 \leqslant i \leqslant \log \delta$, dure tant que la distance du *porteur courant du message* (*i.e.* le noeud courant traité par l'algorithme) à la cible appartient à $(2^{i-1}, 2^i]$. On peut alors montrer qu'en phase i, la probabilité que le j-ème contact longue-distance du porteur courant du message soit à distance strictement inférieure à 2^{i-1} de la cible est supérieure à $1/(c \log n)$, où c est une constante strictement positive, pour tout $1 \leqslant j \leqslant k$. On en déduit que l'espérance du nombre de noeuds visités par $\mathcal{G}$ durant la i-ème phase d'exécution est $O\left((\log n)/k\right)$. En sommant sur toutes les phases, on obtient l'espérance de

longueur totale du chemin de routage : $O\left((\log n \log \delta)/k\right)$.

La clé de cette analyse est la bonne répartition des liens longue-distance sur toutes les échelles de distance, dans le modèle de Kleinberg. Cette répartition garantit qu'en visitant $\Omega\left((\log n)/k\right)$ nouveaux noeuds en espérance, à distance comprise entre x et $x/2$ de la cible du routage, on trouve un contact longue-distance qui est à distance strictement inférieure à $x/2$ de la cible, pour tout x inférieur ou égal à n. La visite gloutonne d'un ensemble de noeuds de taille S allonge de S la longueur du chemin de routage. Nous pouvons chercher à visiter ces noeuds de façon plus compacte, afin de diminuer la longueur du chemin nécessaire pour trouver un contact deux fois plus proche de la cible. Une idée naturelle est de construire une exploration arborescente, c'est-à-dire de visiter un voisinage de noeuds autour du noeud courant en suivant les liens locaux et longue-distance des voisins, de proche en proche. Chaque noeud du modèle ayant un degré sortant constant, on peut penser visiter de cette façon $\Omega\left((\log n)/k\right)$ nouveaux noeuds (*i.e.* non visités) avec une profondeur d'exploration $\Theta\left(\log\left((\log n)/k\right)\right)$. Toutefois, les contacts longue-distance des voisins peuvent être des noeuds précédemment visités par l'algorithme. On rencontre alors des problèmes de dépendance des tirages des liens et les branches de l'arbre ne sont pas disjointes, ce qui réduit le nombre de nouveaux noeuds contenus dans l'ensemble exploré. Dans les sections suivantes, nous allons étudier comment parvenir à résoudre ces problèmes de dépendance pour construire une structure d'exploration compacte *et* contenant suffisamment de nouveaux noeuds.

Principe général de l'algorithme. Pour une source et une cible donnée, notre algorithme va construire des structures d'explorations successives des contacts locaux et longue-distance, dont la profondeur dépend de la distance courante à la cible. À la fin de chaque exploration, le message à transmettre est envoyé de la racine de l'exploration au noeud le plus proche de la cible, parmi les noeuds explorés. Nous faisons l'hypothèse que la découverte d'un lien longue-distance lors de l'exploration nous autorise à renvoyer une faible quantité d'information en sens inverse sur ce lien, ce qui permet à la racine de déterminer le noeud le plus proche de la cible dans l'ensemble exploré. En revanche, le message à transmettre emprunte les liens longue-distance suivant leur orientation. Cette hypothèse peut s'illustrer par le passage d'une lettre dans un réseau social : supposons que Ringo connaît John mais que John ne connaît pas Ringo ; si Ringo envoie une lettre à John en indiquant son adresse au dos de l'enveloppe (découverte du lien longue-distance), John peut alors envoyer une lettre à Ringo. On peut également illustrer ce fait par la navigation dans le réseau des pages web : supposons que la page web de Germaine pointe sur celle du groupe The Beatles, sans que celle-ci ait un hyperlien vers la page de Germaine ; un internaute passant de la page de Germaine à la page des Beatles par l'hyperlien (découverte du lien longue-distance) peut

revenir sur la page de Germaine par la touche « back »du navigateur.

2.2.1 Compromis entre le recoupement et la profondeur d'exploration

Dans cette section, nous étudions les marges d'optimisation offertes par une exploration des contacts locaux et longue-distance à partir d'un noeud. On se donne une source et une cible pour le routage, et l'on note x la distance sous-jacente d'un noeud $\mathbf{x}$ à la cible.

On considère à présent le noeud $\mathbf{x}$ qui est la racine de notre structure d'exploration. Pour faciliter la lecture, nous allons représenter la structure d'exploration par un arbre. Les *branches* sont les chemins constitués de liens locaux et longue-distance des noeuds explorés, l'ensemble des fils d'un noeud est constitué de l'ensemble de ses contacts locaux et longue-distance qui sont plus proches de la cible que la racine.

Nous avons vu que, dans cet arbre, les branches ne sont pas disjointes. On distingue deux cas de recoupements :

-1er cas : le contact d'un noeud est un noeud déjà présent dans la structure d'exploration,

-2ème cas : le contact d'un noeud est à distance strictement supérieure à x de la cible. En effet, un noeud à distance strictement supérieure à x de la cible peut avoir été visité par l'algorithme dans les phases précédentes, son indépendance n'est pas garantie.

Chaque recoupement élimine de la structure arborescente tout le sous-arbre qui y est enraciné. Cela diminue donc le nombre total de noeuds ; toutefois, ces noeuds sont des répétitions, et leur visite est donc inutile dans la recherche d'un contact deux fois plus proche de la cible. Un exemple d'exploration est illustré sur la figure 2.1. Pour augmenter le nombre de noeuds de la structure et compenser les recoupements, on peut augmenter la profondeur de l'exploration. Toutefois, dans le but de diminuer la longueur totale du chemin de routage, la profondeur de l'exploration doit être limitée, puisqu'elle représente la longueur du morceau de chemin que l'on ajoute au chemin courant à chaque phase de division par deux de la distance à la cible. Il s'agit donc d'arriver à un compromis entre une profondeur d'exploration restreinte et un nombre de nouveaux noeuds suffisamment grand.

Dans la suite, nous allons quantifier précisément la probabilité de recoupement des voisins qui nous permet de minorer le nombre de nouveaux noeuds dans l'arbre, pour chaque profondeur, avec une certaine probabilité. Nous verrons que cette méthode seule ne permet pas d'obtenir une profondeur d'exploration suffisamment faible tout en garantissant $\Omega\left((\log n)/k\right)$ nouveaux noeuds dans la structure. Afin d'augmenter le nombre de nouveaux noeuds dans la structure, en

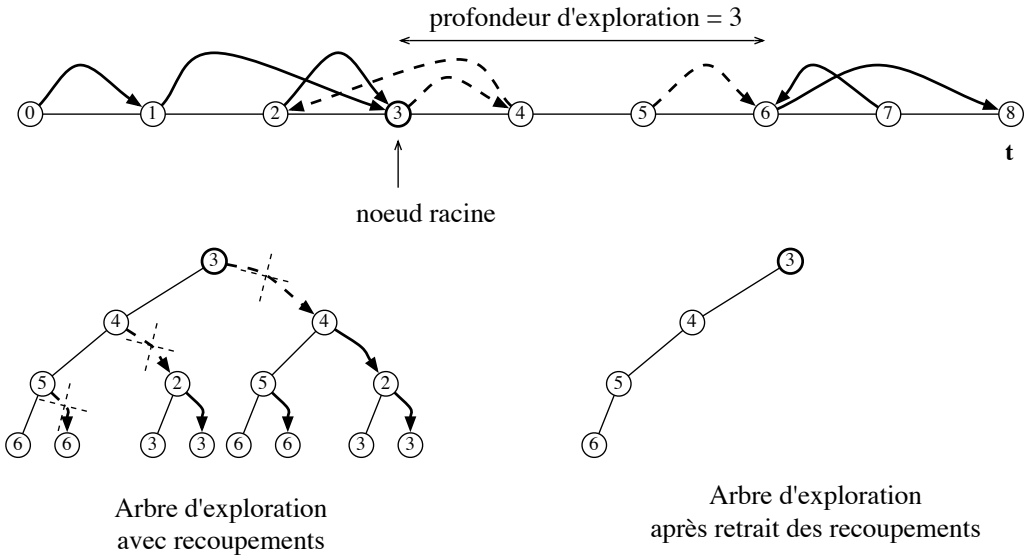


FIG. 2.1 : Arbre d'exploration de profondeur 3 dans un réseau $\mathcal{K}^1_{1,n}$, à partir du noeud racine 3 et que l'on a dirigé vers la cible $\mathbf{t}$. En gras et en pointillés figurent les liens longue-distance, les pointillés correspondent aux recoupements.

augmentant très peu la profondeur de la structure, nous la prolongeons par des chaînes de liens locaux dirigées vers la cible, enracinées aux feuilles de l'arbre (cf. figure 2.2). Tant que l'on garantit que ces chaînes ne recoupent pas les autres noeuds de la structure, l'ajout de ces chaînes multiplie le nombre de nouveaux noeuds présents dans la structure (les feuilles) par la longueur des chaînes.

Définition 2.1 (Chaîne)
On appelle chaîne*, un chemin constitué uniquement de liens locaux.*

2.2.2 *Lien valide* et *zone de sécurité*

Cette section introduit dans notre vocabulaire les termes *lien valide* et *zone de sécurité*, relatifs à la gestion des recoupements, afin de simplifier la description de notre algorithme dans la suite.

Afin de garantir l'indépendance de la structure d'exploration courante vis-à-vis des noeuds visités dans une structure d'exploration précédemment visitée, nous ne considérons que les contacts qui sont strictement plus proches de la cible que leur origine. On appelle *liens valides* les liens associés, définis formellement ci-dessous.

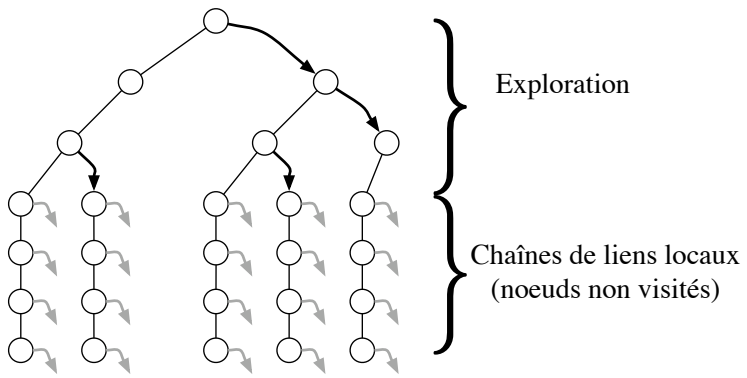


FIG. 2.2 : Exemple d'arbre d'exploration de profondeur 3 avec ses chaînes de liens locaux de longueur 3 attachées aux feuilles, dans un réseau $\mathcal{K}^1_{1,n}$.

Définition 2.2 (Lien valide)
On dit qu'un lien (local ou longue-distance) d'un noeud $\mathbf{u}$ *vers un noeud* $\mathbf{v}$ *est* valide *si* $\mathbf{v}$ *est strictement plus proche de la cible que* $\mathbf{u}$ *selon la distance de Manhattan. On dit alors que* $\mathbf{v}$ *est un* contact valide *(local ou longue-distance) de* $\mathbf{u}$.
On dit qu'un noeud $\mathbf{v}$ *est* à h liens valides *de* $\mathbf{u}$, *s'il existe un chemin de longueur inférieure ou égale à* h *de* $\mathbf{u}$ *vers* $\mathbf{v}$ *composé uniquement de liens valides.*
On dit que $\mathbf{v}$ *est* à h liens locaux valides *de* $\mathbf{u}$ *si ce chemin est uniquement composé de liens locaux valides.*

Supposons que la structure d'exploration est composée de liens locaux et longue-distance et prolongée par des chaînes de liens locaux enracinées en chacune de ses feuilles, comme décrit dans la section précédente.

Afin de garantir que les chemins de liens locaux de l'arbre (dont les chaînes) sont constitués de noeuds distincts, il suffit de ne considérer dans la structure que les contacts longue-distance qui sont à une distance supérieure ou égale à la hauteur de l'arbre des autres noeuds de la structure. On appelle *zone de sécurité* d'un noeud $\mathbf{u}$ l'ensemble des noeuds à distance inférieure à la hauteur de l'arbre de $\mathbf{u}$. On dit que ces noeuds sont *interdits* pour les contacts longue-distance des autres noeuds de la structure. On remarquera que les contacts longue-distance que l'on ne considère pas sont visités par ailleurs dans la structure, puisque ce sont des répétitions.

2.3 Notre algorithme de routage

Dans cette section, nous présentons en détail le fonctionnement de notre algorithme de routage. Nous considérons tout d'abord le cas unidimensionnel pour

la lisibilité. La section 2.5 généralisera cet algorithme, ainsi que tous les résultats obtenus, à une dimension quelconque.

Nous décrivons notre algorithme suivant un parcours en profondeur qui permet de comprendre son fonctionnement de façon simple. Nous verrons par la suite une implémentation de l'algorithme suivant un parcours en largeur, moins efficace en pratique, mais dont l'analyse probabiliste est plus simple.

Description de l'algorithme (selon le parcours en profondeur). Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible respectivement. On suppose que trois fonctions $h_{\max}(x)$, $b_{\max}(x)$ et $g(x)$, ainsi qu'un paramètre stop nous sont donnés. La fonction $h_{\max}(x)$ représente la profondeur d'exploration pour une racine à distance x de $\mathbf{t}$, $b_{\max}(x)$ le nombre de feuilles que l'on cherche à obtenir, $g(x)$ la longueur des chaînes de liens locaux (cf. figure 2.3). Enfin, stop est la distance seuil à $\mathbf{t}$ en dessous de laquelle on exécute l'algorithme glouton de Kleinberg car l'exploration devient inutile.

Les valeurs de ces fonctions seront données dans la suite. La difficulté consiste à choisir les valeurs de ces fonctions de façon à minimiser la longueur du chemin $h_{\max}(x)+g(x)$ tout en garantissant un nombre de feuilles suffisamment grand.

Soit $\mathbf{x}$ le porteur courant du message et π le chemin courant de $\mathbf{s}$ à $\mathbf{x}$:

1. Tant que $|\mathbf{x}-\mathbf{t}| > \mathrm{stop}$:
   - Explorer les noeuds à $h_{\max}(x)$ liens valides de $\mathbf{x}$ suivant un parcours en profondeur, enregistrer dans un ensemble F tous les contacts longue-distance valides visités, mais sauter dans l'exploration tous les contacts longue-distance à distance strictement inférieure à $h_{\max}(x) + g(x)$ d'un noeud de l'ensemble F courant. Le parcours en profondeur est arrêté dès que $|F| = b_{\max}(x)$, même s'il reste des noeuds à parcourir.
   - À chaque fois qu'un noeud $\mathbf{z}$, à exactement $h_{\max}(x)$ liens valides de $\mathbf{x}$, est rencontré, lire les adresses des contacts longue-distance des noeuds à $g(x)$ liens locaux valides de $\mathbf{z}$, et enregistrer dans une variable $\mathbf{y}$ le noeud le plus proche de la cible parmi les noeuds visités et leurs contacts.
   - À la fin de l'exploration en profondeur, router le message de $\mathbf{x}$ à $\mathbf{y}$ dans l'arbre d'exploration et étendre le chemin π jusqu'à $\mathbf{y}$, suivant ces liens.
2. Lorsque $|\mathbf{x}-\mathbf{t}| \leqslant \mathrm{stop}$: appliquer l'algorithme glouton de Kleinberg, *i.e.* transmettre le message au contact de $\mathbf{x}$ le plus proche de $\mathbf{t}$ et étendre le chemin π par ce lien, jusqu'à atteindre $\mathbf{t}$.

La figure 2.3 illustre le principe de la structure visitée durant chaque phase d'exploration : les lignes droites représentent les liens locaux valides et les flèches représentent les liens longue-distance valides ; les noeuds de F sont représentés par des cercles blancs, chacun commence une nouvelle chaîne de liens locaux de

longueur inférieure ou égale à $h_{\max}(x) + g(x)$, dirigée vers la cible. La structure explorée est composée d'un arbre $(k+1)$-aire (partiel) de hauteur $h_{\max}(x)$ étendu par des chaînes de liens locaux de longueur $g(x)$ attachées à ses feuilles.

Pour alléger les notations, on pose $\hat{k} =_{\text{def}} k + 1$.

Les zones de sécurité (cf. section 2.2.2) des noeuds sont respectées, puisqu'on ne considère que des contacts longue-distance à distance supérieure à $h_{\max}(x) + g(x)$ (la profondeur de la structure) de l'ensemble des noeuds déjà explorés F. Cela garantit que les chaînes de liens locaux enracinées en les noeuds de F ne se recoupent pas. Rappelons que les contacts évités sont en fait des répétitions de noeuds existant dans notre structure, cet évitement n'est donc pas une limite sur les performances de l'algorithme.

L'arbre est dessiné dans le plan pour illustrer la structure arborescente, mais il est en fait plongé dans le tore. À la fin d'une phase d'exploration, le chemin est étendu depuis $\mathbf{x}$ vers le noeud $\mathbf{y}$ le plus proche de la cible parmi les noeuds explorés et leurs contacts. Une nouvelle phase d'exploration commence alors depuis le noeud $\mathbf{y}$.

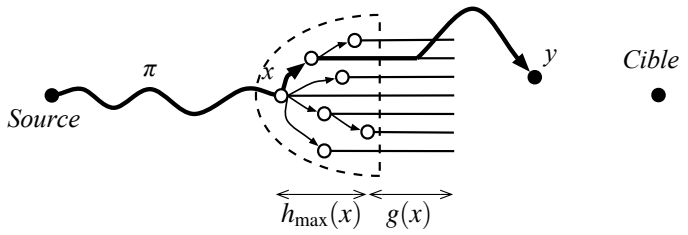


FIG. 2.3 : Extension du chemin π (en gras) à la fin d'une phase d'exploration.

La figure 2.4 illustre la succession des structures d'exploration durant un routage.

L'implémentation selon un parcours en largeur (algorithme 2) que nous analysons dans la section suivante construit la structure arborescente d'exploration niveau par niveau. Même si la partie de l'arbre explorée n'est pas la même que dans l'implémentation selon le parcours en profondeur, l'analyse des deux programmes est identique.

La figure 2.5 illustre les ensembles de noeuds qui correspondent à un niveau intermédiaire h, compris entre 1 et $h_{\max}$, de la construction de la structure. A est l'ensemble des noeuds visités durant la phase d'exploration (sans les chaînes de liens locaux C).

Les liens suivis durant une phase d'exploration forment une structure d'arbre $\hat{k}$-aire sur A, de hauteur h_{stop}. h_{stop} est la hauteur à laquelle le nombre de branches a atteint le seuil $b_{\max}(x)$. L'ensemble des feuilles $B_{h_{\text{stop}}}$ est étendu par $|B_{h_{\text{stop}}}|$ chaînes de liens locaux valides de longueur $g(x)$, enracinées en les noeuds de

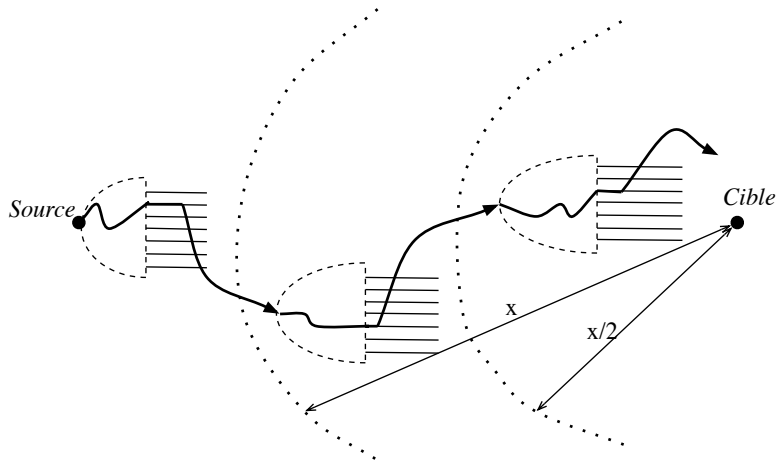


FIG. 2.4 : Routage par explorations successives.

$B_{h_{\text{stop}}}$. L'ensemble B_h est l'ensemble des noeuds au niveau h de l'arbre. A peut aussi être vu comme un ensemble de $|F| = |B_{h_{\text{stop}}}|$ chaînes de longueur inférieure à $h_{\max}(x) + g(x)$, enracinées en les noeuds de F, qui ne se recoupent pas, pointent vers la cible et sont reliées entre elles par des liens longue-distance.

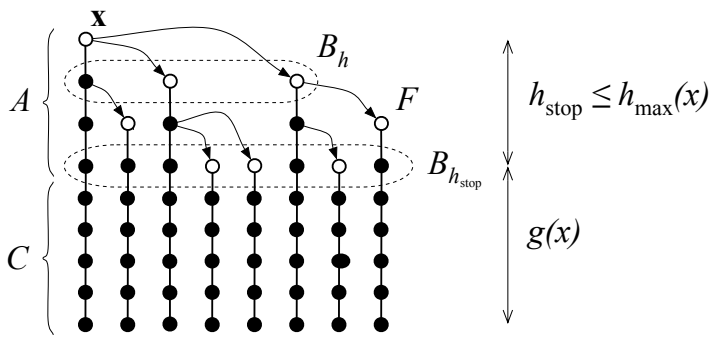


FIG. 2.5 : Les ensembles A, B_h, C et F (noeuds blancs) durant une phase d'exploration.

2.4 Analyse de l'algorithme en dimension 1

Dans cette section, nous analysons l'algorithme 2 sur $\mathcal{K}^1_{k,n}$. L'analyse consiste à étudier la structure d'exploration enracinée en un noeud $\mathbf{x}$, porteur du message, à distance x de la cible $\mathbf{t}$. Nous montrons que pour une profondeur $O(\log n \log\log n / \log x)$ seulement, elle est suffisamment large pour garantir

Algorithme 2 Algorithme de routage selon le parcours en largeur

Entrée : la source $\mathbf{s}$ et la cible $\mathbf{t}$.

Initialisation : $\mathbf{x} \leftarrow \mathbf{s}$.

tant que $|\mathbf{x} - \mathbf{t}| > \text{stop}$ **faire**

Phase d'exploration :

$x \leftarrow |\mathbf{x} - \mathbf{t}|$, $A_0 \leftarrow \{\mathbf{x}\}$, $B_0 \leftarrow \{\mathbf{x}\}$, $F \leftarrow \{\mathbf{x}\}$, $h \leftarrow 0$.

tant que $h < h_{\max}(x)$ et $|B_h| < b_{\max}(x)$ **faire**

$B_{h+1} \leftarrow \varnothing$.

pour chaque $\mathbf{u} \in B_h$ **faire**

$B_{h+1} \leftarrow$ les contacts locaux valides de $\mathbf{u}$.

pour chaque contact longue-distance valide $\mathbf{v}$ de $\mathbf{u}$ **faire**

si $\forall \mathbf{w} \in F,\ |\mathbf{v} - \mathbf{w}| \geqslant h_{\max}(x) + g(x)$ **alors**

$F \leftarrow F \cup \{\mathbf{v}\}$, $B_{h+1} \leftarrow B_{h+1} \cup \{\mathbf{v}\}$.

fin si

fin pour

fin pour

fin tant que

$A_{h+1} \leftarrow A_h \cup B_{h+1}$.

h++.

si $|B_h| > b_{\max}(x)$ **alors**

enlever les $(|B_h| - b_{\max}(x))$ derniers noeuds insérés dans B_h et F.

fin si

$h_{\text{stop}} \leftarrow h$,

$A \leftarrow A_{h_{\text{stop}}-1} \cup B_{h_{\text{stop}}}$. (*Remarque :* $|B_{h_{\text{stop}}}| \leqslant b_{\max}(x)$)

$C \leftarrow \bigcup_{b \in B_{h_{\text{stop}}}} C_b$, où C_b est l'ensemble des noeuds à moins de $g(x)$ liens locaux valides de b.

Phase de retransmission du message :

$\mathbf{x} \leftarrow$ le noeud le plus proche de la cible $\mathbf{t}$ parmi les contacts locaux et longue-distance d'un noeud dans $A \cup C$.

Router le message vers $\mathbf{x}$ dans l'arbre formé par $A \cup C$.

fin tant que

Phase finale (*Algorithme glouton*) : Retransmettre le message au noeud le plus proche de la cible parmi les contacts (locaux ou longue-distance) du porteur courant du message, jusqu'à ce qu'il atteigne la cible $\mathbf{t}$.

l'existence d'un contact dont la distance à $\mathbf{t}$ est inférieure à $\frac{\log \hat{k}}{2k} \cdot x$. Comme cette analyse est indépendante de la façon dont on explore l'arbre, elle s'appliquera également à l'implémentation (plus efficace) selon le parcours en profondeur.

Rappelons que $\mathcal{K}^1_{k,n}$ est un anneau augmenté de $2n+1$ noeuds, numérotés de $-n$ à n. En plus de ses deux voisins sur l'anneau (ses *contacts locaux*), chaque noeud $\mathbf{u}$ est l'origine de k liens orientés supplémentaires ($1 \leqslant k \leqslant \log n$), chacun pointant vers un noeud $\mathbf{v}_j$ (le j-ème contact longue-distance de $\mathbf{u}$) choisi indépendamment suivant la distribution 1-harmonique, *i.e.* avec probabilité $1/(2H_n \cdot |\mathbf{u} - \mathbf{v}_j|)$.

Définition 2.3 (Paramètres pour l'algorithme 2 en dimension 1)
Les valeurs des paramètres pour l'algorithme 2 en dimension 1 sont, pour un noeud $\mathbf{x}$ *à distance* x *de la cible :*

$$\begin{aligned} h_{\max}(x) &= \frac{\log\log x - \log\log\log n}{\log\left(1 + k\log x/(6\log n)\right)}, \\ b_{\max}(x) &= \frac{\log x}{\log\log n}, \\ g(x) &= \frac{\log n \cdot \log\log n}{\log \hat{k} \cdot \log x}, \\ \textit{et} \quad \mathrm{stop} &= k\log^2 n. \end{aligned}$$

On remarquera que ces valeurs nous permettent d'obtenir une profondeur totale d'exploration au plus $h_{\max} + g(x)$, de l'ordre de $\log n \log\log n / \log x$, strictement inférieure à $\log n$ et décroissante avec la distance courante à la cible. Par ailleurs, l'exploration est arrêtée dès que l'on atteint $b_{\max}(x)$ branches ; dans ce cas, le nombre de nouveaux noeuds présents dans la structure est $g(x) \cdot b_{\max}(x) = \log n / \log \hat{k}$, c'est le nombre de noeud minimal à partir duquel on garantit l'existence d'un contact longue distance à distance inférieure à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible avec probabilité constante (lemme 2.4). Nous verrons dans l'analyse que si l'exploration s'arrête pour la seconde raison, c'est-à-dire lorsque l'on atteint la hauteur maximale $h_{\max}(x)$, la structure contient le même nombre de noeuds, à un facteur constant près.

Plan de l'analyse. Nous allons prouver que l'ensemble des chaînes de liens locaux C contient bien $\Omega(\log n / \log \hat{k})$ nouveaux noeuds, avec probabilité constante (lemme 2.7). En appliquant le lemme 2.4, nous pourrons conclure qu'en visitant un nombre constant de structures en espérance, on trouve un contact longue-distance qui est $2k/\log \hat{k}$ fois plus près de la cible que la racine de la première exploration.

Pour prouver le lemme 2.7, nous allons montrer que le nombre de branches de la structure d'arbre de A est suffisamment grand en comparant l'arbre à un processus de branchement stochastique. Pour cela, on minore la probabilité de créer une nouvelle branche dans A (corollaire 2.6). Cela nous permet de minorer la croissance de A par celle d'un processus de branchement classique que l'on sait évaluer.

2.4.1 Analyse d'un processus de branchement générique

Chaque lien local valide d'un noeud de A (il en existe toujours un pour chaque noeud) crée une nouvelle branche vers le niveau suivant dans A. Chaque lien longue-distance d'un noeud de A crée une nouvelle branche vers le niveau suivant si et seulement si :

- il est valide,
- *et* son contact longue-distance n'intersecte aucune des zones de sécurité des noeuds déjà construits.

Le lemme 2.6 donne un minorant, en fonction de x, de la probabilité de cet événement pour un lien longue-distance donné ; on le note α^-. On peut alors minorer la taille de A par celle d'un processus de branchement classique de paramètre α^-, où k branches sont créées indépendamment avec une probabilité α^- et une branche avec probabilité 1 (correspondant à un lien local dans A). Dans cette section, nous étudions la croissance du nombre de branches dans un tel processus.

Définition du processus étudié.
Soit $\alpha \in]0, 1]$ et posons, pour $0 \leqslant l \leqslant k$:

$$\rho_l = \binom{k}{l} \alpha^l (1-\alpha)^{k-l}.$$

On considère le processus de branchement suivant : commencer avec un noeud ; durant le h-ième pas, donner indépendamment, à chaque noeud du niveau $h-1$, exactement $1+l$ fils, avec probabilité ρ_l.

Les lemmes 2.1 et 2.2 démontrent que le nombre de branches au niveau h, dans l'arbre $(k+1)$-aire partiel résultant de ce processus, est supérieur à la moitié de son espérance, avec probabilité constante.

Soit b_h la variable aléatoire pour le nombre de branches au niveau h (*i.e.*, pour le nombre de noeuds au niveau h de l'arbre). Le lemme suivant est une analyse classique de processus de Galton-Watson (voir par exemple [HMRe98]).

Lemme 2.1 *Pour tout $h \geqslant 0$,*

$$\mathbb{E}[b_h] = (1+k\alpha)^h \quad \textit{et} \quad \mathbb{E}[(b_h)^2] \leqslant \left(1 + \frac{1-\alpha}{1+k\alpha}\right)(1+k\alpha)^{2h}.$$

Preuve. On procède par induction. Pour $h = 0$,

$$\mathbb{E}[b_0] = 1 \quad \text{et} \quad \mathbb{E}[b_0] = 1 \leqslant 1 + \frac{1-\alpha}{1+k\alpha}.$$

Supposons maintenant h strictement positif. L'espérance du nombre de fils de chaque noeud au niveau $h-1$ est $(1+k\alpha)$ puisque chaque branche correspondant à un lien longue-distance est créée avec probabilité α. L'espérance du nombre de noeuds au niveau h est donc $(1+k\alpha)\mathbb{E}[b_{h-1}]$. Par hypothèse d'induction, on obtient :

$$\mathbb{E}[b_h] = (1+k\alpha)^h.$$

On évalue à présent l'espérance de $(b_h)^2$. Avec probabilité ρ_l, la racine du processus de branchement a exactement $1+l$ fils ; si l'on note $\beta_0, \beta_1, \ldots, \beta_l$ les variables aléatoires pour le nombre de branches au niveau $h-1$ dans chacun des $1+l$ sous-arbres de la racine, on a :

$$(b_h)^2 = (\sum_{i=0}^{l} \beta_i)^2 = \sum_{i=0}^{l} (\beta_i)^2 + 2\sum_{i=0}^{l}\sum_{j=i+1}^{l} \beta_i\beta_j.$$

Les β_i sont des variables indépendantes, distribuées identiquement à b_{h-1}. Ainsi :

$$\mathbb{E}[b_h^2 \,|\, \text{la racine a exactement } l \text{ fils}] = (l+1)\mathbb{E}[(b_{h-1})^2] + l(l+1)\mathbb{E}[b_{h-1}]^2.$$

En sommant sur l, on obtient :

$$\begin{aligned}\mathbb{E}[(b_h)^2] &= \sum_{l=0}^{k} \binom{k}{l} \alpha^l (1-\alpha)^{k-l} \left((l+1)\mathbb{E}[(b_{h-1})^2] + l(l+1)\mathbb{E}[b_{h-1}]^2\right) \\ &= (1+k\alpha)\mathbb{E}[(b_{h-1})^2] + k\alpha(2+\alpha(k-1))\mathbb{E}[b_{h-1}]^2.\end{aligned}$$

Comme, par induction,

$$\mathbb{E}[(b_{h-1})^2] \leqslant \left(1 + \frac{1-\alpha}{1+k\alpha}\right)(1+k\alpha)^{2h-2} \quad \text{et} \quad \mathbb{E}[b_{h-1}] = (1+k\alpha)^{h-1},$$

on conclut que :

$$\begin{aligned}\mathbb{E}[(b_h)^2] &\leqslant (1+k\alpha)^{2h-2}(1+2k\alpha+k^2\alpha^2+1-\alpha) \\ &= \left(1 + \frac{1-\alpha}{1+k\alpha}\right)(1+k\alpha)^{2h}.\end{aligned}$$

□

La méthode du second moment, introduite à la remarque 3.1 de [JŁR00] pour les variables aléatoires positives, permet d'assurer qu'une variable aléatoire prend des valeurs suffisamment grandes lorsque l'on peut majorer sa variance. La généralisation suivante de ce résultat aux variables pouvant prendre des valeurs négatives nous sera très utile par la suite.

Lemme 2.2 (Méthode du second moment) *Soit X une variable aléatoire réelle, on pose $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$ tels que $X = X^+ - X^-$.*
Si $\mathbb{E}[X^-] \leqslant 2\mathbb{E}[X^+]$, alors :

$$\Pr\{X > 0\} \geqslant \mathbb{E}[X]^2/\mathbb{E}[X^2].$$

On remarquera que la minoration $\mathbb{E}[X]^2/\mathbb{E}[X^2]$ s'exprime aussi comme $\mathbb{E}[X]^2/(\mathbb{E}[X]^2 + \mathrm{Var}(X))$, où $\mathrm{Var}(X)$ est la variance de X.

Preuve. Il s'agit d'une généralisation immédiate de [JŁR00, Remarque 3.1]. Soit $Y = 1\!\!1_{X>0}$ la variable aléatoire qui vaut 1 lorsque $X > 0$ et 0 sinon. On a :

$$\begin{aligned}\mathbb{E}[X]^2 = (\mathbb{E}[X^+] - \mathbb{E}[X^-])^2 &= \mathbb{E}[X^+]^2 - \mathbb{E}[X^-](2\mathbb{E}[X^+] - \mathbb{E}[X^-]) \\ &\leqslant \mathbb{E}[X^+]^2 = \mathbb{E}[XY]^2,\end{aligned}$$

puisque $X^+ = XY$. L'inégalité de Cauchy-Schwarz donne :

$$\mathbb{E}[X]^2 \leqslant \mathbb{E}[XY]^2 \leqslant \mathbb{E}[X^2]\mathbb{E}[Y^2] = \mathbb{E}[X^2]\Pr\{X > 0\},$$

puisque $Y^2 = Y$ et $\mathbb{E}[Y] = \Pr\{X > 0\}$. Cela conclut la preuve. □

On montre à présent qu'avec probabilité constante, le nombre de branches dans le processus de branchement est supérieur à la moitié de son espérance.

Lemme 2.3 *Pour tout h,* $\Pr\{b_h \geqslant \mathbb{E}[b_h]/2\} \geqslant \frac{1}{5}$.

Preuve. On définit une variable aléatoire $X = b_h - \mathbb{E}[b_h]/2$ et l'on étudie $\Pr\{X > 0\}$. On a $\mathbb{E}[X] = \mathbb{E}[b_h]/2 > 0$, donc $\mathbb{E}[X^+] - \mathbb{E}[X^-] > 0$ et la méthode du second moment (lemme 2.2) s'applique : $\Pr\{X > 0\} \geqslant \mathbb{E}[X]^2/\mathbb{E}[X^2]$.
On a :

$$\mathbb{E}[X]^2 = \frac{\mathbb{E}[b_h]^2}{4} \quad \text{et} \quad \mathbb{E}[X^2] = \mathbb{E}[(b_h)^2] - \frac{3}{4}\mathbb{E}[b_h]^2.$$

D'après le lemme 2.1,

$$\mathbb{E}[(b_h)^2] < \left(1 + \frac{1-\alpha}{1+k\alpha}\right)\mathbb{E}[b_h]^2 < 2\mathbb{E}[b_h]^2.$$

Ainsi, $\mathbb{E}[X^2] < \frac{5}{4}\mathbb{E}[b_h]^2$. Finalement, on obtient :

$$\Pr\{X > 0\} \geqslant \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} > \frac{1}{4}\frac{1}{(5/4)} = \frac{1}{5}.$$

□

2.4.2 Analyse de l'algorithme

L'analyse de l'algorithme glouton de [Kle00] partage l'exécution en phases successives de division de la distance à la cible par deux. Nous montrons que l'on peut choisir un facteur de division de la distance croissant avec k. Le nombre de phases d'exécution est alors décroissant avec k. La preuve de ce résultat (lemme 2.4) est inspirée de [Kle00].

Lemme 2.4 *Soit* $\mathbf{t}$ *la cible d'un routage dans* $\mathcal{K}^1_{k,n}$. *Soient* $\gamma > 0$ *et* $1 < x < n$, *il existe une constante* $p'_1 > 0$ *(indépendante de* n, x *et* k*) telle que, pour tout sous-ensemble* Γ *de* $\gamma \cdot \frac{\log n}{\log \hat{k}}$ *noeuds, à distance strictement supérieure à* $\frac{\log \hat{k}}{2k} \cdot x$ *et inférieure ou égale à* x *de la cible, un noeud de* Γ *(au moins) a un contact longue-distance à distance inférieure ou égale à* $\frac{\log \hat{k}}{2k} \cdot x$ *de la cible, avec probabilité au moins* $p'_1 > 0$.

Preuve. Soit $y = \frac{\log \hat{k}}{2k} \cdot x$ et $\mathbf{u}$ un noeud à distance $u \in (y, x]$ de la cible. La probabilité que le j-ème contact longue-distance de $\mathbf{u}$ soit à distance inférieure ou égale à y de la cible est supérieure ou égale à[1] :

$$\begin{aligned}
\frac{1}{2H_n}\sum_{i=u-y}^{u+y}\frac{1}{i} &\geqslant \frac{1}{2H_n}\int_{u-y}^{u+y}\frac{dt}{t} = \frac{1}{2H_n}\ln\left(1 + \frac{2x\log\hat{k}}{2ku - x\log\hat{k}}\right) \\
&\geqslant \frac{1}{2H_n}\ln\left(1 + \frac{\log\hat{k}}{k}\right) \\
&\geqslant \frac{\ln 2}{2}\cdot\frac{\log\hat{k}}{kH_n},
\end{aligned}$$

puisque $u \leqslant x$ et $\ln(1+z) \geqslant z\ln 2$, pour $z \in [0,1]$. La probabilité que tous les $k|\Gamma|$ contacts longue-distance des noeuds de Γ soient à distance strictement supérieure à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible est donc inférieure à :

$$\left(1 - \frac{\ln 2}{2}\cdot\frac{\log\hat{k}}{kH_n}\right)^{\gamma\frac{k\log n}{\log\hat{k}}} \leqslant e^{-\gamma/2} =_{\text{def}} 1 - p'_1 < 1.$$

[1] On a l'égalité pour $|u| \leqslant n/2$.

□

Nous allons à présent minorer la probabilité de non-recoupement, c'est-à-dire la probabilité qu'un lien longue distance donné crée une nouvelle branche dans A.

Lemme 2.5 *Soient* $\mathbf{u}$ *un noeud à distance* u *de la cible* $\mathbf{t}$*,* $\mathbf{v}$ *son* j*-ème contact longue-distance,* Q *un ensemble de* q *noeuds interdits et* r *un entier.*
La probabilité que $\mathbf{v}$ *soit valide et à distance supérieure ou égale à* r *de tout noeud de* Q *est supérieure ou égale à :*

$$\frac{H_{2u-1} - H_{2rq-1}}{2H_n}.$$

Preuve. Soit $\mathcal{E}$ l'événement : « $\mathbf{v}$ est valide et à distance supérieure ou égale à r de tout noeud de Q ». $\mathcal{E}$ est l'événement « $\mathbf{v}$ est valide et n'appartient à aucune des q chaînes de noeuds de longueur $2r$ centrées sur les noeuds de Q ». On minore la probabilité de $\mathcal{E}$ en remarquant que la probabilité que $\mathbf{v}$ soit à distance δ de $\mathbf{u}$ décroît quand δ augmente. Ainsi, la probabilité de $\mathcal{E}$ est minimisée lorsque les noeuds des q chaînes sont tous distincts, dans l'intervalle de noeuds de rayon $u-1$ autour de la cible et aussi proche que possible de $\mathbf{u}$, selon la distance dans la grille. La figure 2.6 illustre cette minoration : les sommets de Q sont représentés en blanc et entourés de leurs zones de sécurité de rayon r sur le schéma supérieur ; le schéma inférieur représente le cas de figure où la probabilité de l'événement $\mathcal{E}$ est minimale car les zones sont plus proches de l'origine du lien.

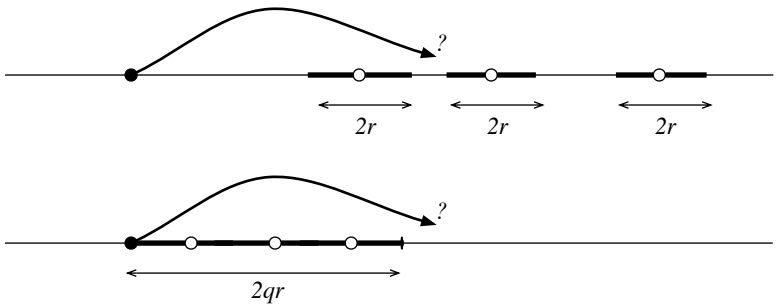


FIG. 2.6 : Minimiser la probabilité de recoupement (lemme 2.5).

Le placement des $2rq$ noeuds interdits sur l'anneau qui correspond à la probabilité minimale est sur la zone qui est à la fois à distance inférieure à u de $\mathbf{t}$ et le plus proche possible de $\mathbf{u}$. Nous devons alors examiner trois cas, selon le placement de $\mathbf{u}$ sur l'anneau. La figure 2.7 illustre les trois cas $u \leqslant n/2$ ou $n/2 \leqslant u \leqslant n - rq$ ou $u \geqslant n - rq$, et la position de la zone de noeuds interdite associée qui minimise la probabilité.

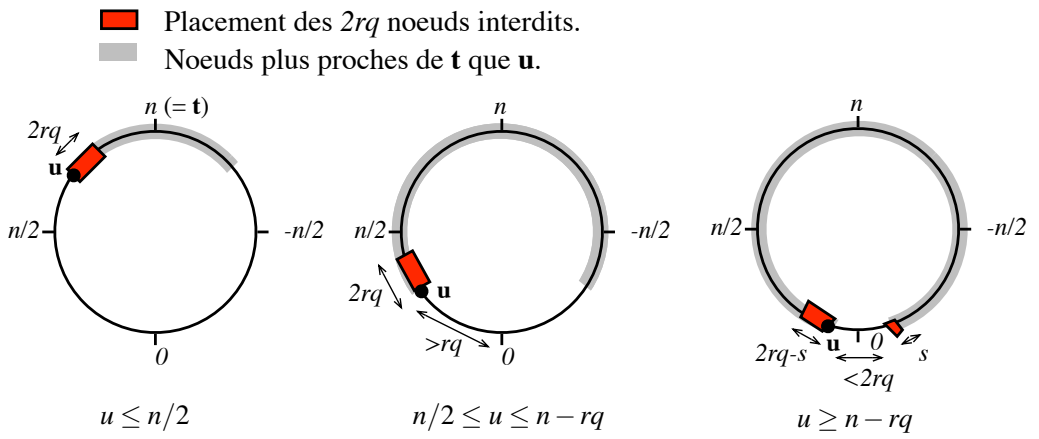


FIG. 2.7 : Illustration de la preuve du lemme 2.5

On conclut que la probabilité de $\mathcal{E}$ est supérieure à la probabilité que $\mathbf{v}$ soit à distance supérieure ou égale à $2rq$ de $\mathbf{u}$ et à distance strictement inférieure à u de la cible. Ainsi :

$$\Pr \mathcal{E} \geqslant \frac{1}{2H_n} \sum_{i=2rq}^{2u-1} \frac{1}{i} = \frac{H_{2u-1} - H_{2rq-1}}{2H_n} \geqslant \frac{H_{2u} - H_{2rq}}{2H_n}.$$

□

On peut déduire de ce lemme un corollaire nous donnant une minoration α^- de la probabilité de créer une branche pour chaque niveau de l'exploration qui pourra être utilisée pour la comparaison avec un processus de branchement classique.

Corollaire 2.6 *Pour nos valeurs de $b_{\max}(x)$, $h_{\max}(x)$ et $g(x)$, il existe une constante n_0 indépendante de x, n et k, telle que si $n \geqslant n_0$, durant chaque phase d'exploration, pour tout noeud $\mathbf{u}$ non encore visité, à distance u strictement supérieure à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible, la probabilité $\alpha_{\mathbf{u}}$ que le j-ème contact longue-distance $\mathbf{v}$ de $\mathbf{u}$ soit valide et à distance $\geqslant h^1_{\max}(x) + g(x)$ de tout ensemble de $b_{\max}(x)$ noeuds est supérieure à :*

$$\frac{\log x}{6 \log n} =_{\text{def}} \alpha^-.$$

Preuve. En appliquant le lemme 2.5 avec $Q = F$ et $r = h_{\max}(x) + g(x)$, on obtient :

$$\alpha_{\mathbf{u}} \geqslant \frac{H_{2u} - H_{2b_{\max}(x)(h_{\max}(x)+g(x))}}{2H_n}.$$

Comme on a :

$$h_{\max}(x) = \frac{\log\log x - \log\log\log n}{\log(1 + kH_x/(6H_n))},$$

et $\log(1 + kz) \geqslant z \log \hat{k}$ pour tout $0 \leqslant z \leqslant 1$, on obtient :

$$h_{\max}(x) \leqslant \frac{6H_n}{H_x} \frac{\log\log x}{\log \hat{k}} \leqslant 6g(x).$$

Et comme $b_{\max}(x)g(x) = \log n$, on a : $2b_{\max}(x)(h_{\max}(x) + g(x)) \leqslant 14 \log n$, ainsi :

$$\alpha_{\mathbf{u}} \geqslant \frac{H_{2u} - H_{14\log n}}{2H_n} \geqslant \log\left(\frac{x(\log \hat{k})/k}{14\log n}\right) \Big/ (2\log n),$$

puisque $u \geqslant x(\log \hat{k})/(2k)$.

On peut alors réécrire $x/k = (x^{2/3}/k) \cdot x^{1/3}$. Comme $x > k\log^2 n$, on a $x^{2/3} > k^{2/3}\log^{4/3} n$. Ainsi : $x/k > ((\log^{4/3} n)/k^{1/3}) \cdot x^{1/3}$. Comme $k \leqslant \log n$, on a $x/k > \log n \cdot x^{1/3}$. On conclut que, pour $n \geqslant n_0$, pour une constante absolue n_0 (indépendante de x, n et k) :

$$\alpha_u > \frac{\log(x^{1/3})}{2\log n} \geqslant \frac{\log x}{6\log n}.$$

□

Le lemme suivant montre qu'à la fin d'une phase d'exploration, avec probabilité constante :

- soit un noeud de $A \cup C$ à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible a déjà été atteint,
- soit l'arbre est suffisamment large pour contenir le nombre requis de noeuds pour pouvoir appliquer le lemme 2.4.

Nous donnons ce lemme pour des valeurs quelconques des paramètres $h_{\max}(x)$ et $g(x)$, ce qui nous permettra de le généraliser directement à la dimension d. Il s'applique en particulier avec les valeurs des paramètres de la dimension 1 en prenant $\beta^- = \alpha^-$.

Lemme 2.7 *Soient $b_{\max}(x) = \log x/\log\log n$ et β^- un minorant de la probabilité que le j-ème contact longue-distance d'un noeud $\mathbf{u}$ à distance comprise entre $\frac{\log \hat{k}}{2k} \cdot x$ et x de la cible soit valide et à distance supérieure ou égale à $h_{\max}(x) + g(x)$ de tout ensemble de $b_{\max}(x)$ noeuds, pour $h_{\max}(x)$ et $g(x)$ deux valeurs satisfaisant :*

$$(1 + k\beta^-)^{h_{\max}(x)} = \frac{\log x}{\log\log n}.$$

Alors, à la fin de toute phase d'exploration, avec probabilité au moins $\frac{1}{5}$:

– *soit il existe un noeud dans A à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible,*
– *soit $|B_{h_{\text{stop}}}| \geqslant \frac{\log x}{2 \log \log n}$.*

Preuve. Soit $\mathcal{E}$ l'événement : « à la fin d'une phase d'exploration, il existe un noeud de A à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible dans la grille, ou $|B_{h_{\text{stop}}}| \geqslant \log x/(2 \log \log n)$ ».

Soit $Z = \{\mathbf{z} : |\mathbf{z} - \mathbf{t}] > \frac{\log \hat{k}}{2k} x\}$ et $\bar{Z}$ son complémentaire.
Durant toute phase d'exploration, comme F contient moins de $b_{\max}(x)$ noeuds, pour tout noeud $\mathbf{u} \in Z$ non encore visité, la probabilité que le j-ème contact longue-distance de $\mathbf{u}$ soit valide et à distance supérieure ou égale à $h_{\max}(x) + g(x)$ de tout noeud de F est supérieure à β^-. Ainsi, tant que l'on considère les noeuds de Z, chacun de leurs contacts longue-distance sera ajouté à B_{h+1} avec probabilité supérieure ou égale à β^-. Dès qu'un noeud $\mathbf{u}$ de $\bar{Z}$ est inséré dans A_h, pour un h donné, la probabilité que, pour un j donné, son j-ème contact longue-distance soit valide et à distance $\geqslant h_{\max}(x) + g(x)$ de tout noeud de l'ensemble F courant dans la grille n'est plus minorée par β^- ; mais l'événement $\mathcal{E}$ est vérifié.

Nous utilisons un argument de couplage probabiliste pour minorer la probabilité de $\mathcal{E}$, en exécutant virtuellement une phase d'exploration sur un réseau fictif, construit à partir du réseau original. Ce réseau est construit de la façon suivante : il a la même grille sous-jacente ; les noeuds de Z ont exactement les même liens que dans le réseau original ; mais on considère une distribution virtuelle de liens pour les noeuds de $\bar{Z}$ telle que pour tout noeud $\mathbf{u}$ non encore visité, la probabilité que son j-ème contact longue-distance soit valide et à distance supérieure ou égale à $h_{\max}(x) + g(x)$ de tout ensemble de noeuds de taille inférieure ou égale à $b_{\max}(x)$ est β^- (on remarquera qu'il n'est pas utile que cette distribution existe réellement). On exécute la phase d'exploration sur ce réseau fictif depuis le même $\mathbf{x}$ que dans le réseau réel, à l'exception du fait que l'on ne l'interrompt pas avant que $h = h_{\max}(x)$. Cela donne trois familles d'ensembles (A'_h), (B'_h) et F', telles que : $A_h \cap Z = A'_h \cap Z$, $B_h \cap Z = B'_h \cap Z$, et $F \cap Z = F' \cap Z$, pour tout $1 \leqslant h \leqslant h_{\text{stop}}$.

Les liens suivis durant la phase d'exploration du réseau fictif définissent une structure d'arbre qui ne se recoupe pas, de hauteur exactement $h^1_{\max}(x)$, sur l'ensemble des noeuds $A' = \cup_h A'_h$, où B'_h est l'ensemble des noeuds au niveau h. Soit $\mathcal{E}'$ l'événement « $|B'_{h_{\max}(x)}| \geqslant \log x/(2 \log \log n)$ ».
On montre à présent que $\Pr \mathcal{E} \geqslant \Pr \mathcal{E}'$:

– Si, dans le réseau original, $A \cap \bar{Z} = \varnothing$, alors $B'_{h_{\text{stop}}} = B_{h_{\text{stop}}}$. Si $h_{\text{stop}} < h_{\max}(x)$, alors

$$|B'_{h_{\max}(x)}| \geqslant |B'_{h_{\text{stop}}}| = |B_{h_{\text{stop}}}| = b_{\max}(x) = \log x / \log \log n,$$

et $\mathcal{E}$ et $\mathcal{E}'$ sont tous les deux vérifiés. Si $h_{\text{stop}} = h_{\max}(x)$, $B_{h_{\text{stop}}} = B'_{h_{\max}(x)}$ et

$\mathcal{E}$ et $\mathcal{E}'$ sont alors équivalents. Ainsi, quel que soit le réseau fictif à l'intérieur de $\bar{Z}$,

$$\Pr\{\mathcal{E} \mid A \cap \bar{Z} = \varnothing\} = \Pr\{\mathcal{E}' \mid A \cap \bar{Z} = \varnothing\}.$$

– Si, dans le réseau original, $A \cap \bar{Z} \neq \varnothing$, alors $\mathcal{E}$ est vérifié, et quel que soit le réseau fictif à l'intérieur de $\bar{Z}$,

$$\Pr\{\mathcal{E} \mid A \cap \bar{Z} \neq \varnothing\} = 1 \geqslant \Pr\{\mathcal{E}' \mid A \cap \bar{Z} \neq \varnothing\}.$$

On minore à présent $\Pr \mathcal{E}'$. L'ensemble $A' = \cup_h A'_h$ est structuré comme un arbre aléatoire de racine $\mathbf{x}$, dans lequel tout noeud $\mathbf{u}$, au niveau h, a indépendamment un nombre aléatoire $1 + l$ de fils (un contact local et l contacts longue-distance), où l est donné par une loi binomiale de paramètres $(k, \alpha_{\mathbf{u}})$, avec $\alpha_{\mathbf{u}} \geqslant \beta^-$. Ainsi, le nombre de noeuds au niveau h, $|B'_h|$, domine stochastiquement la variable aléatoire b_h du nombre de noeuds au h-ième niveau[2] dans le processus de branchement suivant : commencer avec un noeud, au niveau h, chaque noeud au niveau $h - 1$ reçoit, indépendamment, exactement $1 + l$ fils, avec probabilité $\rho_l = \binom{k}{l}(\beta^-)^l(1 - \beta^-)^{k-l}$, où $0 \leqslant l \leqslant k$. D'après le lemme 2.1, on a : $\mathbb{E}[b_h] = (1 + k\beta^-)^h$, puis, par la méthode du second moment (lemme 2.3), avec probabilité au moins $\frac{1}{5}$, $b_h \geqslant \mathbb{E}[b_h]/2$. Ainsi, puisque $(1 + k\beta^-)^{h_{\max}(x)} = \log x / \log\log n$, on conclut que

$$\Pr \mathcal{E} \geqslant \Pr \mathcal{E}' = \Pr\{|B'_{h^1_{\max}(x)}| \geqslant \log x/(2 \log\log n)\} \geqslant \frac{1}{5}.$$

□

En combinant les lemmes 2.6 et 2.7, on obtient le corollaire suivant :

Corollaire 2.8 *Pour $n \geqslant n_0$, à la fin de toute phase d'exploration, avec probabilité au moins $\frac{1}{5}$,*

- *soit il existe un noeud dans A à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible,*
- *soit il y a plus de $\frac{\log n}{2 \log \hat{k}}$ noeuds distincts non encore visités dans C.*

Preuve. D'après le corollaire 2.6, pour les valeurs des paramètres de $h_{\max}(x)$ et $g(x)$ données pour la dimension 1, α^- est un minorant de la probabilité que le j-ème contact longue-distance d'un noeud $\mathbf{u}$, à distance comprise entre $\frac{\log \hat{k}}{2k} \cdot x$ et x de la cible, soit valide et à distance supérieure ou égale à $h_{\max}(x) + g(x)$ de tout ensemble de $b_{\max}(x)$ noeuds. Comme ces valeurs vérifient $(1 + k\alpha^-)^{h_{\max}(x)} = \log x \log\log n$, on peut appliquer le lemme 2.7. Ainsi, avec probabilité au moins $\frac{1}{5}$,

[2] i.e., pour tout z, $\Pr\{|B'_h| \geqslant z\} \geqslant \Pr\{b_h \geqslant z\}$.

– soit il existe un noeud de A à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible,
– soit $|B_{h_{\text{stop}}}| \geqslant \log n/(2 \log \log n)$.

Dans ce dernier cas, puisque $A \cup C$ est composé de $|B_{h_{\text{stop}}}|$ chaînes de longueur inférieure ou égale à $h_{\max}(x) + g(x)$ enracinées en les noeuds de F et puisque les noeuds de F sont à distance $\geqslant h^1_{\max}(x) + g(x)$ les uns des autres dans la grille, il y a exactement $|B_{h_{\text{stop}}}| \cdot g(x) \geqslant \frac{\log n}{2 \log \hat{k}}$ noeuds distincts non encore visités dans C. □

Nous pouvons à présent conclure qu'avec une probabilité constante, on trouve un lien longue-distance qui mène $\frac{2k}{\log \hat{k}}$ fois plus près de la cible.

Proposition 2.9 *Il existe deux constantes $p_1 > 0$ et n_0, indépendantes de n et $\mathbf{x}$, telles que, pour $n \geqslant n_0$, à la fin de toute phase d'exploration, avec probabilité supérieure à p_1, il existe un noeud $\mathbf{u}$ dans $A \cup C$ tel que $\mathbf{u}$ ou l'un de ses contacts longue-distance soit à distance $\leqslant \frac{\log \hat{k}}{2k} \cdot x$ de la cible dans la grille.*

Preuve. D'après le corollaire 2.8, avec probabilité au moins $\frac{1}{5}$, si aucun noeud de A n'est à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible, alors C contient $\frac{\log n}{2 \log \hat{k}}$ noeuds distincts et aucun de leurs contacts longue-distance n'a encore été visité. Dans ce cas, soit il existe un noeud de C à distance inférieure ou égale à $\frac{\log \hat{k}}{2k} \cdot x$ de la cible, soit, d'après le lemme 2.4, avec probabilité supérieure à p'_1, au moins un noeud de C a un contact longue-distance à distance $\leqslant \frac{\log \hat{k}}{2k} \cdot x$ de la cible. Puisque les liens longue-distance de C n'ont pas encore été visités, ces événements sont indépendants et en posant $p_1 = p'_1/5$, on conclut la preuve. □

Ce résultat nous permet de prouver le théorème suivant :

Théorème 2.10 (Analyse de l'algorithme 2)
Pour tout $1 \leqslant k \leqslant \log n$, l'algorithme 2 requiert $\Theta(\log^2 n/\log \hat{k})$ bits de mémoire, et calcule, pour toute paire de noeuds $(\mathbf{s}, \mathbf{t})$ de $\mathcal{K}^1_{k,n}$, un chemin de $\mathbf{s}$ à $\mathbf{t}$ dont l'espérance de la longueur vaut

$$O\left(\log n \left(\frac{\log \log n}{\log(1+k)}\right)^2\right).$$

De plus, il visite $O\left((\log n/\log(1+k))^2\right)$ noeuds en espérance pour calculer ce chemin.

Preuve. Sans perte de généralité, on peut supposer que la cible est $\mathbf{0}$ et que la source est $\mathbf{s}$, à distance s de $\mathbf{0}$. Soient $\mathbf{x}$ le porteur courant du message et x sa distance à la cible. Rappelons tout d'abord qu'à la fin d'une phase d'exploration, l'algorithme sélectionne le noeud le plus proche de la cible parmi les contacts

locaux ou longue-distance de $A \cup C$ et que l'ensemble $A \cup C$ s'étend en direction de la cible ; ainsi, chaque phase d'exploration visite des noeuds qui n'ont jamais été explorés auparavant. Chaque phase d'exploration est indépendante des précédentes.

L'exécution de l'algorithme se décompose en T phases successives de division de la distance courante à la cible par $\frac{2k}{\log \hat{k}}$ ($\hat{k} = k+1$). T est la partie entière de la solution T' de $(\frac{2k}{\log(1+k)})^{T'} = s$. Lorsque l'on atteint une distance à la cible inférieure à stop, on exécute l'algorithme glouton, cela se produit après U divisions de la distance par $\frac{2k}{\log \hat{k}}$, où U est la partie entière de la solution U' de l'équation $(\frac{2k}{\log(1+k)})^{U'} = \mathrm{stop}$. On remarquera que :

$$T \sim \frac{\log s}{\log(1+k)} \quad \text{et} \quad U \sim \frac{3 \log\log n + \log k}{\log(1+k)}.$$

La phase i, $0 \leqslant i \leqslant T$, dure tant que $(\frac{2k}{\log \hat{k}})^{i-1} < x \leqslant (\frac{2k}{\log \hat{k}})^{i}$. On dit qu'une étape d'exploration en phase i *réussit* si elle conduit à une phase $j \leqslant i-1$. Soient Y_i et Z_i les variables aléatoires pour le nombre de noeuds visités durant la phase i et pour la longueur du chemin par lequel le message est routé durant la phase i respectivement.

Supposons que nous soyons en phase i, avec $T \geqslant i > U$, alors $x > \mathrm{stop}$. Selon la proposition 2.9, chaque étape d'exploration réussit avec probabilité supérieure ou égale à p_1. Chaque étape d'exploration visite moins de $\big(h_{\max}(x) + g(x)\big)\, b_{\max}(x) \leqslant 7g(x) b_{\max}(x)$ noeuds et route le message le long d'un chemin de longueur inférieure ou égale à $h_{\max}(x) + g(x) \leqslant 7g(x)$ vers la cible. On obtient :

$$\begin{aligned}
\mathbb{E}[Y_i] &\leqslant \sum_{j \geqslant 1} \Big(\big(j \cdot 7g(x) b_{\max}(x)\big)\, (1-p_1)^{j-1} \Big) \\
&\leqslant \frac{7g(x) b_{\max}(x)}{p_1} \leqslant \frac{7}{p_1} \frac{\log n}{\log \hat{k}} \\
\text{et } \mathbb{E}[Z_i] &\leqslant \frac{7g(x)}{p_1} \leqslant \frac{7}{p_1} \frac{\log n \log\log n}{i \log^2 \hat{k}},
\end{aligned}$$

puisque $\log x \geqslant i \log(1+k)$. Une fois que l'on a atteint une phase $i \leqslant U$, on a $x \leqslant \mathrm{stop} = k \log^2 n$ et l'algorithme exécute l'algorithme glouton $\mathcal{G}$ de Kleinberg. D'après [Kle00], nous savons que, pour une certaine constante γ, l'algorithme glouton calcule alors un chemin de longueur inférieure à

$$\gamma \log n \log(\mathrm{stop})/k \leqslant 3\gamma(\log n \log\log n)/k,$$

en espérance. De plus, il visite moins de $3\gamma(\log n \log\log n)/k$ noeuds en espérance.

Posons $x_i = (\frac{2k}{\log \hat{k}})^i$. On a $\log x_i \geqslant i \log \hat{k}/2$, ainsi :

$$\begin{aligned} h_{\max}(x_i) &\leqslant 6g(x_i) \\ \text{et } g(x_i) &\leqslant \frac{2 \log n \log\log n}{i \log^2 \hat{k}}. \end{aligned}$$

La longueur moyenne du chemin de **s** à **0** calculée par l'algorithme est majorée par :

$$\begin{aligned} \sum_{i=0}^{T} \mathbb{E}[Z_i] &\leqslant \gamma \frac{\log n \log(\text{stop})}{k} + \frac{1}{p_1} \sum_{U<i\leqslant T} (h_{\max}(x_i) + g(x_i)) \\ &\leqslant 3\gamma \, \frac{\log n \log\log n}{k} + \frac{14}{p_1} \frac{\log n \log\log n}{\log^2 \hat{k}} \sum_{U<i\leqslant T} \frac{1}{i} \\ &= O\left(\log n \left(\frac{\log\log n}{\log \hat{k}} \right)^2 \right). \end{aligned}$$

Le nombre moyen de noeuds visités par l'algorithme est majoré par :

$$\begin{aligned} \sum_{i=0}^{T} \mathbb{E}[Y_i] &= \frac{1}{p_1} \sum_{i=0}^{T} (h_{\max}(x_i) + g(x_i)) \, b_{\max}(x_i) \\ &\leqslant \frac{14}{p_1} \frac{\log n \log\log n}{\log^2 \hat{k}} \sum_{i=0}^{T} \frac{\log x_i}{\log\log n} \, \frac{1}{i} \\ &= O\left(\left(\frac{\log n}{\log \hat{k}} \right)^2 \right). \end{aligned}$$

Pour la durée de chaque étape d'exploration, l'implémentation selon le parcours en profondeur de l'algorithme requiert seulement

$$\Theta\left(\log n \cdot (b_{\max}(x) + h_{\max}(x) + g(x))\right) = O(\log^2 n / \log \hat{k})$$

bits de mémoire. En effet, chaque adresse de noeud requiert $O(\log n)$ bits et chaque étape d'exploration requiert seulement le stockage de l'adresse de la cible, de l'adresse des noeuds de F (dont le nombre est inférieur à $b_{\max}(x) = O(\log n / \log\log n) = O(\log n / \log \hat{k})$), ainsi que l'état de la pile durant le parcours en profondeur de $A \cup C$ (dont la profondeur est majorée par $h_{\max}(x) + g(x) = O(\log n / \log \hat{k})$) et le chemin de **x** au meilleur noeud courant **y** parmi les noeuds de $A \cup C$ et leurs contacts. □

2.5 Généralisation à toute dimension

Rappelons que dans un réseau d-dimensionnel, $d > 1$, la grille sous-jacente est le tore d-dimensionnel $\{-n, \ldots, n\}^d$. Chaque noeud $\mathbf{u}$ a k liens longue-distance, le j-ème lien ($1 \leqslant j \leqslant k$) pointant vers un noeud $\mathbf{v}_j$ choisi indépendamment selon la distribution d-harmonique, *i.e.* avec probabilité $1/(Z_{n,d} \cdot |\mathbf{u} - \mathbf{v}|^d)$.

$Z_{n,d}$ est le facteur de normalisation calculé dans le préambule, p. 22, qui satisfait :

$$\frac{2^d}{(d-1)!} H_n \leqslant Z_{n,d} \leqslant \frac{2^d}{(d-1)!} H_n + O(1).$$

L'algorithme pour les réseaux d-dimensionnels. Pour généraliser l'algorithme 2 à toute dimension, il suffit d'adapter les paramètres de l'exploration de la façon suivante, le reste de l'algorithme est inchangé.

Définition 2.4 (Paramètres pour l'algorithme 2 en dimension quelconque)
Les valeurs des paramètres pour l'algorithme 2 en dimension d sont, pour chaque exploration enracinée en un noeud $\mathbf{x}$ à distance x de la cible :

$$h^d_{\max}(x) = \frac{\log\log x - \log\log\log n}{\log\left(1 + ck\left(H_x/H_n\right)\right)},$$
$$\mathrm{stop}_d = k \log^{2^d+1} n,$$

où $c > 0$ est une constante telle que $c \leqslant \frac{1}{2^{d+3}}$. Les fonctions $b_{\max}(x)$ et $g(x)$ sont les mêmes qu'en dimension 1.

Barière *et al.* [BFKK01] ont démontré que le réseau de Kleinberg d-dimensionnel présentait des propriétés de routage similaires au réseau unidimensionnel. En particulier, la visite de $\Omega(\log n/k)$ nouveaux noeuds à distance comprise entre x et $x/2$ de la cible garantit de trouver un contact longue-distance à distance inférieure à $x/2$ de la cible avec probabilité constante. Nous montrons que nous pouvons visiter ces nouveaux noeuds dans une structure d'exploration compacte et obtenir une longueur de chemin de même espérance que celle obtenue en dimension 1, avec le même algorithme. Cela confirme que la dimension a peu d'effets sur le routage décentralisé.

La structure des boules dans le tore d-dimensionnel rend toutefois l'analyse plus difficile. En effet, on ne peut plus minorer la probabilité de recoupement dans la structure d'exploration en considérant que des ensembles de noeuds interdits (zones de sécurité) sont placés sur une chaîne de liens locaux dirigée vers la cible. Dans la sous-section suivante, nous analysons les propriétés géométriques des boules et sphères du tore d-dimensionnel. Cela nous permettra de déterminer le

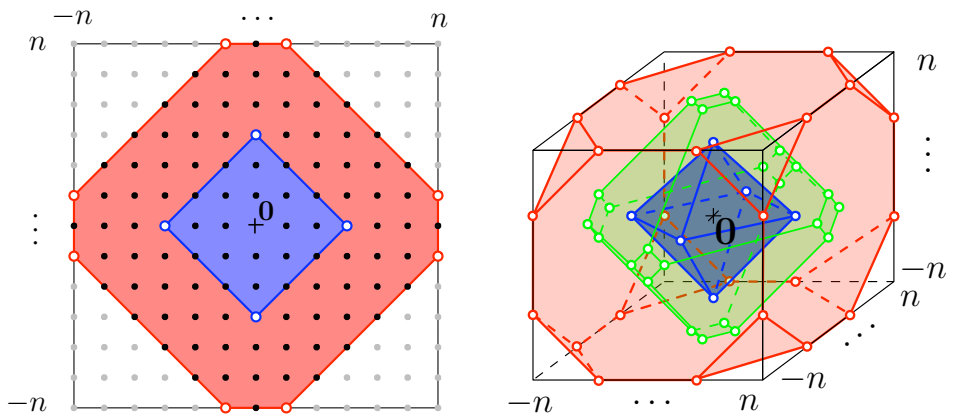


FIG. 2.8 : Deux exemples de coins (points blancs) des boules centrées en $\mathbf{0}$ dans le tore (2D et 3D).

placement d'un ensemble de noeuds interdits qui minimise la probabilité qu'un lien longue-distance les évite.

Par ailleurs, nous exprimons, dans notre analyse, la dépendance des constantes en d, ce qui permet d'évaluer la perte d'efficacité lorsque la dimension augmente. Cette dépendance est rarement explicitée dans la littérature du routage dans le réseau de Kleinberg.

Rappelons que l'on note $\mathcal{S}(\mathbf{u}, r)$ et $\mathcal{B}(\mathbf{u}, r)$ la sphère ℓ_1 et la boule ℓ_1 centrées sur $\mathbf{u}$ et de rayon r. On note respectivement $s(r)$ et $v(r)$ leur cardinalité.

2.5.1 Lemmes géométriques préliminaires

Dans le tore 2-dimensionnel, les sphères ont des formes en *diamant* et présentent des *coins*. Pour un noeud placé sur une sphère centrée sur la cible, la probabilité que son contact longue-distance soit strictement plus proche de la cible est minimale lorsqu'il est placé sur un de ces coins. Nous montrons ici ce résultat en dimension $d \geqslant 1$. Cela nous permettra, dans la sous-section suivante, de minorer la probabilité de recoupement d'une branche en plaçant l'origine du lien sur un coin de la sphère centrée sur la cible, de rayon maximal pour la phase considérée.

Définition 2.5 (Coin d'une sphère)
On dit qu'un noeud $\mathbf{u}$ *est un* coin *de la sphère* $\mathcal{S}(\mathbf{0}, r)$ *si* $\mathbf{u} \in \mathcal{S}(\mathbf{0}, r)$ *et si au plus une des coordonnées de* $\mathbf{u}$ *n'appartient pas à l'ensemble* $\{-min(r, n), 0, min(r, n)\}$ *(voir Fig. 2.8) .*
Les coins d'une sphère $\mathcal{S}(\mathbf{v}, r)$ *centrée en un noeud arbitraire* $\mathbf{v}$ *sont définis par translation du vecteur* $\mathbf{v}$.

On remarquera que lorsque $r \leqslant n$, $\mathcal{S}(\mathbf{v}, r)$ est un « octahèdre d-dimensionnel »et que notre définition de coins correspond à la définition géométrique usuelle.

Lemme 2.11 *Soit* $\mathbf{u}$ *un noeud à distance* u *de la cible. Pour tout* $r \geqslant 0$*, la probabilité que son* j*-ème contact longue-distance appartienne à la boule* $\mathcal{B}(\mathbf{t}, r)$ *est minimale lorsque* $\mathbf{u}$ *est placé en un coin de la sphère* $\mathcal{S}(\mathbf{t}, u)$ *centrée sur la cible* $\mathbf{t}$ *et de rayon* u*.*

Preuve. On peut supposer sans perte de généralité que la cible est $\mathbf{0}$. On considère un sommet $\mathbf{u} = (u_1, \ldots, u_d)$ qui n'est pas un coin de $\mathcal{S}(\mathbf{0}, u)$. Sans perte de généralité, $n > u_1 \geqslant u_2 > 0$. Soit $\mathbf{v} = (u_1 + 1, u_2 - 1, u_3, \ldots, u_d) \in \mathcal{S}(\mathbf{0}, u)$. Soit $\rho_{\mathbf{u}}$ (resp., $\rho_{\mathbf{v}}$) la probabilité que le j-ème contact longue-distance de $\mathbf{u}$ (resp., de $\mathbf{v}$) appartienne à $\mathcal{B}(\mathbf{0}, r)$. Nous allons montrer que $\rho_{\mathbf{v}} \leqslant \rho_{\mathbf{u}}$. En itérant ce procédé un nombre fini de fois, on pousse $\mathbf{v}$ dans un coin de la sphère, ce qui conduit au résultat.

Pour prouver le résultat, on construit une bijection $\phi : \mathcal{B}(\mathbf{0}, r) \to \mathcal{B}(\mathbf{0}, r)$, telle que pour tout $\mathbf{z} \in \mathcal{B}(\mathbf{0}, r)$, $\phi(\mathbf{z})$ soit plus loin de $\mathbf{v}$ que de $\mathbf{u}$, *i.e.* $|\phi(\mathbf{z}) - \mathbf{v}| \geqslant |\phi(\mathbf{z}) - \mathbf{u}|$. En fait, pour la plupart des noeuds $\mathbf{z}$ de la boule $\mathcal{B}(\mathbf{0}, r)$, la distance $|\phi(\mathbf{z}) - \mathbf{v}|$ sera égale à $|\mathbf{z} - \mathbf{u}|$; et pour les autres noeuds $\mathbf{z}$, $|\phi(\mathbf{z}) - \mathbf{v}|$ sera supérieure ou égale à $|\mathbf{z} - \mathbf{u}|$. Puisque la probabilité qu'un noeud $\mathbf{z}$ soit un contact longue-distance de $\mathbf{u}$ décroît lorsque la distance $|\mathbf{z} - \mathbf{u}|$ augmente, on a :

$$\begin{aligned}\rho_{\mathbf{v}} &= \sum_{\mathbf{z} \in \mathcal{B}(\mathbf{0}, r)} \frac{1}{|\mathbf{z} - \mathbf{v}| Z_{n,d}} = \sum_{\mathbf{z} \in \mathcal{B}(\mathbf{0}, r)} \frac{1}{|\phi(\mathbf{z}) - \mathbf{v}| Z_{n,d}} \\ &\leqslant \sum_{\mathbf{z} \in \mathcal{B}(\mathbf{0}, r)} \frac{1}{|\mathbf{z} - \mathbf{u}| Z_{n,d}} = \rho_{\mathbf{u}}.\end{aligned}$$

La construction de ϕ se fait à partir de la translation du vecteur $(1, -1, 0, \ldots, 0)$ sur le tore. L'action de ϕ est illustrée par la figure 2.9. On pose :

$$\Gamma_1 = \{\mathbf{z} \in \mathcal{S}(\mathbf{0}, r) \cup \mathcal{S}(\mathbf{0}, r-1) : z_1 \geqslant 0 \text{ et } z_2 \leqslant 0\}$$

$$\text{et } \Gamma_2 = \{\mathbf{z} \in \mathcal{S}(\mathbf{0}, r) \cup \mathcal{S}(\mathbf{0}, r-1) : z_1 \leqslant 0 \text{ et } z_2 \geqslant 0\}.$$

On remarquera que $\Gamma_1 \cap \Gamma_2 = \varnothing$ et que Γ_1 et Γ_2 sont symétriques l'un de l'autre par rapport à $\mathbf{0}$. On définit $\phi : \mathcal{B}(\mathbf{0}, r) \to \mathcal{B}(\mathbf{0}, r)$ de la façon suivante :

$$\phi : \mathbf{z} \mapsto \begin{cases} (z_1 + 1, z_2 - 1, z_3, z_4, \ldots, z_d) & \text{si } \mathbf{z} \notin \Gamma_1, \\ (z_2, z_1, z_3, z_4 \ldots, z_d) & \text{sinon.} \end{cases}$$

ϕ est bien une bijection puisque : $\mathbf{z} \in \Gamma_1 \Leftrightarrow \phi(\mathbf{z}) \in \Gamma_2$ et la restriction de ϕ à $\mathcal{B}(\mathbf{0}, r) \setminus \Gamma_1$ est une translation vers $\mathcal{B}(\mathbf{0}, r) \setminus \Gamma_2$. En particulier, pour $\mathbf{z} \in$

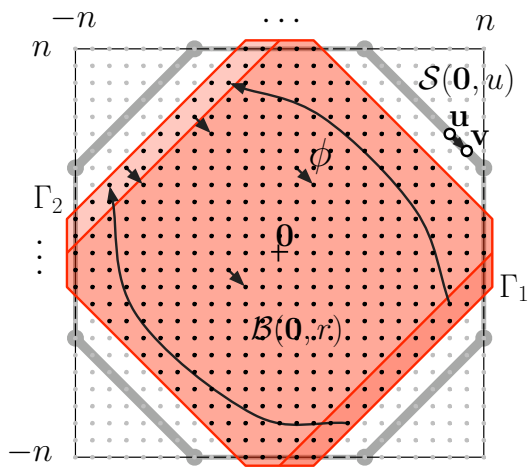


FIG. 2.9 : Γ_1, Γ_2 et l'action de ϕ (représentée par les flèches noires).

$\mathcal{B}(\mathbf{0},r) \setminus \Gamma_1$, on a $|\mathbf{z}-\mathbf{u}| = |\phi(\mathbf{z})-\mathbf{v}|$.
On considère maintenant $\mathbf{z} \in \Gamma_1$ et l'on pose :

$$\Delta = |\phi(\mathbf{z}),\mathbf{v}| - |\mathbf{z}-\mathbf{u}| = |z_2-1-u_1| + |z_1+1-u_2| - |z_1-u_1| - |z_2-u_2|.$$

Puisque $z_1 \geqslant 0$, $z_2 \leqslant 0$ et $0 < u_2 \leqslant u_1$, on a :

$$\Delta = 1 + u_1 - u_2 + |z_1+1-u_2| - |z_1-u_1|.$$

Ainsi,
- si $z_1 \leqslant u_2 - 1$, alors $\Delta = 0$,
- si $u_2 - 1 \leqslant z_1 \leqslant u_1$, $\Delta = 2(1-u_2+z_1) \geqslant 0$,
- et enfin, si $z_1 \geqslant u_1$, $\Delta = 2(1+u_1-u_2) \geqslant 2$.

Δ est donc toujours positif, ce qui conclut la preuve. □

Lemme 2.12 *Pour tout $1 \leqslant x \leqslant \min(r,n)$, le nombre de noeuds de $\mathcal{B}(\mathbf{0},r)$ à distance x d'un coin de $\mathcal{S}(\mathbf{0},r)$ est supérieur à $x^{d-1}/(d-1)!$.*

Preuve. Supposons dans un premier temps que r est inférieur ou égal à n. Sans perte de généralité, on peut supposer que le coin considéré est $\mathbf{c} = (r,0,\ldots,0)$. Soit $\mathbf{c}' = (r - \lfloor x/2 \rfloor, 0, \ldots, 0)$. Pour $x \leqslant r$, tout noeud $\mathbf{y}$ de $\mathcal{S}(\mathbf{c}', \lfloor x/2 \rfloor)$ tel que y_1 est inférieur ou égal à $r - \lfloor x/2 \rfloor$ appartient à $\mathcal{B}(\mathbf{0},r)$ et est à distance x de $\mathbf{c}$ (voir la figure 2.10). Il y a $\frac{1}{2}s(\lfloor x/2 \rfloor)$ noeuds de ce type et $\frac{1}{2}s(\lfloor x/2 \rfloor)$ est supérieur à $x^{d-1}/(d-1)!$.

Supposons à présent que $r > n$, soit $\mu = \lfloor r/n \rfloor$ et $x \leqslant n$. Sans perte de généralité, on peut prendre comme coin $\mathbf{c} = (c_1,\ldots,c_d)$, avec
- $c_i = n$ pour tout $i \leqslant \mu$,

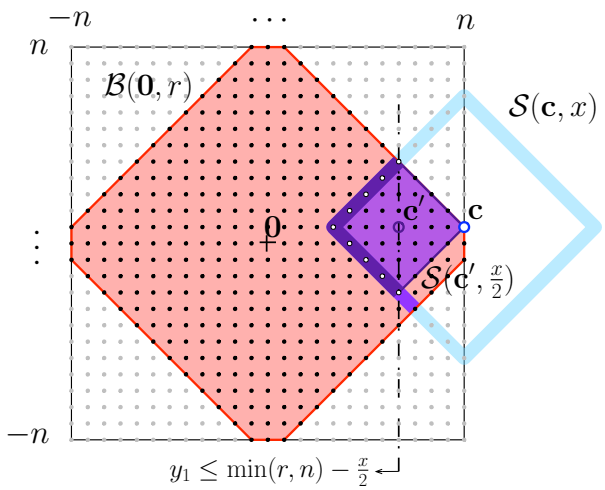


FIG. 2.10 : Illustration de la preuve du lemme 2.12.

- $c_{\mu+1} = r - \mu n$,
- et $c_i = 0$ pour tout $i > \mu + 1$.

Soit $\mathbf{c}'$ l'image de $\mathbf{c}$ par la translation de vecteur $(-\lfloor x/2 \rfloor, 0, \ldots, 0)$. Tout noeud $\mathbf{y}$ de $\mathcal{S}(\mathbf{c}', \lfloor x/2 \rfloor)$ tel que $y_1 \leqslant n - \lfloor x/2 \rfloor$ appartient à $\mathcal{B}(\mathbf{0}, r)$ et est à distance x de $\mathbf{c}$. Encore une fois, il y a $\frac{1}{2}s(x/2) \geqslant x^{d-1}/(d-1)!$ noeuds de ce type.

On en conclut que pour $x \leqslant \min(n, r)$, il y a toujours au moins $x^{d-1}/(d-1)!$ noeuds à distance x de tout coin de $\mathcal{S}(\mathbf{0}, r)$ dans $\mathcal{B}(\mathbf{0}, r)$. $\square$

2.5.2 Analyse de l'algorithme

La preuve suit le même déroulement que l'analyse unidimensionnelle de la section 2.4.2. On commence par montrer qu'avec probabilité constante, parmi $\Omega(\log n / \log \hat{k})$ noeuds, il en existe un dont le contact longue-distance est $\left(\frac{2k}{\log \hat{k}}\right)^{1/d}$ fois plus près de la cible.

Lemme 2.13 *Pour tout $0 < \lambda \leqslant 1/2$, la probabilité qu'un noeud $\mathbf{u}$ à distance u de la cible ait son j-ème contact longue-distance à distance inférieure ou égale à λu de la cible est supérieure à :*

$$\frac{\ln 2}{d^d \, 2^{2d}} \, \frac{\lambda^d}{H_n}.$$

Preuve. Soit p la probabilité à évaluer. Sans perte de généralité on peut supposer que la cible est $\mathbf{0}$. Comme il s'agit de minorer p, on peut supposer que $\mathbf{u}$ est placé sur un coin, d'après le lemme 2.11. Soit $\mathbf{c}$ le plus proche coin de $\mathcal{S}(\mathbf{0}, \lambda u)$ à distance $(1-\lambda)u$ de $\mathbf{u}$. p est minorée par la somme des probabilités que le j-ème

contact longue-distance de $\mathbf{u}$ soit à distance i de $\mathbf{c}$ et dans $\mathcal{B}(\mathbf{0}, \lambda u)$, pour i allant de 1 à $2\lambda u$ (voir la figure 2.11).

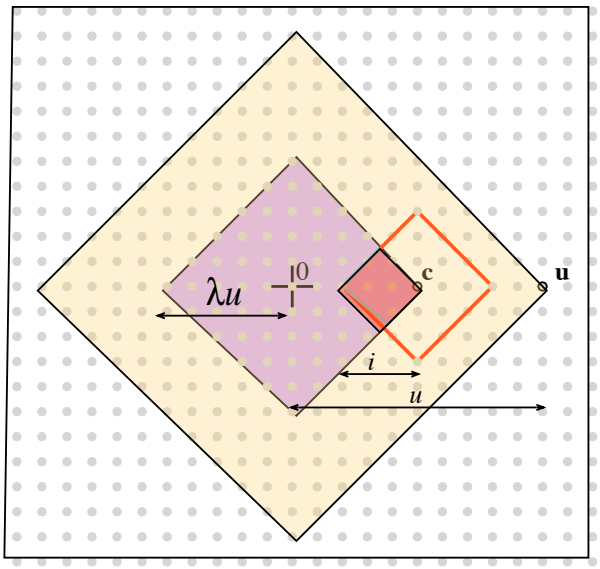


FIG. 2.11 : Illustration de la preuve du lemme 2.13.

Le lemme 2.12 garantit qu'il y a au moins $i^{d-1}/(d-1)!$ noeuds dans $\mathcal{B}(\mathbf{0}, \lambda u)$ à distance i de $\mathbf{c}$, c'est-à-dire à distance inférieure ou égale à $i + (1-\lambda)u$ de $\mathbf{u}$. Par ailleurs, il suffit de sommer les probabilités pour i allant de $\lambda u/(2d)$ à $\lambda u/d$ pour obtenir une minoration (la division par d assure que la valeur est inférieure ou égale à n), on obtient ainsi $i + (1-\lambda)u \leqslant \frac{2d}{\lambda} i$. Cela conduit à :

$$\begin{aligned} p &\geqslant \sum_{i=\lambda u/(2d)}^{\lambda u/d} \frac{i^{d-1}/(d-1)!}{H_{n,d}(i+(1-\lambda)u)^d} \geqslant \frac{\lambda^d}{(2d)^d (d-1)! H_{n,d}} \int_{\lambda u/(2d)}^{\lambda i/d} \frac{dt}{t} \\ &\geqslant \frac{\lambda^d \ln 2}{(2d)^d (d-1)! H_{n,d}} \geqslant \frac{\ln 2}{d^d\, 2^{2d}} \frac{\lambda^d}{H_n}. \end{aligned}$$

□

Corollaire 2.14 *Étant donné $\gamma > 0$, il existe une constante $p'_d > 0$, indépendante de n, x et k telle que, pour tout sous-ensemble Γ de $\gamma \cdot \frac{\log n}{\log \hat{k}}$ noeuds à distance strictement supérieure à $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ et inférieure ou égale à x de la cible, il existe au moins un sommet de Γ dont un des contacts longue-distance est à distance inférieure ou égale à $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ de la cible, avec probabilité au moins p'_d.*

Preuve. Selon le lemme 2.13, avec $\lambda = \left(\frac{\log \hat{k}}{2k}\right)^{1/d}$, la probabilité que le j-ème contact longue-distance d'un noeud $\mathbf{u}$ donné soit à distance inférieure ou égale à λu de la cible est supérieure ou égale à :

$$\frac{\ln 2}{d^d\ 2^{2d+1}} \frac{\log(1+k)}{kH_n}.$$

La probabilité que tous les $k|\Gamma|$ contacts longue-distance des noeuds de Γ soient à distance strictement supérieure à $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ de la cible est donc inférieure à

$$\left(1 - \frac{\ln 2}{d^d\ 2^{2d+1}} \frac{\log(1+k)}{kH_n}\right)^{\gamma \frac{k \log n}{\log(1+k)}} \leqslant e^{-\gamma \log 2/(d^d\ 2^{2d+1})} =_{\text{def}} 1 - p'_d < 1.$$

□

Nous allons à présent minorer la croissance du processus de branchement sous-jacent dans $A \cup C$. Nous commençons par minorer la probabilité de création d'une nouvelle branche par $\alpha_d^- = \Theta(\log x / \log n)$, de façon similaire au cas unidimensionnel (cf. lemme 2.6).

Lemme 2.15 *Il existe une constante $c_1 \geqslant 0$ ne dépendant que de d telle que, pour tout noeud $\mathbf{u}$ à distance u de la cible $\mathbf{t}$, la probabilité que son j-ème contact longue-distance $\mathbf{v}$ soit valide est au moins :*

$$\frac{H_u - c_1}{(d-1)! H_{n,d}}.$$

Preuve. D'après le lemme 2.11, la probabilité que $\mathbf{v}$ soit valide est minimale lorsque $\mathbf{u}$ est un coin de $\mathcal{S}(\mathbf{t}, u)$. D'après le lemme 2.12, le nombre de noeuds $\mathcal{B}(\mathbf{t}, u-1)$ à distance i d'un coin de $\mathcal{S}(\mathbf{t}, u-1)$ est supérieure à $\frac{i^{d-1}}{(d-1)!}$, ainsi la probabilité considérée est supérieure à :

$$\sum_{i=1}^{\min(u,n)} \left(\frac{i^{d-1}}{(d-1)!} \cdot \frac{1}{i^d Z_{n,d}}\right) \geqslant \frac{H_{\min(u,n)}}{(d-1)! Z_{n,d}} \geqslant \frac{H_u - c_1}{(d-1)! Z_{n,d}},$$

pour $c_1 = \ln d + 1$, puisque $u \leqslant nd$. □

Le lemme suivant correspond au lemme 2.5 en dimension 1.

Lemme 2.16 *Soient $1 \leqslant j \leqslant k$, $\mathbf{u}$ un noeud à distance u de la cible, $\mathbf{v}$ son j-ème contact longue-distance, Q un ensemble de q noeuds interdits et r un entier. La*

probabilité que $\mathbf{v}$ *soit valide et à distance supérieure ou égale à* r *de tout noeud de* Q *est supérieure ou égale à :*

$$\frac{H_u - 2^d H_{rq^{1/d}} - c_3}{2^d H_n + c_4},$$

pour deux constantes $c_3, c_4 \geqslant 0$ *qui ne dépendent que de* d.

Preuve. Soit $\mathcal{E}$ l'événement « $\mathbf{v}$ est valide et à distance supérieure ou égale à r de tout noeud de Q ». On peut réécrire l'événement $\mathcal{E}$ comme « $\mathbf{v}$ est valide et n'appartient à aucune des q boules de rayon r centrées sur les noeuds de Q ». On peut alors minorer la probabilité de $\mathcal{E}$ en remarquant que la probabilité que $\mathbf{v}$ soit à distance δ de $\mathbf{u}$ décroît lorsque δ augmente. Ainsi, la probabilité de $\mathcal{E}$ est minimale lorsque tous les noeuds interdits sont aussi près que possible de $\mathbf{u}$, c'est-à-dire lorsqu'ils sont contenus dans une boule centrée sur $\mathbf{u}$ de rayon minimum δ, où $\delta \leqslant c_2 q f^{1/d}$, pour une constante $c_2 > 0$, est la solution de l'équation $v(\delta) = q \cdot v(r)$.

On conclut que :

$$\begin{aligned}\Pr \mathcal{E} &\geqslant \Pr\{\mathbf{v} \in \mathcal{B}(\mathbf{t}, u-1)\} - \Pr\{\mathbf{v} \in \mathcal{B}(\mathbf{u}, \delta)\} \\ &\geqslant \frac{H_u - c_3}{(d-1)! Z_{n,d}} - \frac{H_\delta}{Z_{n,d}},\end{aligned}$$

d'après le lemme 2.15. Ainsi :

$$\Pr \mathcal{E} \geqslant \frac{H_u - 2^d H_{rq^{1/d}} - c_3}{2^d H_n + c_4},$$

pour des constantes c_3 et c_4 qui ne dépendent que de d. □

Le corollaire suivant correspond au corollaire 2.6 en dimension 1.

Corollaire 2.17 *Il existe une constante* $n_d \in \mathbf{N}$*, indépendante de* x*,* n *et* k *telle que, si* $n \geqslant n_d$*, durant toute phase d'exploration, pour tout noeud* $\mathbf{u}$ *qui n'a jamais été visité auparavant et à distance* u *supérieure à* $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ *de la cible, la probabilité* $\alpha_{u,d}$ *que le* j*-ème contact longue-distance* $\mathbf{v}$ *de* $\mathbf{u}$ *soit valide et à distance supérieure ou égale à* $h^d_{\max}(x) + g(x)$ *de tout noeud de* F *est supérieure à* $\alpha_d^- =_{\text{def}} c \log x / \log n$.

Preuve. A tout instant d'une phase d'exploration, F contient moins de $b_{\max}(x)$ noeuds. D'après le lemme 2.16,

$$\alpha_{u,d} \geqslant \frac{H_u - 2^d H_{(h^d_{\max}(x)+g(x)) b_{\max}(x)^{1/d}} - c_3}{2^d H_n + c_4}.$$

Mais, $h^d_{\max}(x) \leqslant \frac{H_n}{H_x}\frac{\log\log x}{c\log\hat{k}}$, donc :

$$h^d_{\max}(x) + g(x) \leqslant \frac{2H_n}{H_x}\frac{\log\log x}{c\log\hat{k}},$$

puisque $0 < c < 1$. Ainsi,

$$(h^d_{\max}(x) + g(x))b_{\max}(x)^{1/d} \leqslant (h^d_{\max}(x) + g(x))b_{\max} \leqslant \frac{2\log n}{c\log\hat{k}}.$$

On obtient :

$$\alpha_{u,d} \geqslant \frac{H_u - 2^d H_{2\log n/c\log\hat{k}} - c_3}{2^d H_n + c_4} \geqslant \frac{\log\left(x\left(\frac{\log\hat{k}}{2k}\right)^{1/d}\frac{c^{2^d}\log^{2^d}\hat{k}}{2^{2^d}\log^{2^d} n}\right) - c_3}{2^{d+1}\log n},$$

pour $n \geqslant n_2$ où n_2 est un entier qui ne dépend que de d.

Mais $x > \mathrm{stop}_d \geqslant k\log^{2^d+1} n$, on peut donc écrire $x \geqslant \sqrt{x}\sqrt{k\log^{2^d+1} n}$. On remarquera aussi que $\log\hat{k} \geqslant 1$.

$$\alpha_{u,d} \geqslant \frac{\log\left(\sqrt{x}\,\frac{\sqrt{k}}{(2k)^{1/d}}\,\frac{c^{2^d}\,\log^{\frac{2^d+1}{2}} n}{2^{2^d}\log^{2^d} n}\right) - c_3}{2^{d+1}\log n}$$

Mais il existe une constante entière n_3 ne dépendant que de d et c telle que, pour $n \geqslant n_3$, $c^{2^d}\,\log^{\frac{2^d+1}{2}} n \geqslant 1$. Ainsi, pour $n \geqslant \max(n_2, n_3)$, et puisque $k \leqslant \log n$, on obtient :

$$\alpha_{u,d} \geqslant \frac{\log x}{2^{d+3}\log n} - \frac{\log\left(2^{2^d+1/d}\,\log^{2^d+1/d-1/2} n\right) + c_3}{2^{d+1}\log n}$$

Puisque le terme négatif est $o(1)$ lorsque n tend vers l'infini, il existe une constante entière n_4 ne dépendant que de d telle que, pour $n \geqslant \max(n_2, n_3, n_4) =_{\mathrm{def}} n_d$,

$$\alpha_{u,d} \geqslant \frac{\log x}{2^{d+3}\log n} \geqslant c\log x/\log n.$$

□

Puisque les paramètres sont choisis de façon à ce que $(1 + k\alpha^-_d)^{h^d_{\max}(x)} = (1 + ck\,\frac{H_x}{H_n})^{h^d_{\max}(x)} = \log x/\log\log n$, nous sommes maintenant dans un cadre exactement identique à celui de la dimension 1. Nous pouvons appliquer le lemme 2.7 avec $\beta^- = \alpha^-_d$. Comme le minorant α^-_d tient compte du nouveau facteur de division de distance à la cible, on remarquera que le lemme peut s'énoncer avec ce nouveau facteur. Nous omettons donc sa preuve qui est identique à la dimension 1.

Lemme 2.18 *Pour $n \geqslant n_d$, à la fin de toute phase d'exploration, avec probabilité au moins $\frac{1}{5}$,*

- *soit il existe un noeud de A à distance inférieure ou égale à $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ de la cible,*
- *soit il y a plus de $\log n/(2\log \hat{k})$ noeuds distincts dans C.*

La proposition suivante est le dernier pas pour pouvoir prouver le théorème 2.10 en dimension $d > 1$, comme la proposition 2.9 en dimension 1. Le cadre étant également identique, elle se déduit des lemmes précédents de façon similaire à la proposition 2.9 en dimension 1 et nous omettons sa preuve.

Proposition 2.19 *Il existe deux constantes $p_d > 0$ et n_d, indépendantes de n et $\mathbf{x}$, telles que, pour $n \geqslant n_d$, à la fin de toute phase d'exploration, avec probabilité $\geqslant p_d$, il existe un noeud $\mathbf{u}$ dans $A \cup C$ tel que $\mathbf{u}$, ou l'un de ses contacts longue-distance, est à distance inférieure ou égale à $x\left(\frac{\log \hat{k}}{2k}\right)^{1/d}$ de la cible.*

Nous pouvons à présent conclure avec la généralisation du théorème 2.10 en dimension $d > 1$.

Théorème 2.20 (Analyse de l'algorithme 2 en dimension quelconque)
Pour toute dimension $d \geqslant 1$ et pour tout $1 \leqslant k \leqslant \log n$, l'algorithme 2 avec les paramètres $h_{\max}(x) = h^d_{\max}(x)$ et $\mathrm{stop} = \mathrm{stop}_d$ requiert $\Theta(\log^2 n/\log \hat{k})$ bits de mémoire et calcule, pour toute paire de noeuds $(\mathbf{s}, \mathbf{t})$ de $\mathcal{K}^d_{k,n}$, un chemin de $\mathbf{s}$ à $\mathbf{t}$ dont l'espérance de la longueur vaut

$$O\left(\log n\left(\frac{\log\log n}{\log(1+k)}\right)^2\right).$$

Il visite $O\left((\log n/\log(1+k))^2\right)$ noeuds en espérance pour calculer ce chemin.

Preuve. Soient T et U les solutions de $\left(\frac{2k}{\log \hat{k}}\right)^{T/d} = s$ et $\left(\frac{2k}{\log \hat{k}}\right)^{U/d} = \mathrm{stop}_d$. On remarquera que :

$$T \sim \frac{d\log s}{\log \hat{k}} \text{ et } U \sim \frac{d\left((2^d+1)\log\log n + \log k\right)}{\log \hat{k}}.$$

Comme en dimension 1, on décompose l'exécution de $\mathcal{A}$ en T phases : l'exécution de la i-ème phase, pour $1 \leqslant i \leqslant T$, dure tant que la distance x du porteur courant du message à la cible vérifie $\left(\frac{2k}{\log \hat{k}}\right)^{(i-1)/d} < x \leqslant \left(\frac{2k}{\log \hat{k}}\right)^{i/d}$. Si Z_i et Y_i représentent respectivement les variables aléatoires pour le nombre de pas et pour

le nombre de noeuds visités en phase $0 \leqslant i \leqslant N$, avec $x_i = \left(\frac{2k}{\log(1+k)}\right)^{i/d}$, on obtient les formules :

$$\sum_{i=0}^{T} \mathbb{E}[Z_i] \leqslant \omega \frac{\log n \log(\mathrm{stop}_d)}{k} + \frac{1}{p_d} \sum_{T<i\leqslant U} h^d_{\max}(x_i) + g(x_i)$$

$$\text{et} \quad \sum_{i=0}^{T} \mathbb{E}[Y_i] = \frac{1}{p_d} \sum_{i=0}^{T} (h^d_{\max}(x_i) + g(x_i))\, b_{\max}(x_i),$$

où $\omega > 0$ est une constante (indépendante de n, k). Avec notre choix de paramètres stop_d et $h^d_{\max}(x)$, on a :

$$h^d_{\max}(x_i) + g(x_i) \leqslant 2 \log n \log\log n / (c \log(x_i) \log \hat{k}),$$

et aussi $\log x_i \geqslant i \log \hat{k}/2d$, on obtient :

$$\begin{aligned}\sum_{i=0}^{T} \mathbb{E}[Z_i] &\leqslant \omega(2^d+1) \frac{\log n \log\log n}{k} + \frac{2d}{cp_d} \frac{\log n \log\log n}{\log^2 \hat{k}} \sum_{U<i\leqslant T} \frac{1}{i}, \\ &\leqslant O\left(\log n \left(\frac{\log\log n}{\log \hat{k}}\right)^2\right) \\ \text{et} \quad \sum_{i=0}^{T} \mathbb{E}[Y_i] &\leqslant O\left(\left(\frac{\log n}{\log(1+k)}\right)^2\right).\end{aligned}$$

Enfin, pour la durée de chaque étape d'exploration, l'implémentation selon le parcours en profondeur de l'algorithme requiert $\Theta\big(\log n \cdot (b_{\max} + h^d_{\max}(x) + g(x))\big)$ bits de mémoire (comme en dimension 1), soit $O\big(\log^2 n / \log \hat{k}\big)$ bits. □

2.6 Remarques générales sur les performances de l'algorithme et applications

Dépendance des constantes en la dimension. Les facteurs 2^d qui apparaissent dans les constantes lorsque $d > 1$ peuvent paraître exagérés. Ils sont en effet obtenus en considérant systématiquement les pires cas dans le but d'un résultat asymptotique en n. Une analyse plus fine, et sans doute très délicate, des cas moyens pourrait améliorer ces facteurs. Cependant, la croissance de la taille des boules dans le tore fait naturellement apparaître des facteurs exponentiels en d qui ne sont pas compensés par la distribution harmonique ; il est donc possible qu'on ne puisse pas obtenir de constante polynomiale en d.

Un algorithme proche de l'optimal. Le réseau $\mathcal{K}^d_{k,n}$ ayant un degré sortant constant $k + 2d$ ($2d$ liens locaux et k liens longue-distance), son diamètre est $\Omega\big(\log n/\log(k + 2d)\big)$. En 2002, Coppersmith *et al.* [CGS02a, CGS02b] ont démontré que, dans un réseau de Kleinberg où les liens longue-distance sont symétriques (non orientés), le diamètre est $\Theta(\log n/\log\log n)$ (l'espérance du degré étant $O(\log n)$ dans la configuration symétrique). Cela prouve que de tels chemins courts existent, toutefois ils n'exhibent pas d'algorithme décentralisé qui les calcule. En 2002, Aspnes *et al.* [ADS02] ont démontré que, sur un anneau régulier et pour une grande classe de distributions d'arcs aléatoires supplémentaires (incluant celle du modèle de Kleinberg en dimension 1), les chemins calculés par l'algorithme glouton sont de longueur $\Omega\big(\log^2 n/(k\log\log n)\big)$ pour $k \leqslant \log n$ liens par noeuds. Martel et Nguyen [MN04] ont démontré par la suite que le diamètre d'un réseau de Kleinberg $\mathcal{K}^d_{k,n}$ est $\Theta(\log n)$ lorsque $k = O(1)$ avec forte probabilité, sans toutefois donner un algorithme décentralisé qui calcule ces chemins optimaux. On peut ainsi dire que notre algorithme est proche de l'optimal en terme de longueur de chemin puisqu'il calcule des chemins de longueur moyenne $O\left(\log n(\log\log n/\log(1 + k))^2\right)$, optimale à un facteur $(\log\log n/\log(1 + k))^2$ près. En particulier, les chemins sont à un facteur $\log\log n$ de l'optimal lorsque le degré sortant est $\Theta(\log n)$. Ce résultat a en particulier démontré que les chemins calculés par l'algorithme glouton n'étaient pas optimaux (le résultat de Martel et Nguyen [MN04] étant ultérieur).

Par la suite, d'autres travaux récents ont proposé des algorithmes de routage décentralisés. Manku et al. [MNW04] proposent l'algorithme décentralisé NoN (*Neighbor-Of-Neighbor*). Cet algorithme, à la base de la preuve de [CGS02b], effectue une exploration de profondeur 2 dans $\mathcal{K}^d_{O(\log n),n}$ sur les liens locaux et longue-distance. Il calcule des chemins de longueur $O(\log n/\log\log n)$ en espérance, qui sont donc optimaux dans ce réseau où $k = \log n$. Toutefois l'algorithme ne s'applique pas aux réseaux de degré sortant $o(\log n)$. Fraigniaud et al. [FGP04] ont pour leur part proposé un algorithme décentralisé où l'on augmente légèrement la mémoire de chaque noeud ($O(\log^2 n)$ bits au lieu de $O(\log n)$) afin de mémoriser l'ensemble de contacts longue-distance de $O(\log n)$ voisins locaux. Cette connaissance est en fait équivalente à une exploration des seuls liens locaux sur un rayon $O(\log^{1/d} n)$, où d est la dimension. Avec cette connaissance, l'algorithme glouton calcule, entre toute paire de noeuds, un chemin de longueur $O\left(\log n\left(\frac{\log n}{k}\right)^{1/d}\right)$ en espérance. Ils montrent également que cette taille d'exploration est optimale, au sens qu'une exploration locale plus large augmente le nombre de noeuds visités sans raccourcir les chemins obtenus. Toutefois, les chemins calculés restent en moyenne plus longs que ceux calculés par notre algorithme, ce qui montre que l'exploration des voisins des contacts longue-distance permet un gain substantiel.

Par ailleurs, Manku *et al.* ont également démontré, dans le lemme 5.2 de [MNW04], que tout algorithme décentralisé doit visiter $\Omega(\log^2 n/k)$ noeuds en espérance pour router un message entre deux sommets à distance n dans $\mathcal{K}^d_{k,n}$, c'est-à-dire la même latence que l'algorithme glouton. Notre algorithme visite en espérance $O\left((\log n/\log(1+k))^2\right)$ noeuds pour router un message entre deux sommets distants de n, et est donc asymptotiquement optimal en terme de latence pour des réseaux de degré sortant constant ; en visitant un nombre de noeuds du même ordre de grandeur que l'algorithme glouton, nous parvenons à diviser par un facteur $\log n/(\log\log n)^2$ la longueur du chemin de routage calculé.

À notre connaissance, il n'existe pas de meilleur algorithme décentralisé sur les réseaux de Kleinberg, c'est-à-dire qui calcule une longueur de chemin plus faible asymptotiquement pour tout degré sortant $O(\log n)$.

Pertinence pour les grands réseaux d'interactions. Concernant la navigation sur les pages web, Kleinberg remarquait dans [Kle02] qu'un individu recherchant une information sur le réseau du web explore de nouvelles pages à l'aide des hyperliens, et revient en arrière par la touche « back » du navigateur, effectuant de cette façon une série d'aller-retours jusqu'à la page désirée. Il s'agit cette fois d'une recherche (algorithme proche du routage puisque l'on construit un chemin) utilisant une faible exploration du réseau. Dans ce cadre, il est pertinent de considérer un algorithme qui, comme celui que nous venons de présenter, effectue une exploration sur des liens orientés et est autorisé à les reprendre en sens inverse lors de l'exploration, le message étant finalement renvoyé selon un chemin qui respecte l'orientation.

En termes de réseaux sociaux, notre algorithme reste réaliste puisque la mémoire requise pour qu'un individu calcule un chemin est faible, $\Theta\big(\log^2 n/\log(1+k)\big)$, et le processus de choix du prochain intermédiaire est simple à implémenter, puisqu'il suffit de sonder un voisinage de connaissances avant d'envoyer le message à la connaissance la mieux placée.

Enfin, l'algorithme que nous venons de présenter calcule des chemins de longueur très courte sans nécessiter de calcul global sur le réseau, il pourrait donc s'avérer très efficace pour le routage des fichiers dans un réseau pair-à-pair. Des travaux récents ont par exemple montré que les résultats de Kleinberg pouvaient permettre d'améliorer les performances du système pair-à-pair Freenet [ZGG02].

Chapitre 3

Charge du réseau

Introduction

Dans le chapitre précédent, nous avons constaté qu'il est possible d'améliorer l'efficacité du routage, dans le réseau petit monde de Kleinberg, grâce à de nouveaux algorithmes décentralisés. Nous nous intéressons, dans ce chapitre, aux conséquences de leur fonctionnement sur un réseau, en particulier en terme de charge des liens.

Nous donnons ici une définition de la charge d'un lien adaptée aux réseaux augmentés aléatoirement.

Définition 3.1 (Charge d'un lien)
Soient un réseau R dont une partie des liens est tirée selon une distribution de probabilité ϕ, un algorithme de routage $\mathcal{A}$ et ℓ un lien de R. On appelle charge *du lien ℓ la probabilité que ce lien soit sur le chemin de routage de $\mathcal{A}$ entre* $\mathtt{s}$ *et* $\mathtt{t}$ *lorsque* $\mathtt{s}$ *et* $\mathtt{t}$ *sont deux noeuds choisis aléatoirement uniformément dans R.*

On remarquera que dans le cas d'algorithmes de routage monotones où un lien ne peut apparaître qu'une seule fois sur le chemin de routage, la charge d'un lien est aussi l'espérance (prise sur ϕ) du nombre de chemins qui le traversent lorsque $\mathcal{A}$ calcule les chemins entre tous les couples de noeuds de R simultanément, divisée par le nombre de couples de noeuds de R.

L'étude de la charge des liens contient deux problématiques :

- la question du choix d'un algorithme qui limite la charge du réseau tout en restant efficace : dans le cas où l'on peut le choisir, comme pour les réseaux pair-à-pair ;
- et le problème de déduire, d'après la charge observée sur un réseau réel, le type d'algorithme qui y est utilisé : on peut par exemple penser à un réseau

social dont on observe la quantité de messages le long d'une connexion donnée et dont on cherche à comprendre le mécanisme de diffusion.

Nous étudions ici la charge induite par deux algorithmes de routage décentralisés sur le modèle de Kleinberg. Nous comparons la charge induite par l'algorithme glouton à celle d'un algorithme utilisant une faible exploration des liens. Plusieurs algorithmes décentralisés utilisant une exploration des liens ont en effet été proposés récemment pour réduire la longueur des chemins calculés et l'on peut s'interroger sur leurs conséquences en terme de charge. Concernant l'algorithme que nous avons présenté dans le chapitre précédent, l'exploration se fait à la fois sur les liens locaux et longue-distance. L'algorithme de Fraigniaud *et al.* [FGP04] explore pour sa part les seuls liens locaux (sur une profondeur $O(\log n)$). Effectuer un choix sur un ensemble de liens longue-distance en fonction de leur longueur peut entraîner une surcharge des liens les plus longs, qui sont par ailleurs les plus rares. Nous cherchons donc à déterminer l'influence d'une exploration locale sur la charge induite sur un lien longue-distance, en fonction de sa longueur.

Les simulations expérimentales qui permettraient d'obtenir une bonne approximation de la charge d'un lien longue-distance de longueur donnée demanderaient de simuler un grand nombre de réseaux de Kleinberg et pour chacun, un grand nombre de routages entre toutes les paires de noeuds. Ce procédé étant très lourd et coûteux en temps de calcul, nous avons choisi une approche analytique du problème nous permettant d'obtenir les valeurs exactes de la charge pour toutes valeurs des paramètres (taille du réseau et longueur du lien). Nous étudierons la charge du réseau à travers les observations graphiques de ces valeurs déterminées analytiquement. En outre, les équations obtenues dans le cadre de l'étude de la charge des liens nous permettrons par la suite de préciser notre étude, en comparant les taux d'utilisation du graphe sous-jacent par les deux algorithmes de routage.

L'approche analytique de la charge met en jeu des équations de récurrence non linéaires qui sont des objets mathématiques encore difficiles à analyser ; c'est pourquoi nous nous restreindrons, dans ce chapitre, à l'étude du modèle de Kleinberg unidimensionnel et unidirectionnel $\vec{\mathcal{K}}_{k,n}$.

Définition 3.2 (Réseau de Kleinberg unidimensionnel et unidirectionnel)
Étant donnés $n, k \geqslant 1$, un réseau de Kleinberg unidimensionnel et unidirectionnel, *noté $\vec{\mathcal{K}}_{k,n}$, est un anneau orienté $\{0, \ldots, n-1\}$ où chaque noeud $\mathbf{u}$, $\mathbf{u} \in \{0, \ldots, n-1\}$, est l'origine d'un lien local pointant vers $\mathbf{u} + 1 \mod n$ et de k liens longue-distance aléatoires tels que le j-ème contact longue distance de $\mathbf{u}$ est $\mathbf{v}$ avec probabilité $1/(\vec{\delta}(\mathbf{u}, \mathbf{v})H_n)$, où $\vec{\delta}(\mathbf{u}, \mathbf{v})$ est la distance orientée de $\mathbf{u}$ vers $\mathbf{v}$ dans l'anneau.*

Par extension, on notera $\mathbf{u} + i$ *le noeud* $\mathbf{v}$ *tel que* $\vec{\delta}(\mathbf{u}, \mathbf{v}) = i$*, pour tout* $1 \leqslant i \leqslant n$*.*

Définition 3.3 (Intervalle de $\mathbf{u}$ à $\mathbf{v}$)
Étant donnés deux noeuds $\mathbf{u}$ *et* $\mathbf{v}$ *de* $\vec{\mathcal{K}}_{k,n}$*, pour* $k, n \geqslant 1$*, on appelle* intervalle de $\mathbf{u}$ à $\mathbf{v}$*, noté* $[\mathbf{u}, \mathbf{v}]$*, le sous-graphe de l'anneau induit par l'ensemble des noeuds compris entre* $\mathbf{u}$ *et* $\mathbf{v}$ *sur la partie de l'anneau orientée de* $\mathbf{u}$ *vers* $\mathbf{v}$*.*
On note $(\mathbf{u}, \mathbf{v}]$ *(resp.* $[\mathbf{u}, \mathbf{v})$*) l'intervalle de* $\mathbf{u}$ *à* $\mathbf{v}$ *privé de* $\mathbf{u}$ *(resp. privé de* $\mathbf{v}$*).*

La figure 3.1 illustre un réseau $\vec{\mathcal{K}}_{1,n}$. Les liens longue-distance sont en gras ; ils ne sont pas tous représentés pour plus de lisibilité. Comme la plupart des protocoles pair-à-pair fonctionnent sur des anneaux (on peut citer Chord [RFH+01] et Tapestry [ZHS+04]), l'étude du seul cas unidimensionnel n'est pas restrictive.

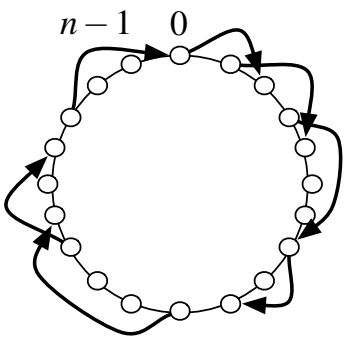


FIG. 3.1 : Réseau $\vec{\mathcal{K}}_{1,n}$.

3.1 Charge de l'algorithme glouton et espérance du nombre de noeuds visités

Dans cette section, nous étudions la probabilité de visite d'un noeud par l'algorithme glouton $\mathcal{G}$ afin d'obtenir une première estimation de la charge d'un lien longue-distance induite par $\mathcal{G}$. Nous considérons le routage d'une source $\mathbf{s}$ vers une cible $\mathbf{t}$. L'argumentation informelle présentée dans cette section est basée sur l'hypothèse, vérifiée expérimentalement, que la charge d'un lien est une fonction unimodulaire de la distance de son origine à $\mathbf{s}$. Nous verrons par la suite que l'expression de la charge d'un lien rend cette hypothèse difficile à démontrer mathématiquement. Afin de clarifier notre argumentation, nous nous restreignons ici au cas où le nombre k de liens longue-distance par noeud vaut 1. On notera $L(\mathbf{x})$ le contact longue-distance du noeud $\mathbf{x}$.

3.1.1 Espérance du nombre de noeuds visités

Dans le chapitre précédent, nous avons vu que l'analyse du routage glouton peut se faire en décomposant son exécution en phases de divisions successives de la distance à la cible par deux. La phase i, pour i variant de 1 à $\log n$, dure tant que la distance courante à la cible est comprise entre 2^i et 2^{i-1}. Kleinberg [Kle00] montre que l'algorithme glouton $\mathcal{G}$ visite $O(\log n)$ noeuds durant chaque phase. Barrière *et al.* [BFKK01] ont montré par la suite que cette borne supérieure était également une borne inférieure. Précisément, pour toutes les phases i de l'exécution, où i est supérieur ou égal à $\log\log n$, l'espérance du nombre de noeuds visités par $\mathcal{G}$ est supérieure ou égale à $c \log n$, où $c > 0$ est une constante absolue (théorème 5 de [BFKK01]). Notons C_i l'ensemble des noeuds dont la distance à la cible est comprise entre 2^{i-1} et 2^i. Manku *et al.* [MNW04] ont démontré que $\mathcal{G}$ visite au moins un noeud de chaque ensemble C_i avec probabilité constante $p > 0$, indépendante de n et i (lemme 5.3 de [MNW04]). En combinant ces trois résultats, on obtient le fait suivant.

Fait 3.1 *Pour tout* $\log\log n \leqslant i \leqslant \log n$*, soit* E_i *l'espérance du nombre de noeuds de* C_i *visités par* $\mathcal{G}$*. Il existe deux constantes absolues* c, c' *telles que :*

$$c \log n \leqslant E_i \leqslant c' \log n.$$

Rappelons que nous considérons le routage entre une source $\mathbf{s}$ et une cible $\mathbf{t}$ données. Étant donné un noeud $\mathbf{x}$, notons $\mathcal{E}_{\text{lien chargé}}$ l'événement « le lien longue-distance de $\mathbf{x}$ appartient au chemin calculé par $\mathcal{G}$ de $\mathbf{s}$ à $\mathbf{t}$ » et $\mathcal{E}_{\text{visité}}$ l'événement « $\mathbf{x}$ est visité par $\mathcal{G}$ durant le routage de $\mathbf{s}$ à $\mathbf{t}$ ». On a :

$$\Pr \mathcal{E}_{\text{lien chargé}} = \Pr\{L(\mathbf{x}) \in (\mathbf{x}, \mathbf{t}]\} \cdot \Pr \mathcal{E}_{\text{visité}}.$$

La probabilité $\Pr\{L(\mathbf{x}) \in (\mathbf{x}, \mathbf{t}]\}$ peut être évaluée directement depuis la distribution des liens. Afin d'estimer la charge d'un lien longue-distance, nous allons donc évaluer, dans la section suivante, la probabilité que son origine soit visitée. On remarquera par ailleurs que la définition de $\mathcal{G}$ implique, sur $\vec{\mathcal{K}}_{1,n}$:

$$\Pr \mathcal{E}_{\text{visité}} = 0 \quad \text{si} \quad \mathbf{x} \notin [\mathbf{s}, \mathbf{t}].$$

Par conséquent, nous ne considérons que les noeuds appartenant à l'intervalle $[\mathbf{s}, \mathbf{t}]$.

3.1.2 Une première estimation de la charge

On peut avoir l'intuition que la probabilité qu'un noeud soit visité par $\mathcal{G}$ décroît avec sa distance à $\mathbf{s}$, puis croît en se rapprochant de $\mathbf{t}$. En effet, pour un noeud

proche de $\mathbf{s}$, la probabilité qu'il soit visité est proche de la probabilité qu'il soit le contact longue-distance de la source, la longueur moyenne d'un lien longue-distance étant $n/\log n$. Mais plus on s'approche de la cible, plus ce phénomène s'atténue. En effet, de plus en plus de contacts longue-distance sont au-delà de $\mathbf{t}$ et ne peuvent être pris par $\mathcal{G}$, ce qui favorise la visite des contacts locaux suivants. À partir de ces observations, nous avons commencé par étudier la charge des liens sous l'hypothèse que la probabilité qu'un noeud soit visité lors du routage de $\mathbf{s}$ à $\mathbf{t}$ est une fonction unimodulaire, décroissante puis croissante, de la distance de $\mathbf{s}$ à ce noeud. Nous verrons que cette hypothèse est vérifiée, expérimentalement, pour le cas $k = 1$.

Hypothèse 3.2 *Il existe un noeud* $\mathbf{y} \in [\mathbf{s}, \mathbf{t}]$ *tel que la probabilité qu'un noeud* $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$ *soit visité par* $\mathcal{G}$ *est une fonction* $f(\mathbf{x})$ *décroissante sur* $[\mathbf{s}, \mathbf{y}]$ *(*i.e. *pour tout* $0 \leqslant i < \vec{\delta}(\mathbf{s}, \mathbf{y})$*,* $f(\mathbf{s} + i) \geqslant f(\mathbf{s} + i + 1)$*) et croissante sur* $[\mathbf{y}, \mathbf{t}]$ *(*i.e. *pour tout* $0 \leqslant i < \vec{\delta}(\mathbf{y}, \mathbf{t})$*,* $f(\mathbf{y} + i) \leqslant f(\mathbf{y} + i + 1)$*).*

Sous cette hypothèse, nous obtenons le lemme suivant qui nous donne une première estimation de la charge induite par $\mathcal{G}$ sur un lien longue-distance.

Lemme 3.3 *Soient* $n \geqslant 1$ *et* $\mathbf{s}, \mathbf{t} \in \vec{\mathcal{K}}_{1,n}$*. Sous l'hypothèse 3.2, il existe* $\mathbf{y}' \in [\mathbf{y}, \mathbf{y} + \vec{\delta}(\mathbf{s}, \mathbf{y})/2]$ *tel que, pour tout noeud* $\mathbf{x} \in [\mathbf{y}', \mathbf{t} - \log n]$*, la probabilité que* $\mathbf{x}$ *appartienne au chemin calculé par* $\mathcal{G}$ *de* $\mathbf{s}$ *à* $\mathbf{t}$ *est :*

$$\Pr \mathcal{E}_{\textit{visité}} = \Theta\left(\frac{\log n}{\vec{\delta}(\mathbf{x}, \mathbf{t})}\right),$$

et la probabilité que son lien longue-distance appartienne au chemin de routage est :

$$\Pr \mathcal{E}_{\textit{lien chargé}} = \Theta\left(\frac{H_{\vec{\delta}(\mathbf{x},\mathbf{t})}}{\vec{\delta}(\mathbf{x}, \mathbf{t})}\right).$$

Preuve. On considère la première phase i_0 à partir de laquelle la probabilité f qu'un noeud soit visité devient croissante avec sa distance à la source (i_0 existe d'après l'hypothèse 3.2) :

$$i_0 = \max\{i \mid 2^i \leqslant \vec{\delta}(\mathbf{y}, \mathbf{t})\}.$$

Supposons $\log\log n \leqslant i \leqslant i_0$. D'après le fait 3.1, l'espérance du nombre de noeuds visités par $\mathcal{G}$ dans C_i est supérieure à $c \log n$ et inférieure à $c' \log n$, pour deux constantes absolues $c, c' > 0$. Par ailleurs, l'espérance du nombre de noeuds visités dans C_i est aussi la somme, sur tous les noeuds $\mathbf{u}$ de C_i, des probabilités qu'ils soient visités, par linéarité de l'espérance. On remarquera également que

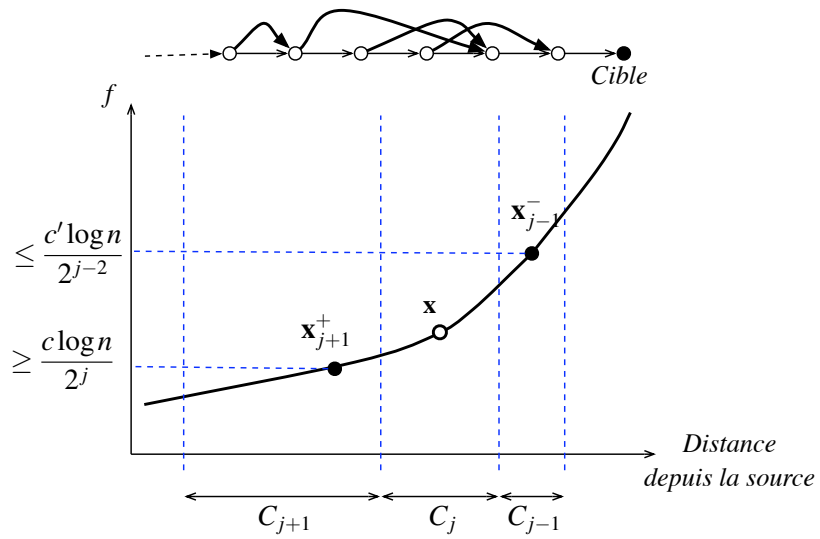


FIG. 3.2 : Illustration de la preuve du lemme 3.3.

$C_i = [\mathbf{t} - 2^i, \mathbf{t} - 2^{i-1}]$. Comme f croît sur l'intervalle C_i d'après l'hypothèse 3.2, il existe deux noeuds $\mathbf{x}_i^-, \mathbf{x}_i^+ \in C_i$ tels que :

$$\forall \mathbf{u} \in [\mathbf{t} - 2^i, \mathbf{x}_i^-], f(\mathbf{u}) \leqslant \frac{c' \log n}{2^{i-1}},$$
$$\text{et } \forall \mathbf{u} \in [\mathbf{x}_i^+, \mathbf{t} - 2^{i-1}], f(\mathbf{u}) \geqslant \frac{c \log n}{2^{i-1}}.$$

En effet, la somme des $f(\vec{\delta}(\mathbf{s}, \mathbf{u}))$ sur tous les noeuds $\mathbf{u}$ de C_i est $\Theta(\log n)$ (c'est l'espérance du nombre de noeuds visités) et il y a 2^{i-1} noeuds $\mathbf{u}$ dans cet ensemble C_i. Ce raisonnement est illustré sur la figure 3.2.
Soit j tel que $\mathbf{x} \in C_j$. Supposons dans un premier temps que $\log \log n < j < i_0$. Par monotonie de f, on a :

$$f(\mathbf{x}_{j+1}^+) \leqslant f(\mathbf{x}) \leqslant f(\mathbf{x}_{j-1}^-),$$
$$\text{ainsi } \frac{c \log n}{2^j} \leqslant f(\mathbf{x}) \leqslant \frac{c' \log n}{2^{j-2}},$$
$$\text{c'est-à-dire } \frac{c \log n}{2\vec{\delta}(\mathbf{x}, \mathbf{t})} \leqslant f(\mathbf{x}) \leqslant \frac{4c' \log n}{\vec{\delta}(\mathbf{x}, \mathbf{t})}.$$

Ce raisonnement nécessite d'avoir deux autres ensembles C_{j-1} et C_{j+1} qui respectent l'hypothèse de monotonie, ce qui n'est pas le cas lorsque $j = i_0$. Toutefois, lorsque $j = i_0$, on peut refaire ce raisonnement pour les noeuds situés après $\mathbf{x}_{i_0}^+ =_{\text{def}} \mathbf{y}'$ puisque l'on dispose alors d'une borne inférieure dans C_{i_0}.

On obtient donc l'équivalent asymptotique de $\Pr \mathcal{E}_{\text{visité}}$ pour tout noeud de $[\mathbf{y}', \mathbf{t} - \log n]$. On tire directement l'équivalent de $\Pr \mathcal{E}_{\text{lien chargé}}$ en remarquant que $\Pr\{L(\mathbf{x}) \in (\mathbf{x}, \mathbf{t}]\} = H_{\vec{\delta}(\mathbf{x},\mathbf{t})}/H_n$. $\square$

On tire de ce lemme un corollaire immédiat sur la probabilité qu'un lien longue-distance donné de longueur ℓ appartienne au chemin de routage.

Corollaire 3.4 *Soient $n \geqslant 1$ et $\mathbf{s}, \mathbf{t} \in \vec{\mathcal{K}}_{1,n}$. Sous l'hypothèse 3.2, il existe $\mathbf{y}' \in [\mathbf{y}, \mathbf{y} + \vec{\delta}(\mathbf{y}, \mathbf{s})/2]$ tel que, pour tout lien longue-distance donné d'origine $\mathbf{x} \in [\mathbf{y}', \mathbf{t} - \log n]$ et de longueur $\ell \geqslant 1$, la probabilité qu'il appartienne au chemin de routage est*

$$\Theta\Big(\frac{\log n}{\delta(\mathbf{x}, \mathbf{t})}\Big) \text{ si } 1 \leqslant \ell \leqslant \vec{\delta}(\mathbf{x}, \mathbf{t}), \text{ et } 0 \text{ sinon,}$$

où $\mathbf{y}'$ est donné par le lemme 3.3.

Dans la section suivante, nous analysons précisément les équations de récurrence qui régissent la probabilité qu'un noeud soit visité.

3.2 Charge de deux algorithmes décentralisés

Nous comparons ici la charge induite par deux algorithmes de routage décentralisés. Nous allons déterminer leur expression analytique pour pouvoir ensuite observer leurs variations graphiquement. Le premier est l'algorithme glouton $\mathcal{G}$ (décrit p. 21). Nous avons ensuite choisi un algorithme de routage décentralisé $\mathbf{E}_\omega$ où il existe une exploration locale des liens longue-distance afin de mettre en valeur l'impact d'une telle exploration. Il s'agit d'une version simplifiée de l'algorithme présenté dans [FGP04], choisie dans le but de simplifier l'analyse.

L'algorithme $\mathbf{E}_\omega$, de paramètre $\omega \geqslant 1$ (algorithme 3), procède par explorations successives des contacts longue-distance des ω voisins locaux immédiatement à droite du porteur courant du message.

Parmi les contacts longue-distance explorés, le *meilleur* est celui qui est le plus proche de $\mathbf{t}$ dans $[\mathbf{s}, \mathbf{t}]$; on dit qu'il s'agit du *meilleur* lien longue-distance de la fenêtre d'exploration. À la fin d'une exploration, le chemin de routage courant est étendu par la chaîne de liens locaux qui mène à l'origine du meilleur lien longue-distance et par ce meilleur lien longue-distance. S'il n'existe pas de meilleur lien longue-distance, le chemin est simplement étendu par la chaîne de liens locaux de longueur ω, on dit alors que l'exploration est *en échec*.

La figure 3.3 illustre le fonctionnement de $\mathbf{E}_\omega$: le schéma supérieur représente l'exploration de la fenêtre (noeuds gris) et le schéma inférieur, le prolongement π du chemin courant (en gras).

Algorithme 3 Algorithme $\mathbf{E}_\omega$ de routage par exploration locale

Entrée : la source s, la cible $\mathbf{t}$, la largeur de fenêtre $\omega \geqslant 1$.

Initialisation : $\mathbf{x} \leftarrow \mathrm{s}$.

tant que $\vec{\delta}(\mathbf{x}, \mathbf{t}) > \omega$ **faire**

$F \leftarrow (\mathbf{x}, \mathbf{x} + \omega]$ (F est la *fenêtre d'exploration*).
$A \leftarrow \{L(\mathbf{y}) \,|\, \mathbf{y} \in F,\ L(\mathbf{y}) \in [\mathbf{x} + \omega + 1, \mathbf{t}]\}$.
si $A \neq \emptyset$, **alors**
$\mathbf{x} \leftarrow L(\mathbf{y})$, le *meilleur* contact de A (choix uniforme en cas d'égalité).
sinon $\mathbf{x} \leftarrow \mathbf{x} + \omega + 1$ (exploration *en échec*).
fin si
fin tant que

Phase finale : Suivre les contacts locaux jusqu'à la cible $\mathbf{t}$.

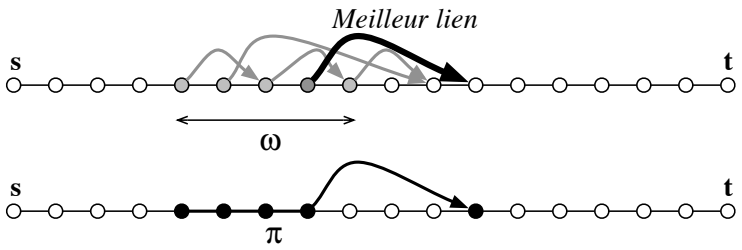


FIG. 3.3 : Fonctionnement de l'algorithme $\mathbf{E}_\omega$.

Le fait suivant nous sera très utile pour condenser les notations des probabilités dans l'analyse des deux algorithmes. C'est une conséquence directe de la distribution des liens longue-distance sur $\vec{\mathcal{K}}_{k,n}$.

Fait 3.5 *Soient $k, n \geqslant 1$. Pour tous $1 \leqslant j \leqslant k$ et $1 \leqslant a \leqslant b \leqslant n$, la probabilité que le j-ème contact longue-distance du noeud $\mathbf{v} \in \vec{\mathcal{K}}_{k,n}$ soit de longueur supérieure ou égale à a et inférieure ou égale à b est :*

$$\frac{H_b}{H_n} \quad \textit{si} \quad a = 1$$
$$\textit{et} \quad \frac{H_b - H_{a-1}}{H_n} \quad \textit{sinon.}$$

Preuve. La probabilité qu'un lien longue-distance donné soit de longueur ℓ est $1/(\ell H_n)$ dans $\vec{\mathcal{K}}_{k,n}$. En sommant sur ℓ allant de a à b, on obtient le résultat. □

Dans la suite, nous omettrons généralement le paramètre n dans le nom des fonctions considérées pour alléger l'écriture.

3.2.1 Charge induite par l'algorithme glouton

Dans cette section, nous calculons la charge induite par l'algorithme $\mathcal{G}$ sur un lien longue-distance.

3.2.1.1 Distance critique source-cible

Avant de calculer la charge d'un lien longue-distance, nous étudions dans cette section le cas critique où $\mathcal{G}$ n'utilise aucun lien longue-distance dans son chemin de routage.

Si $\mathbf{s}$ et $\mathbf{t}$ sont trop proches, tous les liens longue-distance des noeuds de $[\mathbf{s}, \mathbf{t}]$ ont une forte probabilité de dépasser la cible. La proposition suivante démontre qu'il existe ainsi une distance minimale source-cible en dessous de laquelle $\mathcal{G}$ n'emprunte aucun lien longue-distance dans son chemin de routage, avec forte probabilité.

Proposition 3.6 *Soient $k, n \geqslant 1$ et $\mathbf{s}, \mathbf{t} \in \vec{\mathcal{K}}_{k,n}$. Si*

$$\vec{\delta}(\mathbf{s}, \mathbf{t}) \leqslant \frac{\log n}{k(\log\log n)^2},$$

alors le chemin de routage calculé par $\mathcal{G}$ de $\mathbf{s}$ à $\mathbf{t}$ est composé uniquement de liens locaux avec probabilité $1 - o(1)$ (lorsque n tend vers l'infini).

Preuve. Notons $m = \vec{\delta}(\mathbf{s}, \mathbf{t})$ et p la probabilité considérée, *i.e.* p est la probabilité que que tous les contacts longue-distance des noeuds de $[\mathbf{s}, \mathbf{t}]$ soient au-delà de $\mathbf{t}$. Soit $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$, d'après le fait 3.5, pour tout $1 \leqslant j \leqslant k$, la probabilité que $L_j(\mathbf{x})$ soit au delà de $\mathbf{t}$ est :

$$1 - \frac{H_{\vec{\delta}(\mathbf{x},\mathbf{t})}}{H_n} \geqslant 1 - \frac{H_m}{H_n}.$$

Chacun des k liens longue-distance de chacun des m noeuds de $[\mathbf{s}, \mathbf{t}]$ étant tiré indépendamment, on en déduit :

$$p \geqslant \left(1 - \frac{H_m}{H_n}\right)^{km} \geqslant \left(1 - \frac{H_{\log n/(k(\log\log n)^2)}}{H_n}\right)^{k\left(\log n/(k(\log\log n)^2)\right)}$$

Soit encore :

$$\begin{aligned}\log p &\geqslant \frac{\log n}{(\log\log n)^2}\left(-\frac{\log\log n}{\log n} + o\left(\frac{\log\log n}{\log n}\right)\right)\\ &\geqslant -\frac{1}{\log\log n} + o\left(\frac{1}{\log\log n}\right).\end{aligned}$$

Finalement,

$$p \geqslant 1 - \frac{1}{\log\log n} = 1 - o(1).$$

□

3.2.1.2 Probabilité de visite d'un noeud

Comme nous l'avons remarqué dans la section 3.1, une première étape pour quantifier la charge des liens induite par l'algorithme glouton entre une source et une cible données consiste à déterminer la probabilité qu'un noeud de l'intervalle $[\mathbf{s}, \mathbf{t}]$ soit visité. Les probabilités sont calculées sur tous les réseaux aléatoires de Kleinberg possibles.

Étant donnés $\mathbf{s}$ et $\mathbf{t}$, pour tout $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$, on note $g\Big(\vec{\delta}(\mathbf{s}, \mathbf{x}), \vec{\delta}(\mathbf{s}, \mathbf{t})\Big)$ la probabilité que $\mathbf{x}$ appartienne au chemin de routage calculé par $\mathcal{G}$ de $\mathbf{s}$ à $\mathbf{t}$.

Proposition 3.7 *Soient $k, n \geqslant 0$ et $\mathbf{s}, \mathbf{t} \in \vec{\mathcal{K}}_{k,n}$, avec $\vec{\delta}(\mathbf{s}, \mathbf{t}) = m$. Pour tout noeud $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$ où l'on note $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$ on a :*

$$g(x, m) = 1 \quad \textit{si} \quad x = 0,$$
$$g(x, m) = \sum_{i=1}^{x} (g(x-i, m) \cdot p_{x-i}) + \left(1 - \frac{H_{m-x+1}}{H_n}\right)^k g(x-1, m) \quad \textit{sinon},$$

avec :

$$p_{x-i} = \left(1 - \frac{H_{m-x+i} - H_i}{H_n}\right)^k - \left(1 - \frac{H_{m-x+i} - H_{i-1}}{H_n}\right)^k.$$

Preuve. Soit ϵ_x l'événement « le noeud $\mathbf{x}$ appartient au chemin de routage ». On remarquera tout d'abord que $x = 0$ correspond à $\mathbf{x} = \mathbf{s}$, comme la source est toujours sur le chemin de routage, on a $g(0, m) = 1$.

Supposons $x \geqslant 1$. Soit $\mathbf{y} \in [\mathbf{s}, \mathbf{x})$ son prédécesseur immédiat sur le chemin. Alors :

- soit $\mathbf{x}$ est un contact longue-distance de $\mathbf{y}$ et le lien longue-distance appartient au chemin (événement ϵ_x^1),
- soit $\mathbf{y} = \mathbf{x} - 1$ et le lien local de $\mathbf{y}$ à $\mathbf{x}$ appartient au chemin (événement ϵ_x^2).

On a $\epsilon_x = \epsilon_x^1 \cup \epsilon_x^2$, où les deux événements ϵ_x^1 et ϵ_x^2 sont disjoints. Ainsi, $\Pr \epsilon_x = \Pr \epsilon_x^1 + \Pr \epsilon_x^2$.

La probabilité de ϵ_x^2 est la probabilité que $\mathbf{y}$ appartienne au chemin et ait tous ses contacts longue-distance au-delà de $\mathbf{t}$. Comme l'appartenance de $\mathbf{y}$ au chemin

calculé par $\mathcal{G}$ est indépendante de ses liens longue-distance, on obtient en utilisant le fait 3.5 :

$$\Pr \epsilon_x^2 = \big(1 - \frac{H_{m-x+1}}{H_n}\big)^k g(x-1, m).$$

Pour calculer la probabilité de l'événement ϵ_x^1, on somme sur toutes les positions possibles de $\mathbf{y}$ la probabilité que $\mathbf{y}$ soit visité, qu'il ait $\mathbf{x}$ pour contact longue-distance et que $\mathcal{G}$ emprunte ce lien longue-distance. Comme $\mathbf{x}$ a un unique prédécesseur sur le chemin, ce sont des événements deux-à-deux disjoints et l'on peut bien faire la somme de ces probabilités.

Soit p_{x-i} la probabilité que le noeud $\mathbf{y} = \mathbf{x} - i$ ait j liens longue-distance d'extrémité $\mathbf{x}$ et ses $k - j$ autres contacts longue-distance hors de l'intervalle $[\mathbf{x}, \mathbf{t}]$, pour tous les $1 \leqslant j \leqslant k$ possibles. On remarquera que dans ce cas de figure, si $\mathcal{G}$ visite $\mathbf{y}$, le noeud visité suivant est $\mathbf{x}$. Cet événement est indépendant de ϵ_{x-i}, on a donc :

$$\Pr \epsilon_x^1 = \sum_{0<i\leqslant x} (\Pr \epsilon_{x-i} \cdot p_{x-i}),$$

où $\Pr \epsilon_{x-i} = g(x-i, m)$.

On évalue enfin p_{x-i} en sommant sur toutes les positions possibles des j liens longue-distance de $\mathbf{y}$ de longueur égale à i (et donc d'extrémité $\mathbf{x}$) parmi ses k liens. Pour chaque j, il y a $\binom{k}{j}$ positions possibles pour les liens de longueur i. La probabilité d'avoir j liens de longueur i est $1/(iH_n)^j$ et celle d'avoir $k - j$ liens dont l'extrémité n'appartient pas à $[\mathbf{x}, \mathbf{t}]$ est $(1 - (H_{m-x+i} - H_{i-1})/H_n)^{k-j}$ en utilisant le fait 3.5. On obtient :

$$p_{x-i} = \sum_{j=1}^{k} \binom{k}{j} \left(\frac{1}{iH_n}\right)^j \left(1 - \frac{H_{m-x+i} - H_{i-1}}{H_n}\right)^{k-j}.$$

En ajoutant le terme $j = 0$, on reconnaît l'expression du binôme $(1 - (H_{m-x+i} - H_i)/H_n)^k$. Puisque $H_i = H_{i-1} + 1/i$, cela démontre le résultat. □

Nous avons à présent une équation de récurrence non linéaire définissant de façon exacte la probabilité de visite d'un noeud. Toutefois les équations de récurrence non linéaires sont encore mal connues et se révèlent difficiles à analyser. En particulier, l'unimodularité est une propriété difficile à obtenir, et qu'on ne sait démontrer que sur des cas très particuliers. La programmation de cette récurrence nous permet cependant de tracer ses valeurs et d'observer ses variations pour toutes valeurs des paramètres m et n. Les courbes ont une grande stabilité par rapport à la taille de l'anneau n, on peut donc se restreindre à une petite taille d'anneau tout en observant des tracés représentatifs des valeurs supérieures de n. En choisissant par exemple un anneau de taille 50 ou 100, nous parvenons à

des tracés représentatifs des anneaux plus grands, tout en conservant un temps de calcul très faible.

Nous avons choisi de représenter le tracé de $g(x, m)$ pour $m = n - 1$ et pour différents degrés sortants $k \in \{1, \ldots, 6\}$ (figure 3.4 et 3.5 pour son tracé en échelle semi-logarithmique). Dans le cas où $m = n - 1$, tous les noeuds sont susceptibles d'être visités. Pour des valeurs inférieures de m, la probabilité qu'un lien longue-distance dépasse $\mathbf{t}$ augmente mais elle est renormalisée uniformément pour tous les noeuds de l'intervalle $[\mathbf{s}, \mathbf{t}]$ et l'on obtient une courbe similaire. Le tracé en échelle semi-logarithmique nous permet d'observer les variations plus précisément.

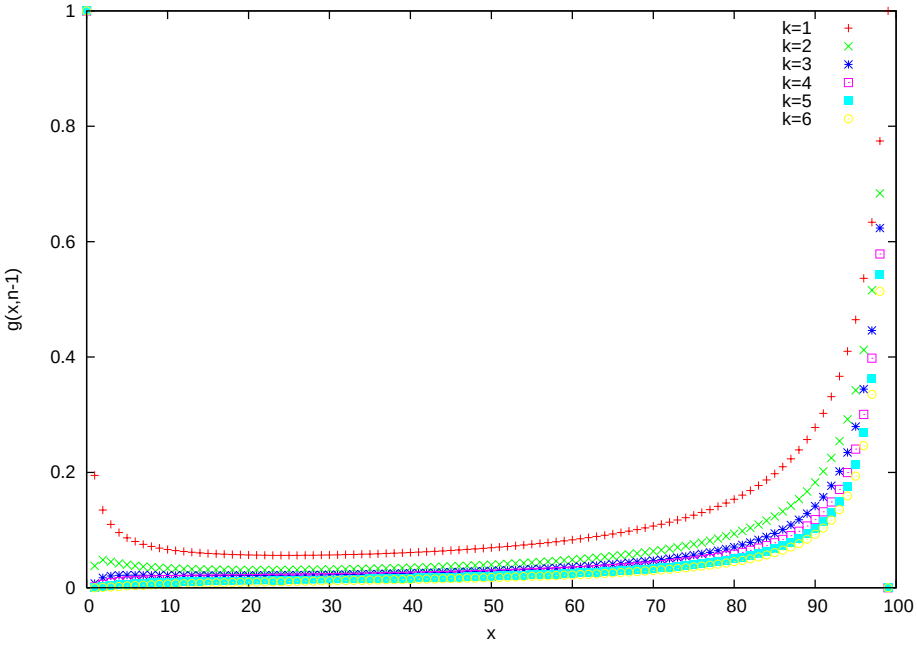


FIG. 3.4 : Variations de $g(x, n - 1)$ en fonction de x pour $n = 100$.

Observations des tracés. Lorsqu'il n'y a qu'un lien longue-distance par noeud ($k = 1$) la courbe décroît assez rapidement jusqu'à atteindre un minimum, puis devient croissante à partir de ce minimum (figures 3.4 et 3.5). Nous avons pu observer expérimentalement une propriété remarquable de la position de ce minimum sur $[\mathbf{s}, \mathbf{t}]$. La figure 3.6 représente les différentes positions x du minimum de la courbe $g(x, n - 1)$ lorsque l'on fait varier n de 1 à 10^4. On observe une droite de pente $1/4$ (figure 3.6) qui suggère que la probabilité minimale est toujours atteinte au quart de l'intervalle $[\mathbf{s}, \mathbf{t}]$ lorsque $k = 1$. Toutefois, la distribution des liens longue-distance fait intervenir H_n qui varie peu lorsque n varie de 1 à

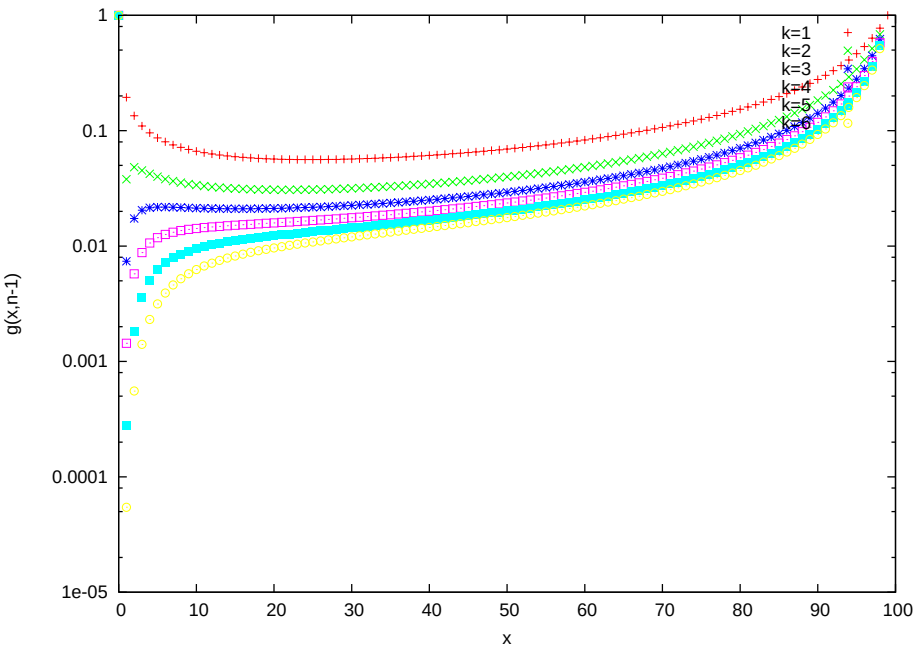


FIG. 3.5 : Variations de $g(x, n-1)$ en fonction de x pour $n = 100$ en échelle semi-logarithmique.

10^4 ; la valeur $1/4$ peut donc cacher une dépendance en n (par exemple, la pente pourrait en fait s'exprimer comme $1/(4 + H_n/n^2)$, mais ce type de dépendance est invisible pour $n \leqslant 10^4$).

Lorsque $k > 1$, la partie croissante de la courbe est assez similaire au cas $k = 1$, mais la partie décroissante atteint un minimum dès $x = 1$ et remonte rapidement avec un point d'inflexion. Pour des valeurs intermédiaires de k ($k = 2$ et 3), la fonction présente même deux extrema locaux. Cela peut s'expliquer en remarquant que lorsque $m = n - 1$, $\mathcal{G}$ utilise toujours un lien longue-distance de s, et lorsque $k > 1$, la probabilité que le lien pris soit long est plus grande. La probabilité de visiter un noeud très proche de s devient donc de plus en plus faible et la courbe s'infléchit de plus en plus (lorsque k augmente), en passant par une étape intermédiaire où le minimum, qui existe pour $k = 0$, est encore présent.

Les tracés observés présentent un comportement assez complexes pour les petites valeurs de x qui justifie notre approche graphique. On remarquera que ces tracés corroborent l'hypothèse 3.2 lorsque $k = 1$.

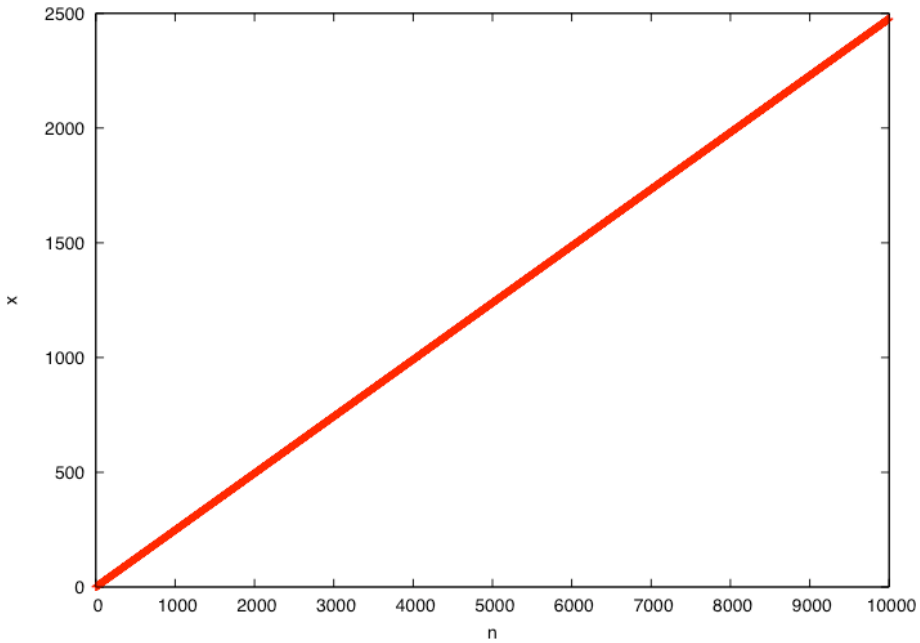


FIG. 3.6 : Valeurs de x au minimum de la courbe $g(x, n-1)$ pour $k = 1$ et $1 \leqslant n \leqslant 10000$

3.2.1.3 Charge d'un lien de longueur $\ell \geqslant 1$.

Nous sommes à présent en mesure d'exprimer la charge induite par l'algorithme glouton $\mathcal{G}$ sur un lien longue-distance de longueur $\ell \geqslant 1$ et d'origine donnée.

Proposition 3.8 *Pour tout $k, n \geqslant 1$, soit $load^{\mathcal{G}}(k, \ell)$ la charge d'un lien longue-distance, de longueur $\ell \geqslant 1$ et d'origine $\mathbf{x}$ induite par $\mathcal{G}$ dans $\vec{\mathcal{K}}_{k,n}$. On a :*

$$load^{\mathcal{G}}(k,\ell) = \frac{1}{n^2}\sum_{m=\ell}^{n-1}\sum_{x=0}^{m-\ell}\left[g(x,m)\left(\left(1-\frac{H_{m-x}-H_\ell}{H_n}\right)^{k-1} + \sum_{i=1}^{k-1}\binom{k-1}{i}\left(1-\frac{H_{m-x}-H_\ell}{H_n}\right)^{k-1-i}\left(\frac{1}{\ell H_n}\right)^i\frac{1}{i}\right)\right].$$

Preuve. Nous calculons l'espérance du nombre de chemins traversant le lien d'origine $\mathbf{x}$ et de longueur ℓ lorsque l'on choisit une source et une cible aléatoirement uniformément. On considère le routage entre chacun des n^2 couples $(\mathbf{s}, \mathbf{t})$ du réseau. On remarquera que, pour un lien d'origine $\mathbf{x}$ et de longueur $\ell \geqslant 1$ fixées,

les seuls couples $(\mathbf{s}, \mathbf{t})$ source-cible pour lesquels le chemin de routage est susceptible de contenir le lien sont ceux qui vérifient $\mathbf{x} \in [\mathbf{s}, \mathbf{t} - \ell]$. Notons $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$ et $m = \vec{\delta}(\mathbf{s}, \mathbf{t})$, on suppose à présent $m \geqslant \ell + x$ (les autres valeurs de m ont une contribution nulle dans l'espérance).

La probabilité que le lien longue-distance considéré soit sur le chemin calculé entre $\mathbf{s}$ et $\mathbf{t}$ est exactement la probabilité que $\mathbf{x}$ appartienne au chemin (probabilité $g(x, m)$) *et* que les $k - 1$ autres liens longue-distance de $\mathbf{x}$ soient de longueur inférieure à ℓ ou strictement supérieure à $m - x$ (avec un choix aléatoire uniforme en cas d'égalité). Lorsqu'il n'y a pas de cas d'égalité, d'après le fait 3.5, le lien est choisi avec probabilité :

$$p_1(x, m) = \Big(1 - \frac{H_{m-x} - H_\ell}{H_n}\Big)^{k-1}.$$

En cas d'égalité avec i autres liens longue-distance de longueur ℓ, pour $1 \leqslant i \leqslant k$, le lien considéré est choisi avec probabilité uniforme $1/i$. Il y a alors $\binom{i}{k-1}$ positions possibles pour ces liens de longueur ℓ parmi les $k - 1$ liens autres que celui considéré. La probabilité d'avoir i liens de longueur ℓ est $1/(\ell H_n)^i$; et la probabilité d'avoir $k - 1 - i$ liens de longueur strictement inférieure à ℓ ou qui vont au-delà de $\mathbf{t}$ est $(1 - (H_{m-x} - H_\ell)/H_n)^{k-1-i}$. Le lien considéré est donc choisi avec probabilité :

$$p_2(x, m) = \sum_{i=1}^{k-1} \binom{i}{k-1} \Big(1 - \frac{H_{m-x} - H_\ell}{H_n}\Big)^{k-1-i} \Big(\frac{1}{\ell H_n}\Big)^i \frac{1}{i}.$$

La probabilité que le lien appartienne au chemin est finalement :

$$g(x, m)\big(p_1(x, m) + p_2(x, m)\big).$$

Notons $X_{(\mathbf{s},\mathbf{t})}$ la variable aléatoire qui vaut 1 si le chemin de routage calculé par $\mathcal{G}$ utilise le lien considéré et 0 sinon. $X_{(\mathbf{s},\mathbf{t})}$ est conditionnée par le choix de $(\mathbf{s}, \mathbf{t})$ comme couple source-cible pour le routage. Néanmoins, $X_{(\mathbf{s},\mathbf{t})}$ est déterminée de façon unique par les paramètres x et m, et son espérance vaut

$$\begin{aligned} \mathbb{E}[X_{(\mathbf{s},\mathbf{t})}] &= g(x, m)\,(p_1(x, m) + p_2(x, m)) \quad \text{si} \quad m \geqslant \ell + x \\ &= 0 \quad \text{sinon.} \end{aligned}$$

Comme la source et la cible sont choisies de façon aléatoire uniforme, tous les couples $(\mathbf{s}, \mathbf{t})$ ont une probabilité $1/n^2$ d'être choisis pour le routage. Par linéarité de l'espérance, on en déduit :

$$\text{load}^{\mathcal{G}}(k, \ell) = \frac{1}{n^2} \sum_{m=\ell}^{n-1} \sum_{x=0}^{m-\ell} g(x, m)\,(p_1(x, m) + p_2(x, m))\,.$$

□

Observations des tracés. Pour pouvoir comparer les tracés avec différentes valeurs de k, nous devons renormaliser les valeurs. En effet, lorsque l'on multiplie le nombre de liens longue-distance par noeud par un facteur $c > 0$, on diminue la probabilité qu'un lien soit parcouru par un chemin. Comme il y a c fois plus de liens de longueur ℓ en espérance qu'auparavant, on renormalise le nombre de chemins de routage entre chaque paire source-cible en multipliant la charge par c afin de faciliter la comparaison des courbes en fonction de ℓ pour les différentes valeurs de k. Ce facteur de normalisation est celui qui a donné les meilleurs résultats. La figure 3.7 (3.8 pour le tracé en échelle semi-logarithmique) représente les tracés des charges $\text{load}^{\mathcal{G}}(k, \ell)$ en fonction de ℓ pour un anneau de $n = 50$ noeuds et différents degrés sortants $k \in \{1, 3, 4, 5\}$.

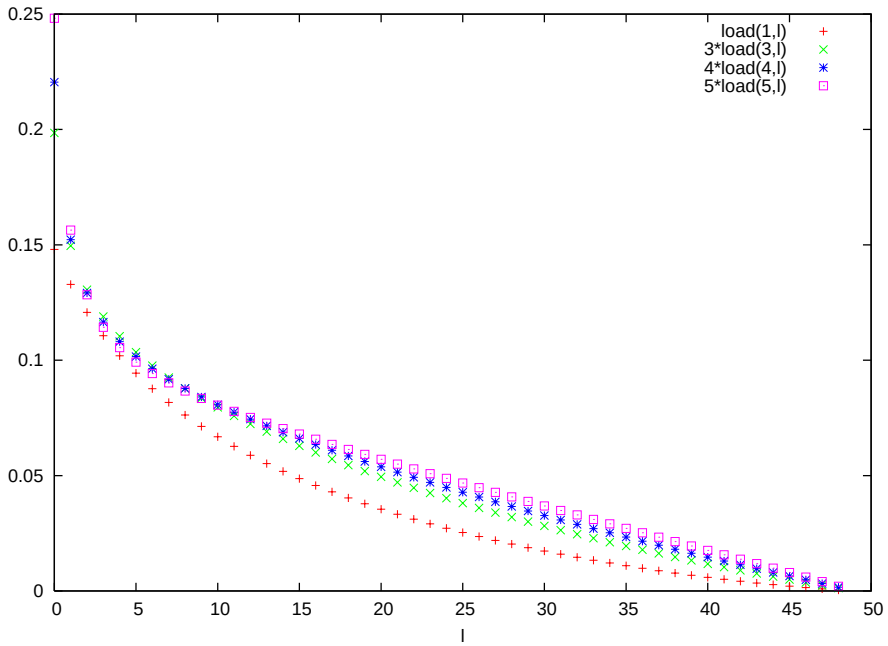


FIG. 3.7 : $\text{load}^{\mathcal{G}}(k, \ell)$ pour $n = 50$ et $k \in \{1, 3, 4, 5\}$.

On observe une charge décroissante en ℓ, c'est-à-dire que le ratio de l'espérance du nombre de chemins traversant un lien sur le nombre total de chemins est décroissante avec sa longueur. Cette décroissance est assez lente, sauf pour les petites valeurs de ℓ ($\ell \leqslant 5$). Dans le cas $k = 1$, on observe une décroissance quasi linéaire en échelle semi-logarithmique à partir d'une longueur de lien $\ell = 4$. Pour des degrés $k > 1$, la décroissance en échelle semi-logarithmique n'est plus linéaire. Ces observations expérimentales montrent que

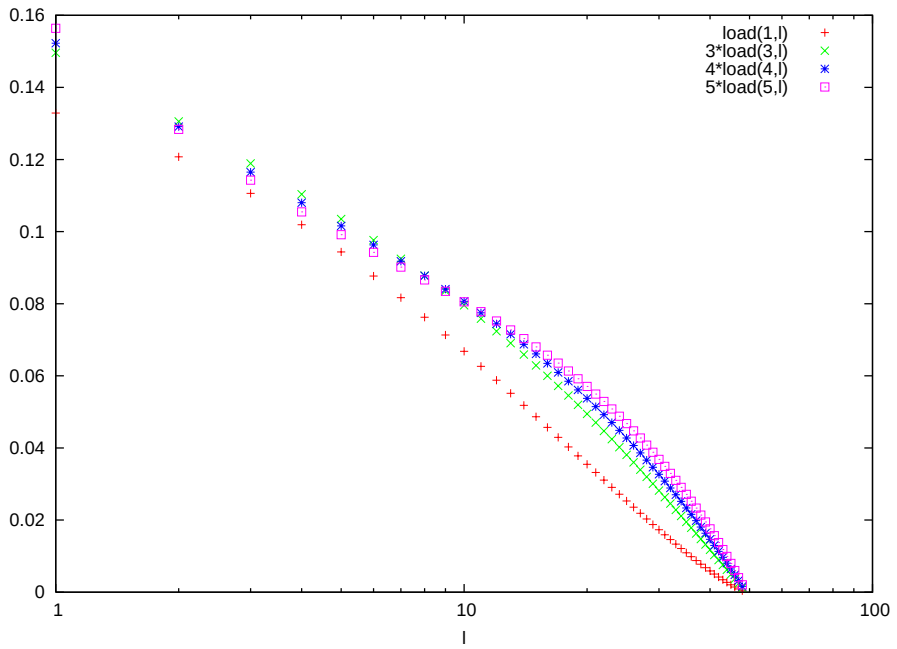


FIG. 3.8 : load$^{\mathcal{G}}(k,\ell)$ pour $n = 50$ et $k \in \{1,3,4,5\}$ en échelle semi-logarithmique.

$\mathcal{G}$ utilise surtout les liens longue-distance très courts (chargés à plus de 10%) qui sont aussi les plus fréquents. Cependant, les liens de grande longueur (supérieure à $n/2$) sont utilisés de façon non négligeable et le sont d'autant plus qu'on augmente le degré sortant k ; il s'agit donc d'une distribution à queue lourde.

3.2.2 Analyse de l'algorithme d'exploration

Pour analyser la charge induite par l'algorithme $\mathbf{E}_\omega$ sur un lien longue-distance donné L, nous cherchons à décomposer l'événement « L appartient au chemin de routage » en événements disjoints correspondant aux différentes étapes possibles précédant la visite de son origine, comme dans l'analyse de $\mathcal{G}$. Cette analyse se révèle plus difficile pour l'algorithme $\mathbf{E}_\omega$. En effet, un lien longue-distance sera emprunté s'il est le meilleur de sa fenêtre d'exploration F ; celle-ci peut débuter en chacune des $\omega + 1$ positions à gauche de son origine. La position de la racine de la fenêtre d'exploration correspond en outre à l'achèvement d'une exploration précédente ; il s'agit du contact longue-distance du dernier lien longue-distance suivi *ou* du dernier noeud de la fenêtre précédente, si celle-ci ne contenait aucun lien longue-distance valide pour le routage (fenêtre en échec).

Par ailleurs, dans le cas de l'algorithme $\mathbf{E}_\omega$, la probabilité de visite d'un noeud

n'apporte pas suffisamment d'information pour le calcul de la charge. Nous calculons donc directement la probabilité qu'un lien longue-distance soit emprunté en fonction de sa longueur, de sa distance à la source et de la distance entre la source et la cible.

3.2.2.1 Charge d'un lien longue-distance pour une source et une cible données

Pour cette section, on se donne $n, \omega \geqslant 1$ ainsi qu'une source $\mathbf{s}$ et une cible $\mathbf{t}$ pour le routage de $\mathbf{E}_\omega$ dans $\vec{\mathcal{K}}_{1,n}$ et l'on note $m = \vec{\delta}(\mathbf{s}, \mathbf{t})$. Du fait de la difficulté de l'analyse de la charge pour cet algorithme, nous nous restreignons au cas $k = 1$.

Nous commençons par calculer d'une part la probabilité qu'une fenêtre d'exploration F soit en échec (lemme 3.9) et d'autre part celle qu'un lien longue-distance soit le meilleur d'une fenêtre donnée (lemme 3.10). En calculant ces deux probabilités pour toutes les positions possibles d'une fenêtre, nous pourrons calculer la probabilité qu'un lien longue-distance de longueur et d'origine données appartienne au chemin calculé par $\mathbf{E}_\omega$ entre $\mathbf{s}$ et $\mathbf{t}$ (proposition 3.11).

Lemme 3.9 *Soit* $\mathbf{x}$ *le porteur courant du message, on note* $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$. *Pour tout* $0 \leqslant x \leqslant m - \omega - 1$, *la probabilité* $Echec(x, m)$ *que la fenêtre d'exploration* F *enracinée en* $\mathbf{x}$ *soit en échec est :*

$$Echec(x, m) = \left(1 - \frac{H_{m-\omega-x}}{H_n}\right) \prod_{y=0}^{\omega-1} \left(1 - \frac{H_{m-y-x} - H_{\omega-y}}{H_n}\right).$$

Preuve. La fenêtre $F = [\mathbf{x}, \mathbf{x} + \omega]$ est en échec si et seulement si tous les contacts longue-distance des noeuds de F sont hors de l'intervalle $[\mathbf{s} + x + \omega + 1, \mathbf{t}]$. Soit encore : si pour tout noeud $\mathbf{y}$ de F, à distance $y \leqslant \omega$ de $\mathbf{x}$, son lien longue-distance est de longueur supérieure stricte à $m - y - x$ ou inférieure stricte à $\omega - y + 1$. D'après le fait 3.5, la probabilité de cet événement est égale à :

$$\left(1 - \frac{H_{m-y-x} - H_{\omega-y}}{H_n}\right) \text{ si } y \neq \omega,$$
$$\left(1 - \frac{H_{m-\omega-x}}{H_n}\right) \text{ si } y = \omega.$$

Les liens longue-distance des noeuds de F étant tirés indépendamment, la probabilité de l'intersection de ces événements est le produit de ces probabilités. □

Lemme 3.10 *Soit* $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$ *l'origine d'un lien longue-distance* L *de longueur* ℓ*, on note* $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$*. La probabilité conditionnelle* $Meilleur(i, x, \ell, m)$ *que* L *appartienne au chemin calculé par l'algorithme* $\mathbf{E}_\omega$ *de* $\mathbf{s}$ *à* $\mathbf{t}$*, sachant qu'une fenêtre d'exploration est enracinée en* $\mathbf{y} = \mathbf{x} - i$ *avec* $0 \leqslant i \leqslant \omega$*, vaut :*

$$Meilleur(i, x, \ell, m) = 0 \quad \textit{si} \quad \ell \leqslant \omega - i \textit{ ou } \ell > \omega - x,$$

$$Meilleur(i, x, \ell, m) = \sum_{\alpha=0}^{\omega} \Big[\sum_{j_1 < \cdots < j_\alpha \in \{0,\ldots,\omega\} \setminus \{i\}} \frac{1}{\alpha+1} \prod_{q=1}^{\alpha} \left(\frac{1}{(\ell + i - j_q) H_n} \right) \times \prod_{j \in \{0,\ldots,\omega\} \setminus \{i, j_1, \ldots, j_\alpha\}} \left(1 - \frac{H_{m-x+i-j} - H_{i-j+\ell-1}}{H_n} \right) \Big] \quad \textit{sinon.}$$

Preuve. L'événement considéré ne dépend que de l'ensemble des longueurs des liens longue-distance des noeuds de la fenêtre d'exploration $[\mathbf{x} - i, \mathbf{x} - i + \omega]$.

L appartient au chemin de routage s'il est le meilleur de la fenêtre. Cela signifie que son contact longue-distance $L(\mathbf{x})$ appartient à $[\mathbf{x} - i + \omega + 1, \mathbf{t}]$ *et* est plus proche de $\mathbf{t}$ que les autres contacts de la fenêtre qui appartiennent à $[\mathbf{x} - i + \omega + 1, \mathbf{t}]$. Si $\ell \leqslant \omega - i$ ou $\ell > \omega - x$, $L(x)$ n'appartient pas à $[\mathbf{x} - i + \omega + 1, \mathbf{t}]$ et la probabilité est nulle.

Supposons maintenant que $L(x)$ appartient à $[\mathbf{x} - i + \omega + 1, \mathbf{t}]$. On peut décomposer cet événement en une union de $\omega + 1$ événements disjoints $\mathcal{E}_\alpha$, $0 \leqslant \alpha \leqslant \omega$, en fonction du nombre de liens ex-aequo avec L. Pour tout $0 \leqslant \alpha \leqslant \omega$, $\mathcal{E}_\alpha$ est l'événement « α noeuds de $[\mathbf{x} - i, \mathbf{x} - i + \omega]$, distincts de $\mathbf{x}$, ont le même contact longue-distance $L(\mathbf{x})$ que $\mathbf{x}$, aucun des $\omega - \alpha$ autres noeuds de la fenêtre n'a son contact longue-distance dans $[L(\mathbf{x}), \mathbf{t}]$, *et* L est choisi comme meilleur parmi les ex-aequo ». On remarquera que la probabilité que L soit choisi uniformément lorsqu'il y a α autres liens ex-aequo est $1/(\alpha + 1)$.

Chaque $\mathcal{E}_\alpha$ se décompose à son tour en une somme d'événements disjoints $\mathcal{E}_{j_1,\ldots,j_\alpha}$, chacun correspondant à la restriction de $\mathcal{E}_\alpha$ à un des placements possibles $0 \leqslant j_1 < \ldots < j_\alpha \leqslant \omega$, $j_q \neq i$ pour tout $1 \leqslant q \leqslant \alpha$, des α origines $\mathbf{x}_{j_q}$ des liens ex-aequo dans la fenêtre, où $\vec{\delta}(\mathbf{s}, \mathbf{x}_{j_q}) = \mathbf{x} - i + j_q$. La figure 3.9 illustre ces notations dans le cas où $\alpha = 2$; les noeuds de la fenêtre sont gris clair et les ex-aequo sont gris foncé.

Comme il s'agit d'événements disjoints, on a :

$$Meilleur(i, x, \ell, m) = \sum_{\alpha=0}^{\omega} \sum_{j_1 < \cdots < j_\alpha \in \{0,\ldots,\omega\} \setminus \{i\}} \frac{1}{\alpha+1} \cdot \Pr \mathcal{E}_{j_1,\ldots,j_\alpha}.$$

L'événement $\mathcal{E}_{j_1,\ldots,j_\alpha}$ est l'intersection des événements e_q : « $L(\mathbf{x} - i + j_q) = L(\mathbf{x})$ », pour tous les indices q tels que $1 \leqslant q \leqslant \alpha$, et des événements e'_j :« $L(j)$ est

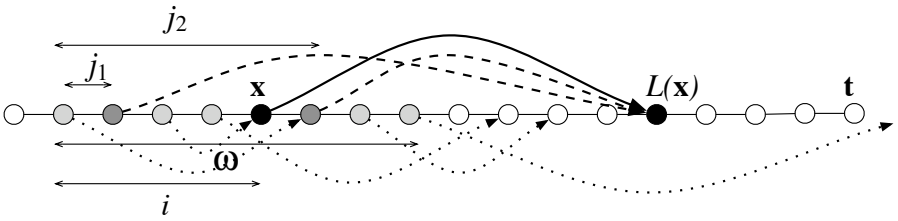


FIG. 3.9 : Illustration du lemme 3.10.

en dehors de l'intervalle $[L(\mathbf{x}), \mathbf{t}]$ » pour tous les autres indices $j \in \{0, \ldots, \omega\} \setminus \{i, j_0, \ldots, j_\alpha\}$. Par définition de la distribution, on a, pour tout $1 \leqslant q \leqslant \alpha$:

$$\Pr e_q = \frac{1}{(\ell + i - j_q)H_n},$$

et, à l'aide du fait 3.5, on obtient, pour tout $j \in \{0, \ldots, \omega\} \setminus \{i, j_0, \ldots, j_\alpha\}$:

$$\Pr e'_j = 1 - \frac{H_{m-x+i-j} - H_{i-j+\ell-1}}{H_n}.$$

Comme les liens sont tirés de façon indépendante, on peut faire le produit de ces probabilités, ce qui conduit au résultat. □

Proposition 3.11 *Soit* $\mathbf{x}$ *l'origine d'un lien longue-distance* L *de longueur* ℓ *et notons* $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$. *Soit* $Racine(r, m)$ *la probabilité que le noeud* $\mathbf{r}$, *avec* $r = \vec{\delta}(\mathbf{s}, \mathbf{r})$, *soit la racine d'une fenêtre d'exploration de* $\mathbf{E}_\omega$ *lors du routage de* $\mathbf{s}$ *vers* $\mathbf{t}$.

La probabilité $e(x, \ell, m)$ *que* L *appartienne au chemin calculé par* $\mathbf{E}_\omega$ *de* $\mathbf{s}$ *à* $\mathbf{t}$ *est :*

$$\begin{aligned}
&e(x, \ell, m) = 0 \textit{ si } x \in \{m - \ell + 1, \ldots, n\},\\
&e(x, \ell, m) = Meilleur(x, x, \ell, m) \textit{ si } x \in \{0, \ldots, \omega\},\\
&e(x, \ell, m) = \sum_{i=0}^{\omega} \Big[Racine(x - i, m) \cdot Meilleur(i, x, \ell, m)\Big]\\
&\qquad\qquad\qquad\qquad \textit{si } x \in \{\omega + 1, \ldots, m - 1\},
\end{aligned}$$

où l'on a, pour tout $0 \leqslant r \leqslant m - \omega$ *:*

$$\begin{aligned} Racine(r,m) &= Racine(r-\omega,m) \cdot Echec(r-\omega,m) \\ &\qquad + \sum_{j=0}^{r-1} \frac{e(j,r-j,m)}{(r-j)H_n} \textit{ si } r \geqslant \omega, \\ &= 0 \textit{ si } r \in \{1,\ldots,\omega-1\}, \\ &= 1 \textit{ si } r = 0. \end{aligned}$$

Preuve. Si x est strictement supérieur à $m - \ell$, $L(\mathbf{x})$ est au-delà de $\mathbf{t}$ et L n'appartient jamais au chemin de routage ; la probabilité est donc nulle.

Si $1 \leqslant x \leqslant \omega$, $\mathbf{x}$ appartient à la première fenêtre d'exploration de $\mathbf{E}_\omega$, enracinée en $\mathbf{s}$. Par définition de la fonction $Meilleur$ du lemme 3.10, la probabilité que L soit sur le chemin est donc $Meilleur(x,x,m)$.

Supposons à présent $x \in \{\omega+1,\ldots,m-\ell\}$. On note $\mathcal{E}_{x,\ell,m}$ l'événement associé à la probabilité $e(x,\ell,m)$ et $\mathcal{R}_{r,m}$ l'événement associé à la probabilité $Racine(r,m)$. L appartient au chemin de routage si et seulement si $\mathbf{x}$ appartient à une fenêtre d'exploration du routage et s'il est le meilleur lien longue-distance de cette fenêtre. On peut alors décomposer $e(x,\ell,m)$ en probabilités conditionnelles selon la position de la racine de l'exploration à laquelle $\mathbf{x}$ appartient :

$$e(x,\ell,m) = \sum_{i=0}^{\omega} \Pr\left(\mathcal{E}_{x,\ell,m} \,|\, \mathcal{R}_{x-i,m}\right)) \cdot Racine(x-i,m).$$

D'après la définition de la fonction $Meilleur$ du lemme 3.10, on a :

$$\Pr\{\mathcal{E}_{x,\ell,m} \,|\, \mathcal{R}_{x-i,m}\} = Meilleur(i,x,\ell,m).$$

Il reste à évaluer $Racine(x-i,m)$. Le noeud $\mathbf{y} = \mathbf{x} - i$ est la racine d'une fenêtre d'exploration de $\mathbf{E}_\omega$ si :

- soit c'est l'extrémité d'un lien longue-distance du chemin de routage (événement $\mathcal{R}^1_{x-i,m}$) ;
- soit le noeud $\mathbf{y} - \omega$ est la racine d'une fenêtre d'exploration et cette exploration est en échec (événement $\mathcal{R}^2_{x-i,m}$).

Comme ce sont des événements disjoints, on a :

$$Racine(x-i,m) = \Pr \mathcal{R}^1_{x-i,m} + \Pr \mathcal{R}^2_{x-i,m}.$$

On peut décomposer $\mathcal{R}^1_{x-i,m}$ en l'union de $x-i$ événements disjoints correspondant aux $x-i$ positions j possibles de l'origine du lien longue-distance du chemin de routage dont $\mathbf{y}$ est le contact longue-distance, $0 \leqslant j \leqslant x-i-1$. Ils sont chacun de probabilité $e(j,x-i-j,m)$, pour $0 \leqslant j \leqslant x-i-1$, puisque la longueur

d'un tel lien est $x - i - j$. Par ailleurs, la probabilité que ce lien soit de longueur $x - i - j$ est $1/((x - i - j)H_n)$, on obtient :

$$\Pr \mathcal{R}^1_{x-i,m} = \sum_{j=0}^{x-i-1} \frac{e(j, x - i - j, m)}{(x - i - j)H_n}.$$

D'après la définition de $Echec$ donnée par le lemme 3.9, on a :

$$\Pr \mathcal{R}^2_{x-i,m} = Racine(x - i - \omega, m) \cdot Echec(x - i - \omega, m),$$

puisque l'événement que $\mathbf{x} - i - \omega$ soit la racine d'une fenêtre d'exploration est indépendant du tirage des liens longue distance situés sur $[\mathbf{x} - i - \omega, \mathbf{t}]$.

Enfin, on initialise la récurrence sur $Racine$ grâce aux ω premières valeurs qui ne sont pas aléatoires puisque :

- $\mathbf{s}$ est toujours la racine de la première fenêtre d'exploration ($Racine(0, m) = 1$) ;
- aucun noeud de cette fenêtre n'est susceptible d'être la racine d'une fenêtre suivante durant l'exécution de $\mathbf{E}_\omega$ ($Racine(r, m) = 0$ pour tout $r \in \{1, \ldots, \omega\}$).

□

3.2.2.2 Charge d'un lien longue-distance de longueur ℓ

L'expression de la probabilité calculée dans la proposition 3.11 nous permet d'en déduire facilement la charge induite par l'algorithme $\mathbf{E}_\omega$ d'un lien d'origine et de longueur données.

Proposition 3.12 *Pour tout $n \geqslant 1$, soit $load^{\mathbf{E}_\omega}(\ell)$ la charge induite par $\mathbf{E}_\omega$ dans $\vec{\mathcal{K}}_{1,n}$ d'un lien longue-distance de longueur $\ell \geqslant 1$ et d'origine $\mathbf{x}$. On a :*

$$load^{\mathbf{E}_\omega}(\ell) = \frac{1}{n^2} \sum_{m=\ell}^{n-1} \sum_{x=0}^{m-\ell} e(x, \ell, m).$$

Preuve. On considère le routage simultané entre tous les n^2 couples $(\mathbf{s}, \mathbf{t})$ du réseau. Soit un couple $(\mathbf{s}, \mathbf{t})$ et notons $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$ et $m = \vec{\delta}(\mathbf{s}, \mathbf{t})$. Alors, $e(x, k, \ell, m)$ est l'espérance de la variable aléatoire qui vaut 1 si le lien longue distance de $\mathbf{x}$ appartient au chemin de routage de $\mathbf{s}$ vers $\mathbf{t}$ et 0 sinon. Par linéarité de l'espérance, il suffit de sommer les valeurs $e(x, k, \ell, m)$ sur tous les couples (x, m) possibles pour obtenir l'espérance du nombre de chemins utilisant le lien considéré. En remarquant que les contributions des couples $(\mathbf{s}, \mathbf{t})$ tels que $x > m - \ell$ ou $m < \ell$ sont nulles, on obtient le résultat. □

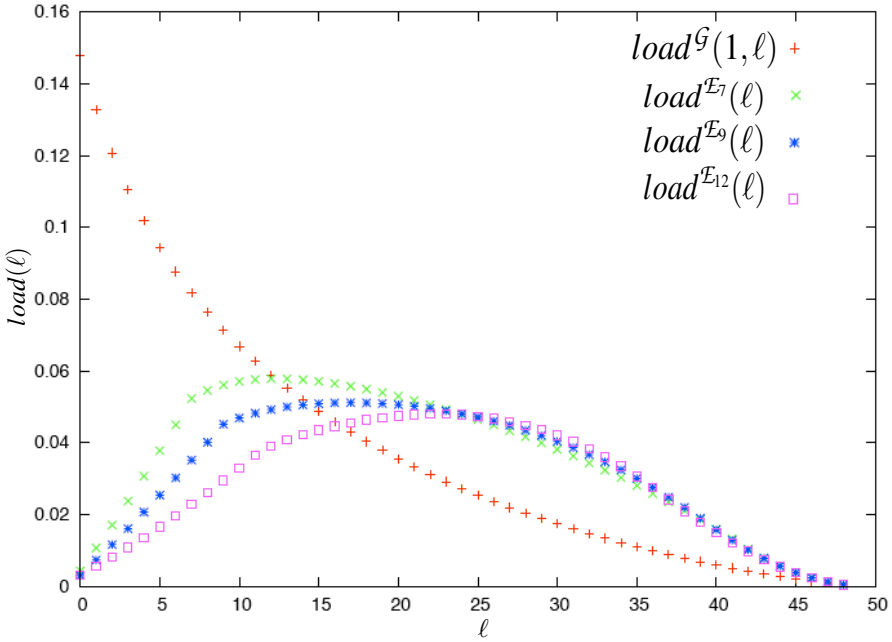


FIG. 3.10 : $\text{load}^{G}(\ell)$ et $\text{load}^{\mathbf{E}_\omega}(\ell)$ pour $n = 50$ et $\omega \in \{7, 9, 12\}$.

Observation des tracés. La figure 3.10 présente les tracés de la charge d'un lien longue-distance en fonction de sa longueur pour les algorithmes $\mathcal{G}$ et $\mathbf{E}_\omega$. Nous avons représenté des tracés de $\text{load}^{\mathbf{E}_\omega}(\ell)$ pour des valeurs assez grandes de ω afin d'accentuer les différences de charge en comparaison de l'algorithme glouton. Pour des valeurs inférieures de ω, les courbes pour $\mathcal{G}$ et $\mathbf{E}_\omega$ sont plus proches, mais la charge induite par $\mathbf{E}_\omega$ présente le même comportement que celui décrit ici.

On observe une rupture de pente de la charge induite par $\mathbf{E}_\omega$ lorsque $\ell = \omega$. On peut expliquer cette rupture en observant que si le lien considéré est de longueur inférieure à ω, il ne peut appartenir à aucun des chemins de routage dont son origine est la source ; ce phénomène disparaît à partir de la valeur ω. La variation de la charge induite par $\mathbf{E}_\omega$, croissante puis décroissante, peut s'expliquer par deux effets opposés de l'accroissement de la longueur ℓ :

- le lien augmente sa probabilité d'être le meilleur d'une fenêtre d'exploration,
- mais le nombre de chemins susceptibles d'utiliser le lien diminue.

On observe également que plus la taille de la fenêtre d'exploration augmente, plus la charge se déplace vers les liens les plus longs, du fait du choix d'un meilleur lien parmi un échantillon plus grand.

On constate que la charge induite par l'algorithme à exploration locale $\mathbf{E}_\omega$ se répartit de façon très différente à celle de $\mathcal{G}$. En effet, pour l'algorithme à exploration, la charge se répartit davantage sur toutes les longueurs de liens, alors que l'algorithme glouton charge sensiblement plus les liens les plus courts. Précisément, on observe que la plupart des longueurs de liens sont chargés à 4% pour l'algorithme à exploration, tandis que l'algorithme glouton charge les liens de faible longueur à plus de 10%. Cette meilleure répartition serait un avantage si les longueurs de liens étaient réparties uniformément ; mais les liens longs sont plus rares que les liens courts dans $\vec{\mathcal{K}}_{1,n}$. Il y a donc un risque de surcharge des liens les plus longs, dû au choix d'un meilleur lien parmi un échantillon.

On conclut de l'observation de ces courbes que l'observation expérimentale de la charge sur un réseau réel peut donner des informations sur le routage qui y est pratiqué sous l'hypothèse d'une certaine distribution des liens. Par exemple, dans le cadre de la compréhension des réseaux sociaux, l'observation expérimentale de la différence de fréquence d'utilisation des connexion entre individus proches et entre individus éloignés (selon un critère qui peut être géographique ou professionnel par exemple) peut donner des indications sur le type d'algorithme utilisé pour le routage (d'information, ou de lettres par exemple). En particulier, si ces deux types de connexion sont chargés de façon équivalente, l'hypothèse du routage glouton doit être écartée si l'on suppose que la distribution des longueurs des connexions est proche de la distribution harmonique. Cette étude n'a pas encore été menée à notre connaissance, elle permettrait de répondre à cette question.

Par ailleurs, dans la perspective de la construction de réseaux informatiques décentralisés, comme les réseaux pair-à-pair, ces observations posent la question de la modification de la distribution des liens. En effet, l'algorithme à exploration locale améliore les longueurs des chemins de routage, ce qui est un paramètre crucial pour ces réseaux, mais il risque de surcharger les liens très longs car ils sont rares. Une question est donc de déterminer si l'on peut modifier la distribution pour augmenter le nombre de liens très longs (de longueur supérieure à $n/2$) tout en conservant la propriété de petit monde navigable et le gain d'efficacité de l'algorithme à exploration.

3.3 Autres paramètres de comparaison

Afin de préciser notre comparaison des deux algorithmes de routage $\mathcal{G}$ et $\mathbf{E}_\omega$, nous étudions, dans cette section, deux autres paramètres qui sont le taux d'utilisation de l'anneau sous-jacent et la longueur des chemins entre des couples $(\mathbf{s}, \mathbf{t})$ proches (à distance très inférieure à n).

Les liens longue-distance jouent un rôle essentiel dans la réduction exponen-

tielle du diamètre de l'anneau, c'est pourquoi il est intéressant de quantifier l'utilisation de ces raccourcis dans les chemins de routage. Par ailleurs, cela permet d'étudier une éventuelle surcharge du réseau provoquée par la sur-utilisation de l'anneau sous-jacent si ces raccourcis sont très peu utilisés.

Il est connu que la longueur des chemins calculés par $\mathbf{E}_\omega$ est asymptotiquement plus courte que celle des chemins calculés par $\mathcal{G}$ lorsque la dimension est strictement supérieure à 1, pour une source et une cible à distance n (puisque cet algorithme est une version simplifiée de [FGP04]). Nous comparons ici les performances en dimension 1, sur toutes les distances possibles source-cible pour un taille d'anneau fixée, et pour la distance source-cible maximale en faisant varier la taille de l'anneau.

3.3.1 Taux moyen de liens locaux utilisés dans un routage

Nous avons vu dans la proposition 3.6 qu'il existe une distance critique source-cible en deçà de laquelle les liens longue-distance ne sont pas utilisés par l'algorithme glouton. Nous étudions ici précisément le taux de liens locaux utilisés dans les chemins de routage de $\mathcal{G}$ et $\mathbf{E}_\omega$ en fonction de la distance séparant la source de la cible. Pour cela, nous calculons le ratio de l'espérance du nombre de liens locaux utilisés sur un chemin, sur l'espérance du nombre de liens du chemin.

Définition 3.4
Soient $k, n \geqslant 1$ et $\mathcal{A}$ un algorithme de routage sur $\vec{\mathcal{K}}_{k,n}$. Pour tout $m \in \{1, \ldots, n\}$, le taux de liens locaux de $\mathcal{A}$ pour m, *noté $R^{\mathcal{A}}_{local}(k, m)$, est le ratio de l'espérance du nombre de liens locaux sur le chemin de routage calculé par $\mathcal{A}$ entre une source et une cible à distance m, sur l'espérance du nombre de liens de ce chemin.*

Pour toute cette section, on se donne $n \geqslant 1$ et deux noeuds $\mathbf{s}, \mathbf{t} \in \vec{\mathcal{K}}_{k,n}$ qui sont la source et la cible du routage, et l'on note $m = \vec{\delta}(\mathbf{s}, \mathbf{t})$.

Pour $\mathcal{G}$, l'espérance du nombre de liens d'un chemin de routage est $\Theta(\log^2 n)$ [Kle00] ; c'est également la somme sur tous les noeuds de l'intervalle $[\mathbf{s}, \mathbf{t}]$ de leur probabilité d'être visité. On rappelle qu'étant donné un noeud $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$, $g(\vec{\delta}(\mathbf{s}, \mathbf{x}), m)$ désigne la probabilité que $\mathbf{x}$ soit visité par $\mathcal{G}$ lors du routage de $\mathbf{s}$ à $\mathbf{t}$.

Proposition 3.13

$$R^{\mathcal{G}}_{local}(k,m) = \frac{\displaystyle\sum_{x=0}^{m-1} g(x,m)\left(1-\frac{H_{m-x}}{H_n}\right)^k}{\displaystyle\sum_{y=0}^{m-1} g(y,m)} = \frac{\displaystyle\sum_{x=0}^{m-1} g(x,m)\left(1-\frac{H_{m-x}}{H_n}\right)^k}{\Theta(\log^2 n)}.$$

Preuve. Pour $\mathcal{G}$, l'espérance du nombre de liens sur le chemin de routage est aussi l'espérance du nombre de noeuds, soit $\sum_{y=0}^{m-1} g(y,m)$, par linéarité de l'espérance.

Soit $\mathbf{x}$ un noeud de $[\mathbf{s},\mathbf{t}]$, on note $x = \vec{\delta}(\mathbf{s},\mathbf{x})$. La probabilité que son lien local (sortant) appartienne au chemin est la probabilité que le noeud $\mathbf{x}$ soit visité par $\mathcal{G}$ *et* que ses k contacts longue-distance n'appartiennent pas à $[\mathbf{x}+1,\mathbf{t}]$. En utilisant le fait 3.5 et en remarquant l'indépendance des événements, cette probabilité vaut, pour tout $0 \leqslant x \leqslant m-1$:

$$g(x,m)(1-H_{m-x}/H_n)^k.$$

On conclut en remarquant que l'espérance du nombre de liens locaux utilisés par l'algorithme est la somme de ces probabilités pour x variant de 0 à $m-1$, par linéarité de l'espérance. □

Pour déterminer le taux de liens locaux de l'algorithme $\mathbf{E}_\omega$, nous devons tout d'abord déterminer la probabilité qu'un lien local, d'origine donnée entre $\mathbf{s}$ et $\mathbf{t}$, appartienne au chemin.

On rappelle que, pour tout noeud $\mathbf{x}$, où l'on note $x = \vec{\delta}(\mathbf{s},\mathbf{x})$, $Racine(x,m)$ est la probabilité que $\mathbf{x}$ soit la racine d'une fenêtre d'exploration et $Echec(x,m)$ est la probabilité qu'une fenêtre enracinée en $\mathbf{x}$ soit en échec. Enfin, $Meilleur(x,\ell,m)$ est la probabilité que le lien longue-distance de $\mathbf{x}$ appartienne au chemin de routage, conditionnée par le fait que sa longueur soit ℓ.

Proposition 3.14 *Soient $k=1$ et $\mathbf{x} \in [\mathbf{s},\mathbf{t}]$, et notons $x = \vec{\delta}(\mathbf{s},\mathbf{x})$. La probabilité $e^{local}(x,m)$ que le lien local sortant de $\mathbf{x}$ appartienne au chemin calculé par $\mathbf{E}_\omega$ de $\mathbf{s}$ à $\mathbf{t}$ est :*

$$e^{local}(x,m) = \textit{Echec}(0,m) + \sum_{j=x+1}^{\omega}\sum_{\ell=\omega-j+1}^{m-j}\frac{1}{\ell H_n}Meilleur(j,j,\ell,m) \textit{ si } x \leqslant \omega,$$

$$e^{local}(x,m) = \sum_{i=0}^{\omega} Racine(x-i,m)\Big(\textit{Echec}(x-i,m) + \sum_{j=x+1}^{x+\omega-i}\sum_{\ell=\omega-i-j+x+1}^{m-j}\frac{1}{\ell H_n}Meilleur(i+j-x,j,\ell,m)\Big) \textit{ sinon.}$$

Preuve. Le lien local de $\mathbf{x}$ appartient au chemin de routage si et seulement s'il appartient à une fenêtre d'exploration qui est en échec *ou* dont le meilleur lien est au-delà de $\mathbf{x}$.

Supposons $x \leqslant \omega$. Le noeud $\mathbf{x}$ appartient alors à la première fenêtre d'exploration enracinée en $\mathbf{s}$. La probabilité que cette fenêtre soit en échec est $Echec(0, m)$. Si elle n'est pas en échec, on décompose l'événement en événements disjoints correspondant aux positions possibles du meilleur lien tel que le lien local de $\mathbf{x}$ soit sur le chemin. Les positions possibles du meilleur lien sont les noeuds de $[\mathbf{s}+x+1, \mathbf{s}+\omega]$. Pour chacun de ces noeuds $\mathbf{s}+j$, où $j \in \{1+x, \ldots, \omega\}$, la probabilité qu'il soit meilleur est la somme sur toutes les longueurs ℓ possibles qu'il soit de longueur ℓ (probabilité $1/(\ell H_n)$) *et* meilleur (probabilité $Meilleur(j, j, \ell, m)$). Pour chaque $\mathbf{s} + j$, les longueurs ℓ possibles du lien longue-distance sont celles qui permettent que $L(\mathbf{s} + j)$ appartienne à l'intervalle $[\mathbf{s} + \omega + 1, \mathbf{t}]$, soit encore $\ell \in \{\omega - j + 1, \ldots, m - j\}$. En sommant ces probabilités (ce sont des événements disjoints), on obtient l'expression de $e^{\text{local}}(x, m)$ pour $x \leqslant \omega$.

Supposons maintenant $x > \omega$. On décompose l'événement en le conditionnant selon les ω positions possibles $\mathbf{x} - i$ de la racine de la fenêtre d'exploration à laquelle $\mathbf{x}$ appartient, pour $i \in \{0, \ldots, \omega\}$. Chaque position $\mathbf{x} - i$ a une probabilité $Racine(x - i, m)$ de débuter une fenêtre d'exploration. L'événement conditionné par l'existence d'une racine d'exploration en $\mathbf{x} - i$ se décompose alors de façon similaire au cas précédent, par translation de vecteur $-i$. □

L'expression de cette probabilité nous permet de calculer le taux de liens locaux de $\mathbf{E}_\omega$ de façon similaire à l'algorithme glouton.

Proposition 3.15 *Pour $k = 1$,*

$$R_{local}^{\mathbf{E}_\omega}(m) = \frac{\sum_{x=0}^{m-1} e^{local}(x, m)}{\displaystyle\sum_{x=0}^{m-1} \left(e^{local}(x, m) + \sum_{\ell=1}^{m-x} \frac{e(x, \ell, m)}{\ell H_n} \right)}.$$

Preuve. Pour tout noeud $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$, notons $X_{\mathbf{x}}$ la variable aléatoire qui vaut 1 si le chemin utilise le lien local sortant de $\mathbf{x}$ et 0 sinon. Notons $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$. On a $\Pr(X_{\mathbf{x}}) = e^{local}(x, m)$. Par linéarité de l'espérance on obtient que l'espérance du nombre de liens locaux utilisés est $\sum_{x=0}^{m-1} e^{local}(x, m)$.

Le nombre de liens du chemin de routage de l'algorithme $\mathbf{E}_\omega$ est aussi le nombre de noeuds qu'il visite. Pour l'obtenir, on calcule donc la probabilité que le noeud $\mathbf{x}$ soit visité, pour tout $\mathbf{x} \in [\mathbf{s}, \mathbf{t}]$. Notons $x = \vec{\delta}(\mathbf{s}, \mathbf{x})$. Le noeud $\mathbf{x}$ est visité :

– si son lien local appartient au chemin (probabilité $e^{local}(x, m)$),

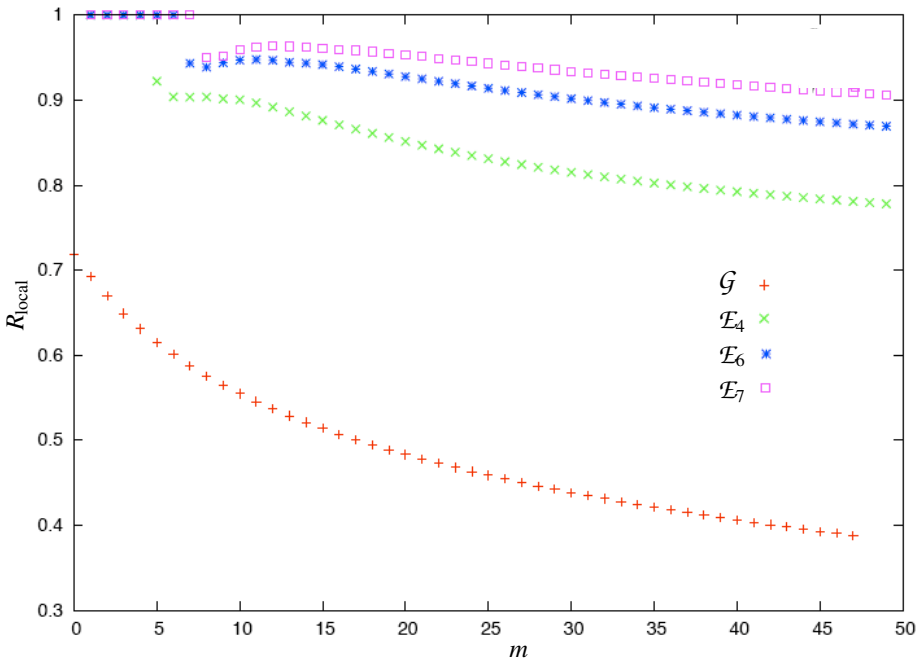


FIG. 3.11 : $R^{\mathcal{G}}_{\text{local}}(1, m)$ et $R^{\mathbf{E}_\omega}_{\text{local}}(m)$ pour $k = 1$, $\omega \in \{4, 6, 7\}$ et $n = 50$.

– *ou* si son lien longue-distance est de longueur $\ell \leqslant x - m$ (probabilité $1/(\ell H_n)$ pour chaque lien) et appartient au chemin (probabilité $e(x, \ell, m)$ pour chaque lien).

Dans le deuxième cas, seules les longueurs de lien ℓ inférieures ou égales à $m - x$ contribuent car les autres ont une probabilité nulle d'appartenir au chemin. On peut faire la somme des probabilités de ces deux cas car ce sont des événements disjoints. Par linéarité de l'espérance, on somme ces probabilités sur tous les noeuds $\mathbf{x}$ pour obtenir l'espérance du nombre de liens du chemin de routage. □

Observation des tracés. La figure 3.11 (et la figure 3.12 pour le tracé en échelle semi-logarithmique) représente les valeurs des taux de liens locaux pour un anneau de 50 noeuds et différentes profondeurs d'exploration $\omega \in \{4, 6, 7\}$ en fonction de m.

Lorsque m est inférieur ou égal à ω, l'algorithme $\mathbf{E}_\omega$ n'utilise aucun lien longue-distance puisque la cible est contenue dans sa première fenêtre d'exploration ; c'est pourquoi on observe un taux constant égal à 1 pour ces valeurs de m. On observe que ce taux est par la suite d'autant plus élevé que ω est grand. En effet, alors que la position du meilleur lien varie peu lorsqu'on augmente ω de 1, la longueur de la chaîne de liens locaux à suivre pour l'atteindre augmente en

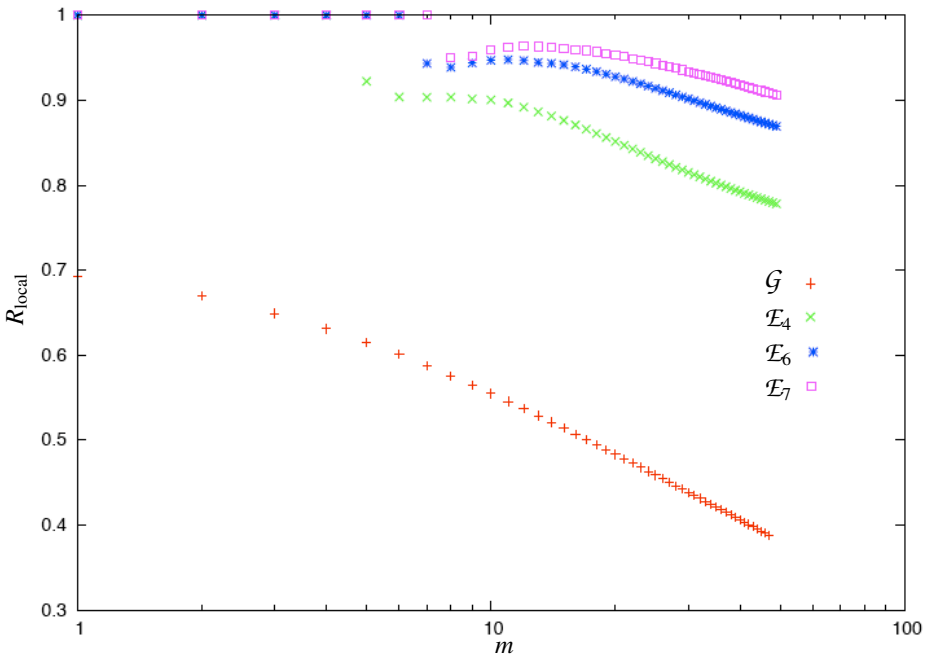


FIG. 3.12 : $R^{\mathcal{G}}_{\text{local}}(1,m)$ et $R^{\mathbf{E}_\omega}_{\text{local}}(m)$ pour $k = 1$, $\omega \in \{4,6,7\}$ et $n = 50$ en échelle semi-logarithmique.

moyenne de $1/2$.

En comparant les taux des deux algorithmes, on observe une utilisation bien plus importante de l'anneau sous-jacent par l'algorithme $\mathbf{E}_\omega$ sur les tracés de la figure 3.11. Sur le tracé en échelle semi-logarithmique (figure 3.12), on observe que $R^{\mathcal{G}}_{\text{local}}(1,m)$ décroît de façon quasi linéaire, suggérant une dépendance en $\log(1/m)$.

En combinant l'observation de la charge et celle du taux de liens locaux, on peut émettre l'hypothèse que l'algorithme $\mathbf{E}_\omega$ utilise des liens longue-distance très longs qui lui permettent de raccourcir le chemin de routage, mais qu'ils sont atteints en utilisant un grand nombre de liens locaux. Par ailleurs, l'algorithme glouton utilise des raccourcis moins longs, mais les atteint par des liens longue-distance de faible longueur plutôt que par des liens locaux.

3.3.2 Comparaison des longueurs de chemin

Fraigniaud *et al.* [FGP04] montrent que l'exploration des liens locaux sur une profondeur $O(\log^{1/d} n)$ permet de réduire l'espérance de la longueur du chemin de routage, entre une source et une cible à distance n, de $\Theta(\log^2 n)$ à $O(\log^{1+1/d} n)$,

dans un réseau de Kleinberg de dimension $d \geqslant 1$. Dans cette section, nous comparons les espérances des longueurs des chemins des algorithmes $\mathcal{G}$ et $\mathbf{E}_\omega$ sur l'anneau ($d = 1$) en étudiant en particulier le cas de couples source-cible rapprochés.

Soient Longueur$^{\mathbf{E}_\omega}(m)$ et Longueur$^{\mathcal{G}}(m)$ les espérances respectives de la longueur de chemin calculé par $\mathbf{E}_\omega$ et $\mathcal{G}$.

Proposition 3.16

$$\textit{Longueur}^{\mathcal{G}}(m) = \sum_{x=0}^{m-1} g(x,m) = \Theta(\log m \log n)$$

et

$$\begin{aligned}\textit{Longueur}^{\mathbf{E}_\omega}(m) &= m \quad \textit{si} \quad m \leqslant \omega, \\ &= \sum_{x=0}^{m-1} \Big(e^{local}(x,m) + \sum_{\ell=1}^{m-x} \frac{e(x,\ell,m)}{\ell H_n} \Big) \quad \textit{sinon.}\end{aligned}$$

Preuve. L'espérance de la longueur du chemin de routage est égale à l'espérance de son nombre de liens empruntés. Cette espérance ayant été calculée pour obtenir le taux de liens locaux, on se reportera à la preuve de la proposition 3.13 pour l'algorithme $\mathcal{G}$ et à celle de la proposition 3.15 pour l'algorithme $\mathbf{E}_\omega$. □

Observation des tracés. La figure 3.13 représente les espérances des longueurs de chemin des deux algorithmes sur un anneau de taille fixée ($n = 100$) pour toutes les distances possibles séparant la source de la cible. De façon naturelle, on observe que ce sont les chemins entre deux noeuds éloignés (m proche de n) qui bénéficient le plus de l'exploration locale de $\mathbf{E}_\omega$. Pour des distances m plus faibles, le gain en longueur obtenu grâce à la sélection de *meilleurs* liens longue-distance ne compense pas le grand nombre de liens locaux suivis sur les fenêtres d'exploration (dont la largeur est ici assez grande). On remarquera de façon intéressante que Longueur$^{\mathbf{E}_5}(m)$ n'est pas une fonction monotone de m pour n fixé, et qu'il existe une distance source-cible critique, dépendante de ω, pour laquelle le chemin calculé est plus long qu'entre les noeuds les plus éloignés de l'anneau ($m = n - 1$).

La figure 3.14 compare les longueurs de chemin des deux algorithmes obtenues pour les distances source-cible maximales ($m = n - 1$) en fonction de la taille de l'anneau n variant de 1 à 200. On observe que, dès que l'anneau dépasse 55 noeuds, la longueur de chemin calculée par $\mathbf{E}_\omega$ est toujours plus courte que celle calculée par $\mathcal{G}$; la croissance de la courbe Longueur$^{\mathbf{E}_3}(n-1)$ est en outre moins forte que celle de Longueur$^{\mathcal{G}}(n-1)$.

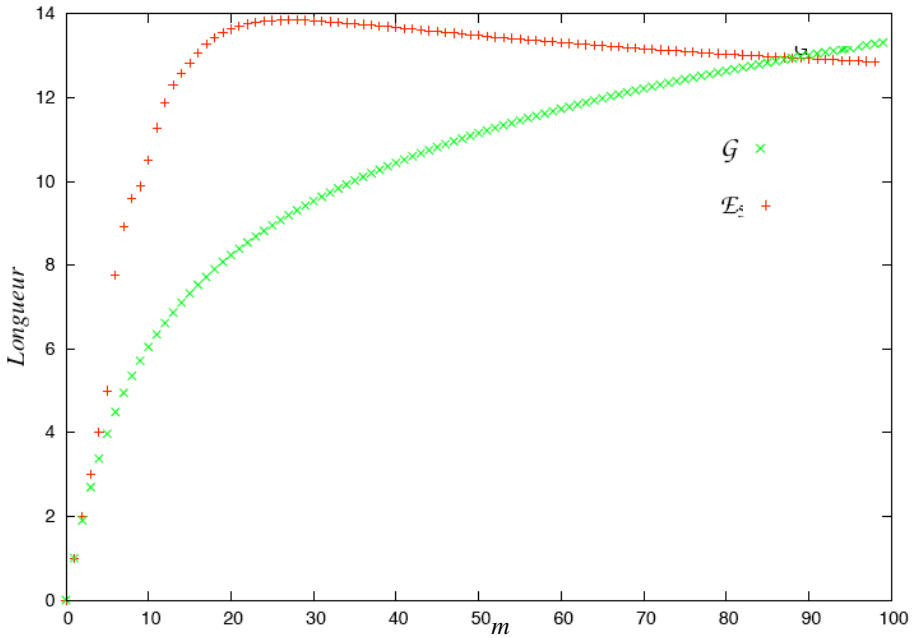


FIG. 3.13 : Longueur$^{\mathbf{E}_5}(m)$ et Longueur$^{\mathcal{G}}(m)$ pour $k = 1$ et $1 \leqslant m \leqslant n = 100$.

Ainsi, l'algorithme à exploration locale $\mathbf{E}_\omega$ est efficace en terme de longueur de chemin pour les couples source-cible très éloignés, tandis que les couples source-cible séparés d'une distance de l'ordre de ω sont peu rapprochés par cette méthode. Il est intéressant de remarquer que les longueurs de chemin pour les couples les plus éloignés ($m = n - 1$) sont plus courtes que celles de l'algorithme glouton à partir d'un anneau de seulement 55 noeuds. Alors qu'on ne disposait que d'une comparaison asymptotique des espérances des longueurs de chemin auparavant, il apparaît que le gain d'efficacité de $\mathbf{E}_\omega$ sur $\mathcal{G}$ est obtenu même à partir d'une taille de réseau très petite.

3.4 Conclusion et perspectives

Dans ce chapitre, nous avons comparé le fonctionnement de deux algorithmes décentralisés sur le réseau de Kleinberg unidimensionnel et unidirectionnel qui sont représentatifs des deux types d'algorithmes décentralisés étudiés jusqu'ici sur ce modèle de petit monde. Il s'agit, à notre connaissance, de la première analyse de la charge du réseau de Kleinberg.

L'analyse de la charge induite et de la longueur des chemins montre que l'utilisation d'une exploration locale des liens longue-distance construit des chemins

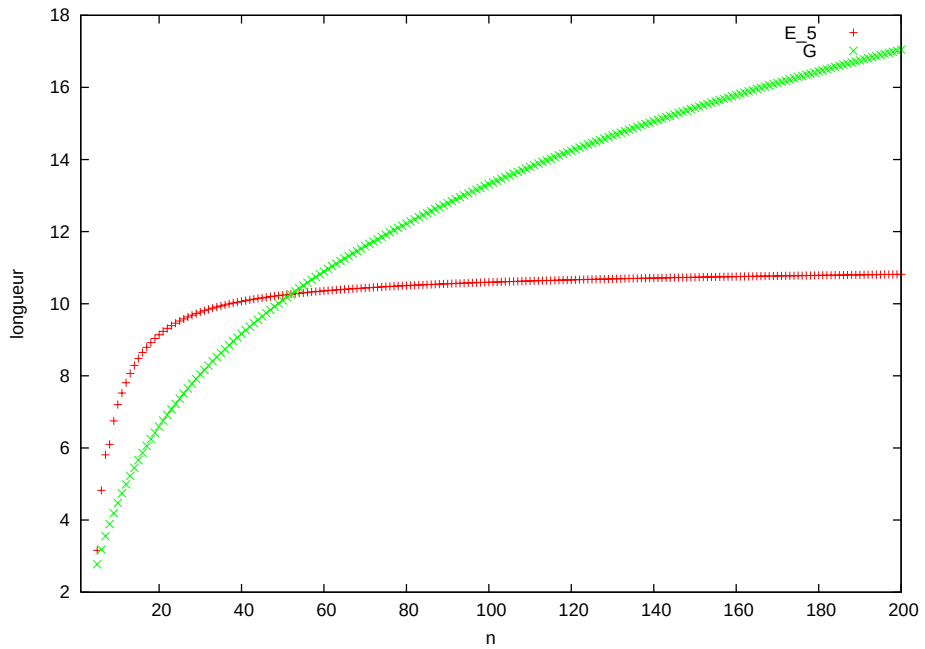


FIG. 3.14 : Longueur$^{\mathbf{E}_3}(n-1)$ et Longueur$^{\mathcal{G}}(n-1)$ pour $k = 1$ et $n \in \{1, \ldots, 200\}$.

contenant une part importante de liens locaux ; elle montre aussi que ces chemins contiennent des liens longue-distance assez longs par rapport à la taille du réseau. Entre une source et une cible très éloignées, ces chemins sont plus courts que ceux calculés par l'algorithme glouton même pour un réseau de très petite taille (environ 60 noeuds) ; mais ils surchargent les liens les plus longs, qui sont les plus rares, et l'anneau sous-jacent. Toutefois, nous n'avons pas analysé d'algorithme où l'exploration se fait à la fois sur les liens locaux et longue-distance, comme celui que nous avons présenté dans le chapitre précédent. Ce type d'exploration atténue *a priori* la charge de l'anneau sous-jacent. De même, il serait intéressant de reprendre l'analyse pour une taille de fenêtre d'exploration non constante qui s'adapte à la distance courante à la cible. En effet, plus la distance restant à parcourir est réduite, moins la marge d'optimisation est grande ; il est donc naturel de penser qu'une taille de fenêtre d'exploration décroissante avec la distance à la cible produirait de meilleurs résultats. C'est également le cas de l'algorithme que nous avons présenté dans le chapitre précédent.

Dans la perspective de construire, à partir de ce modèle, des réseaux informatiques décentralisés où le routage est efficace, ces observations posent la question du choix de la distribution des liens longue-distance. Si l'on veut bénéficier du gain en longueur de chemin de routage apporté par un algorithme à exploration

locale, cela pose la question de l'existence d'une distribution qui crée suffisamment de liens très longs pour éviter leur surcharge, tout en conservant la propriété de petit monde navigable.

Enfin, les observations de la charge de routage d'un modèle de petit monde peuvent apporter des indications pour la compréhension de la dynamique des réseaux réels. Gladwell a par exemple émis l'hypothèse que l'effet petit monde observé dans les réseaux sociaux serait dû à l'existence d'un petit nombre d'individus particulièrement bien connectés au reste du réseau et par qui tous les chemins passeraient [Gla98]. Si cette hypothèse est vraie, l'observation expérimentale du nombre de messages qui transitent entre les individus d'un réseau doit faire apparaître des connexions beaucoup plus chargées que la majorité des autres connexions. Comme nous l'avons remarqué précédemment, l'observation expérimentale de la charge d'un réseau réel apporte également des indications sur le type d'algorithme de routage utilisé sur ce réseau, sous l'hypothèse d'une distribution des liens proche de celle du réseau de Kleinberg. Notre étude de la charge sur ce réseau nous permet d'affirmer que si l'on observe une charge similaire des liens courts et des liens les plus longs, l'hypothèse de l'utilisation de l'algorithme glouton doit être écartée ; en effet, celui-ci charge davantage les liens courts, si l'on suppose que la distribution des liens du réseau est proche du modèle de Kleinberg.

Deuxième partie

Aspect structurel de l'effet petit monde

Chapitre 4

Problématique de la *petit-mondisation*

Introduction

Depuis qu'il est possible de collecter des données statistiques sur les grands réseaux réels, le graphe aléatoire uniforme d'Erdös-Rényi [ER59] s'est révélé être un modèle assez éloigné de ces réseaux, puisque des propriétés statistiques qui semblent pertinentes pour ces réseaux (la distribution des degrés, le coefficient de *clustering* et la navigabilité) y sont très différentes. Ces observations suggèrent la présence d'une structure sous-jacente dans ces réseaux, qui n'est pas simplement aléatoire uniforme et qui donne naissance à des propriétés particulières. Dans cette partie, nous étudions l'aspect structurel de la propriété de petit monde navigable, en cherchant à exhiber les caractéristiques structurelles minimales des graphes présentant cette propriété. Dans ce chapitre, nous étudions les limites du modèle de Kleinberg qui nous ont amené à la problématique de la *petit-mondisation*. Ces limites seront levées dans le chapitre suivant, où nous introduirons deux nouveaux modèles de graphes petits mondes qui le généralisent.

Il est intuitif de supposer qu'un réseau social ne se construit pas de façon aléatoire uniforme. En effet, les relations sociales se construisent plus facilement entre individus d'une communauté sociale similaire (même corps de métier, même lieu d'habitation) qu'entre individus choisis au hasard. Ces constructions spécifiques peuvent avoir une influence sur la facilité de naviguer à travers le réseau. On peut prendre l'exemple de la simulation de routage dans des réseaux sociaux récemment effectuée par Adamic et Adar [AA04], sur des réseaux reconstruits sur la base de carnets d'adresses électroniques et de fichiers mémorisant les échanges de messages. Les auteurs ont appliqué la stratégie gloutonne sur ces

réseaux et ont observé que la propriété de petit monde navigable était vérifiée. Les chemins les plus courts ont été observés sur le réseau reconstruit à partir des *e-mails* de l'entreprise *HP Labs* qui présente une structure hiérarchique, héritée de la structure hiérarchique professionnelle, mettant en valeur l'influence d'une structure sous-jacente dans les réseaux sociaux qui peut favoriser l'effet petit monde.

Barabasi et Albert ont montré que la construction dynamique d'un réseau selon une règle d'attachement préférentiel aux noeuds présents les mieux connectés produisait un réseau dont la distribution des degrés suivait une loi de puissance [BA99]. De façon analogue, la propriété de petit monde navigable peut provenir d'un processus d'ajout de raccourcis, basé sur une structure sous-jacente préexistante, comme l'ensemble des communautés professionnelles dans un réseau social par exemple.

Le modèle de petit monde de Kleinberg [Kle00] a montré que l'augmentation d'un tore régulier par des arcs aléatoires, dont la distribution est étroitement liée à la métrique de ce tore, produisait un graphe petit monde navigable. Ce fait pose une question essentielle pour la compréhension du phénomène petit monde : n'y a-t-il que des métriques spécifiques (comme le tore régulier muni de la norme ℓ_1) qui peuvent être transformées en petit monde par l'ajout de liens longue-distance ? Dans les chapitres suivants, nous montrerons qu'il existe une grande classe de graphes qui peuvent être transformés en petits mondes par l'ajout de liens longue-distance. On parle alors de processus de *petit-mondisation*.

4.1 Vers une généralisation du modèle de Kleinberg

4.1.1 Les généralisations pour les réseaux sociaux

Le réseau de Kleinberg est le premier modèle à reproduire la propriété de petit monde navigable et il présente l'avantage d'être très simple, ce qui en fait un bon outil pour l'analyse de cette propriété. Toutefois, sa structure sous-jacente de tore régulier, qui représente la connaissance globale partagée par tous les noeuds, paraît peu vraisemblable pour un réseau réel. Si l'on fait l'hypothèse que la connaissance globale d'un réseau social est l'ensemble des distances géographiques, on peut par exemple penser que celles-ci ne sont pas réparties selon une grille régulière. Des travaux récents se sont attachés à produire de nouveaux modèles de petit monde navigable, également construits à partir d'une structure sous-jacente que l'on augmente par des arcs aléatoires, pour chercher un modèle plus réaliste.

Le premier d'entre eux est la généralisation proposée par Kleinberg en 2002 [Kle02]. La structure sous-jacente est une structure de *groupes*, qui sont des ensembles de sommets qui respectent des contraintes de croissance et de recouvre-

ment. Chaque noeud $\mathbf{u}$ est l'origine de $O(\log n)$ arcs aléatoires, chacun pointant vers un autre noeud $\mathbf{v}$ avec une probabilité inversement proportionnelle à la taille du plus petit groupe qui contient $\mathbf{u}$ et $\mathbf{v}$. Kleinberg montre que l'algorithme de routage glouton (défini page 21) calcule des chemins de longueur polylogarithmique en le nombre de sommets en espérance dans ce modèle. Comme ce modèle ne contient aucune arête dans sa structure sous-jacente, il nécessite un degré sortant au moins logarithmique en le nombre de sommets pour garantir sa connexité. Le concept de groupe auquel fait référence ce modèle provient de la littérature des problèmes de propagation de rumeurs dans les réseaux sociaux, pour lesquels il est particulièrement adapté. Ce modèle a été conçu dans l'objectif d'un modèle plus réaliste des réseaux sociaux ; ce n'est en revanche pas un bon outil dans la perspective de construction de réseaux informatiques du fait de son degré sortant non constant. Ce degré sortant en fait également un modèle peu réaliste pour les réseaux réels dont on observe une distribution des degrés suivant une loi de puissance, puisque leur degré moyen est constant.

Au même moment, Watts *et al.* [WDN02] ont proposé un autre modèle de petit monde basé sur une structure arborescente, qui est en fait incluse dans la notion de groupe du modèle de [Kle02]. Il s'agit également d'un modèle dédié aux réseaux sociaux, l'arborescence modélisant la hiérarchie présente dans les connaissances sociales. Par exemple, la communauté des infirmières et celle des médecins sont incluses dans la communauté des personnels de santé : dans l'arborescence, ce sont deux fils du noeud qui représente la communauté des personnels de santé. Dans leur modèle, ils ajoutent z liens longue-distance par noeuds, où z est la taille maximale d'une communauté-feuille de l'arborescence et ils choisissent la longueur de chaque lien selon une distribution exponentielle. Cette longueur ℓ déterminée, ils choisissent uniformément un noeud parmi une communauté-feuille dont ℓ est la distance hiérarchique à l'origine du lien. Ils montrent alors, par des simulations expérimentales, que l'algorithme de routage glouton, avec la connaissance de la hiérarchie, trouve des chemins de longueur négligeable devant la taille du graphe.

4.1.2 Caractéristiques et limites structurelles du modèle

Nous avons vu que le tore d-dimensionnel augmenté d'un nombre constant de liens longue-distance par noeuds, distribués selon la distribution d-harmonique, est un petit monde navigable. Kleinberg [Kle00] puis Barrière *et al.* [BFKK01] ont montré que toute distribution s-harmonique des liens longue-distance telle que $s > 0$ est différent de la dimension du tore sous-jacent échoue à produire un graphe petit monde navigable. Ce résultat renforce l'impression que la navigabilité du modèle de Kleinberg est très spécifique à la topologie de tore régulier. Lorsque s appartient à $]d, 2d[$, Martel et Nguyen [MN04, MN05] ont démontré que

le diamètre du réseau était pourtant toujours polylogarithmique, toutefois aucun algorithme décentralisé ne peut trouver ces chemins courts.

Pour comprendre la spécificité de la distribution de liens d-harmonique sur le tore régulier d-dimensionnel, il faut remarquer que l'ensemble des noeuds à distance égale à $r \geqslant 1$ d'un noeud $\mathbf{u}$ donné (la sphère $\mathcal{S}_{\mathbf{u}}(r)$) contient $\Theta(r^{d-1})$ noeuds dans ce tore. Ainsi, si l'on ajoute un lien aléatoire sortant au noeud $\mathbf{u}$, dont la destination est le noeud $\mathbf{v}$ avec probabilité $1/(|\mathbf{u}-\mathbf{v}|^d Z_{n,d})$, pour tout $r \geqslant 1$, la probabilité que la destination soit à distance r de $\mathbf{u}$ est :

$$\sum_{\mathbf{v}\in\mathcal{S}_{\mathbf{u}}(r)} \frac{1}{r^d Z_{n,d}} = \Theta\left(\frac{1}{r Z_{n,d}}\right).$$

Nous avons vu dans les préliminaires que le facteur de normalisation $Z_{n,d}$ est $\Theta(\log n)$ (voir p. 22). On obtient alors que la probabilité que la destination de $\mathbf{u}$ soit à distance comprise entre r et $2r$ de $\mathbf{u}$ est indépendante de r :

$$\Theta\left(\frac{H_{2r}-H_r}{\log n}\right) = \frac{\Theta(1)}{\log n}.$$

On obtient donc un graphe où, quelle que soit la dimension du tore sous-jacent et pour tout $r \geqslant 1$, chaque contact longue-distance a une probabilité identique d'appartenir à chaque couronne de noeuds compris à distance entre $2^i \cdot r$ et $2^{i+1} \cdot r$ de son origine, pour $0 \leqslant i \leqslant \log n - 1$. Les liens sont donc *bien répartis* pour effectuer une recherche efficace de la cible par dichotomie. Au contraire, lorsque l'exposant s est strictement supérieur à d, les liens ont tendance à se condenser autour de leur origine et le graphe ne contient pas suffisamment de liens longs. Lorsque l'exposant s est strictement inférieur à d, les liens sont trop éloignés de leur origine, ce qui ralentit les dernières étapes d'un routage et ne permet pas non plus d'obtenir des routes polylogarithmiques.

La distribution des liens étant identique pour tous les noeuds, le raisonnement précédent repose sur l'homogénéité des tailles des sphères dans le tore régulier $n \times \ldots \times n$. Les sphères ont en effet des tailles qui ne dépendent que de leur rayon et non de leur centre, cela permet à la distribution harmonique de bien répartir les liens longue-distance de tous les noeuds. Cette homogénéité n'est plus présente si les côtés du tore sont de longueurs très différentes. La figure 4.1 représente par exemple le cas du tore de dimension 2 étiré le long d'un côté.

De façon intéressante, la propriété suivante montre que cette grille allongée ne peut pas être augmentée en un petit monde navigable par aucune distribution harmonique de liens longue-distance lorsque qu'un des côtés est une puissance constante de l'autre.

Soient $s > 0$ et G le tore $n^\alpha \times n$, où $\alpha > 1$, augmenté d'un arc aléatoire sortant par noeud $\mathbf{u}$ tel que la probabilité que sa destination soit $\mathbf{v}$ est proportionnelle à $1/|\mathbf{u}-\mathbf{v}|^s$.

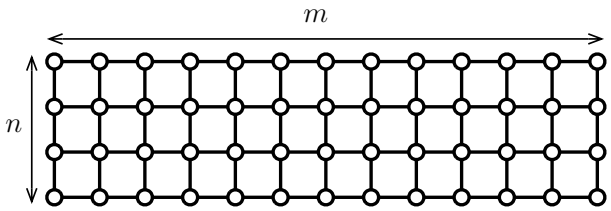


FIG. 4.1 : Grille allongée ($m \gg n$).

Proposition 4.1 *Quel que soit $s > 0$, l'espérance de la longueur du chemin de routage entre deux noeuds choisis aléatoirement uniformément dans G est supérieure à un polynôme en n, pour tout algorithme de routage décentralisé.*

Preuve. La preuve de cette borne inférieure repose sur les mêmes arguments que la preuve des bornes inférieures de [Kle00] dans le cas où les côtés du tores sont de même longueur pour les exposants différents de 2. Cependant, pour la grille allongée, le cas $s = 2$ fait également partie des exposants qui concentrent les liens autour des noeuds. Nous distinguons donc deux cas : $s < 2$ et $s \geqslant 2$.

• *Supposons $s < 2$.* Nous montrons que le facteur de normalisation de la distribution des liens d'un sommet donné est supérieur à un polynôme. Pour tout noeud $\mathbf{u} \in G$, le facteur de normalisation $Z_{\mathbf{u}}$ est :

$$Z_{\mathbf{u}} = \sum_{\mathbf{v} \neq \mathbf{u}} \frac{1}{|\mathbf{u} - \mathbf{v}|^s} \geqslant \sum_{k=1}^{n/2} \frac{k}{k^s} \geqslant c_0 \cdot n^{2-s},$$

où c_0 est une constante strictement positive indépendante de n.

Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible du routage respectivement. Soit $\gamma = (2-s)/3 < 1$, nous montrons que tout algorithme décentralisé n'utilise que des liens locaux, avec probabilité constante, pour atteindre la cible en partant d'une distance supérieure à n^γ de la cible.

On note $\mathcal{B}(\mathbf{t}, n^\gamma)$ la boule de centre $\mathbf{t}$ et de rayon n^γ. La probabilité qu'un noeud $\mathbf{u}$ à l'extérieur de $\mathcal{B}(\mathbf{t}, n^\gamma)$ ait un contact longue-distance dans $\mathcal{B}(\mathbf{t}, n^\gamma)$ est inférieure à :

$$\frac{|\mathcal{B}(\mathbf{t}, n^\gamma)|}{c_0 \cdot n^{2-s}} \leqslant c_1 \frac{n^{2\gamma}}{n^{2-s}},$$

où c_1 est une constante strictement positive indépendante de n. Par la borne de l'union, on obtient que la probabilité qu'un noeud sur le chemin emprunté par l'algorithme glouton à partir d'une distance, en partant d'une distance supérieure à n^γ, ait un contact longue-distance dans $\mathcal{B}_{\mathbf{t}}(n^\gamma)$, est inférieure à :

$$c_1 \frac{n^{3\gamma}}{n^{2-s}} \leqslant c_1.$$

La probabilité que la distance $|\mathbf{s}-\mathbf{t}|$ soit strictement supérieure à n^γ, lorsqu'on choisit $\mathbf{s}$ et $\mathbf{t}$ aléatoirement uniformément, est supérieure à $1-n^{2\gamma}/n^{\alpha+1} \geqslant c_2$, où c_2 est une constante strictement positive indépendante de n.

La probabilité que la distance $|\mathbf{s}-\mathbf{t}|$ soit supérieure à n^γ *et* que l'algorithme ne visite pas de contact longue-distance dans $\mathcal{B}(\mathbf{t}, n^\gamma)$ durant n^γ pas, est donc supérieure à $c_1 \cdot c_2$ et l'espérance du nombre de pas de l'algorithme conditionnée par cet événement est supérieure à n^γ. L'espérance totale est donc minorée par $\Omega(n^\gamma)$.

• *Supposons maintenant que* $s \geqslant 2$. Nous allons montrer que le graphe ne possède pas suffisamment de liens longue-distance suffisamment longs, *i.e.* de longueur polynomiale. On remarquera que le nombre de points à distance r d'un noeud $\mathbf{u}$ est inférieur à $C_0 \cdot n$ dès que $r \geqslant n$, pour une constante $C_0 > 0$ indépendante de n. Comme $\alpha > 1$, il existe une constante $q > 1$ telle que $n^\alpha/q > n$. Pour tout noeud $\mathbf{u}$, on a donc

$$b_{\mathbf{u}}(n^\alpha/q) \leqslant n^2 + \sum_{r=n}^{n^\alpha/q} C_0 n = n^2 + C_0 n(n^\alpha/q - n).$$

Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible du routage respectivement, choisies aléatoirement et uniformément. La probabilité que la distance $|\mathbf{s}-\mathbf{t}|$ soit supérieure à n^α/q est donc supérieure à :

$$1 - \frac{n^2 + C_0 n(n^\alpha/q - n)}{n^{\alpha+1}} \geqslant C_1,$$

où C_1 est une constante strictement positive et indépendante de n. On considère deux cas : $s > \alpha+1$ et $2 \leqslant s \leqslant \alpha+1$.

–*Supposons* $s > \alpha+1$. Étant donné un noeud $\mathbf{u}$, la probabilité que son contact longue-distance soit de longueur supérieure ou égale à n est inférieure ou égale à :

$$\sum_{r=n}^{\infty} \frac{C_0 \cdot n}{r^s} \leqslant \frac{C_2}{n^{s-2}},$$

où $C_2 > 0$ est une constante indépendante de n. La probabilité de visiter un noeud dont le lien longue-distance est de longueur supérieure ou égale à n durant moins de $n^{\alpha-1}/C_2$ pas est donc inférieure à $1/n^{s-2-\alpha+1} \leqslant C_3$ pour une constante $C_3 > 0$ indépendante de n, puisque $s > \alpha+1$. Ainsi, avec probabilité supérieure à une constante, l'algorithme se rapproche de $\mathbf{t}$ d'une distance inférieure à $n^{\alpha-1}/C_2 \cdot n = n^\alpha/C_2$ durant $n^{\alpha-1}/C_2$ pas. Comme $|\mathbf{s}-\mathbf{t}| > 2n^\alpha/C_2$ avec probabilité constante, l'espérance totale du nombre de pas est donc minorée par $\Omega(n^{\alpha-1})$.

–*Supposons* $2 \leqslant s \leqslant \alpha+1$. Posons $\gamma = (\alpha+1)/s$. La probabilité que le contact longue-distance d'un noeud donné soit de longueur supérieure ou égale à

n^γ est inférieure ou égale à :

$$\sum_{r=n^\gamma}^{\infty} \frac{C_0 \cdot n}{r^s} \leqslant \frac{C_2}{n^{\gamma(s-1)-1}}.$$

Posons $\beta = \gamma(s-1) - 1$. La probabilité de visiter un noeud dont le lien longue-distance est de longueur supérieure ou égale à n^γ durant moins de $n^\beta/2C_2$ pas est donc inférieure à $1/(2C_2)$. Ainsi, avec probabilité supérieure à une constante, l'algorithme se rapproche de $\mathbf{t}$ d'une distance inférieure à $n^\gamma \cdot n^\beta/2C_2 = n^\alpha/(2C_2)$ durant $n^\beta/2C_2$ pas. Comme $|\mathbf{s} - \mathbf{t}| > n^\alpha/C_2$ avec probabilité constante, l'espérance totale du nombre de pas est donc minorée par $\Omega(n^\beta)$. $\square$

Le déséquilibre des longueurs des côtés de la grille allongée la fait apparaître comme une grille 2-dimensionnelle lorsque qu'on observe un ensemble de noeuds sur un rayon inférieur à n, alors qu'elle se rapproche davantage d'un anneau lorsqu'on observe un ensemble de noeuds à distance $\Omega(n^\alpha)$ les uns des autres. On peut dire qu'elle a deux dimensions *apparentes* différentes (1 et 2) selon la granularité à laquelle on l'observe. C'est ce saut de dimension qui met en défaut les distributions s-harmoniques.

L'exemple de la grille allongée suggère que le paramètre de la croissance des sphères et des boules dans le graphe est plus pertinent que la distance sous-jacente pour construire une distribution de liens longue-distance qui rend le graphe petit monde navigable. Dans le chapitre suivant, nous définirons un processus d'augmentation de graphe en graphe petit monde navigable par l'ajout d'un nombre constant de liens longue-distance aléatoires par noeud, en exprimant la distribution des liens comme une fonction de la taille des boules dans la métrique du graphe que l'on augmente.

4.1.3 Extension aux grilles infinies

Dans cette section, nous montrons que le modèle original de Kleinberg peut s'étendre aux grilles infinies en modifiant légèrement la distribution des liens longue-distance pour s'assurer que le facteur de normalisation reste borné.

Définition 4.1 (Réseau de Kleinberg infini)
Etant donnés $k, d \geqslant 1$ et $0 < \varepsilon < 1$, le réseau de Kleinberg infini, *noté $\mathcal{K}_{k,\infty}^{d,\varepsilon}$, est la grille régulière infinie $\mathbb{Z}^d$, augmentée de k liens longue-distance aléatoires sortants par noeud, tel que pour tout noeud $\mathbf{u}$, la probabilité que son j-ème contact longue-distance soit le noeud $\mathbf{v} \neq \mathbf{u}$ est proportionnelle à :*

$$\frac{1}{|\mathbf{u} - \mathbf{v}|^d \log^{1+\varepsilon}(1 + |\mathbf{u} - \mathbf{v}|)}.$$

Proposition 4.2 *L'algorithme de routage glouton $\mathcal{G}$ calcule, entre tout couple de noeuds $(\mathbf{s}, \mathbf{t})$ de $\mathcal{K}_{k,\infty}^{d,\varepsilon}$, un chemin dont l'espérance de la longueur est :*

$$O\left(\frac{1}{\varepsilon} \cdot \frac{\log^{2+\varepsilon} |\mathbf{s} - \mathbf{t}|}{k}\right).$$

Remarque : seule la dépendance significative en ε est exprimée dans l'ordre de grandeur, les dépendances précises en d et ε sont données dans la preuve.

Preuve. Notons Z_d^∞ le facteur de normalisation de la loi. Pour tout noeud $\mathbf{u}$, on a :

$$Z_d^\infty = \sum_{\mathbf{v} \neq \mathbf{u}} \frac{1}{|\mathbf{u} - \mathbf{v}|^d \log^{1+\varepsilon}(1 + |\mathbf{u} - \mathbf{v}|)}.$$

Dans $\mathbb{Z}^d$, pour tout $r \geqslant 1$, le nombre de points à distance r de $\mathbf{u}$ est :

$$|\mathcal{B}_{\mathbf{u}}(r)| = \frac{2^d}{(d-1)!} r^{d-1} + \xi(r),$$

où $\xi(r)$ est un polynôme positif de degré $d - 2$. On obtient :

$$Z_d^\infty \leqslant \sum_{r \geqslant 2} \left(\frac{2^d}{(d-1)!} \frac{1}{r \log^{1+\varepsilon} r} + \frac{\xi(r)}{r^d \log^{1+\varepsilon} r} \right).$$

Comme $\xi(r)/(r^d \log^{1+\varepsilon} r) = o(1/r^2)$ et $1/(r \log^{1+\varepsilon} r)$ sont des termes généraux de séries convergente, $Z_d^\infty < \infty$. De plus :

$$Z_d^\infty \leqslant \frac{2^d}{(d-1)!} \frac{1}{\varepsilon}.$$

Nous pouvons à présent suivre les lignes de l'analyse de l'algorithme glouton sur le réseau de Kleinberg fini pour l'analyser sur le réseau de Kleinberg infini. Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible du routage, respectivement. On divise l'exécution de l'algorithme en $\log |\mathbf{s} - \mathbf{t}|$ phases. Pour tout i tel que $1 \leqslant i \leqslant \log |\mathbf{s} - \mathbf{t}|$, la phase i dure tant que la distance courante à la cible appartient à $(2^{i-1}, 2^i]$.

Supposons que l'algorithme est en phase i. Soit $\mathbf{x}$ le porteur courant du message. Soit $\mathcal{E}$ l'événement :« $L(\mathbf{x})$ appartient à $\mathcal{B}_{\mathbf{t}}(2^{i-1})$ ». Comme la distribution des liens longue-distance est décroissante avec la distance, $\Pr \mathcal{E}$ est minorée par la probabilité que le contact longue-distance $L(\mathbf{y})$ d'un noeud $\mathbf{y} \in \mathcal{S}_{\mathbf{t}}(2^i)$ donné, appartienne à $\mathcal{B}_{\mathbf{t}}(2^{i-1})$. Le lemme 2.11, démontré au chapitre 2, affirme que, dans un réseau de Kleinberg fini de dimension d, la probabilité qu'un noeud $\mathbf{y}$, à distance D d'un noeud $\mathbf{t}$, ait son j-ème contact longue-distance dans $\mathcal{B}_{\mathbf{t}}(D/2)$, est minimale lorsque $\mathbf{y}$ est un coin de la sphère. Nous pouvons appliquer ce lemme ici avec

$D = 2^i$ en nous restreignant au réseau fini induit par les noeuds de $\mathcal{B}_{\mathbf{t}}(2^i)$. Supposons que $i \geqslant 2$. Il reste à évaluer la probabilité que $L(\mathbf{y})$ appartienne à $\mathcal{B}_{\mathbf{t}}(2^{i-1})$ sachant que $\mathbf{y}$ est un coin de la sphère $\mathcal{S}_{\mathbf{t}}(2^i)$, pour obtenir une minoration de $\Pr \mathcal{E}$. Comme nous nous sommes restreints à un réseau de Kleinberg fini, nous pouvons également utiliser le lemme 2.12 démontré au chapitre 2. Il affirme que le nombre de noeuds de $\mathcal{B}_{\mathbf{t}}(2^{i-1})$ à distance r d'un coin de $\mathcal{B}_{\mathbf{t}}(2^{i-1})$ est supérieur ou égal à $r^{d-1}/(d-1)!$. Par conséquent, le nombre de noeuds de $\mathcal{B}_{\mathbf{t}}(2^{i-1})$, à distance $r + 2^{i-1}$ du coin $\mathbf{y}$ de $\mathcal{B}_{\mathbf{t}}(2^i)$, est supérieur ou égal à $r^{d-1}/(d-1)!$. En sommant sur ces noeuds, on obtient que la probabilité que $L(\mathbf{y})$ appartienne à $\mathcal{B}_{\mathbf{t}}(2^{i-1})$ est :

$$\begin{aligned}
&\frac{1}{Z_d^\infty} \sum_{\mathbf{u} \in \mathcal{B}_{\mathbf{t}}(2^{i-1})} \frac{1}{|\mathbf{u}-\mathbf{y}|^d \log^{1+\varepsilon}(|\mathbf{u}-\mathbf{y}|+1)} \\
&\geqslant \frac{1}{Z_d^\infty} \sum_{r=1}^{2^i} \frac{r^{d-1}/(d-1)!}{(r+2^{i-1})^d \log^{1+\varepsilon}(r+2^{i-1})}.
\end{aligned}$$

Comme il s'agit d'une minoration, on peut choisir de ne considérer, dans la somme, que les indices r supérieurs à 2^{i-1}. On obtient alors $r + 2^{i-1} \leqslant 2r$ et l'on remarquera que $\log^{1+\varepsilon}(2r) \leqslant (2\log r)^{1+\varepsilon}$. Nous obtenons le minorant suivant :

$$\frac{1}{2^{d+1+\varepsilon}(d-1)!Z_d^\infty} \sum_{r=2^{i-1}}^{2^i} \frac{1}{r \log^{1+\varepsilon} r}.$$

Par ailleurs, on a :

$$\begin{aligned}
\sum_{r=2^{i-1}}^{2^i} \frac{1}{r\log^{1+\varepsilon} r} &\geqslant \int_{2^{i-1}}^{2^i+1} \frac{dr}{r\log^{1+\varepsilon} r} \\
&= \frac{1}{\varepsilon}\left(\frac{1}{\log^\varepsilon 2^{i-1}} - \frac{1}{\log^\varepsilon(2^i+1)}\right) = \frac{1}{\varepsilon}\left(\frac{1}{(i-1)^\varepsilon} - \frac{1}{i^\varepsilon}\right) \\
&= \frac{1}{\varepsilon}\frac{1}{i^\varepsilon}\left(\frac{1}{(1-1/i)^\varepsilon} - 1\right) \geqslant \frac{1}{i^{\varepsilon+1}}.
\end{aligned}$$

Ainsi l'espérance du nombre de pas exécutés par l'algorithme avant de trouver un contact à distance inférieure à 2^{i-1} de $\mathbf{t}$, depuis $\mathbf{x}$, est inférieure ou égal à :

$$2^{d+\varepsilon+1}(d-1)!Z_d^\infty \cdot \frac{i^{\varepsilon+1}}{k},$$

puisqu'il y a k liens longue-distance indépendants par noeud. On en déduit la majoration suivante de l'espérance de la longueur du chemin de routage de $\mathbf{s}$ à $\mathbf{t}$:

$$2 + 2^{d+\varepsilon+1}(d-1)!Z_d^\infty \cdot \frac{1}{k} \sum_{i=2}^{\log|\mathbf{s}-\mathbf{t}|} i^{\varepsilon+1} = O\left(\frac{1}{\varepsilon} \cdot \frac{\log^{2+\epsilon}|\mathbf{s}-\mathbf{t}|}{k}\right).$$

□

4.2 Métriques de graphes et petits mondes

Dans cette section, nous présentons le cadre formel des graphes sur lesquels nous allons construire des processus de *petit-mondisation* dans les chapitres suivantes.

Dans toute la suite, on ne considérera que des graphes infinis de degré maximum Δ, une constante fixée, ou des graphes finis de degré moyen Δ', une constante fixée. Notre processus de *petit-mondisation* se base sur l'ajout d'arcs aléatoires à une structure sous-jacente connue. Cela motive les définitions suivantes.

La définition suivante rappelle que l'on peut toujours associer la métrique des plus courts chemins à un graphe. Par abus de langage, on parlera simplement de *métrique* pour désigner un espace métrique.

Définition 4.2 (Métrique d'un graphe)
Étant donné un graphe $H = (V, E)$, on appelle métrique du graphe *H l'ensemble V muni de la distance distance δ_H, où pour tout $(\mathbf{u}, \mathbf{v}) \in V \times V$, $\delta_H(\mathbf{u}, \mathbf{v})$ est la distance de $\mathbf{u}$ à $\mathbf{v}$ dans H.*

Définition 4.3 (Graphe augmenté aléatoirement)
On appelle graphe augmenté aléatoirement *un graphe $G = (V, E, E')$, tel que l'ensemble d'arcs E' a été choisi aléatoirement selon une distribution qui ne dépend que de δ_H, la distance dans H, où $H = (V, E)$ est un graphe connexe appelé* graphe de base.
(V, δ_H) est appelée métrique sous-jacente *de G.*

On remarquera en particulier que le modèle de Kleinberg $\mathcal{K}_n^d$ est un graphe augmenté aléatoirement, son graphe de base étant le tore d-dimensionnel et l'ensemble E' étant l'ensemble des liens longue-distance. Le graphe de base d'un graphe aléatoire augmenté représente, dans notre modèle, la connaissance globale possédée par tous les noeuds. Un algorithme de routage décentralisé a donc la connaissance globale du graphe de base, mais n'a la connaissance d'un arc aléatoire qu'en visitant son origine (voir p.19 la définition du routage décentralisé).

Dans la suite, nous préciserons le graphe concerné dans les notations des boules et sphères en cas d'ambiguïté. On notera $\mathcal{B}_{H,\mathbf{u}}(r)$ et $\mathcal{S}_{H,\mathbf{u}}(r)$ respectivement, la boule et la sphère de rayon r et de centre $\mathbf{u}$ dans le graphe $H = (V, E)$, et $b_{H,\mathbf{u}}(r)$ et $s_{H,\mathbf{u}}(r)$ leur cardinal.

Considérons un graphe aléatoirement augmenté $G = (V, E, E')$ de graphe de base $H = (V, E)$. Dans un graphe augmenté de degré borné, la distance d'un

noeud $\mathbf{v}$ dans $\mathcal{B}_{H,\mathbf{u}}(r)$ à $\mathbf{u}$ est au moins $\Omega(\log b_{H,\mathbf{u}}(r))$ dans le graphe augmenté G. Cela motive la définition suivante qui étend la définition de petit monde aux graphes infinis aléatoirement augmentés.

Définition 4.4 (Petit-monde navigable)
Un graphe infini aléatoirement augmenté G, ayant H pour graphe de base, est un petit monde navigable *s'il existe un algorithme décentralisé utilisant la métrique sous-jacente δ_H qui, étant donné deux noeuds $\mathbf{u}$ et $\mathbf{v}$, calcule un chemin de $\mathbf{u}$ à $\mathbf{v}$ dans G en visitant un nombre de noeuds polylogarithmique en $b_{H,\mathbf{v}}(\delta_H(\mathbf{u},\mathbf{v}))$, en espérance.*

On peut adapter la définition aux familles infinies de graphes finis.

Définition 4.5 (Famille petit monde navigable)
Une famille de graphes finis aléatoirement augmentés $(G_i)_{i\in\mathcal{I}}$, où l'on note H_i le graphe de base de G_i, est une famille petit monde navigable *s'il existe un polynôme uniforme p et un algorithme décentralisé utilisant la métrique sous-jacente δ_{H_i} sur G_i qui, pour chaque $i \in \mathcal{I}$ et toute paire de noeuds $\mathbf{u}$ et $\mathbf{v}$ dans G_i, calcule un chemin de $\mathbf{u}$ à $\mathbf{v}$, en visitant un nombre de noeuds dont l'espérance est inférieure à $p(\log b_{H_i,\mathbf{v}}(\delta_{H_i}(\mathbf{u},\mathbf{v})))$.*

On remarquera que certains graphes sont intrinsèquement des petits mondes navigables et ne nécessitent pas d'augmentation. Par exemple, tout graphe $H = (V, E)$ dont $b_{\mathbf{v}}(r)$ est minoré uniformément par une fonction $b(r)$ exponentielle en r pour tout $\mathbf{v}$ est un petit monde navigable. En effet, l'algorithme glouton qui utilise la métrique (V, δ_H) de ce graphe, calcule un chemin de longueur $r = \delta_H(\mathbf{u},\mathbf{v})$ entre deux noeuds à distance r et $r = \text{polylog}(b(r))$ dans ce cas. L'exemple le plus simple de cette situation est un arbre k-aire infini (avec $k \geqslant 3$), où les boules de rayon r sont de taille $\Theta((k-1)^r)$. Au contraire, les graphes dont les boules croissent polynomialement (*i.e.* tels que $b(r) = O(r^c)$ pour tout r, pour une constante c) ne sont pas intrinsèquement des petits mondes navigables puisque la longueur du chemin optimal calculé par l'algorithme glouton entre deux noeuds à distance r est $r \gg \text{polylog}(b(r)) = \text{polylog}(r)$.

4.2.1 Graphes *petit-mondisables*

Nous introduisons ici les néologismes *petit-mondisation*, *petit-mondiser* et *petit-mondisable*, dont la sonorité n'est pas heureuse, mais qui simplifient la lecture. Une *petit-mondisation* est un processus qui augmente une famille de graphe en petits mondes navigables via l'ajout de liens aléatoires en chaque noeud, suivant une certaine distribution (cf. définition 4.6). Dans la plupart des graphes

réels, on observe un nombre d'arêtes linéaire en le nombre de sommets, c'est par exemple le cas pour le graphe d'Internet. Une construction basée sur l'ajout d'un nombre non constant d'arcs en moyenne par noeud n'est donc pas réaliste en terme de modélisation des graphes réels. Les *petit-mondisations* que nous présentons dans le chapitre suivant ajoutent exactement *un* lien sortant par noeud. Par ailleurs, l'ajout d'un nombre constant $k > 1$ de liens sortants change peu les résultats, comme les preuves le feront clairement apparaître.

La destination $L(\mathbf{u})$ du lien longue-distance sortant de $\mathbf{u}$ est choisie aléatoirement selon une distribution qui donne un poids égal à tous les noeuds équidistants de $\mathbf{u}$ dans le graphe de base. Notons δ la métrique sous-jacente, pour chaque noeud $\mathbf{u}$, il existe ainsi une fonction $f_{\mathbf{u}}$ telle que chaque noeud $\mathbf{v} \neq \mathbf{u}$ a une probabilité proportionnelle à $f_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v}))$ d'être $L(\mathbf{u})$. Le facteur de normalisation $Z_{\mathbf{u}}$ est donné par :

$$Z_{\mathbf{u}} = \sum_{\mathbf{v} \neq \mathbf{u}} f_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v})) = \sum_{r>0} s_{\mathbf{u}}(r) \cdot f_{\mathbf{u}}(r) = \sum_{r>0} (b_{\mathbf{u}}(r) - b_{\mathbf{u}}(r-1))\, f_{\mathbf{u}}(r).$$

Définition 4.6 (Graphe infini *petit-mondisable*)
Soit H un graphe infini. Le graphe augmenté de H selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u} \in H}$ est le graphe H augmenté en ajoutant un lien longue-distance aléatoire sortant à chaque noeud $\mathbf{u}$ de H selon $f_{\mathbf{u}}$, i.e. la destination du lien est $\mathbf{v}$ avec une probabilité proportionnelle à $f_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v}))$.

On dit qu'un graphe infini H est petit-mondisable *selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u} \in H}$ si le graphe augmenté de H selon la famille $(f_{\mathbf{u}})_{\mathbf{u} \in H}$ est un petit monde navigable.*

On peut étendre la définition aux familles infinies de graphes finis.

Définition 4.7 (Famille infinie de graphes finis *petit-mondisable*)
Soit $(H_i)_{i \in \mathcal{I}}$ une famille infinie de graphes finis. La famille augmentée de $(H_i)_{i \in \mathcal{I}}$ selon la famille de distributions $(f_{i,\mathbf{u}})_{\mathbf{u} \in H_i, i \in I}$ est la famille $(G_i)_{i \in \mathcal{I}}$ où G_i est obtenu en ajoutant un lien longue-distance aléatoire sortant à chaque noeud $\mathbf{u}$ de H_i selon $f_{i,\mathbf{u}}(r)$, i.e. la destination est $\mathbf{v}$ avec une probabilité proportionnelle à $f_{i,\mathbf{u}}(\delta_{H_i}(\mathbf{u}, \mathbf{v}))$.

On dit qu'une famille infinie de graphes finis $(H_i)_{i \in \mathcal{I}}$ est petit-mondisable *selon la famille de distributions $(f_{i,\mathbf{u}})_{\mathbf{u} \in H_i, i \in I}$ si la famille augmentée de $(Hi)_{i \in \mathcal{I}}$ selon la famille $(f_{i,\mathbf{u}})_{\mathbf{u} \in H_i, i \in I}$ est une famille petit monde navigable.*

Par abus de langage, nous dirons que l'on *petit-mondise* un graphe lorsqu'on l'augmente aléatoirement pour obtenir un petit monde navigable (*i.e.* lorsqu'on lui applique une *petit-mondisation*).

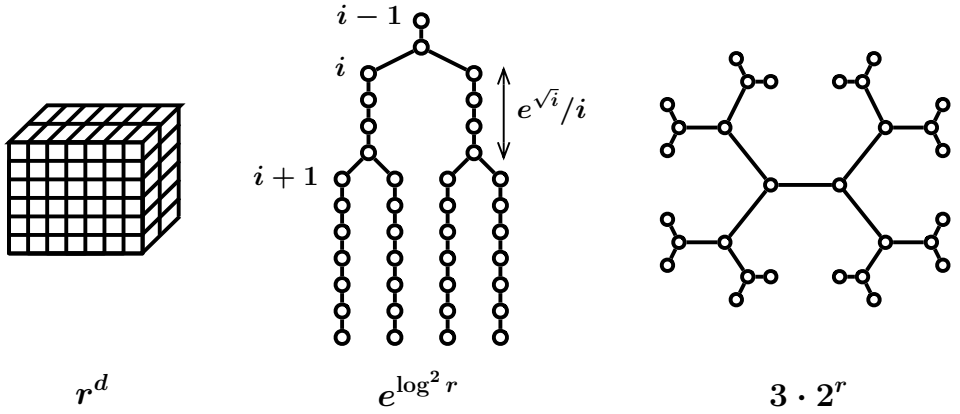


FIG. 4.2 : Illustration de trois types de graphes et de l'ordre de grandeur de la taille d'une boule de rayon r pour chacun.

Quels sont les graphes *petit-mondisables* ? Nous savons que les grilles carrées d-dimensionnelles $n \times \cdots \times n$ sont *petit-mondisables*. La croissance régulières de leurs boules, en $\Theta(r^d)$, permet à la distribution d-harmonique de bien répartir des raccourcis sur toutes les échelles de distance. Nous savons également que les arbres k-aires complets sont des petits mondes navigables, sans nécessiter d'augmentation, grâce à une croissance de boule exponentielle $\Theta\left((k-1)^r\right)$. Mais l'on peut également construire des graphes moins réguliers, avec des tailles de boules très différentes. La figure 4.2 illustre par exemple un graphe dont la croissance de boules est intermédiaire entre celle des grilles régulières et celles des arbres réguliers. Est-il possible de *petit-mondiser* tous les graphes ? Pour les métriques de graphes générales, nous ne disposons pas d'information sur la densité d'un ensemble de noeuds à une distance donnée, ce qui rend difficile la définition d'une distribution aléatoire de raccourcis bien répartis. Dans le chapitre suivant, nous nous abordons ce problème et nous montrons qu'il existe deux grandes classes de graphes qui peuvent être *petit-mondisées* : les graphes *à croissance modérée* et les graphes admettant un plongement de faible distorsion dans $(\mathbb{R}^d, \ell_1)$, où la dimension d est bornée.

Chapitre 5

Processus de *petit-mondisation*

Introduction

Dans ce chapitre, nous proposons deux nouveaux modèles de petits mondes navigables qui généralisent le modèle original de Kleinberg. Le premier modèle, décrit dans la section 5.2, est basé sur l'augmentation d'une classe de graphe dont la métrique des distance respecte certaines contraintes de croissance. On augmente les graphes par un lien longue-distance par noeud selon une distribution dépendant de la taille des boules de la métrique. Le second modèle, décrit dans la section 5.6, utilise une méthode différente pour le calcul de la distribution, en utilisant un plongement de faible distorsion de la métrique sous-jacente dans un espace euclidien de faible dimension. La *petit-mondisation* est alors garantie pour la classe des graphes dont la métrique sous-jacente admet un tel plongement.

5.1 Graphes à croissance modérée

Dans cette section, nous définissons une grande classe de graphes infinis (et de familles infinies de graphes finis) pour laquelle nous obtenons des modèles d'augmentation qui résultent en des petits mondes navigables, *i.e.* des *petit-mondisations*.

Définition 5.1 (Graphe à croissance β-modérée)
Étant donné $\beta \geqslant 0$, un graphe H, fini ou infini, de degré borné Δ, est un graphe à croissance β-modérée *s'il existe une constante $c_0 > 0$, telle que, pour tout noeud $\mathbf{u}$ de H et tout $r > 0$, on a :*

$$b_{\mathbf{u}}(2r) \leqslant c_0 (\log r)^{\beta}\, b_{\mathbf{u}}(r).$$

On peut étendre naturellement cette définition aux familles infinies de graphes finis.

Définition 5.2 (Famille infinie de graphes finis à croissance β-modérée)
Étant donné $\beta \geqslant 0$, une famille infinie de graphes finis $(H_i)_{i\in\mathcal{I}}$, de degré moyen uniformément borné Δ, est une famille à croissance β-modérée *s'il existe une constante uniforme $c_0 > 0$, telle que pour tout i, tout noeud $\mathbf{u}$ de H_i et tout $r > 0$, on a :*

$$b_{H_i,\mathbf{u}}(2r) \leqslant c_0(\log r)^\beta\, b_{H_i,\mathbf{u}}(r).$$

Cette classe de graphe contient la classe des graphes à *croissance bornée*[1] [KR02], qui sont les graphes à croissance 0-modérée (*i.e.* toute boule est inférieure à c fois la taille de la boule de même centre et de rayon moitié, pour une constante $c > 0$ uniforme).

Dimension intrinsèque. Il existe plusieurs dénominations de la classe des graphes à croissance bornée dans la littérature : on parle également de graphes à *croissance restreinte*[2] [KR02], ou encore de graphes de *dimension de grille bornée*[3] [GKL03]. Ces dénominations font référence à la faible « dimension » associée à cette classe. Précisément, Krauthgamer et Lee [KL03] ont récemment démontré que la croissance des boules caractérise la *dimension intrinsèque* d'un graphe, au sens que le taux de croissance ρ, tel que toute boule de rayon r contient $O(r^\rho)$ noeuds, est équivalent à la dimension d minimale pour laquelle le graphe est un sous-graphe de $\mathbb{Z}^d_\infty$. Pour bien comprendre à quoi correspond notre définition de graphe à croissance β-modérée, on peut traduire sa contrainte en une contrainte sur la dimension intrinsèque du graphe. On définit donc la *dimension intrinsèque* au noeud $\mathbf{u}$ par la fonction $d_\mathbf{u}$ telle que, pour tout $r > 0$, $b_\mathbf{u}(r) = r^{d_\mathbf{u}(r)}$, *i.e.* $d_\mathbf{u}(r) = \log_r b_\mathbf{u}(r)$. Pour un graphe à croissance β-modérée, on a, pour tout $r > 0$,

$$b_\mathbf{u}(r) \leqslant c_0 \log^\beta(r/2) b_\mathbf{u}(r/4) \leqslant \ldots \leqslant c_0^{\log r} \log^{\beta \log r}(r/2) b_\mathbf{u}(1).$$

Soit encore :

$$b_\mathbf{u}(r) \leqslant r^{c_0 + \beta \log\log(r/2)}.$$

On en déduit :

$$d_\mathbf{u}(r) \leqslant c_0 + \beta \log\log(r/2).$$

La dimension intrinsèque d'un graphe à croissance β-modérée ne peut donc pas augmenter de façon trop brutale. Nous obtenons également une borne $O(r^{O(\beta \log\log r)})$ sur la taille de toute boule de rayon r.

[1] *Bounded growth graphs* en anglais.
[2] *Restricted growth graphs* en anglais.
[3] *Bounded grid dimension* en anglais.

5.2 *Petit-mondisation* des graphes à croissance modérée

le théorème suivant, qui est le résultat principal de ce chapitre, donne un processus pour *petit-mondiser* les graphes à croissance modérée. Nous distinguons trois cas : les graphes infinis, les familles de graphes infinis et les graphes finis.

Théorème 5.1 (*Petit-mondisation* des graphes à croissance β-modérée)

(i) Pour tout $0 < \varepsilon < 1$, tout graphe infini H à croissance β-modérée, $\beta \geqslant 0$, est petit-mondisable *selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u}\in H}$ définie par :*

$$f_{\mathbf{u}}(r) = \frac{1}{b_{\mathbf{u}}(r)\log^{1+\beta+\varepsilon}(r+1)} \quad \textit{pour tout } r > 0.$$

L'algorithme glouton calcule, pour tout couple de noeuds $(\mathbf{s},\mathbf{t})$ à distance m, un chemin de $\mathbf{s}$ à $\mathbf{t}$ dans le graphe augmenté, de longueur $O\big(c'_\varepsilon \cdot (\log m)^{2+4\beta+\varepsilon}\big)$ en espérance, où $c'_\varepsilon = 2^{4\beta}/(\varepsilon(2+4\beta))$.

(ii) Pour tout $0 < \varepsilon < 1$, toute famille infinie de graphes finis $(H_i)_{i\in\mathcal{I}}$ à croissance β-modérée, $\beta \geqslant 0$, est petit-mondisable *selon la famille de distributions $(f_{i,\mathbf{u}})_{\mathbf{u}\in H_i,\, i\in I}$ définie par :*

$$f_{i,\mathbf{u}}(r) = \frac{1}{b_{H_i,\mathbf{u}}(r)\log^{1+\beta+\varepsilon}(r+1)} \quad \textit{pour tout } r > 0.$$

L'algorithme glouton calcule, pour tout couple de noeuds $(\mathbf{s},\mathbf{t})$ à distance m, un chemin de $\mathbf{s}$ à $\mathbf{t}$ dans la famille augmentée, de longueur $O\big(c'_\varepsilon \cdot (\log m)^{2+4\beta+\varepsilon}\big)$ en espérance, où $c'_\varepsilon = 2^{4\beta}/(\varepsilon(2+4\beta))$.

(iii) Tout graphe fini H, de n noeuds, à croissance β-modérée, $\beta \geqslant 0$, augmenté selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u}\in H}$ définie par :

$$f_{\mathbf{u}}(r) = \frac{1}{b_{H,\mathbf{u}}(r)} \quad \textit{pour tout } r > 0,$$

est un graphe où l'algorithme glouton calcule, entre toute paire de noeuds à distance m, un chemin dans le graphe augmenté de longueur $O\left((2^{3\beta}/\beta^2)\,(\log n)^{\beta+1}\,(\log m)^{3\beta+1}\right)$ en espérance.

D'autres résultats montrant la possibilité de *petit-mondiser* d'autres classes de graphes ont été démontrés simultanément [Sli05, Fra05], on trouvera une discussion comparative de ces résultats dans la section 5.7.

La partie (iii) du théorème n'est pas une *petit-mondisation* au sens de la définition 4.4 puisque la longueur d'un chemin de $\mathbf{s}$ à $\mathbf{t}$ ne dépend pas uniquement de la taille de la boule $b_{\mathbf{s}}(\delta(\mathbf{s},\mathbf{t}))$ mais aussi de la taille n du graphe. Toutefois, ces graphes finis sont *petit-mondisables* au sens de la définition via le cas (ii), nous ajoutons donc cette augmentation par une autre distribution dans le seul but de donner une alternative selon que le choix le plus efficace est d'exprimer la longueur en fonction de n (lorsque la contante c'_ε est grande) ou seulement en fonction de $b_{\mathbf{s}}(\delta(\mathbf{s},\mathbf{t}))$ (lorsque que la plupart des chemins à calculer sont entre des paires proches).

Pour prouver le théorème 5.1, nous montrons tout d'abord que le facteur de normalisation de la distribution est borné, puis nous analysons l'algorithme de routage glouton sur les graphes augmentés.

On remarquera que, pour prouver que la famille $(H_i)_{i\in\mathcal{I}}$ est *petit-mondisable*, il suffit d'obtenir des bornes uniformes, *i.e.* indépendantes de i, sur l'espérance de la longueur du chemin calculée par l'algorithme glouton, valables pour tout graphe fini H_i. Par conséquent, les preuves sont analogues pour les cas (i) et (ii) des graphes finis et des familles infinies de graphes finis respectivement. Il suffit donc de démontrer le cas (i) avec des bornes sur la longueur des chemins indépendantes du graphe pour démontrer le cas (ii).

Preuve du théorème 5.1 (i). Soit H un graphe infini, de degré borné, à croissance β-modérée, de métrique δ_H. Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible du routage respectivement. On considère l'algorithme de routage glouton $\mathcal{G}$ dans le graphe augmenté aléatoirement par $f_{\mathbf{u}}$, pour chaque noeud $\mathbf{u}$.

L'argument principal de l'analyse de Kleinberg, dont cette preuve est inspirée, est que parmi un nombre polylogarithmique de noeuds à distance comprise entre r et $r/2$ de $\mathbf{t}$, pour $r \geqslant 2$, il en existe un, avec probabilité constante, dont le contact longue-distance est à distance inférieure à $r/2$ de la cible. De ce résultat, on tire que l'espérance de la longueur de chemin est polylogarithmique. Nous utilisons ici un argument similaire, en le modifiant de façon à ce que le majorant puisse être exprimé uniquement en fonction de la métrique originale δ_H (et non de la taille totale du graphe).

Nous devons tout d'abord vérifier que les facteurs de normalisation $Z_{\mathbf{u}} = \sum_{\mathbf{v}\in V} f_{\mathbf{u}}(\delta_H(\mathbf{u},\mathbf{v}))$ sont bornés, pour que la distribution soit bien définie. Comme

il y a $s_{\mathbf{u}}(r)$ noeuds à distance r de $\mathbf{u}$ dans H, on peut écrire :

$$\begin{aligned}
Z_{\mathbf{u}} &= \sum_{r\geqslant 1} s_{\mathbf{u}}(r)\cdot f_{\mathbf{u}}(r) = \sum_{r\geqslant 1} \frac{s_{\mathbf{u}}(r)}{b_{\mathbf{u}}(r)\log^{1+\beta+\varepsilon}(r+1)} \\
&= 1 + \sum_{i\geqslant 0}\sum_{r=2^i+1}^{2^{i+1}} \frac{b_{\mathbf{u}}(r)-b_{\mathbf{u}}(r-1)}{b_{\mathbf{u}}(r)\log^{1+\beta+\varepsilon}(r+1)} \\
&\leqslant 1 + \sum_{i\geqslant 0} \frac{1}{b_{\mathbf{u}}(2^i+1)\log^{1+\beta+\varepsilon}(2^i+2)} \sum_{r=2^i+1}^{2^{i+1}} (b_{\mathbf{u}}(r)-b_{\mathbf{u}}(r-1)) \\
&\leqslant 1 + (1-\frac{1}{b_{\mathbf{u}}(2)}) + \sum_{i\geqslant 1} \frac{b_{\mathbf{u}}(2^{i+1})-b_{\mathbf{u}}(2^i)}{b_{\mathbf{u}}(2^i+1)\cdot i^{1+\beta+\varepsilon}}.
\end{aligned}$$

Or, comme H est à croissance β-modérée, on a :

$$b_{\mathbf{u}}(2^{i+1}) - b_{\mathbf{u}}(2^i) \leqslant (c_0\log^{\beta}(2^i)-1)b_{\mathbf{u}}(2^i) \leqslant c_0 i^{\beta} b_{\mathbf{u}}(2^i+1).$$

On obtient :

$$Z_{\mathbf{u}} \leqslant 1 + \left(1-\frac{1}{b_{\mathbf{u}}(2)}\right) + \sum_{i\geqslant 1}\frac{c_0}{i^{1+\varepsilon}} \leqslant 2 + \frac{c_0}{\varepsilon} =_{\text{def}} Z_{\varepsilon}.$$

Nous analysons à présent l'espérance de la longueur du chemin calculé par l'algorithme glouton. Soit un entier $r \geqslant 2$ et un noeud $\mathbf{u}$ tel que $r/2 < \delta_H(\mathbf{u},\mathbf{t}) \leqslant r$ et notons $L(\mathbf{u})$ son contact longue-distance. Nous donnons une minoration de $\Pr\{\delta_H(L(\mathbf{u}),\mathbf{t}) \leqslant r/2\}$, la probabilité que $L(\mathbf{u})$ appartienne à $\mathcal{B}_{\mathbf{t}}(r/2)$. Comme $f_{\mathbf{u}}$ est une fonction décroissante et $\mathcal{B}_{\mathbf{t}}(r/2) \subseteq \mathcal{B}_{\mathbf{u}}(3r/2)$, chaque noeud de $\mathcal{B}_{\mathbf{t}}(r/2)$ a une probabilité d'être $L(\mathbf{u})$ supérieure à $f_{\mathbf{u}}(3r/2)/Z_{\mathbf{u}}$. Comme par ailleurs, $\mathcal{B}_{\mathbf{u}}(3r/2) \subseteq \mathcal{B}_{\mathbf{t}}(5r/2)$, nous pouvons minorer $f_{\mathbf{u}}(3r/2)$ en fonction de $b_{\mathbf{t}}(5r/2)$ (cf. figure 5.2) :

$$f_{\mathbf{u}}(3r/2) \geqslant \frac{1}{b_{\mathbf{t}}(5r/2)\log^{1+b+\varepsilon}(1+3r/2)}.$$

Cette minoration, combinée à la croissance β-modérée, nous permet d'obtenir une minoration de la probabilité considérée qui ne dépend que de $\mathbf{t}$ et r :

$$\begin{aligned}
\Pr\{\delta_H(L(\mathbf{u}),\mathbf{t}) \leqslant r/2\} &\geqslant \frac{1}{Z_{\varepsilon}\log^{1+\beta+\epsilon}(1+3r/2)}\,\frac{b_{\mathbf{t}}(r/2)}{b_{\mathbf{t}}(5r/2)} \\
&\geqslant \left(c_0^3 Z_{\varepsilon}\log^{1+\beta+\varepsilon}(1+3r/2)\log^{3\beta}(5r/4)\right)^{-1} \\
&\geqslant \left(2^{1+4\beta+\varepsilon}c_0^4 Z_{\varepsilon}\log^{1+4\beta+\varepsilon}(r)\right)^{-1}
\end{aligned}$$

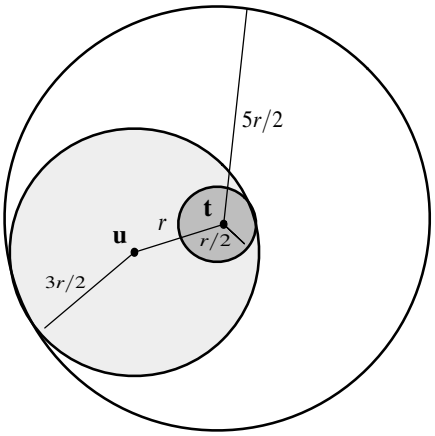


FIG. 5.1 : Illustration de l'inclusion des boules pour la preuve du théorème 5.1.

Nous revenons à présent à la question initiale de la longueur du chemin glouton de $\mathbf{s}$ à $\mathbf{t}$. Comme pour l'analyse du réseau de Kleinberg original, on divise l'exécution de $\mathcal{G}$ en $\log|\mathbf{s}-\mathbf{t}|$ phases successives de division par deux de la distance à la cible. La phase i dure tant que la distance du noeud courant à $\mathbf{t}$, dans δ_H, appartient à $(2^{i-1}, 2^i]$. La discussion précédente prouve que lorsque l'algorithme est en phase i, la probabilité de quitter cette phase pour une phase $j < i$ est supérieure à $\Omega(i^{-\gamma}2^{-\gamma}\varepsilon)$ en chaque noeud, où $\gamma = 1 + 4\beta + \varepsilon$. Ainsi, pour tout $i \in \{1, \ldots, \log|\mathbf{s}-\mathbf{t}|\}$, l'algorithme glouton visite au plus $O(\varepsilon^{-1}i^{\gamma}2^{\gamma})$ noeuds en espérance durant la phase i. Par linéarité de l'espérance, l'espérance de la longueur du chemin calculé par l'algorithme glouton de $\mathbf{s}$ à $\mathbf{t}$ est

$$O\left(\frac{2^{4\beta}}{\varepsilon(2+4\beta)}\left(\log\delta_H(\mathbf{s},\mathbf{t})\right)^{2+4\beta+\varepsilon}\right).$$

Finalement, l'espérance de la longueur du chemin de $\mathbf{s}$ à $\mathbf{t}$ est polylogarithmique en $m = \delta_H(\mathbf{s},\mathbf{t})$ et *a fortiori* en $b_{\mathbf{t}}(\delta_H(\mathbf{s},\mathbf{t}))$. Le graphe augmenté aléatoirement est donc bien un petit monde navigable. □

Preuve du théorème 5.1 (iii). La preuve est essentiellement identique à celle des graphes infinis, à l'exception que l'absence du facteur logarithmique dans la distribution fait intervenir la taille du graphe dans la longueur du chemin.

Le facteur de normalisation de la distribution au noeud $\mathbf{u}$ est :

$$\begin{aligned} Z_{\mathbf{u}} &= \sum_{r=1}^{n} s_{\mathbf{u}}(r) \cdot f_{\mathbf{u}}(r) = \sum_{r=1}^{n} \frac{s_{\mathbf{u}}(r)}{b_{\mathbf{u}}(r)} \\ &= 1 + \sum_{i=0}^{\log n - 1} \sum_{r=2^i+1}^{2^{i+1}} \frac{b_{\mathbf{u}}(r) - b_{\mathbf{u}}(r-1)}{b_{\mathbf{u}}(r)}. \end{aligned}$$

En utilisant le même argument de croissance modérée que dans la preuve du cas (i), on obtient :

$$Z_{\mathbf{u}} \leqslant 1 + \left(1 - \frac{1}{b_{\mathbf{u}}(2)}\right) + \sum_{i=1}^{\log n - 1} c_0\, i^{\beta} \leqslant 2 + \frac{c_0}{\beta+1} \log^{\beta+1} n =_{\text{def}} Z_n.$$

Pour la lisibilité, on remarquera qu'il existe $n_0 > 0$ tel que pour tout n supérieur à n_0, $Z_n \leqslant (2c_0/(\beta+1)) \log^{\beta+1} n$.

On analyse à présent la longueur du chemin calculé par l'algorithme glouton entre une source $\mathbf{s}$ et une cible $\mathbf{t}$, en considérant un entier $r > 1$ et un noeud $\mathbf{u}$ tel que $r/2 < \delta_H(\mathbf{u}, \mathbf{t}) \leqslant r$. Par un argument identique à celui utilisé dans la preuve pour le cas infini, on a, pour tout $n \geqslant n_0$:

$$\begin{aligned} \Pr\{\delta_H(L(\mathbf{u}), \mathbf{t}) \leqslant r/2\} &\geqslant \frac{1}{Z} \frac{b_{H,\mathbf{t}}(r/2)}{b_{H,\mathbf{t}}(5r/2)} \geqslant \frac{\beta+1}{2c_0 \log^{\beta+1} n} \frac{1}{c_0^3 \log^{3\beta}(5r/4)} \\ &\geqslant \frac{\beta+1}{2^{3\beta+1} c_0^4 \log^{\beta+1} n} \frac{1}{\log^{3\beta}(r)}. \end{aligned}$$

En menant une analyse par phases de division par deux de la distance à la cible, identique au cas infini, on obtient une espérance de longueur de chemin inférieure à :

$$\begin{aligned} &\frac{2^{3\beta+1} c_0^4 \log^{\beta+1} n}{\beta+1} \sum_{i=0}^{\log \delta_H(\mathbf{s},\mathbf{t})} i^{3\beta} \leqslant \frac{2^{3\beta+1} c_0^4 \log^{\beta+1} n}{(\beta+1)(3\beta+1)} \left(\log \delta_H(\mathbf{s}, \mathbf{t})\right)^{3\beta+1} \\ &= O\left(\frac{2^{3\beta}}{\beta^2} (\log n)^{\beta+1} \left(\log \delta_H(\mathbf{s}, \mathbf{t})\right)^{3\beta+1}\right). \end{aligned}$$

□

Remarques. On remarquera que, pour le tore régulier, qui correspond au réseau de Kleinberg infini, on a $\beta = 0$, puisque les boules sont de croissance polynomiale et la dimension intrinsèque est constante. L'espérance de la longueur de chemin de l'algorithme glouton dans un graphe *petit-mondisé* par le théorème 5.1 reste donc

identique à celle obtenue dans le réseau de Kleinberg, *i.e.* $O\big((1/\varepsilon)\log^{2+\varepsilon} m\big)$ entre deux noeuds distants de m dans le cas général (cas (i) et (ii)) et $O(\log n \log m)$ dans le tore de côté n lorsqu'on ne fait pas intervenir ε (cas (iii)).

On notera que le facteur logarithmique $\log^{1+\beta+\varepsilon}(\delta(\mathbf{u},\mathbf{v}))$ de la distribution n'est pas contraignant en soi et est introduit pour la généralisation aux graphes infinis et aux familles infinies de graphes finis, ainsi que pour obtenir une longueur de chemin qui ne dépend que de la distance entre les noeuds source et cible.

Ce théorème montre que la propriété de petit monde navigable n'est pas propre aux tores réguliers augmentés et donne un processus permettant d'obtenir une large classe de graphes *petit-mondisables*. En particulier, il traite le cas du tore allongé et de tous les graphes ayant des tailles de boules $b(r)$ croissant comme $r^{\alpha \log\log r}$, $\alpha > 0$, ou plus lentement. La classe des graphes à croissance modérée, incluse dans celle des graphes à croissance β-modérée, est par ailleurs considérée comme un bon modèle de la structure des grands réseaux informatiques, comme le réseau Internet ou les réseaux pair-à-pair, qui sont peu denses [KR02]. Cette classe a été particulièrement étudiée récemment dans le cadre de la localisation distribuée d'objets (*i.e.* la construction de tables de hachage distribuées des réseaux pair-à-pair). On pourra se référer à [HKK04, PRR99].

Dans le cadre des réseaux de routage, la classe des graphes de Cayley a également été particulièrement étudiée (voir par exemple [Zho04]). Rappelons qu'un graphe de Cayley est un graphe défini par un groupe G, ayant un nombre fini de générateurs $g_1, \ldots, g_k$. Les sommets sont les éléments de G et il existe arête entre x et y si et seulement s'il existe un générateur $g_i \in G$ tel que $x = g_i y$, ou $y = g_i x$. Notre théorème nous permet d'affirmer que tous les graphes de Cayley connus sont *petit-mondisables*. En effet,

- soit leur expansion est polynomiale et ils sont couverts par le théorème 5.1 avec $\beta = 0$,
- soit leur diamètre est polylogarithmique (expansion exponentielle ou presque exponentielle, c'est-à-dire $\Omega(2^{\Theta(r^d)})$ pour $d > 0$) et ils sont intrinsèquement petits mondes.

En effet, l'existence de groupes de taille de boule *intermédiaire* (super-polynomiale ou sous-exponentielle) est encore indéterminée. plus précisément, la question de l'existence d'un groupe de taille de boule $b(r)$ super-polynomiale mais inférieure à $e^{\sqrt{r}}$ est toujours ouverte. On pourra par exemple se reporter à [Gri90] et [Bar02] pour un état de l'art de ce sujet central en géométrie algébrique.

Si nous savons à présent *petit-mondiser* une grande classe de graphes, le théorème 5.1 ne caractérise toutefois pas tous les graphes *petit-mondisables* et la question de l'existence de graphes qui ne sont pas *petit-mondisables* est actuellement ouverte. La discussion précédente indique par ailleurs que la recherche de graphes qui ne seraient pas *petit-mondisables* doit se faire parmi d'autres classes de graphes que les graphes de Cayley.

Concernant l'émergence de l'effet petit monde dans les réseaux sociaux, la distribution inversement proportionnelle à la taille des boules semble refléter la répartition des contacts sociaux non géographiques. En effet, Liben-Nowell *et al.* [LNJK$^+$05] ont étudié expérimentalement un réseau social et ont séparé les connaissances en deux types : connaissances « géographiques » et « non géographiques ». Ils montrent que les connaissances « non géorgraphiques » sont réparties selon la distribution inverse des tailles de boules, c'est-à-dire que l'individu X a d'autant plus de chances de connaître l'individu Y si le nombre d'individus plus proches de X que Y est faible.

Produit cartésien de graphes *petit-mondisables*. Un fait remarquable de la propriété de petit monde du modèle de Kleinberg est sa relative indépendance en la dimension. En effet, l'espérance de la longueur des chemins calculés par l'algorithme glouton est la même lorsque la métrique sous-jacente est un anneau que lorsque c'est un tore de grande dimension. La grille de grande dimension étant un produit cartésien d'anneaux, on peut penser que la propriété présente dans l'anneau est conservée par l'application du produit cartésien.

Définition 5.3 (Produit cartésien)
Le produit cartésien $H = F \times G$ *de deux graphes non orientés* F *et* G *est le graphe* (V_H, E_H) *où* $V_H = V_F \times V_G$ *et* $E_H = \{\{(f,g),(f,g')\} \mid \{g,g'\} \in E_G, f \in V_F\} \cup \{\{(f,g),(f',g)\} \mid g \in V_G, \{f,f'\} \in E_F\}$.

Le théorème 5.1 confirme cette clôture par produit cartésien de la *petit-mondisation* pour la nouvelle classe des graphes à croissance β-modérée. En effet, le produit cartésien H de deux graphes G_1 et G_2 à croissance β_1- et β_2-modérée respectivement est un graphe à croissance $3(\beta_1 + \beta_2)$-modérée puisque :

$$b_{G_1,\mathbf{u}_1}(r/2)b_{G_2,\mathbf{u}_2}(r/2) \leqslant b_{H,(\mathbf{u}_{G_1},\mathbf{u}_{G_2})}(r) \leqslant b_{G_2,\mathbf{u}_2}(r)b_{G_1,\mathbf{u}_2}(r),$$

pour tout couple de noeuds $(\mathbf{u}_1, \mathbf{u}_2) \in G_1 \times G_2$.

Cette remarque, combinée au théorème 5.1, nous permet de donner une méthode alternative simple, pour obtenir la généralisation du graphe de Kleinberg de dimension 1 aux tores de dimension supérieure, que l'on peut voir comme des produits cartésiens de tores de dimension 1. Par exemple, en conséquence, tout tore $C_{n_1} \times C_{n_2} \times \ldots \times C_{n_l}$ peut être augmenté aléatoirement en un petit monde navigable, pour tout $\varepsilon > 0$, dans lequel l'algorithme glouton calcule des chemins dont l'espérance de la longueur est $O((1/\varepsilon)\log^{2+\varepsilon}(\max_i n_i))$. Enfin, cette conservation de la *petit-mondisation* par le produit cartésien renforce l'impression que la propriété petit monde d'un réseau peut reposer sur plusieurs sous-structures différentes.

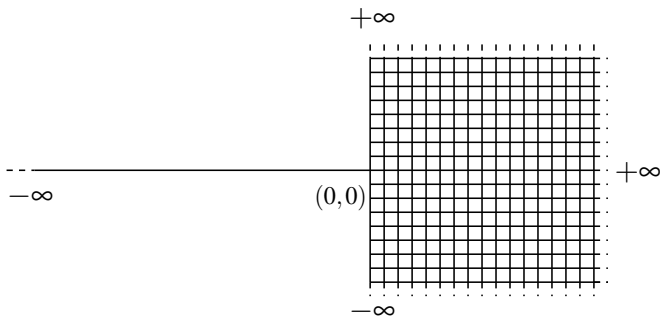


FIG. 5.2 : Exemple de la *tapette à mouches infinie*, *petit-mondisable*, mais qui n'est pas à croissance modérée.

5.3 Une caractérisation non exhaustive des graphes *petit-mondisables*

Dans la section précédente, nous avons étendu le modèle de Kleinberg [Kle00], en démontrant qu'une grande classe de graphes est *petit-mondisable*. Toutefois, le théorème 5.1 laisse ouverte la question de déterminer si l'on peut *petit-mondiser* tous les graphes. La condition de croissance modérée est encore trop restrictive puisque nous pouvons exhiber des exemples de graphes qui ne la satisfont pas et sont pourtant *petit-mondisables*. Dans l'exemple illustré sur la figure 5.2, le graphe est composé de deux parties distinctes de dimension intrinsèque différentes. Il s'agit d'une chaîne semi-infinie connectée à une grille semi-infinie de dimension 2. Nous appelons ce graphe *tapette à mouche infinie*. Le graphe a pour ensemble de sommets $(\mathbb{Z}^-_{<0} \times \{0\}) \bigcup (\mathbb{Z}^+_{\geqslant 0} \times \mathbb{Z})$; deux noeuds sont adjacents si leur distance de Manhattan est égale à 1. Il n'existe alors pas de constante β telle que ce graphe soit à croissance β-modérée, car pour tout noeud $\mathbf{u}$ sur la chaîne, à distance m de la grille, on a : $b_{\mathbf{u}}(m) = 2m$, mais $b_{\mathbf{u}}(2m) = \Theta(m^2) + 3m$. Pourtant, on peut *petit-mondiser* ce graphe en l'augmentant par la distribution $1/(r^2 \log^{1+\varepsilon}(r+1))$ sur la partie bidimensionnelle et par la distribution $1/(r \log^{1+\varepsilon}(r+1))$ sur la partie unidimensionnelle, pour un $\varepsilon > 0$ quelconque. D'après notre extension du modèle de Kleinberg aux graphes infinis, nous savons en effet qu'en augmentant le graphe de cette façon, l'algorithme glouton trouve des chemins de longueur polylogarithmique en espérance entre tous couples de noeuds dans chaque sous-graphe et par conséquent entre tous couples de noeuds du graphe. Cependant, nous verrons dans la section 5.6 un deuxième processus de *petit-mondisation* permettant de traiter l'exemple de la tapette à mouche infinie.

5.4 D'autres distributions sont-elles possibles ?

Après avoir déterminé une grande classe de graphe pouvant être augmentée en petit monde navigable, nous renforçons ce résultat en présentant, dans cette section, une borne inférieure polynomiale sur l'espérance de la longueur du chemin calculé par un algorithme décentralisé arbitraire lorsque l'on modifie l'exposant sur la taille de boule de la distribution. Ce résultat généralise celui de Kleinberg [Kle00] et Barrière *et al.* [BFKK01] qui démontrent que l'augmentation par un nombre constant de liens longue-distance par noeud, dans le tore régulier d-dimensionnel, selon une distribution s-harmonique, pour s différent de d, rend le graphe augmenté non-navigable pour tout algorithme décentralisé.

Théorème 5.2

(i) Soit $\alpha > 1$. Pour tout $\varepsilon > 0$, $\beta \geqslant 0$ et tout graphe infini H à croissance β-modérée, $\beta \geqslant 0$, augmenté selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u}\in H}$ définie par :

$$f_{\mathbf{u}}(r) = \frac{1}{\big(b_{\mathbf{u}}(r)\big)^{\alpha} \log^{1+\beta+\varepsilon}(r+1)},$$

il existe une constante $c > 0$ telle que l'espérance de la longueur du chemin calculé par tout algorithme décentralisé entre tout couple de noeuds distants de m, est supérieure à $c \cdot m^{1-1/\alpha}$.

(ii) Soit $\alpha < 1$. Pour tout $k > 1$ et tout tout graphe fini H à croissance β-modérée, $\beta \geqslant 0$, de n noeuds, augmenté par selon la famille de distributions $(f_{\mathbf{u}})_{\mathbf{u}\in H}$ définie par :

$$f_{\mathbf{u}}(r) = \frac{1}{\big(b_{\mathbf{u}}(r)\big)^{\alpha}},$$

il existe une constante $c' > 0$ telle que l'espérance de la longueur du chemin calculé par tout algorithme décentralisé entre tout couple de noeuds, est supérieure à $c' \cdot n^{1/(\log\log n)^2}$.

Dans le cas $\alpha < 1$, nous ne démontrons la borne inférieure que pour les graphes finis, car la distribution n'est pas définie pour les graphes infinis, comme la preuve le fera apparaître (même si l'on ajoutait un facteur logarithmique au dénominateur). Dans le cas $\alpha > 1$, on remarquera que le terme logarithmique de la distribution n'est pas significatif, puisque son rôle est simplement d'obtenir un facteur de normalisation constant pour que la distribution soit bien définie sur les graphes infinis. Le résultat démontré pour les graphes infinis s'étend de façon immédiate aux familles infinies de graphes finis à croissance modérée, comme nous l'avons remarqué dans les sections précédentes.

Preuve du théorème 5.2 (i). Soit H un graphe infini à croissance β-modérée. On peut tout d'abord vérifier que cette distribution est toujours bien définie en remar-

quant que la constante de normalisation $Z_{\mathbf{u}}$, pour chaque noeud $\mathbf{u}$, est strictement inférieure à celle de la distribution $1/(b_{\mathbf{u}}(r)\log^{1+\beta+\varepsilon}(r+1))$ dont on a vérifié la convergence dans la preuve du théorème 5.1 et pour laquelle il existe une majoration constante. Cette majoration est par ailleurs indépendante de $\mathbf{u}$, on note Z_ε la constante uniforme telle que pour tout noeud $\mathbf{u}$, $Z_{\mathbf{u}} < Z_\varepsilon$.

On considère à présent un algorithme de routage décentralisé $\mathcal{A}$ quelconque, selon la définition 1.10 donnée dans le préambule. Nous allons montrer qu'avec cette distribution, les liens longue-distance sont trop courts pour permettre au routage décentralisé d'être efficace.

Pour tout noeud $\mathbf{u}$ et tout $\ell > 1$, la probabilité que le j-ème lien longue-distance de $\mathbf{u}$ soit de longueur supérieure à ℓ est :

$$\begin{aligned}
\Pr\{\delta_H(\mathbf{u}, L(\mathbf{u})) \geqslant \ell\} &= \frac{1}{Z_\varepsilon}\sum_{r\geqslant\ell}\frac{s_{\mathbf{u}}(r)}{\big(b_{\mathbf{u}}(r)\big)^\alpha\log^{1+b+\varepsilon}(r+1)} \\
&= \frac{1}{Z_\varepsilon}\sum_{i\geqslant\log(\ell-1)}\sum_{r=2^i+1}^{2^{i+1}}\frac{b_{\mathbf{u}}(r)-b_{\mathbf{u}}(r-1)}{\big(b_{\mathbf{u}}(r)\big)^\alpha\log^{1+\beta+\varepsilon}(r+1)} \\
&\leqslant \frac{1}{Z_\varepsilon}\sum_{i\geqslant\log(\ell-1)}\frac{\sum_{r=2^i+1}^{2^{i+1}}\big(b_{\mathbf{u}}(r)-b_{\mathbf{u}}(r-1)\big)}{\big(b_{\mathbf{u}}(2^i+1)\big)^\alpha\log^{1+\beta+\varepsilon}(2^i+1)}.
\end{aligned}$$

Par la croissance β-modérée de H, on a :

$$\sum_{r=2^i+1}^{2^{i+1}}\big(b_{\mathbf{u}}(r)-b_{\mathbf{u}}(r-1)\big) = b_{\mathbf{u}}(2^{i+1})-b_{\mathbf{u}}(2^i) \leqslant 2^\beta c_0 i^\beta b_{\mathbf{u}}(2^i).$$

On obtient :

$$\Pr\{\delta_H(\mathbf{u}, L(\mathbf{u})) \geqslant \ell\} \leqslant \frac{2^\beta c_0}{Z_\varepsilon}\sum_{i\geqslant\log(\ell-1)}\frac{1}{\big(b_{\mathbf{u}}(2^i)\big)^{\alpha-1}i^{1+\varepsilon}}.$$

On a par ailleurs $b_{\mathbf{u}}(2^i) \geqslant 2^i$ par connexité, ainsi :

$$\begin{aligned}
\Pr\{\delta_H(\mathbf{u}, L(\mathbf{u})) \geqslant \ell\} &\leqslant \frac{2^\beta c_0}{Z_\varepsilon}\sum_{i\geqslant\log(\ell-1)}\frac{1}{(2^{\alpha-1})^i} \leqslant \frac{2^\beta c_0}{Z_\varepsilon}\int_{\log(\ell-1)-1}^{\infty}\frac{dt}{(2^{\alpha-1})^t} \\
&\leqslant \frac{2^\beta c_0}{Z_\varepsilon(\alpha-1)\ln 2}\,\frac{1}{(\ell-1)^{\alpha-1}}
\end{aligned}$$

Soit m la distance séparant la source et la cible du routage. La probabilité de visiter un noeud dont le lien longue-distance est de longueur supérieure à $(m+1)^{1/\alpha}$ parmi les $qm^{(\alpha-1)/\alpha}$ premiers noeuds visités, $q < 1$ est donc inférieure à :

$$\frac{2^\beta c_0}{Z_\varepsilon(\alpha-1)\ln 2}\,\frac{qm^{(\alpha-1)/\alpha}}{(m^{1/\alpha})^{\alpha-1}} \leqslant \frac{q2^\beta c_0}{Z_\varepsilon(\alpha-1)\ln 2}.$$

Il existe q_0 tel que pour $q \leqslant q_0$, on a $(q2^\beta c_0)/(Z(\alpha-1)\ln 2) < 1$, on suppose à présent $q < q_0$.

Soit $\mathcal{E}$ l'événement : « parmi les $qm^{(\alpha-1)/\alpha}$ premiers noeuds visités, on ne visite aucun noeud dont le lien longue-distance est de longueur supérieure à $(m+1)^{1/\alpha}$ ». Soit X la variable aléatoire pour la longueur du chemin calculé par $\mathcal{A}$. Si l'événement $\mathcal{E}$ se produit, l'algorithme se rapproche de la cible d'une distance au plus $(1/4)(m+1)^{1/\alpha} \cdot m^{(\alpha-1)/\alpha} < 2^{1/\alpha}qm$, durant les $qm^{(\alpha-1)/\alpha}$ premiers pas. Il existe q_1 tel que pour $q \leqslant q_1$ on ait $(q2^{1/\alpha}) < 1$, dans ce cas, l'algorithme n'atteint donc pas encore la cible. On suppose à présent $q \leqslant \min(q_0, q_1)$. On a donc : $\mathbb{E}[X \mid \mathcal{E}] \geqslant (1/q)m^{(\alpha-1)/\alpha}$. On obtient finalement :

$$\mathbb{E}[X] \geqslant \mathbb{E}[X \mid \mathcal{E}] \cdot \Pr \mathcal{E} \geqslant qm^{(\alpha-1)/\alpha} \cdot \left(1 - \frac{q2^\beta c_0}{Z(\alpha-1)\ln 2}\right) = \Omega(m^{(\alpha-1)/\alpha}).$$

□

Preuve du théorème 5.2 (ii). Soit H un graphe fini, de n noeuds, à croissance β-modérée. Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible pour le routage respectivement. La probabilité que $\mathbf{s}$ n'appartienne pas à $\mathcal{B}_{\mathbf{t}}(n^{1/(\log\log n)^2})$ est :

$$1 - \frac{b_{\mathbf{t}}(n^{1/(\log\log n)^2})}{n} \geqslant 1 - \frac{n^{2\beta/\log\log n}}{n},$$

pour n supérieur à une constante $n_0 > 0$, puisque la croissance β-modérée de H impose que :

$$b_{\mathbf{t}}\left(n^{1/(\log\log n)^2}\right) \leqslant n^{\dfrac{c_0 + \beta\log\log(n^{1/(\log\log n)^2})}{(\log\log n)^2}} \leqslant n^{2\beta/\log\log n}.$$

Comme $2\beta/\log\log n = o(1)$, il existe une constante $n_1 \geqslant n_0$ telle que, pour $n \geqslant n_1$, $1 - n^{2\beta/\log\log n - 1} \geqslant \frac{1}{4}$. La source a donc une probabilité constante d'être hors de la boule de rayon $n^{1/(\log\log n)^2}$ centrée sur la cible pour $n \geqslant n_1$. Nous allons montrer que tout algorithme décentralisé met un temps $n^{1/(\log\log n)^2}$ en espérance pour atteindre cette boule.

Pour tout noeud $\mathbf{u}$, le facteur de normalisation de la distribution est :

$$\begin{aligned} Z_{\mathbf{u}} &= \sum_{\mathbf{v}\neq\mathbf{u}} \frac{1}{(b_{\mathbf{u}}(\delta(\mathbf{u},\mathbf{v}))^\alpha} = \sum_{r=1}^{n} \frac{s_{\mathbf{u}}(r)}{(b_{\mathbf{u}}(r))^\alpha} \\ &\geqslant \sum_{r=1}^{n} \left((b_{\mathbf{u}}(r))^{1-\alpha} - (b_{\mathbf{u}}(r-1))^{1-\alpha}\right) \geqslant (b_{\mathbf{u}}(n))^{1-\alpha} - 1 \end{aligned}$$

Comme $f_{\mathbf{u}}$ est inférieure à 1 pour tout noeud $\mathbf{u}$, la probabilité qu'un noeud $\mathbf{u}$ ait son contact longue-distance dans la boule $\mathcal{B}_{\mathbf{t}}(n^{1/(\log\log n)^2})$ est inférieure à :

$$\frac{b_{\mathbf{t}}(n^{1/(\log\log n)^2})}{Z_{\mathbf{u}}} \leqslant \frac{b_{\mathbf{t}}(n^{1/(\log\log n)^2})}{(b_{\mathbf{u}}(n))^{1-\alpha}-1} \leqslant \frac{n^{2\beta/\log\log n}}{n^{1-\alpha}-1}.$$

On considère un algorithme décentralisé $\mathcal{A}$ quelconque et l'événement $\mathcal{E}$: « $\mathcal{A}$ visite un noeud dont le contact longue-distance appartient à $\mathcal{B}_{\mathbf{t}}(n^{1/(\log\log n)^2})$ durant les $n^{1/(\log\log n)^2}$ premiers pas ». Comme les liens sont tirés indépendamment, on a, par la borne de l'union :

$$\Pr\mathcal{E} \leqslant kn^{1/(\log\log n)^2}\,\frac{n^{2\beta/\log\log n}}{n^{1-\alpha}-1}.$$

Or, il existe une constante $n_2 > 0$ telle que, pour n supérieur à n_2, on ait $1-\alpha-(2\beta/\log\log n)-1/(\log\log n)^2 > 0$. Pour n supérieur à $\max(n_1, n_2)$, la probabilité que la source n'appartienne pas à $\mathcal{B}_{\mathbf{t}}(n^{1/(\log\log n)^2})$ et que $\mathcal{E}$ ne se produise pas est donc supérieure à une constante $C > 0$, indépendante de n. Soit $\mathcal{E}'$ cet événement. L'espérance de la longueur du chemin conditionnée par $\mathcal{E}'$ est supérieure à $n^{1/(\log\log n)^2}$, car pour être inférieure, l'algorithme doit trouver un contact longue-distance dans $\mathcal{B}_{\mathbf{t}}(n^{1/(\log\log n)^2})$ ce qui ne se produit pas avant $n^{1/(\log\log n)^2}$ pas par hypothèse. Soit X la variable aléatoire pour la longueur de chemin calculée par $\mathcal{A}$, on a donc :

$$\mathbb{E}[X] \geqslant \mathbb{E}[X \mid \mathcal{E}']\Pr\mathcal{E}' \geqslant Cn^{1/(\log\log n)^2}.$$

□

5.5 Un modèle qui peut engendrer (presque) toutes les distributions de degrés

Une des principales critique faites au modèle de Kleinberg est qu'il ne modélise par la distribution des degrés suivant une loi de puissance qui est observée sur de nombreux graphes réels, comme le graphe d'Internet [FFF99]. Pourtant, sur ce modèle comme sur celui que nous venons de présenter, il suffit de faire une modification mineure pour obtenir cette distribution des degrés et plus généralement, toute distribution de degrés qui assure un degré supérieur ou égal à 1 avec probabilité constante. C'est en fait le cas pour tout processus de *petit-mondisation* construit en ajoutant 1 lien par noeud.

Le théorème suivant présente l'ensemble des distributions de degrés possibles pour la *petit-mondisation* et montre ainsi que la propriété de petit monde navigable ne semble pas corrélée à une distribution des degrés spécifique.

Théorème 5.3
Soit H un graphe petit-mondisable *selon la famille de distribution $(f_{\mathbf{u}})_{\mathbf{u}\in H}$ (en ajoutant un lien par noeud) et π une distribution sur $\mathbb{N}$ telle qu'il existe une constante $c > 0$ telle que $\sum_{k\geqslant 1}\pi(k) \geqslant c$. Le graphe H augmenté en ajoutant à chaque noeud $\mathbf{u}$, k liens indépendants distribués selon $f_{\mathbf{u}}$ avec probabilité $\pi(k)$ est un petit monde navigable.*

Preuve. La probabilité qu'un noeud $\mathbf{u}$ donné ait au moins un lien longue-distance sortant est supérieure à c par hypothèse, par conséquent, il suffit de multiplier par c les probabilités qui portent sur un lien longue-distance dans la preuve de la *petit-mondisation* du théorème 5.1 pour obtenir le résultat. □

En particulier, soit π une distribution des degrés suivant une loi de puissance de paramètre α, *i.e.* $\pi(k) = 1/(Z_\alpha k^\alpha)$ pour $k \in \mathbb{N}^\star$, où Z est le facteur de normalisation distribution : $Z_\alpha = \sum_{k\geqslant 1}(1/k^\alpha)$. Pour $\alpha > 1$, on a $Z_\alpha < \infty$, la probabilité que le degré sortant d'un noeud donné soit supérieur ou égal à 1 est alors supérieur à une constante. L'exposant de la loi de puissance observé dans le graphe d'Internet est compris entre 2 et 3 [FFF99], le modèle que nous avons présenté dans la section 5.2 peut donc produire une telle distribution de degrés tout en produisant un petit monde navigable.

5.6 Une autre approche : la *petit-mondisation* par plongement

Dans cette section, nous présentons une nouvelle approche pour la *petit-mondisation* de graphe via le plongement du graphe initial dans $\mathbb{R}^d$. L'étude du modèle de Kleinberg infini (section 4.1.3, p. 105) a montré que $\mathbb{Z}^d$ était *petit-mondisable*. Nous montrons que si l'on sait plonger un graphe dans $\mathbb{R}^d$, pour une faible dimension d et avec une distorsion restreinte des distances, nous pouvons le *petit-mondiser* via l'utilisation de la distribution d-harmonique dans le plongement.

Définition 5.4 (Plongement)
Étant donnés deux espaces métriques (E, δ) et (E', δ'), on appelle plongement *de E dans E' toute fonction σ telle qu'il existe deux constantes $0 < \alpha < 1$ et $\beta > 1$ telles que pour tout couple $(\mathbf{u}, \mathbf{v}) \in E^2$,*

$$\alpha \cdot \delta(\mathbf{u}, \mathbf{v}) \leqslant \delta'(\sigma(\mathbf{u}), \sigma(\mathbf{v})) \leqslant \beta \cdot \delta(\mathbf{u}, \mathbf{v}).$$

α et β sont appelés respectivement contraction *et* expansion *de σ, et $\gamma =_{\text{def}} \beta/\alpha$ est la* distorsion *de σ.*

Rappelons que tout graphe définit de façon naturelle un espace métrique, dont les éléments sont les sommets du graphe et dont la distance entre deux points correspond à la longueur du plus court chemin les reliant dans le graphe.

Nous définissons à présent notre nouveau processus d'augmentation d'un graphe en un petit monde (algorithme 4).

Algorithme 4 Algorithme de *petit-mondisation* par plongement

Entrée : Un graphe $G = (V, E)$, un plongement σ de G dans $(\mathbb{R}^d, \ell_1)$ et une constante $\varepsilon > 0$.

Étape 1 **:**

pour chaque $\mathbf{u} \in V$ **faire**

tirer un point-destination $\tau_{\mathbf{u}}$ dans $\mathbb{R}^d$, avec une densité de probabilité proportionnelle à

$$\frac{1}{(2 + |\sigma(\mathbf{u}) - \tau|)^d \log^{2+\varepsilon}(2 + |\sigma(\mathbf{u}) - \tau|)},$$

sur tout $\tau \in \mathbb{R}^d$.

fin pour

Étape 2 **:**

pour chaque $\mathbf{u} \in V$ **faire**

$L(\mathbf{u}) \leftarrow \sigma^{-1}(\tau)$, où τ est le point de $\sigma(G)$ le plus proche de $\tau_{\mathbf{u}}$ dans $\mathbb{R}^d$.

fin pour

Dans l'algorithme 4, on suppose qu'il existe un plongement particulier σ du graphe G dans $\mathbb{R}^d$. L'étape 1 associe un point-destination à chaque point de $\mathbb{R}^d$ choisi selon la distribution qui *petit-mondise* le modèle de Kleinberg $\mathcal{K}^d_\infty$. Pour chaque noeud $\mathbf{u}$ du graphe, il existe alors un point-destination dans $\mathbb{R}^d$ (celui de son représentant $\sigma(\mathbf{u})$ dans le plongement). Si ce point destination est l'image d'un noeud $\mathbf{v}$ par σ, le contact longue-distance $L(\mathbf{u})$ de $\mathbf{u}$ est simplement $\mathbf{v}$. Sinon, $L(\mathbf{u})$ est le noeud $\mathbf{w}$ de $\sigma(G)$ qui est le plus proche du point-destination de $\sigma(\mathbf{u})$ dans $\mathbb{R}^d$. On obtient ainsi un ensemble de liens dans le graphe de base, qui ont été distribués en fonction des distances dans le plongement.

Pour que ces liens augmentent le graphe de base en un petit monde navigable, il est nécessaire que les distances entre les noeuds plongés dans $\mathbb{R}^d$ ne soient pas trop différentes des distances dans le graphe de base, afin de conserver les propriétés de petit monde du graphe $\mathbb{R}^d$ augmenté. Le théorème suivant affirme que l'algorithme 4 augmente le graphe de base en un petit monde lorsque le plongement a une distorsion γ limitée et que la dimension d n'est pas trop grande,

précisément γ^d doit être polylogarithmique.

Théorème 5.4 (*Petit-mondisation* par plongement)
Pour toute constante $\varepsilon > 0$ et tout graphe G, de n noeuds, tel qu'il existe un plongement σ de G dans $(\mathbb{R}^d, \ell_1)$, de distorsion γ, l'algorithme de routage glouton calcule, entre tout couple de noeuds du graphe aléatoirement augmenté obtenu en appliquant l'algorithme 4 à (G, σ, ε), un chemin de longueur :

$$O\left(\frac{2^{3d}}{\varepsilon}\gamma^{d+1+\varepsilon}\log^{2+\varepsilon} n\right),$$

en espérance.

Preuve. On considère $\mathbf{s}$ et $\mathbf{t}$ deux noeuds de G, qui sont la source et la cible du routage respectivement. On décompose l'exécution du routage glouton en phases : la phase i dure tant que la distance du noeud courant à $\mathbf{t}$ appartient à $(2^{i-1}, 2^i]$, pour i variant de 1 à $\log \delta(\mathbf{s}, \mathbf{t})$, où δ est la distance dans G. On note respectivement α et β la contraction et l'expansion de σ. Quitte à renormaliser le plongement, on peut supposer qu'il est non-contractant, *i.e.* que la contraction vaut 1.

Supposons que le noeud courant $\mathbf{x}$ est à distance $x = \delta(\mathbf{x}, \mathbf{t})$ de $\mathbf{t}$. Soit $X = |\sigma(\mathbf{x}) - \sigma(\mathbf{t})|$. Comme σ est de distorsion γ, on a : $\delta(\mathbf{x}, \mathbf{t}) \leqslant |\sigma(\mathbf{x}) - \sigma(\mathbf{t})| \leqslant \gamma \cdot \delta(\mathbf{x}, \mathbf{t})$, *i.e.* $x \leqslant X \leqslant \gamma \cdot x$.

Soit $\mathcal{E}$ l'événement : « $\tau_{\mathbf{x}}$ est à distance inférieure à $X/(4\gamma)$ de $\sigma(\mathbf{t})$ ». Si $\mathcal{E}$ se produit, alors $|\sigma(L(\mathbf{x})) - \sigma(\mathbf{t})| \leqslant X/(2\gamma)$ puisque :

$$|\sigma(L(\mathbf{x})) - \sigma(\mathbf{t})| \leqslant |\sigma(L(\mathbf{x})) - \tau_{\mathbf{x}}| + |\tau_{\mathbf{x}} + \sigma(\mathbf{t})| \leqslant 2|\tau_{\mathbf{x}} + \sigma(\mathbf{t})| \leqslant X/(2\gamma),$$

puisque $\sigma(L(\mathbf{x}))$ est plus proche de $\tau_{\mathbf{x}}$ que $\sigma(\mathbf{t})$ (au sens large). On obtient :

$$\delta(L(\mathbf{x}) - \mathbf{t}) \leqslant |\sigma(L(\mathbf{x})) - \sigma(\mathbf{t})| \leqslant X/(2\gamma) \leqslant x/2.$$

Nous évaluons à présent la probabilité de $\mathcal{E}$. Dans $\mathbb{R}^d$, le facteur de normalisation de la distribution est le même en tout point, au point $\mathbf{0}$ de coordonnées nulles, il vaut :

$$Z^{\infty}_{\mathbb{R}^d} = \int_{\tau \in \mathbb{R}^d,\, \tau \neq \mathbf{0}} \frac{d\tau}{(|\tau| + 2)^d \log^{1+\varepsilon}(|\tau| + 2)}.$$

Pour évaluer cette intégrale nous utilisons le volume $s_{\mathbb{R}^d}(r)$ d'une hypersphère de rayon r dans $\mathbb{R}^d$. On remarquera que $s(r-1) \leqslant s_{\mathbb{R}^d}(r) \leqslant s(r+1)$, où $s(r)$ est le volume d'une hypersphère de rayon r dans $\mathbb{Z}^d$. Nous avons évalué ce volume dans le préambule, page 23, on en déduit : $c'\left(2^d/(d-1)!\right)\, r^{d-1} \leqslant s_{\mathbb{R}^d}(r) \leqslant c\left(2^d/(d-1)!\right)\, r^{d-1}$, où $c > 1$ et $c' < 1$ sont des constantes. On obtient :

$$Z^{\infty}_{\mathbb{R}^d} \leqslant c\,\frac{2^d}{(d-1)!}\int_{r>0} \frac{dr}{(r+2)\log^{1+\varepsilon}(r+2)} \leqslant \frac{c}{\ln 2}\,\frac{2^d}{(d-1)!}\,\frac{1}{\varepsilon}.$$

Nous utilisons à présent les minorants obtenus pour une grille d-dimensionnelle dans le chapitre 2 : nous pouvons généraliser les arguments des preuves à $\mathbb{R}^d$ en faisant tendre le maillage vers l'infini et nous adaptons les valeurs des minorants à l'aide du volume d'une hypersphère donné plus haut. D'après le lemme 2.11, la probabilité de $\mathcal{E}$ est minimale lorsque $\sigma(\mathbf{x})$ est placé sur un coin de la sphère $\mathcal{S}_{\sigma(\mathbf{t})}(X)$. Lorsque $\sigma(\mathbf{x})$ est un coin de la sphère, en utilisant les même arguments que pour le lemme 2.12, il y au moins $\mathbf{s}(i/2)/2 \geqslant c'(2^d/(d-1)!)i^{d-1}$ noeuds dans $\mathcal{B}_{\sigma(\mathbf{t})}(\lambda X)$ à distance inférieure ou égale à $i+(1-\lambda)X$ de $\sigma(\mathbf{x})$ pour i allant de 1 à λX, pour tout $0 < \lambda < 1$. Comme nous cherchons une minoration, on peut sommer les probabilités pour i allant de $\lambda X/2$ à λX seulement et dans ce cas $i+(1-\lambda)X+2 \leqslant (2/\lambda)i$. On obtient :

$$\begin{aligned}
\Pr \mathcal{E} &\geqslant c' \frac{2^d/(d-1)!}{Z^{\infty}_{\mathbb{R}^d}} \int_{\lambda X/2}^{\lambda X} \frac{i^{d-1} di}{(i+(1-\lambda)X+2)^d \log^{1+\varepsilon}(i+(1-\lambda)X+2)} \\
&\geqslant c' \ln 2\, \varepsilon \left(\frac{\lambda}{2}\right)^{d+1+\varepsilon} \int_{\lambda X/2}^{\lambda X} \frac{dr}{r \log^{1+\varepsilon} r} \\
&\geqslant c' \ln 2 \left(\frac{\lambda}{2}\right)^{d+1+\varepsilon} \left(\frac{1}{(\log(\lambda X)-1)^{\varepsilon}} - \frac{1}{(\log(\lambda X))^{\varepsilon}}\right) \\
&\geqslant c' \ln 2 \left(\frac{\lambda}{2}\right)^{d+1+\varepsilon} \frac{\varepsilon}{\log^{1+\varepsilon}(\lambda X)}
\end{aligned}$$

En prenant $\lambda = 1/(4\gamma)$, on obtient :

$$\Pr \mathcal{E} \geqslant c' \ln 2 \frac{\varepsilon}{2^{3d+1+\varepsilon}\, \gamma^{d+1+\varepsilon}} \frac{1}{\log^{1+\varepsilon}(X/\gamma)}.$$

Finalement : $\Pr \mathcal{E} \geqslant f(\varepsilon, d, \gamma)/\log^{1+\varepsilon}(x)$, où $f(\varepsilon, d, \gamma)$ décrit la constante calculée ci-dessus.

L'espérance du nombre de noeuds visités par l'algorithme glouton dans G, en phase i est donc inférieure à $i^{1+\varepsilon}/f(\varepsilon, d, \gamma)$ et la longueur totale du chemin est inférieure à :

$$\frac{1}{f(\varepsilon, d, \gamma)} \int_{i=1}^{\delta(\mathbf{s},\mathbf{t})} i^{1+\varepsilon} \leqslant \frac{1}{c' \ln 2} \frac{2^{3d+1+\varepsilon}}{\varepsilon(2+\varepsilon)} \gamma^{d+1+\varepsilon} \log^{2+\varepsilon}(\delta(\mathbf{s},\mathbf{t})),$$

On obtient finalement une espérance de longueur de chemin :

$$O\left(\frac{2^{3d}}{\varepsilon}\, \gamma^{d+1+\varepsilon} \log^{2+\epsilon} n\right).$$

□

Comme un graphe aléatoirement augmenté dans lequel l'algorithme glouton calcule des chemins de longueur polylogarithmique en espérance est un petit monde navigable, on obtient le corollaire suivant.

Corollaire 5.5 *Tout graphe G de n noeuds pour lequel il existe un plongement dans $(\mathbb{R}^d, \ell_1)$ de distorsion γ tel que $\gamma^d = polylog(n)$ est* petit-mondisable *par l'algorithme 4.*

La preuve est immédiate.

L'exemple de la tapette à mouche infinie, qui n'était pas couvert par la première caractérisation de graphes *petit-mondisables*, est couvert par ce corollaire puisque ce graphe est naturellement plongé dans $\mathbb{Z}^2$, isométriquement (*i.e.* avec distorsion 1). De façon plus générale, les graphes admettant par exemple un plongement de distorsion constante dans $\mathbb{R}^{O(\log\log n)}$, ou un plongement de distorsion polylog(n) dans $\mathbb{R}^{O(1)}$ sont *petit-mondisables*. Toutefois, cette nouvelle caractérisation n'est pas exhaustive et ne répond pas à la question de l'existence de graphes qui ne seraient pas *petit-mondisables*.

Ce problème est relié aux problèmes généraux de plongements de métriques finies dans un espace euclidien de faible dimension qui ont reçu une attention particulière récemment. La majorité des travaux de ce domaine portent sur des plongements de métriques générales. En 1985, Bourgain [Bou85] a démontré un théorème important qui donne la construction explicite d'un plongement de distorsion $O(\log n)$ de toute métrique de n points dans $(\mathbb{R}^{O(\log n)}, \ell_1)$. Ce résultat a permis de créer de nouveaux algorithmes d'approximation, car la résolution d'un problème est souvent plus simple dans un espace euclidien de dimension finie et le facteur d'approximation reste faible. Les travaux sur les plongements de métriques ont donc trouvé de nombreuses applications algorithmiques et combinatoires. Pour le corollaire 5.5, il serait intéressant d'avoir une caractérisation des métriques de graphe *spécifiques* qui satisfont les hypothèses de l'existence du plongement. On peut voir cette question comme un problème de type Ramsey pour les métriques (cf. [BLMN03]). Toutefois, les résultats qui portent sur le compromis entre la dimension et la distorsion d'un plongement dans ℓ_1 se sont essentiellement concentrés sur les métriques générales, et une dimension logarithmique était généralement suffisamment petite pour les problèmes algorithmiques qui ont motivés ces résultats. Ces résultats ne nous permettent donc pas de caractériser les graphes correspondant à l'hypothèse du corollaire 5.5. C'est seulement très récemment que les plongements dans les espaces euclidiens de dimension constante ont connu un véritable essor, du fait du développement des grands réseaux décentralisés. En effet, une problématique de ces réseaux est de pouvoir évaluer les distances entre les noeuds sans avoir à parcourir le graphe, pour avoir une complexité réaliste en regard du très grand nombre de noeuds. En plongeant

le graphe dans un espace ℓ_p de dimension constante avec une distorsion limitée, la distance entre deux noeuds peut être simplement évaluée en calculant la norme ℓ_p entre les deux coordonnées des noeuds dans le plongement. Un résultat récent de Abraham *et al.* [ABC+05] propose en particulier un plongement qui permet de plonger toute métrique dans ℓ_p avec une distorsion et une dimension constante lorsqu'on autorise une fraction constante des arêtes à être arbitrairement distordues. Ce résultat est très prometteur pour une extension de notre méthode de *petit-mondisation* par plongement de tout graphe.

5.7 Discussion

Dans ce chapitre, nous avons montré, par deux méthodes de *petit-mondisation* différentes, qu'il existe une grande classe de graphes qui peuvent être transformés en petits mondes navigables via l'ajout d'un nombre constant de liens aléatoires. Le premier processus (section 5.2) montre que ces liens peuvent être distribués en fonction de la croissance des boules du graphe et le second (section 5.6) donne une caractérisation liée à l'existence d'un plongement de faible distorsion dans un espace euclidien de faible dimension.

D'autres résultats concernant la compréhension des propriétés structurelles de la navigabilité sont parus simultanément et démontrent également la possibilité d'augmenter des graphes spécifiques en petits mondes navigables. Slivinks [Sli05] a ainsi démontré indépendamment que toutes les métriques de dimension doublante bornée, qui contiennent les métriques de croissance bornée, étaient *petit-mondisables* via l'ajout d'un nombre polylogarithmique de liens par noeud. La dimension doublante d'un graphe est l'exposant α tel que toute boule de rayon r peut être couverte par 2^α boules de rayon $r/2$. Il s'agit d'une problématique légèrement différente dans le sens où l'on suppose la donnée d'une métrique de départ et non d'un graphe de départ, il n'y a donc aucune arête entre les noeuds avant la *petit-mondisation*. On doit alors tenir compte de deux paramètres : le nombre de noeuds n et le diamètre de la métrique Δ qui est le ratio de la plus grande distance sur la plus petite distance. L'auteur utilise le fait qu'il existe une mesure doublante[4] pour toute métrique de dimension doublante bornée, cela signifie grossièrement qu'il existe une assignation de poids aux noeuds qui donne à la métrique l'apparence d'une métrique de croissance bornée (on pourra se reporter au chapitre 13 de [Hei01]). Le processus de *petit-mondisation* ajoute alors $O(2^\alpha \log n \log \Delta)$ liens aléatoires par noeud (soit $O(2^\alpha \log^2 n)$ en partant d'un graphe à croissance bornée) et produit un graphe dans lequel un algorithme de routage décentralisé calcule des chemins de longueur $O(2^\alpha \log n)$ en espérance.

[4]une mesure doublante pour un espace métrique est une mesure μ telle qu'il existe une constante $c > 0$ telle que, pour tout noeud $\mathbf{u}$ et tout rayon $r > 0$, on a $\mu(b_{\mathbf{u}}(2r)) \leqslant c \cdot \mu(b_{\mathbf{u}}(r))$.

Ce résultat élargit donc la classe des graphes *petit-mondisables*. Slivkins ne donne pas d'implémentation de cette *petit-mondisation*, mais on remarquera qu'elle requiert le calcul d'une mesure de dimension doublante pour la métrique (le meilleur algorithme est actuellement en $O(n \log n)$ [HM05]), puis le parcours de toutes les boules du graphe (selon une implémentation naïve).

D'autre part, Fraigniaud [Fra05] a démontré que tout graphe fini de *largeur arborescente* bornée ou de *cordalité* bornée, pouvait être augmenté en petit monde via l'ajout d'un lien par noeud. La largeur arborescente d'un graphe G représente grossièrement l'entier minimal k tel que le graphe peut être décomposé en un arbre dont les noeuds sont des sous-graphes de G de diamètre au plus k. La cordalité est la taille maximum d'un cycle sans corde dans G. Le processus de *petit-mondisation* proposé commence par faire une décomposition arborescente du graphe sur laquelle est calculée la distribution des liens-longue-distance. Cette *petit-mondisation* permet d'obtenir une espérance de longueur de chemin $O(\log^2 n)$ pour l'algorithme glouton.

Toutes les *petit-mondisations* proposées distribuent les liens longue-distance selon une loi monotone en la distance sous-jacente. Flammini *et al.* [FMNP05] ont récemment montré une borne inférieure $\Omega(\log^2 n)$ pour l'espérance de la longueur de chemin de l'algorithme glouton pour toutes les distributions monotones. Si les nouveaux modèles de petit monde permettent de mieux comprendre les propriétés structurelles des petits mondes et éventuellement d'en tirer parti pour construire des réseaux informatiques où le routage est efficace, on ne peut donc pas espérer construire de modèle où l'algorithme glouton serait plus efficace que dans le modèle de Kleinberg original en ajoutant des liens selon une distribution monotone.

Les processus de *petit-mondisation* que nous avons présentés dans ce chapitre et ceux de [Fra05, Sli05] ne donnent pas de caractérisation exhaustive et la question de l'existence de graphes non *petit-mondisables* reste ouverte. Le processus de *petit-mondisation* par plongement nous laisse néanmoins suspecter que tout graphe peut être *petit-mondisé*. En effet, un résultat récent de Abraham *et al.* [ABC$^+$05] montre qu'il est possible de plonger, avec une distorsion constante, toute métrique dans $\mathbb{R}^d$, où la dimension d est une constante, si l'on autorise une fraction constante des arêtes à avoir une distorsion arbitraire. Ce résultat, combiné à notre étude d'un processus de *petit-mondisation* par plongement, nous laisse suspecter que tout graphe pourrait être augmenté en un petit monde navigable. Néanmoins, pour dépasser l'erreur induite par les plongements laissant une fraction des distances arbitrairement distordues, un degré sortant polylogarithmique semble nécessaire afin de garantir une couverture de toutes les échelles de distances du graphe par les liens aléatoires ajoutés.

Sous l'hypothèse de la validité de cette conjecture, la caractéristique essentielle de l'existence de la navigabilité dans un graphe serait donc davantage due

à la relation existante entre la distribution des raccourcis et la métrique des distances sous-jacente, qu'à une propriété spécifique de cette métrique, ce qui expliquerait l'omniprésence de cette propriété. Par ailleurs, nous devrions alors rechercher l'origine de l'émergence de l'effet petit monde dans un processus naturel de construction des raccourcis. Dans cette perspective, le chapitre suivant présente une implémentation distribuée de *petit-mondisation*, où chaque noeud choisit son contact longue-distance sans connaître le reste du graphe.

Chapitre 6

Petit-mondisation distribuée

Introduction

Dans ce chapitre, nous nous intéressons à l'émergence de l'effet petit monde dans un réseau via un processus naturel d'ajout de raccourcis.

Si les *petit-mondisations* proposées dans le chapitre 5 et celles de Fraigniaud [Fra05] et Slivkins [Sli05] permettent bien d'augmenter un réseau en un petit monde navigable, elles requièrent néanmoins la connaissance de la quasi-totalité du réseau pour le calcul des liens longue-distance. [Fra05] nécessite notamment de calculer une décomposition arborescente du réseau, [Sli05] nécessite de calculer une mesure doublanteainsi que la taille de toutes les boules de la métrique, enfin le processus de *petit-mondisation* des graphes à croissance modérée que nous avons présenté dans le chapitre 5 nécessite également le calcul de toutes les tailles de boules. Ces processus requièrent donc un précalcul centralisé sur le réseau qui n'est pas envisageable dans le cadre d'un grand réseau informatique décentralisé, ou d'un réseau social. Dans ce chapitre, nous proposons un schéma entièrement distribué pour augmenter les graphes à croissance bornée en petits mondes navigables par l'ajout d'un unique lien par noeud. Nous utilisons pour cela une méthode d'échantillonnage multi-niveaux qui nous permet d'éviter les calculs centralisés des distributions des liens longue-distance aléatoires. En décrivant un schéma distribué pour la *petit-mondisation*, nous donnons un exemple de processus où, via un ensemble d'opérations locales, l'effet petit monde émerge dans un réseau. Pour décrire le fonctionnement de l'algorithme de routage glouton, nous supposons que les noeuds du réseau sous-jacent (avant augmentation) peuvent évaluer leur distance les uns par rapport aux autres, on suppose donc qu'il existe un oracle de comparaison des distances. Formellement, nous nous intéressons au problème suivant :

Problème 6.1
Étant donné un réseau, décrire un algorithme distribué efficace qui ajoute un lien par noeud tel que l'algorithme glouton, en utilisant un oracle répondant à la question "$\delta(\mathbf{u}, \mathbf{w}) \geqslant \delta(\mathbf{v}, \mathbf{w})$?" en temps constant, calcule des chemins de longueur polylogarithmique en espérance entre toute paire de noeuds du graphe augmenté.

6.1 Application aux réseaux informatiques décentralisés

La *petit-mondisation* distribuée permet de réduire le diamètre apparent[1] d'un réseau réel, en effectuant un précalcul qui ajoute un ensemble de liens (lorsqu'il s'agit d'un réseau virtuel). Les réseaux pair-à-pair utilisent par exemple un réseau virtuel de tables de hachage distribuées. Ce réseau permet de naviguer d'un pair à un autre, en utilisant les liens virtuels des tables de hachage. Dans ce contexte, l'ajout d'un nouveau lien par noeud de façon distribuée revient à ajouter une entrée par table de hachage. La *petit-mondisation* devient donc particulièrement intéressante lorsqu'elle n'ajoute qu'un lien par noeud.

Les architectures de réseau pair-à-pair les plus récentes [RD01, FG03, MNR02, SMK+01] construisent directement un réseau garantissant un routage efficace. Toutefois, il peut être difficile de maintenir dynamiquement une topologie précise. Dans un cadre plus général, le réseau peut être construit à partir d'une topologie arbitraire, c'est le cas par exemple des réseaux pair-à-pair non structurés [CSWH00]. Dans ce contexte, on peut imaginer que les pairs diffusent un message pour démarrer un processus de *petit-mondisation* lorsque le routage est trop lent.

L'hypothèse de l'existence d'un oracle de comparaison des distances pour l'algorithme glouton est réaliste pour de nombreuses classes de graphes. Les grilles de coordonnées discrètes, les hypercubes, les graphes de Delaunay et de Yao (appelés aussi Θ-graphes) offrent un tel oracle par le biais de leur propre étiquetage de noeuds. Le routage glouton a été très étudié sur ces topologies [BM04]. Par ailleurs, un tel oracle peut également être construit par un précalcul sur le réseau. Par exemple, des techniques sophistiquées et peu coûteuses, basées sur les spanneurs [TZ05], ou l'étiquetage de distance [GPPR01], permettent de calculer rapidement les distances exactes et approchées entre deux noeuds. Une autre approche consiste à construire des tables de routage compactes (voir [GP03, TZ05, AGM+04] pour des résultats sur les graphes arbitraires). Pour les graphes à croissance bornée [AM05], ou de dimension doublante

[1] Rappelons que le diamètre apparent d'un réseau est l'espérance de la longueur du plus long des plus courts chemins qui peuvent être calculés par un algorithme décentralisé.

bornée [Sli05], il est possible de construire des tables de routage de taille polylogarithmique en chaque noeud de façon à ce que les longueurs des chemins soient optimales à un facteur $1 + \epsilon$.

6.2 Modèle étudié

Dans la suite, nous allons considérer des réseaux non pondérés à *croissance bornée*, c'est-à-dire que pour tout noeud $\mathbf{u}$ et tout rayon $r > 0$, le ratio $b_{\mathbf{u}}(2r)/b_{\mathbf{u}}(r)$ est borné par une constante uniforme c, appelée *taux d'expansion*. Nos résultats s'étendent par ailleurs aux réseaux dont le taux d'expansion est borné par un polylogarithme du nombre de noeuds, les performances sont toutefois dégradées par ce facteur.

Notre schéma distribué s'applique aux réseaux *synchrones*, c'est-à-dire qu'il existe une horloge globale, connue de tous les noeuds, et qu'à chaque pulsation de l'horloge, chaque noeud est autorisé à effectuer une action. On appelle *ronde* la période de temps entre deux pulsations de l'horloge. On suppose que les noeuds communiquent par transmission de messages. Si Δ est le degré maximal du réseau, dans le modèle Δ-port, un noeud peut envoyer, durant chaque ronde, un nombre constant de messages à ses voisins. On suppose que les noeuds ont des identifiants uniques et que chaque noeud connaît les identifiants de ses voisins. L'existence d'un oracle répondant aux spécifications du problème 6.1 est contenue dans l'hypothèse que les tables de routage des noeuds permettent de répondre aux requêtes de comparaison de distance. Enfin, on suppose que chaque noeud a la connaissance de la taille n du réseau[2], et les identifiants de ses voisins. La quantité de mémoire par noeud est donc au moins $\Omega(\Delta \log n)$ bits.

Les principaux critères d'évaluation de l'efficacité des algorithmes distribués sont la complexité en temps (nombre de rondes), la complexité en communication (en nombre de messages) et la quantité de mémoire requise par noeud. Nous ne prenons pas en compte la complexité en temps des calculs locaux entre chaque ronde (ils sont en $O(\Delta)$ et donc dominés par les complexités en temps des communication entre les noeuds).

6.3 *Petit-mondisation* par échantillonnage multi-niveaux

Dans cette section, nous décrivons notre processus de *petit-mondisation* distribuée par échantillonnage multi-niveaux. Les détails de l'implémentation dis-

[2]Dans le cas où cette hypothèse est trop restrictive, des méthodes de précalcul peuvent être utilisées pour en obtenir une approximation, voir par exemple [HM03].

tribuées seront développés dans la section 6.4.

Les nouveaux processus de *petit-mondisation* tentent de répartir des raccourcis à toutes les échelles de distances, pour de grandes classes de graphe. La distribution des raccourcis est donc intimement liée à la métrique sous-jacente et nécessite de connaître la quasi-totalité du graphe. En effet, les *petit-mondisations* du chapitre précédent ainsi que de [Sli05] requièrent non seulement la connaissance des tailles de boules de chaque noeud (pour calculer la distribution des longueurs des liens) mais aussi les identifiants des noeuds de ces boules (pour pouvoir récupérer l'identifiant du contact longue-distance tiré). Le principe de notre schéma distribué est d'éviter ce calcul des tailles de boules par tous les noeuds, de façon à ce que chaque noeud n'explore que $O(\log n)$ noeuds en moyenne. Pour cela, nous construisons un échantillon multi-niveaux du réseau, les différents niveaux couvrant les différentes échelles de distance. Les noeuds de l'échantillon, appelés *leaders*, seront alors chargés du tirage des liens longue-distance des autres noeuds dont ils sont responsables.

6.3.1 Le processus de *petit-mondisation* distribué

Description de la *petit-mondisation* distribuée (Algorithme 5, p. 139). L'idée principale est que chaque noeud $\mathbf{u}$ ne calcule pas lui-même son lien longue-distance, il choisit seulement une échelle de longueur aléatoire $\ell = 2^i$, en choisissant i uniformément dans $\{\lceil \log(2c \log n) \rceil, \ldots, \lceil \log D \rceil\}$ et demande à un noeud *leader* $\mathbf{v}$, à distance inférieure à ℓ de $\mathbf{u}$, de tirer son contact longue-distance de $\mathbf{u}$ uniformément dans sa boule $\mathcal{B}_{\mathbf{v}}(3\ell)$. Un tel noeud $\mathbf{v}$ est appelé le *leader de niveau* i pour $\mathbf{u}$. Seuls les *leaders* de niveau i auront à explorer leur boule de rayon $3 \cdot 2^i$; tous les autres noeuds se référeront à eux pour calculer leur contact longue-distance.

Pour permettre un tel procédé, chaque noeud doit avoir un *leader* de niveau i à distance inférieure à 2^i, pour tous les niveaux i. Cela garantit que le tirage effectué par le *leader* est une bonne approximation du tirage qu'aurait effectué le noeud dans sa propre boule de rayon $3 \cdot 2^i$. Cela impose donc une densité suffisante des *leaders* de niveau i par rapport aux autres noeuds. D'autre part, pour éviter de surcharger le réseau, nous devons garantir que l'ensemble des *leaders* n'est pas trop dense. En effet, chaque *leader* $\mathbf{v}$ de niveau i va explorer $b_{\mathbf{v}}(3 \cdot 2^i)$ noeuds, nous souhaitons donc garantir que la taille totale de l'exploration induite par ces noeuds n'excède pas $O(n \log n)$ au total.

La sélection des *leaders* de niveau i est effectuée par l'algorithme d'échantillonnage aléatoire LEADER (Algorithme 6, p. 140). Nous décrivons à présent son fonctionnement et nous montrons qu'il effectue bien un échantillonnage avec la probabilité désirée, c'est-à-dire que chaque noeud $\mathbf{u}$ est inséré dans l'ensemble des *leaders* de niveau i avec probabilité $(2c \log n)/b_{\mathbf{u}}(2^i)$,

Algorithme 5 PETIT-MONDISATION DISTRIBUÉE

Dans notre schéma, les leaders *de niveau i sont en charge du tirage des contacts longue-distance à distance $O(2^i)$.*

1. Soit $k_0 = \lceil \log(2c \log n) \rceil$.
   Pour tout $i \in \{k_0, \ldots, \lceil \log D \rceil\}$, chaque noeud $\mathbf{u}$ est sélectionné pour être *leader* de niveau i avec probabilité $(2c \log n)/b_{\mathbf{u}}(2^i)$, indépendamment ; cette étape est effectuée en exécutant, pour tout $i \in \{k_0, \ldots, \lceil \log D \rceil\}$, l'algorithme LEADER$(\mathbf{u}, i)$ (Algorithme 6) en parallèle sur tous les noeuds $\mathbf{u} \in V$.
2. Puis, chaque *leader* de niveau i informe tous les noeuds, à distance inférieure ou égale à 2^i, qu'il est leur *leader* de niveau i (les cas d'égalités sont rompus arbitrairement).
3. S'il existe un noeud $\mathbf{u}$ qui n'a pas reçu de *leader* de niveau i, pour un niveau i, retourner au pas 1 (on dit alors que la phase d'échantillonnage a *échoué*.).
4. Chaque *leader* de niveau i explore sa boule de rayon $3 \cdot 2^i$.
5. Chaque noeud $\mathbf{u}$ du réseau choisit alors uniformément i dans $\{k_0, \ldots, \lceil \log D \rceil\}$ et fait une requête de contact longue-distance à son *leader* $\mathbf{v}$ de niveau i ; $\mathbf{v}$ choisit alors uniformément le contact longue-distance de $\mathbf{u}$ uniformément dans sa boule $\mathcal{B}_{\mathbf{v}}(3 \cdot 2^i)$.

indépendamment.

Sélection des *leaders* de niveau i (Algorithme 6). L'algorithme LEADER$(\mathbf{u}, i)$ décide si un noeud $\mathbf{u}$ est un *leader* de niveau i. Une fois exécuté sur tous les noeuds, il assure qu'avec forte probabilité, tout noeud est à distance inférieure à 2^i d'un *leader* de niveau i (Lemme 6.2). On note S_i l'ensemble aléatoire des *leaders* de niveau i.

Le lemme suivant montre que l'algorithme LEADER offre bien la probabilité de sélection désirée pour l'étape 1 de l'algorithme 5.

Lemme 6.1 *Pour tout $i \in \{k_0, \ldots, \lceil \log D \rceil\}$ et pour tout noeud $\mathbf{u} \in V$, $\Pr\{\mathbf{u} \in S_i\} = (2c \log n)/b_{\mathbf{u}}(2^i)$ et les tirages sont indépendants.*

Preuve. Tout d'abord, l'indépendance des tirages est assurée par le fait que chaque noeud $\mathbf{u}$ exécute sa procédure LEADER$(\mathbf{u}, i)$indépendamment des autres noeuds du réseau.

Soit $i \in \{k_0, \ldots, \lceil \log D \rceil\}$ et $\mathbf{u} \in V$. Le noeud $\mathbf{u}$ est inséré dans S_i si et seulement s'il a réussi tous les tirages aléatoires jusqu'à $k = i$. $\mathbf{u}$ survit au pre-

Algorithme 6 LEADER$(\mathbf{u}, i)$

L'algorithme LEADER$(\mathbf{u}, i)$ *procède par des phases d'inondation successives de rayons croissants.*

$k_0 := \lceil \log(2c \log n) \rceil$.
$k := k_0 + 1$.
Explorer la boule $\mathcal{B}_{\mathbf{u}}(2^{k_0})$.
$\mathbf{u}$ meurt avec probabilité $1 - 2c \log n / b_{\mathbf{u}}(2^{k_0})$.
tant que $\mathbf{u}$ est vivant et $k \leqslant i$ **faire**
    Explorer $\mathcal{B}_{\mathbf{u}}(2^k)$.
    $\mathbf{u}$ meurt avec probabilité $1 - b_{\mathbf{u}}(2^{k-1})/b_{\mathbf{u}}(2^k)$.
    $k := k + 1$.
fin tant que
si $\mathbf{u}$ est vivant **alors**
    $\mathbf{u}$ est un *leader* de niveau i.
fin si

mier tirage avec probabilité $2c \log n / b_{\mathbf{u}}(2^{k_0})$. Il survit au suivant avec probabilité $b_{\mathbf{u}}(2^{k_0})/b_{\mathbf{u}}(2^{k_0+1})$, etc. Le noeud $\mathbf{u}$ est *vivant* à la fin de l'exécution s'il a réussi tous les tirages indépendants jusqu'à la valeur $k = i$, soit avec une probabilité :

$$\frac{2c \log n}{b_{\mathbf{u}}(2^{k_0})} \times \frac{b_{\mathbf{u}}(2^{k_0})}{b_{\mathbf{u}}(2^{k_0+1})} \times \cdots \times \frac{b_{\mathbf{u}}(2^{i-1})}{b_{\mathbf{u}}(2^i)} = \frac{2c \log n}{b_{\mathbf{u}}(2^i)}.$$

□

Le lemme suivant montre que la phase d'échantillonnage réussit avec forte probabilité.

Lemme 6.2 *La probabilité que la phase d'échantillonnage de l'algorithme 5 échoue est inférieure à* $\log n / n$.

Preuve. Soit un noeud $\mathbf{u}$ et soit $i \in \{\lceil \log(2c \log n) \rceil, \ldots, \lceil \log D \rceil\}$. La probabilité qu'un noeud $\mathbf{v}$ appartienne à S_i est $2c \log n / b_{\mathbf{v}}(2^i)$. Par inclusion, tout noeud de $\mathbf{v}$ dans $\mathcal{B}_{\mathbf{u}}(2^i)$ satisfait : $b_{\mathbf{v}}(2^i) \leqslant b_{\mathbf{u}}(2^{i+1})$. Par hypothèse de croissance modérée, on a $b_{\mathbf{u}}(2^{i+1}) \leqslant c\, b_{\mathbf{u}}(2^i)$. La probabilité qu'aucun noeud de S_i ne soit dans $\mathcal{B}_{\mathbf{u}}(2^i)$ est donc inférieure à :

$$\left(1 - \frac{2c \log n}{c\, b_{\mathbf{u}}(2^i)}\right)^{b_{\mathbf{u}}(2^i)} \leqslant \frac{1}{n^2}.$$

Par la borne de l'union, la probabilité que cela arrive pour au moins un niveau i est donc inférieure à $\log D / n^2$ (puisqu'il y a moins de $\log D$ niveaux). Finalement,

par la borne de l'union une nouvelle fois, la probabilité que cet échec se produise pour au moins un noeud est inférieure à $\log D/n \leqslant \log n/n$. □

Deviner le taux d'expansion c. On remarquera que la connaissance de la valeur exacte du taux d'expansion c n'est en réalité pas requise pour exécuter les algorithmes LEADER et PETIT-MONDISATION DISTRIBUÉE. En effet, on peut initialiser l'exécution en choisissant $c = 1$ et, si la phase d'échantillonnage échoue, on ré-exécute l'algorithme avec $c := 2c$ jusqu'à ce que l'algorithme réussisse. D'après le lemme 6.2, avec forte probabilité, seulement $\lceil \log c \rceil$ exécutions sont nécessaires avant d'obtenir un succès. Par conséquent, nous supposons à présent c connu.

6.3.2 Analyse de l'algorithme de routage glouton dans le graphe augmenté

Nous analysons l'algorithme de routage glouton dans un réseau de croissance bornée augmenté par l'algorithme PETIT-MONDISATION DISTRIBUÉE, *sous l'hypothèse que la phase d'échantillonnage a réussi*. Nous savons d'après le lemme 6.2 que cela se produit avec forte probabilité. En démontrant que l'algorithme glouton calcule des chemins de longueur polylogarithmique en espérance, nous montrons que le graphe augmenté est bien un petit monde navigable. Nous commençons par minorer, dans le lemme suivant, la probabilité qu'un noeud $\mathbf{v}$ soit le contact longue-distance, $L(\mathbf{u})$, d'un noeud donné $\mathbf{u}$, en fonction de leur distance mutuelle $\delta(\mathbf{u}, \mathbf{v})$ dans le graphe d'origine.

Lemme 6.3 *Pour tout $\mathbf{v} \neq \mathbf{u}$ dans V tels que $\delta(\mathbf{u}, \mathbf{v}) \geqslant 2c \log n$,*

$$\Pr\{L(\mathbf{u}) = \mathbf{v}\} \geqslant \frac{1}{c^2 \log D} \frac{1}{b_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v}))}.$$

Preuve. Puisque $d(\mathbf{u}, \mathbf{v}) \geqslant 2c \log n$, soit $i \in \{k_0, \ldots, \lceil \log D \rceil\}$ tel que $2^i \leqslant d(\mathbf{u}, \mathbf{v}) < 2^{i+1}$. Notons $\mathrm{leader}(\mathbf{u}, i)$ le *leader* de niveau i de $\mathbf{u}$. On remarquera que $\delta(\mathbf{v}, \mathrm{leader}(\mathbf{u}, i)) < 3 \cdot 2^i$ par l'inégalité triangulaire. La probabilité que $L(\mathbf{u}) = \mathbf{v}$ est ainsi supérieure à la probabilité que $\mathbf{u}$ ait choisi le niveau i et que $\mathbf{v}$ ait ensuite été choisi uniformément parmi les noeuds de $\mathcal{B}_{\mathrm{leader}(\mathbf{u},i)}(3 \cdot 2^i)$. Le niveau i est choisi avec probabilité au moins $1/\log D$. Par définition du *leader* de niveau i, on a $\delta(\mathbf{u}, \mathrm{leader}(\mathbf{u}, i)) \leqslant 2^i$, on obtient :

$$b_{\mathrm{leader}(\mathbf{u},i)}(3 \cdot 2^i) \leqslant b_{\mathbf{u}}(4 \cdot 2^i) \leqslant b_{\mathbf{u}}(4\delta(\mathbf{u}, \mathbf{v})).$$

Or, par la croissance modérée, $b_{\mathbf{u}}(4\delta(\mathbf{u}, \mathbf{v})) \leqslant c^2 b_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v}))$, le résultat en découle. □

Théorème 6.4 (Petit-mondisation distribuée)
Dans tout réseau de croissance bornée, augmenté par l'algorithme PETIT-MONDISATION DISTRIBUÉE, *l'algorithme de routage glouton calcule un chemin de longueur* $O(\log D \log m + \log n)$ *en espérance, entre toute paire de noeuds à distance mutuelle* m *dans le réseau de départ.*

Preuve. Soient $\mathbf{s}$ et $\mathbf{t}$ la source et la cible du routage, à distance mutuelle m. On divise l'exécution de l'algorithme glouton en $\lceil \log m \rceil$ phases : l'algorithme est en phase j, $1 \leqslant j \leqslant \lceil \log m \rceil$, tant que la distance courante à la destination $\mathbf{t}$ appartient à $(2^{j-1}, 2^j]$.

Supposons que l'algorithme est en phase j et que $\mathbf{u}$ est le porteur courant du message. Supposons $j \geqslant k_0 + 2$, tout noeud $\mathbf{v}$ de $\mathcal{B}_{\mathbf{t}}(2^{j-2})$ est à distance au moins $2^{j-2} \geqslant 2c \log n$ de $\mathbf{u}$. D'après le lemme 6.3, tout noeud $\mathbf{v}$ dans $\mathcal{B}_{\mathbf{t}}(2^{j-2})$ est le contact longue-distance de $\mathbf{u}$ avec probabilité au moins $1/(c^2 \log D \cdot b_{\mathbf{u}}(\delta(\mathbf{u}, \mathbf{v})))$; de plus $\delta(\mathbf{u}, \mathbf{v}) \leqslant 5 \cdot 2^{j-2}$. Ainsi, le contact longue-distance de $\mathbf{u}$ appartient à $\mathcal{B}_{\mathbf{t}}(2^{j-2})$ avec probabilité supérieure à :

$$\frac{1}{c^2 \log D} \frac{b_{\mathbf{t}}(2^{j-2})}{b_{\mathbf{u}}(5 \cdot 2^{j-2})},$$

Par inclusion, $b_{\mathbf{u}}(5 \cdot 2^{j-2}) \leqslant b_{\mathbf{t}}(10 \cdot 2^{j-2})$ et par croissance modérée, $b_{\mathbf{t}}(10 \cdot 2^{j-2}) \leqslant c^4 b_{\mathbf{t}}(2^{j-2})$. La probabilité de sortir de la phase j est donc supérieure à $1/(c^6 \log D)$ à chaque pas de cette phase. L'espérance du nombre de pas dans chaque phase $j \geqslant k_0 + 2$ est donc $O(\log D)$. Dès que $j < k_0 + 2$, on borne simplement l'espérance du nombre de pas restant par $2^{k_0+1} = O(\log n)$. Finalement, en sommant sur les $\lceil \log m \rceil$ phases, on obtient une longueur de chemin totale de $O(\log D \log m + \log n)$ en espérance. □

6.4 Implémentations décentralisées et performances

Les seules étapes de l'algorithme PETIT-MONDISATION DISTRIBUÉE qui nécessitent une implémentation distribuée non triviale sont l'étape d'exploration de LEADER et la génération aléatoire des contacts longue-distance. Pour l'étape d'exploration du *leader*, nous définissons comment enregistrer les explorations successives sur des rayons croissants à l'aide d'arbres de plus courts chemins. La génération des contacts longue-distance consiste à router chaque requête au *leader* approprié, à choisir un noeud aléatoire dans la boule du *leader* et à renvoyer le contact au noeud demandeur.

6.4.1 L'étape d'exploration de LEADER

Cette procédure explore des boules de plus en plus grandes autour de chaque noeud. Chaque appel de LEADER$(\mathbf{u}, i)$ lance les explorations successives des boules $\mathcal{B}_{\mathbf{u}}(2^{k_0}), \ldots, \mathcal{B}_{\mathbf{u}}(2^i)$, de rayon croissant. Si $\mathbf{u}$ survit à tous les tests probabilistes, il explore la boule $\mathcal{B}_{\mathbf{u}}(3 \cdot 2^i)$ qui sera utilisée, par la suite, pour le tirage des contacts longue-distance.

Les boules sont explorées selon un parcours en profondeur. Pour simplifier la transmission des messages à l'intérieur de chaque boule explorée, on construit un arbre des plus courts chemins $\mathcal{T}_{\mathbf{u},i}$ au cours de l'exploration. On maintient la structure d'arbre en stockant, en chaque noeud, sa mère et ses filles (les cas d'égalité sont rompus arbitrairement). Si $\mathbf{u}$ meurt au $k$ième test probabiliste, $\mathcal{T}_{\mathbf{u},i}$ couvre la boule $\mathcal{B}_{\mathbf{u}}(2^k)$. Si $\mathbf{u}$ survit à tous les tests probabilistes, il informe tous les noeuds de $\mathcal{B}_{\mathbf{u}}(2^i)$, en utilisant la structure d'arbre $\mathcal{T}_{\mathbf{u},i}$, qu'il est leur *leader* de niveau i. $\mathcal{T}_{\mathbf{u},i}$ couvre alors la boule $\mathcal{B}_{\mathbf{u}}(3 \cdot 2^i)$ et $\mathbf{u}$ demande à chacune de ses filles de calculer et d'enregistrer récursivement les tailles de leurs sous-arbres dans $\mathcal{T}_{\mathbf{u},i}$. Ces tailles de sous-arbre seront utilisées par la suite pour tirer un noeud uniformément dans la boule $\mathcal{B}_{\mathbf{u}}(3 \cdot 2^i)$ sur demande.

Le lemme suivant précise l'analyse de notre échantillon en montrant que l'ensemble des *leaders* n'est pas trop dense. Cela sera utile pour l'analyse des implémentations.

Lemme 6.5 *Avec forte probabilité, tout noeud* $\mathbf{u}$ *appartient à* $O(\log n \log D)$ *arbres* $\mathcal{T} \in \{\mathcal{T}_{\mathbf{v},i}\}_{\mathbf{v} \in V, i \leqslant \lceil \log D \rceil}$.

Preuve. Soit $\mathbf{u} \in V$ et $k \geqslant k_0$. Pour tout noeud $\mathbf{v}$ et $j \geqslant k$, la probabilité que $\mathbf{v}$ explore au moins jusqu'au rayon 2^k, durant la phase de sélection des *leaders* de niveau i, est $2c \log n / b_{\mathbf{v}}(2^k)$. Soit $X_{\mathbf{v},j}$ la variable aléatoire égale à 1 si $\mathbf{v}$ explore un rayon au moins 2^k durant la sélection des *leaders* de niveau i et 0 sinon. On a : $\mathbb{E}[X_{\mathbf{v},j}] = 2c \log n / b_{\mathbf{v}}(2^k)$, si $j \geqslant k$ et $\mathbb{E}[X_{\mathbf{v},j}] = 0$ sinon.

Le noeud $\mathbf{u}$ appartient à tous les arbres enracinés en un sommet $\mathbf{v}$ de $\mathcal{B}_{\mathbf{u}}(2^k)$ qui ont exploré jusqu'au rayon 2^k. Pour tous les noeuds $\mathbf{v}$ de $\mathcal{B}_{\mathbf{u}}(2^k)$, on a $b_{\mathbf{v}}(2^k) \leqslant b_{\mathbf{u}}(2^{k+1}) \leqslant c\, b_{\mathbf{u}}(2^k)$ par la croissance modérée. De plus, $b_{\mathbf{v}}(2^k) \geqslant b_{\mathbf{v}}(2^{k+1})/c$ par la croissance modérée. Ainsi, pour tous les noeuds $\mathbf{v}$ de $\mathcal{B}_{\mathbf{u}}(2^k)$,

$$\mathbb{E}[X_{\mathbf{v},j}] \in \left[\frac{2 \log n}{b_{\mathbf{u}}(2^k)}, \frac{2c^2 \log n}{b_{\mathbf{u}}(2^k)}\right].$$

Soit $X = \sum_{\mathbf{v} \in \mathcal{B}_{\mathbf{u}}(2^k)} \sum_j X_{\mathbf{v},j}$, on a $\mathbb{E}[\sum_j X_{\mathbf{v},j}] \in [2 \log n, 2c^2 \log n \log D]$ par linéarité de l'espérance. Puisque les variables $X_{\mathbf{v},j}$ sont indépendantes, on peut utiliser une borne de Chernoff (voir par exemple [HMRe98]) pour majorer la pro-

babilité que leur somme soit supérieure à $4\mathbb{E}[X]$:

$$\Pr\{X > 4\mathbb{E}[X]\} \leqslant \left[\frac{1}{e}(e/4)^4\right]^{\mathbb{E}[X]} \leqslant \frac{1}{e^{2\log n}} \leqslant \frac{1}{n^2}.$$

Ainsi, avec probabilité supérieure à $1-1/n^2$, il y a moins de $O(\log n \log D)$ arbres $\mathcal{T}$ enracinés en $\mathcal{B}_{\mathbf{u}}(2^k)$ qui traversent $\mathbf{u}$. On remarquera que l'on peut conduire le même argument à toutes les échelles $k \leqslant \log D$. Par la borne de l'union, la probabilité qu'il existe k tel qu'il y ait plus de $\log n$ arbres de profondeur supérieur ou égale à 2^k, enracinés en $\mathcal{B}_{\mathbf{u}}(2^k)$, est inférieure à $\log D/n^2$. Ainsi, avec probabilité supérieure à $1 - \log D/n^2$, il y a moins de $O(\log n \log D)$ arbres qui traversent $\mathbf{u}$. En appliquant une nouvelle fois la borne de l'union, on obtient que cette propriété est vraie pour tous les noeuds $\mathbf{u} \in V$ avec probabilité supérieure à $1 - \log D/n$. □

Proposition 6.6 *La phase d'échantillonnage de l'algorithme* PETIT-MONDISATION DISTRIBUÉE *(algorithme 5) requiert :*

- $O(D)$ *rondes,*
- *l'échange de* $O(n\Delta \log n(\log D)^2)$ *messages au total en espérance, et de* $O(\Delta \log n \log D)$ *messages par ronde et par noeud avec forte probabilité,*
- *une mémoire de taille* $O(\Delta \log n \log D)$ *dans chaque noeud avec forte probabilité.*

Preuve. Le nombre de rondes nécessaires pour exécuter LEADER$(\mathbf{u}, i)$ est au plus $6 \cdot 2^i + 1$ dans le cas où les tirages aléatoires ont réussi jusqu'au rayon $3 \cdot 2^i$, puisque les messages doivent descendre dans les feuilles selon ce rayon, puis être renvoyés à la racine. Un noeud $\mathbf{u}$ donné exécute LEADER$(\mathbf{u}, i)$ pour i de k_0 à $\lceil \log D \rceil$, ce qui nécessite donc au plus $O(D)$ rondes. Comme les exécutions se font en parallèle, le nombre total de rondes est $O(D)$.

Pendant l'exécution de LEADER$(\mathbf{u}, i)$, chaque noeud interne de l'arbre d'exploration envoie au plus Δ messages, au plus 3 fois. Avec probabilité inférieure à $1/b_{\mathbf{u}}(2^k)$, l'arbre d'exploration se termine au rayon 2^k et contient $b_{\mathbf{u}}(2^k)$ noeuds, ce qui induit $O(\Delta b_{\mathbf{u}}(2^k))$ messages. L'espérance du nombre de message envoyés par $\mathbf{u}$ pour le niveau i est donc $O(i\Delta)$ et donc $O(\Delta(\log D)^2)$ au total pour $\mathbf{u}$, pour tous les niveaux. Par linéarité de l'espérance, l'espérance du nombre total de messages est $O(n\Delta(\log D)^2)$. Par ailleurs, durant une ronde, chaque noeud $\mathbf{v}$ envoie au plus Δ messages par arbre $\mathcal{T}$ dans $\{\mathcal{T}_{\mathbf{u},i}\}_{\mathbf{u}\in V, i\leqslant\lceil \log D\rceil}$ qui le traverse. D'après le lemme 6.5, chaque noeud est traversé par au plus $O(\log n \log D)$ tels arbres avec forte probabilité, le nombre de messages par ronde et par noeud est donc $O(\Delta \log n \log D)$ avec forte probabilité.

Enfin, un noeud $\mathbf{u}$ doit mémoriser suffisamment d'information pour permettre l'exécution d'un parcours en profondeur pour chaque arbre qui le traverse durant

l'exécution de LEADER$(\mathbf{v}, i)$ sur tous les noeuds $\mathbf{v}$ et tous les niveaux i. Cela nécessite une mémoire de taille $O(\Delta)$ pour chaque arbre. D'après le lemme 6.5, tout noeud est traversé par au plus $O(\log n \log D)$ arbres avec forte probabilité, le résultat en découle. □

6.4.2 L'étape des requêtes de contact longue-distance

Durant cette étape, chaque noeud $\mathbf{u}$ choisit $i \in \{k_0, \ldots, \lceil \log D \rceil\}$ de façon aléatoire uniforme et fait une requête de contact longue-distance à son *leader* de niveau i. Nous proposons deux implémentations pour cette étape, en fonction de la taille mémoire disponible en chaque noeud.

Schéma (a) : minimisation du nombre de messages. Dans cette première implémentation, nous cherchons à minimiser le nombre total de messages qui transitent dans le réseau, en supposant que les noeuds disposent d'une mémoire importante, de taille $O(n)$. Les requêtes envoyées à chaque *leader* $\mathbf{u}$ de niveau i sont routées de façon synchrone depuis les feuilles de $\mathcal{T}_{\mathbf{u},i}$ vers $\mathbf{u}$, le long de $\mathcal{T}_{\mathbf{u},i}$. Chaque noeud interne de l'arbre agrège les requêtes qu'il reçoit de ses filles en un unique message, puis le retransmet à sa mère accompagné de sa propre requête (éventuelle). Soit x le nombre de requêtes pour $\mathbf{u}$ dans tout le réseau, en tant que *leader* de niveau i. Une fois que $\mathbf{u}$ a reçu toutes ses requêtes (après 2^i rondes), $\mathbf{u}$ choisi x entiers $q_1, \ldots, q_x$ aléatoires dans $\{1, \ldots, b_{\mathbf{u}}(3 \cdot 2^i)\}$ de façon uniforme et lance les recherches en ordre préfixe dans $\mathcal{T}_{\mathbf{u},i}$ des q_1-ème, $\ldots$, q_x-ème noeuds $\mathbf{v}_1, \ldots, \mathbf{v}_x$ (en utilisant les tailles des sous-arbres enregistrées en chaque noeud). Les contacts longue-distance $\mathbf{v}_1, \ldots, \mathbf{v}_x$ sont alors renvoyés au *leader* $\mathbf{u}$ (en utilisant le même schéma d'agrégation de messages qu'avant), puis renvoyés à leur destination : le noeud qui en a fait la requête (en utilisant l'oracle, ou simplement le numéro de la destination dans l'ordre préfixe de $\mathcal{T}_{\mathbf{u},i}$, comme avant).

Proposition 6.7 *Le schéma **(a)** requiert :*
- *$O(D)$ rondes,*
- *$O(\log n \log D)$ messages par noeud et par ronde avec forte probabilité, et un nombre total de messages $O(n \log n \log D)$ en espérance,*
- *une mémoire de taille $O(n)$ en chaque noeud.*

Preuve. Dans ce schéma, les *leaders* de niveau $\log D$ peuvent avoir jusqu'à n requêtes à enregistrer, dans le pire cas où tous les noeuds ont choisi ce niveau. Cela nécessite donc une taille mémoire disponible linéaire.

Le nombre de rondes nécessaires à l'exécution de l'étape des requêtes de contacts longue-distance est au plus quatre fois le nombre de rondes nécessaires pour exécuter le plus grand parcours en profondeur, c'est-à-dire $O(D)$.

Concernant le nombre de messages, chaque arbre $\mathcal{T}_{\mathbf{u},i}$ enraciné en un *leader* $\mathbf{u}$ de niveau i génère au plus $4b_{\mathbf{u}}(2^i)$ messages. Pour évaluer le nombre total de messages générés, on peut considérer que chaque *leader* $\mathbf{u}$ de niveau i induit au plus $4b_{\mathbf{u}(2^i)}$ messages et que les autres noeuds n'en induisent pas. L'espérance du nombre total de messages générés durant l'exécution est donc inférieur à :

$$\sum_{\mathbf{u}\in V}\sum_{i=k_0+2}^{\log D} 4b_{\mathbf{u}}(2^i)\Pr\{\mathbf{u}\in S_i\} = O(n\log n\log D).$$

Par ailleurs, chaque noeud $\mathbf{v}$ envoie $O(1)$ messages par ronde pour chaque arbre dans $\{\mathcal{T}_{\mathbf{v},i}\}_{\mathbf{v}\in V, i\leqslant\lceil\log D\rceil}$ auquel il appartient. D'après le lemme 6.5, $\mathbf{v}$ appartient à au plus $O(\log n\log D)$ tels arbres avec forte probabilité. Le nombre de messages par noeud et par ronde est donc $O(\log n\log D)$ avec forte probabilité. □

Schéma (b) : taille mémoire polylogarithmique. Dans le schéma précédent, les *leaders* de haut niveau sont les goulots d'étranglement du réseau et nécessitent une taille mémoire disponible qui peut n'être pas réaliste pour de nombreuses applications. Nous allons à présent décrire un schéma qui ne requiert qu'une mémoire polylogarithmique en n. On remarquera tout d'abord qu'on ne peut pas parvenir à une mémoire polylogarithmique en chaque noeud si le nombre de rondes est $o(n)$. En effet, si l'on suppose que le schéma s'exécute en R rondes, même si les requêtes aux *leaders* sont soigneusement ordonnancées, il existe au moins une ronde durant laquelle un *leader* du niveau maximal reçoit n/R requêtes provenant de n/R noeuds. Pour parvenir à une mémoire de taille $O(\log^a n)$ *en chaque noeud*, pour une constante $a > 0$, R doit donc être de l'ordre de $\Omega(n/\log^a n)$. Nous présentons ici une nouvelle implémentation de l'étape des requêtes de contacts longue-distance qui permet d'effectuer la *petit-mondisation* en $O(n)$ rondes avec une mémoire polylogarithmique pour tous les noeuds, avec forte probabilité.

Nous montrons qu'il est possible d'ordonnancer les requêtes de contacts longue-distance à l'intérieur d'un arbre enraciné en un *leader* de façon à ce que le *leader* ne reçoive qu'une nouvelle requête par ronde. Pour cela, chaque *leader* $\mathbf{u}$ de niveau i crée un jeton $\mathrm{OK}_{\mathbf{u},i}$ qui traverse la boule $\mathcal{B}_{\mathbf{u}}(2^i)$ le long de l'arbre $\mathcal{T}_{\mathbf{u},i}$, c'est-à-dire qu'à chaque ronde, le jeton passe au noeud suivant dans l'ordre préfixe. À chaque fois qu'un noeud $\mathbf{v}$ reçoit le jeton, il envoie requête (éventuelle) de contact longue-distance pour $\mathbf{u}$ à sa mère dans $\mathcal{T}_{\mathbf{u},i}$. Par ailleurs, tout noeud qui reçoit une requête de contact longue-distance pour $\mathbf{u}$ la retransmet à sa mère dans $\mathcal{T}_{\mathbf{u},i}$. Cet ordonnancement est illustré sur la figure 6.4.2. À chaque fois que $\mathbf{u}$ reçoit une requête de contact longue-distance, il choisit un entier aléatoire q dans $\{1,\ldots,b_{\mathbf{u}}(3\cdot 2^i)\}$ de façon uniforme, puis lance une recherche préfixe du noeud $\mathbf{v}_q$ correspondant au nombre q dans $\mathcal{B}_{\mathbf{u}}(3\cdot 2^i)$, le long de $\mathcal{T}_{\mathbf{u},i}$. Le noeud $\mathbf{v}_q$ enregistre tous les identifiants des noeuds *demandeurs*, c'est-à-dire, les noeuds qui

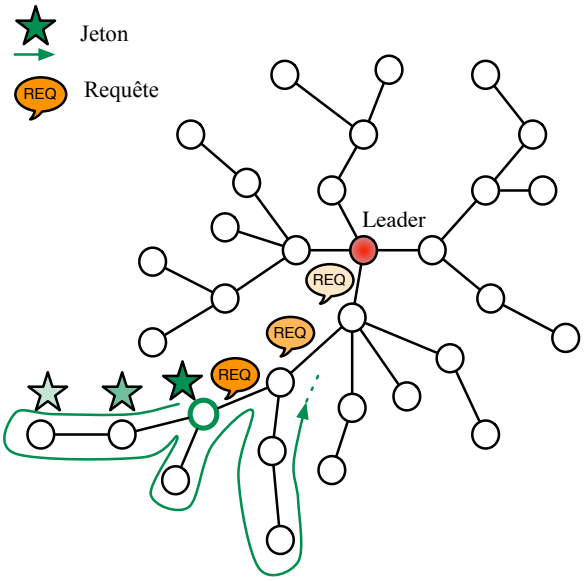


FIG. 6.1 : Illustration de l'ordonnancement des requêtes vers un *leader*.

l'ont demandé comme contact longue-distance. Lorsque le jeton $\mathrm{OK}_{\mathbf{u},i}$ est revenu à $\mathbf{u}$, $\mathbf{u}$ crée un nouveau jeton $\mathrm{GO}_{\mathbf{u},i}$ qui traverse $\mathcal{B}_{\mathbf{u}}(3 \cdot 2^i)$ dans l'ordre préfixe le long de $\mathcal{T}_{\mathbf{u},i}$. À chaque fois qu'un noeud $\mathbf{v}$ reçoit ce nouveau jeton, il renvoie son identifiant à tous ses noeuds *demandeurs*, un par un à chaque ronde, jusqu'à ce qu'il n'y en ait plus. Il passe ensuite le jeton au prochain noeud dans l'ordre préfixe.

Cet algorithme garantit que, dans $\mathcal{T}_{\mathbf{u},i}$, le nombre de messages envoyés simultanément par un noeud est majoré par une constante. D'après le lemme 6.5, aucun noeud n'appartient à plus de $O(\log n \log D)$ tels arbres, avec forte probabilité. La mémoire disponible permet aux noeuds de stocker ces $O(\log n \log D)$ messages.

Le lemme suivant garantit que le nombre d'identifiants de noeuds *demandeurs* enregistrés dans un noeud est au plus $O(\log n)$ et tient donc dans la mémoire de taille polylogarithmique que nous nous sommes fixée. Il s'agit en fait d'un résultat classique de probabilités sur les tirages dans les urnes, puisque le choix de k contacts longue-distance par k noeuds demandeurs est équivalent à k tirages aléatoires uniformes indépendants dans un même ensemble.

Lemme 6.8 *Tout noeud* $\mathbf{v}$ *a au plus* $8c^2 \log n$ *noeuds* demandeurs*, avec forte probabilité.*

Preuve. Considérons un noeud $\mathbf{v}$. Soit $X_{\text{dem}}(\mathbf{v})$ la variable aléatoire du nombre de noeuds *demandeurs* de $\mathbf{v}$ dans tout le réseau. Le nombre total de demandeurs

est la somme des demandeurs de $\mathbf{v}$ sur chaque arbre $\mathcal{T}_{\mathbf{u},i}$ tel que $\delta(\mathbf{u},\mathbf{v}) \leqslant 3 \cdot 2^i$ et $\mathbf{u}$ est un *leader* de niveau i. Nous allons manipuler une variable aléatoire $X^\star_{\text{dem}}(\mathbf{v})$ qui domine stochastiquement $X_{\text{dem}}(\mathbf{v})$. On se place dans un cadre virtuel où, pour chaque noeud $\mathbf{u}$ et chaque niveau i, tous les noeuds $\mathbf{w}$ de $\mathcal{B}_{\mathbf{u}}(2^i)$ tirent deux nombres aléatoirement uniformément : l'un entre k_0+2 et $\lceil \log D \rceil$ et l'autre entre 1 et $b_{\mathbf{u}}(3 \cdot 2^i)$. Pour un noeud $\mathbf{u}$ et un niveau i, on note, pour tout $\mathbf{w} \in \mathcal{B}_{\mathbf{u}}(2^i)$, $\mathbb{1}_{\mathbf{u},i,\mathbf{w}}$ la variable aléatoire indicatrice qui vaut 1 si $\mathbf{u}$ est un leader de niveau i et $\mathbf{w}$ a tiré le couple de nombres $(i, v_{\mathbf{u},i})$, où $v_{\mathbf{u},i}$ est le numéro de $\mathbf{v}$ dans l'ordre préfixe dans $\mathcal{B}_{\mathbf{u}}(3 \cdot 2^i)$ (par exemple). Alors, la variable aléatoire $X^\star_{\text{dem}}(\mathbf{v})$ définie par :

$$X^\star_{\text{dem}}(\mathbf{v}) = \sum_{i=k_0+2}^{\lceil \log D \rceil} \sum_{\mathbf{u} \in \mathcal{B}_{\mathbf{v}}(3 \cdot 2^i)} \sum_{\mathbf{w} \in \mathcal{B}_{\mathbf{u}}(2^i)} \mathbb{1}_{\mathbf{u},i,\mathbf{w}},$$

domine stochastiquement $X_{\text{dem}}(\mathbf{v})$. En effet, dans le cadre réel, un noeud ne tire un contact que pour un seul *leader*, alors que dans ce cadre virtuel, en faisant tirer une paire de nombres par tous les noeuds associés à un leader, on augmente leur probabilité de tirer le numéro $v_{\mathbf{u},i}$, qui correspond au tirage uniforme du noeud $\mathbf{v}$. On remarquera que :

$$\Pr\{\mathbb{1}_{\mathbf{u},i,\mathbf{w}} = 1\} = \frac{2c \log n}{b_{\mathbf{u}}(2^i)} \frac{1}{\lceil \log D \rceil - k_0 - 1} \frac{1}{b_{\mathbf{u}}(3 \cdot 2^i)}.$$

Par ailleurs, pour tout $\mathbf{u} \in \mathcal{B}_{\mathbf{v}}(3 \cdot 2^i)$, $b_{\mathbf{v}}(6 \cdot 2^i)/c \leqslant b_{\mathbf{u}}(3 \cdot 2^i) \leqslant b_{\mathbf{v}}(6 \cdot 2^i)$, par inclusion et croissance modérée. Par conséquent, par linéarité de l'espérance,

$$2c \log n \frac{b_{\mathbf{v}}(3 \cdot 2^i)}{b_{\mathbf{v}}(6 \cdot 2^i)} \leqslant \mathbb{E}[X^\star_{\text{dem}}(\mathbf{v})] \leqslant 2c^2 \log n \frac{b_{\mathbf{v}}(3 \cdot 2^i)}{b_{\mathbf{v}}(6 \cdot 2^i)},$$

c'est-à-dire : $\mathbb{E}[X^\star_{\text{dem}}(\mathbf{v})] \in [2 \log n, 2c^2 \log n]$, en utilisant la croissance modérée. Comme les variables indicatrices sont bornées et indépendantes, nous pouvons utiliser une borne de Chernoff :

$$\Pr\{X^\star_{\text{dem}}(\mathbf{v}) > 4\mathbb{E}[X^\star_{\text{dem}}(\mathbf{v})]\} \leqslant [\frac{1}{e}(e/4)^4]^{\mathbb{E}[X^\star_{\text{dem}}(\mathbf{v})]} \leqslant \frac{1}{e^{2 \log n}} \leqslant \frac{1}{n^2}.$$

Ainsi le noeud $\mathbf{v}$ doit enregistrer moins de $8c^2 \log n$ noeuds *demandeurs* avec une probabilité supérieure à $1 - 1/n^2$, puisque $X^\star_{\text{dem}}(\mathbf{v})$ domine stochastiquement $X_{\text{dem}}(\mathbf{v})$. Par la borne de l'union sur tous les noeuds $\mathbf{v}$, chaque noeud doit enregistrer au plus $8c^2 \log n$ demandeurs avec forte probabilité (supérieure à $1 - 1/n$). □

Le lemme suivant borne le nombre de messages retransmis par un noeud lors d'une ronde.

Lemme 6.9 *Soit* $\mathbf{u}$ *un* leader *de niveau* i*. Pour toute ronde du schéma* ***(b)****, chaque noeud* $\mathbf{v} \in \mathcal{T}_{\mathbf{u},i}$ *envoie ou reçoit au plus* 3 *messages dans* $\mathcal{T}_{\mathbf{u},i}$*.*

Preuve. La transmission des messages est organisée par les jetons $\mathrm{OK}_{\mathbf{u},i}$ et $\mathrm{GO}_{\mathbf{u},i}$ qui sont situés sur un noeud, au plus, par ronde. La réception du jeton $\mathrm{OK}_{\mathbf{u},i}$ indique que le noeud courant est autorisé à envoyer ou recevoir sa requête vers $\mathbf{u}$ lors des prochaines rondes.

Soient $\mathbf{v}$ et $\mathbf{w}$ deux noeuds de $\mathcal{T}_{\mathbf{u},i}$ et supposons que $\mathbf{w}$ reçoit le jeton $\mathrm{OK}_{\mathbf{u},i}$ avant $\mathbf{v}$. Si $\mathbf{w}$ est un ancêtre de $\mathbf{v}$ dans $\mathcal{T}_{\mathbf{u},i}$, $\mathbf{w}$ sera autorisé à envoyer sa requête avant $\mathbf{v}$ et les deux requêtes ne peuvent donc pas se trouver dans le même noeud. Lorsque le jeton $\mathrm{OK}_{\mathbf{u},i}$ quitte un noeud $\mathbf{z}$, le jeton $\mathrm{OK}_{\mathbf{u},i}$ visite le sous-arbre de $\mathcal{T}_{\mathbf{u},i}$ enraciné en $\mathbf{z}$. Les requêtes retransmises par $\mathbf{z}$ ne peuvent donc pas se superposer aux requêtes suivantes.

Tant que le jeton $\mathrm{OK}_{\mathbf{u},i}$ n'est pas revenu à $\mathbf{u}$, l'étape de requête de contact longue-distance d'un noeud $\mathbf{w}$ peut être décomposée en deux étapes : (1) le routage vers le *leader* $\mathbf{u}$ et (2) le routage de $\mathbf{u}$ vers un noeud aléatoire $\mathbf{v}_x$. Comme $\mathbf{u}$ reçoit au plus une requête par ronde et comme le routage peut se faire suivant les plus courts chemins, chaque noeud de $\mathcal{T}_{\mathbf{u},i}$ reçoit au plus une requête de sa mère par ronde. Finalement, chaque noeud de $\mathcal{T}_{\mathbf{u},i}$ reçoit ou envoie au plus 3 messages par ronde. On peut conduire la même analyse pour le retour du contact longue-distance vers son demandeur, organisé par le jeton $\mathrm{GO}_{\mathbf{u},i}$. $\square$

Proposition 6.10 *Le schéma* ***(b)*** *requiert :*

- *au plus* $4n$ *rondes,*
- $O(\log n \log D)$ *messages par noeud et par ronde avec forte probabilité, et un nombre total de messages* $O(nD \log n)$ *en espérance,*
- *une mémoire de taille* $O(\Delta \log n \log D)$ *en chaque noeud avec forte probabilité.*

Preuve. Le nombre de rondes nécessaires pour exécuter le schéma **(b)** est la durée de l'ordonnancement des requêtes, organisé par les jetons. Cet ordonnancement est basé sur un parcours en profondeur. Pour chaque arbre $\mathcal{T}_{\mathbf{u},i}$, la traversée du jeton est exécutée en $b_{\mathbf{u}}(2^i)$ rondes. La première phase de l'étape des requêtes de contacts longue-distance est donc exécutée en $b_{\mathbf{u}}(2^i) + 2^i \leqslant 2b_{\mathbf{u}}(2^i)$ rondes. Pour la seconde phase (remontée des contacts), nous devons ajouter un délai supplémentaire dû au nombre de noeuds de $\mathcal{T}_{\mathbf{u},i}$ qui ont choisi $\mathbf{u}$ comme *leader* pour leur requête. Cette phase est exécutée en au plus $2b_{\mathbf{u}}(3 \cdot 2^i)$ rondes. L'ordonnancement est donc achevé en $2b_{\mathbf{u}}(2^i) + 2b_{\mathbf{u}}(3 \cdot 2^i) \leqslant 4n$ rondes pour tout *leader* $\mathbf{u}$ de tout niveau i.

D'après le lemme 6.9, chaque noeud $\mathbf{v}$ envoie ou reçoit au plus 3 par ronde pour chaque arbre $\{\mathcal{T}_{\mathbf{u},i}\}_{\mathbf{u} \in V, 1 \leqslant \lceil \log D \rceil}$ auquel il appartient. Or, d'après le

lemme 6.5, $\mathbf{v}$ appartient à au plus $O(\log n \log D)$ tels arbres avec forte probabilité. On en déduit que le nombre de messages par noeud et par ronde est $O(\log n \log D)$.

Nous nous intéressons à présent au nombre de messages engendrés dans un arbre $\mathcal{T}_{\mathbf{u},i}$. Comme chaque jeton fait un parcours en profondeur, il y a au plus $2b_{\mathbf{u}}(3 \cdot 2^i)$ messages dédiés aux jetons. Le nombre de requêtes provenant de $\mathcal{T}_{\mathbf{u}}(2^i)$ qui sont envoyées au *leader* $\mathbf{u}$ est au plus $b_{\mathbf{u}}(2^i)$. Chaque requête engendre au plus $4 \cdot 3 \cdot 2^i$ messages. Pour compter le nombre total de messages, on peut considérer que chaque *leader* de niveau i induit au plus $O(2^i b_{\mathbf{u}}(2^i))$ messages et que les autre noeuds n'en induisent pas. Soit $\mathbf{v} \in V$. La probabilité que $\mathbf{v}$ soit un *leader* de niveau i est $2c \log n / b_{\mathbf{v}}(2^i)$, l'espérance du nombre de messages induits par $\mathbf{v}$ est donc :

$$\sum_{i=k_0+2}^{\lceil \log D \rceil} O(2^i \log n)) = O(D \log n).$$

Par linéarité de l'espérance, l'espérance du nombre total de messages engendrés par le schéma **(b)** est donc $O(nD \log n)$.

Concernant la mémoire, d'après le lemme 6.5, chaque noeud requiert une taille mémoire $O(\Delta \log n \log D)$ avec forte probabilité pour stocker ses mères et ses filles pour tous les arbres qui le traversent. d'après le lemme 6.8, la mémoire nécessaire pour stocker les identifiants des noeuds demandeurs est au plus $8c^2 \log n$ avec forte probabilité, le résultat en découle. □

6.4.3 Performances

Nous pouvons à présent donner les performances globales de l'algorithme PETIT-MONDISATION DISTRIBUÉE en fonction de l'implémentation choisie pour l'étape des requêtes de contacts longue-distance. Il suffit pour cela de combiner les résultats des propriétés 6.6, 6.7 et 6.10.

Théorème 6.11
Tout réseau synchrone de n noeuds, à croissance bornée, de diamètre D et de degré maximum Δ peut être petit-mondisé *de façon distribuée, via l'ajout d'un lien par noeud,*

- *en $O(n)$ rondes, avec une taille mémoire $O(\Delta \log n \log D)$ en chaque noeud avec forte probabilité, $O(\Delta \log n \log D)$ messages par noeud et par ronde avec forte probabilité, et un nombre total de messages $O(nD \log n)$ en espérance ;*
- ou *en $O(D)$ rondes, avec une taille mémoire $O(n)$, $O(\Delta \log n \log D)$ messages par noeud et par ronde avec forte probabilité, et un nombre total de messages $O(n \log n \log D)$ en espérance.*

6.5 Conclusion et perspectives

Dans ce chapitre, nous avons présenté le premier schéma distribué, à notre connaissance, pour *petit-mondiser* une grande classe de graphes. Ce schéma est un premier pas vers la compréhension de l'émergence de l'effet petit monde via un processus d'ajout de raccourcis, propre à chaque noeud. On peut toutefois penser que la requête à un *leader* pour choisir une nouvelle connaissance n'est pas réaliste dans le cadre des réseaux sociaux. Une perspective intéressante serait de parvenir à répartir des contacts longue-distance sur toutes les échelles de distance sans utiliser de *leader* de haut niveau.

Notre schéma décentralisé s'applique en premier lieu dans le cadre d'un grand réseau informatique décentralisé. Il serait intéressant d'étendre notre schéma à un réseau dynamique, où l'on autorise l'ajout et le retrait de noeuds dans le temps, comme c'est le cas pour les réseaux pair-à-pair. On peut penser que tant que la propriété de croissance modérée est respectée par l'ajout et le retrait de noeuds, les mises à jour peuvent se faire localement. Lorsqu'un noeud disparaît, les liens longue-distance dont il était l'extrémité pourraient être re-dirigés vers un voisin, qui reprendrait également son statut de *leader* éventuel. Lorsqu'un nouveau noeud est inséré, il pourrait tirer son contact longue-distance à l'aide des *leaders* existants. Lorsque le nombre de noeuds insérés et retirés devient trop important, une nouvelle sélection des *leaders* serait sans doute nécessaire.

Enfin, par la répartition des contacts longue-distance sur toutes les échelles de distance, le réseau *petit-mondisé* doit présenter une bonne tolérance aux pannes qu'il serait intéressant d'analyser.

Conclusion générale

Au terme de ce manuscrit, nous avons pu isoler plusieurs composantes de l'effet petit monde. Ce sujet de recherche a été l'objet de nombreux travaux pendant la durée de cette thèse et nos résultats ont participé à la meilleure compréhension que nous avons, aujourd'hui, de ce phénomène.

Notre étude de l'aspect dynamique de l'effet petit monde nous a conduit à construire un nouvel algorithme de routage décentralisé, qui respecte les caractéristiques de la navigabilité dans les petits mondes, tout en calculant, dans le modèle de Kleinberg, des chemins de longueur asymptotiquement plus courte que les autres algorithmes existants, à notre connaissance. Ce résultat a permis de donner une nouvelle majoration $O(\log n(\log\log n)^2)$ du diamètre apparent du modèle, c'est-à-dire de l'espérance de la longueur du plus long des plus courts chemins calculables par un algorithme décentralisé, la majoration précédente étant $O(\log^2 n)$. Nous savons aujourd'hui que le diamètre du modèle est $\Theta(\log n)$ [MN04], mais la question du diamètre apparent reste ouverte. Ce résultat a également contribué à la réflexion sur le type d'algorithmes de routage utilisés dans les réseaux réels où l'on observe l'effet petit monde, comme les réseaux sociaux. Notre algorithme explore en effet les liens d'un voisinage sur une faible profondeur, ce qui semble une modélisation des comportements humains plus raisonnable qu'un simple algorithme glouton, dans le cadre d'une transmission de message analogue à l'expérience de Milgram. Enfin, dans le cadre du routage dans les nouveaux grands réseaux informatiques décentralisés, comme les réseaux pair-à-pair, notre algorithme peut trouver une application pratique puisque qu'il calcule des chemins bien plus courts, ce qui est très utile lorsque la taille des fichiers à transférer est importante.

Nous avons ensuite étendu notre étude de l'aspect dynamique de l'effet petit monde en comparant la charge induite par les deux types d'algorithmes les plus étudiés récemment sur ce modèle sur le réseau de Kleinberg : l'algorithme glouton et un algorithme explorant un voisinage local. Cette étude a mis en valeur une charge nettement différenciée sur les liens les plus longs du réseau. Du point de vue du routage dans les réseaux informatique, cette étude a montré qu'une exploration d'un voisinage de liens locaux risque de provoquer une surcharge du réseau

au niveau des liens les plus longs et du graphe sous-jacent. Il serait intéressant d'étudier la charge induite par notre algorithme de routage, plus difficile à analyser, qui n'explore pas seulement les liens locaux et atténuerait *a priori* cet effet. Cette étude montre également qu'il est possible de discriminer les deux types d'algorithmes étudiés via l'observation de la charge sur les liens du réseau. Dans la perspective de déterminer le type d'algorithme utilisé en pratique dans les réseaux d'interactions réels, cette méthode pourrait apporter de nouvelles informations via l'observation de données statistiques sur un ensemble de connexions.

L'étude des caractéristiques structurelles du seul modèle de petit monde navigable existant au début de cette thèse, celui de Kleinberg, nous a permis de généraliser son principe en un processus d'augmentation aléatoire d'une grande classe de graphes en petits mondes. Nous avons présenté deux méthodes d'augmentation : la première distribue les liens en fonction de la taille des boules du graphe, la seconde les distribue en fonction d'un plongement existant du graphe sur un espace euclidien de faible dimension. Combiné à d'autres processus d'augmentation pour d'autres classes de graphes qui ont été découverts simultanément, cela amène à penser que la propriété de navigabilité tient moins à une structure très spécifique, qu'à l'existence d'un ensemble de raccourcis *bien répartis* par rapport à la métrique sous-jacente. Ce point contient une question ouverte essentielle qui est l'existence d'un processus d'augmentation de *tout* graphe en un petit monde. Nous suspectons une réponse positive à cette question. Par ailleurs, nous avons mis en lumière l'absence de corrélation directe entre la distribution des degrés et la propriété de petit monde navigable, puisque notre modèle permet d'engendrer une distribution de degrés quasi arbitraire.

Ces premières conclusions sur l'analyse structurelle du modèle laissent penser que l'émergence de l'effet petit monde proviendrait d'un processus spécifique de construction de raccourcis. Partant de cette observation, nous avons exhibé un schéma entièrement distribué qui permet d'augmenter une grande classe de graphes en petits mondes navigables, via l'ajout d'un lien par noeud. Ce résultat est une avancée vers la définition d'un processus d'augmentation *naturel* et *local* en chaque noeud qui produirait l'effet petit monde. Il porte en outre des perspectives d'applications pour les grands réseaux informatiques décentralisés, en permettant de réduire le diamètre apparent pour un coût réduit. Ce schéma est cependant peu réaliste sous sa forme actuelle dans le cadre d'un réseau social.

La propriété de petit monde est aujourd'hui suffisamment comprise pour que l'on puisse tirer parti de sa reproduction sur des réseaux construits, comme les réseaux informatiques, à l'aide des modèles et des algorithmes décentralisés existants. Toutefois, l'émergence de cette propriété dans les réseaux réels est encore largement inconnue. La question de l'existence d'un processus universel qui transformerait tout graphe en un petit monde navigable est une question ouverte fon-

damentale pour la compréhension du phénomène. Si la réponse est positive, elle porte la question de la construction dynamique des liens dans les réseaux réels, qui font émerger l'effet petit monde : comment voit-on apparaître des raccourcis bien répartis sur les distances de départ, par exemple dans un réseau social ? Dans le cas où tous les graphes ne seraient pas *petit-mondisables*, il serait intéressant de déterminer la classe de graphe maximale qui peut être transformée en petit monde et de comprendre sa signification pour les graphes réels. La question de l'augmentation d'un graphe en un petit monde par l'ajout de raccourcis porte par ailleurs une perpective de recherche duale : comment peut-on décomposer, dans un réseau réel, les liens qui jouent le rôle de raccourcis (liens longue-distance) des autre liens ?

Une des richesses de la recherche sur les réseaux d'interactions, et en particulier sur les petits mondes, est qu'elle pose des questions de compréhension de la réalité, qui posent en retour de nouvelles questions pour l'informatique théorique. Nous avons ainsi vu se développer parallèlement la recherche sur les algorithmes de routage décentralisé dans les réseaux informatiques et la recherche sur l'existence de chemins courts dans les réseaux sociaux (cf. première partie). De même, c'est la recherche sur la construction de grands réseaux informatiques décentralisés reproduisant les caractéristiques petit monde qui a motivé les recherches récentes sur les plongements de graphes dans des espaces métriques de faible dimension (cf. deuxième partie). Le croisement des connaissances de ces domaines de recherche pourra peut-être permettre de répondre aux questions soulevées dans cette conclusion.

Table des figures

1 Exemple de grand réseau d'interactions : le réseau des pages web qui contient plusieurs millions de noeuds [BKM+00]. . . . 6
2 Modèle de Watts et Strogatz 13

1.1 Exemple de réseau de Kleinberg $\mathcal{K}^2_{1,n}$. 21

2.1 Arbre d'exploration de profondeur 3 dans un réseau $\mathcal{K}^1_{1,n}$, à partir du noeud racine 3 et que l'on a dirigé vers la cible $\mathbf{t}$. En gras et en pointillés figurent les liens longue-distance, les pointillés correspondent aux recoupements. 31
2.2 Exemple d'arbre d'exploration de profondeur 3 avec ses chaînes de liens locaux de longueur 3 attachées aux feuilles, dans un réseau $\mathcal{K}^1_{1,n}$. 32
2.3 Extension du chemin π (en gras) à la fin d'une phase d'exploration. 34
2.4 Routage par explorations successives. 35
2.5 Les ensembles A, B_h, C et F (noeuds blancs) durant une phase d'exploration. 35
2.6 Minimiser la probabilité de recoupement (lemme 2.5). 42
2.7 Illustration de la preuve du lemme 2.5 43
2.8 Deux exemples de coins (points blancs) des boules centrées en $\mathbf{0}$ dans le tore (2D et 3D). 51
2.9 Γ_1, Γ_2 et l'action de ϕ (représentée par les flèches noires). 53
2.10 Illustration de la preuve du lemme 2.12. 54
2.11 Illustration de la preuve du lemme 2.13. 55

3.1 Réseau $\vec{\mathcal{K}}_{1,n}$. 65
3.2 Illustration de la preuve du lemme 3.3. 68
3.3 Fonctionnement de l'algorithme $\mathbf{E}_\omega$. 70
3.4 Variations de $g(x, n-1)$ en fonction de x pour $n = 100$. 74
3.5 Variations de $g(x, n-1)$ en fonction de x pour $n = 100$ en échelle semi-logarithmique. 75

3.6 Valeurs de x au minimum de la courbe $g(x, n-1)$ pour $k = 1$ et $1 \leqslant n \leqslant 10000$ 76
3.7 $\text{load}^{\mathcal{G}}(k, \ell)$ pour $n = 50$ et $k \in \{1, 3, 4, 5\}$. 78
3.8 $\text{load}^{\mathcal{G}}(k, \ell)$ pour $n = 50$ et $k \in \{1, 3, 4, 5\}$ en échelle semi-logarithmique. 79
3.9 Illustration du lemme 3.10. 82
3.10 $\text{load}^{G}(\ell)$ et $\text{load}^{\mathbf{E}_\omega}(\ell)$ pour $n = 50$ et $\omega \in \{7, 9, 12\}$. 85
3.11 $R^{\mathcal{G}}_{\text{local}}(1, m)$ et $R^{\mathbf{E}_\omega}_{\text{local}}(m)$ pour $k = 1$, $\omega \in \{4, 6, 7\}$ et $n = 50$. . . 90
3.12 $R^{\mathcal{G}}_{\text{local}}(1, m)$ et $R^{\mathbf{E}_\omega}_{\text{local}}(m)$ pour $k = 1$, $\omega \in \{4, 6, 7\}$ et $n = 50$ en échelle semi-logarithmique. 91
3.13 $\text{Longueur}^{\mathbf{E}_5}(m)$ et $\text{Longueur}^{\mathcal{G}}(m)$ pour $k = 1$ et $1 \leqslant m \leqslant n = 100$. 93
3.14 $\text{Longueur}^{\mathbf{E}_3}(n-1)$ et $\text{Longueur}^{\mathcal{G}}(n-1)$ pour $k = 1$ et $n \in \{1, \ldots, 200\}$. 94

4.1 Grille allongée ($m \gg n$). 103
4.2 Illustration de trois types de graphes et de l'ordre de grandeur de la taille d'une boule de rayon r pour chacun. 111

5.1 Illustration de l'inclusion des boules pour la preuve du théorème 5.1. 118
5.2 Exemple de la *tapette à mouches infinie*, *petit-mondisable*, mais qui n'est pas à croissance modérée. 122

6.1 Illustration de l'ordonnancement des requêtes vers un *leader*. . . 147

Table des Algorithmes

1 Algorithme de routage glouton 21
2 Algorithme de routage selon le parcours en largeur 36
3 Algorithme $\mathbf{E}_\omega$ de routage par exploration locale 70
4 Algorithme de *petit-mondisation* par plongement 128
5 PETIT-MONDISATION DISTRIBUÉE 139
6 LEADER$(\mathbf{u}, i)$ 140

Index

H_n, 22
$L_j(\mathbf{u})$, 22
Z_d^∞, 106
$Z_{n,d}$, 50
$\mathcal{B}(\mathbf{u}, r)$, 23
$\mathcal{K}_{k,n}^d$, 22
$\vec{\mathcal{K}}_{k,n}$, 64
$\mathcal{S}(\mathbf{u}, r)$, 23
$\hat{k}$, 34
$b_{\mathbf{u}}(r)$, 23
$s_{\mathbf{u}}(r)$, 23

Algorithme de routage, 19
    avec exploration, 36
    avec exploration locale, 70
    décentralisé, 19
    Glouton, 21
Arête, 11, 17
Arc, 18

Boule ℓ_1, 23

Charge d'un lien, 63
Chemin, 18
    de routage, 19
Cible, 19
Coefficient de *clustering*, 7
Coin d'une sphère, 51
Composante connexe, 18
Contact, 18
    local, 22
    longue-distance, 22
Contraction d'un plongement, 127
Cordalité, 133
Croissance
    bornée, 114
    restreinte, 114
Croissance modérée, 113

Destination, 19
Diamètre, 18
    apparent, 27
Dimension
    de grille bornée, 114
    doublante, 132
    intrinsèque, 114
Distance, 18, 19
    de Manhattan, 20
    sous-jacente, 19
Distorsion d'un plongement, 127
Distribution des degrés, 7
Domination Stochastique, 46

Expansion d'un plongement, 127
Expérience de Milgram, 8
Exploration, 29

Fenêtre d'exploration, 69

Graphe, 11
    à croissance β-modérée, 113
    à croissance bornée, 114
    à croissance restreinte, 114
    augmenté, 19
    augmenté aléatoirement, 108
    connexe, 18
    de base, 108
    de Cayley, 120
    non orienté, 17

orienté, 18
petit-mondisable, 110
sommet-transitif, 22

Intervalle de noeuds, 65

Largeur arborescente, 133
Latence, 19
Lien, 11
local, 22
longue-distance, 22
valide, 31
Loi de puissance, 7

Méthode du second moment, 40
Métrique
de graphe, 108
sous-jacente, 108
Mesure doublante, 132
Modèle
d'Albert et Barabasi, 12
d'Erdös et Rényi, 11
de Kleinberg, 13
de Watts et Strogatz, 12

Noeud, 11

Pair-à-pair, 10
Petit-monde, 10
navigable, 9, 109
Petit-mondisable, 110
Petit-mondisation, 108, 110
des graphes à croissance modérée, 115
distribuée, 138
par plongement, 128
Petit-mondiser, 110
Phase d'exécution, 28
Plongement, 127
contraction, 127
distorsion, 127
expansion, 127
Processus de branchement, 38
Produit cartésien, 121

Réseau de Kleinberg, 22
infini, 105
unidimensionnel et unidirectionnel, 64
Réseau synchrone, 137
Ronde, 137
Routage décentralisé, 18, 19

Sommet, 11
Source, 19
Sphère ℓ_1, 23

Tapette à mouches infinie, 122
Taux d'expansion, 137
Taux de liens locaux, 87
Tore régulier, 21

Voisin, 18

Zone de sécurité, 31

Bibliographie

[AA04] L. Adamic and E. Adar. How to search a social network. Tech. report HP Labs, Palo Alto, 2004.

[AB02] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Review of modern physics*, 74(47–97), 2002.

[ABC+05] I. Abraham, Y. Bartal, T-H. H. Chan, K. Dhamdhere, A. Gupta, J. Kleinberg, O. Neiman, and A. Slivkins. Metric embeddings with relaxed guarantees. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005. To appear.

[ACKM05] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling ; or, power-law degree distributions in regular graphs. In *Proceedings of the 37th ACM Symposium on Theory of Computing (STOC)*, 2005. To appear.

[ADS02] J. Aspnes, Z. Diamadia, and G. Shah. Fault-tolerant routing in peer-to-peer systems. In *Proceedings of the ACM 21st Symposium on Principles of Distributed Computing (PODC)*, volume 31, pages 223–232, 2002.

[AGM+04] Ittai Abraham, Cyril Gavoille, Dahlia Malkhi, Noam Nisan, and Mikkel Thorup. Compact name-independent routing with minimum stretch. In 16th *Annual ACM Symposium on Parallel Algorithms and Architecture (SPAA)*, pages 20–24, 2004.

[AJB99] R. Albert, H. Jeong, and A.-L. Barabási. The diameter of the world wide web. *Nature*, 401(130–131), 1999.

[AM05] Ittai Abraham and Dahlia Malkhi. Name independent routing for growth bounded networks. In 17th *Annual ACM Symposium on Parallel Algorithms and Architecture (SPAA)*, pages 49–55, jul 2005.

[BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(509–512), 1999.

[Bar02] L. Bartholdi. Groups of intermediate growth. preprint oai :arXiv.org :math/0201293 (20040622), 2002.

[BFKK01] L. Barrière, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient routing in networks with long range contacts. In *LNCS Proceedings of 15th International Symposium on Distributed Computing (DISC '01)*, volume 2180, pages 270–284, 2001.

[BKM+00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33 :309–320, 2000.

[BLMN03] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric ramsey-type phenomena. *Annals Maths.*, 2003. To appear.

[BM04] P. Bose and P. Morin. Online routing in triangulations. *SIAM J. Comput.*, 33(4) :937–951, 2004.

[Bol85] B. B. Bollobás. *Random Graphs*. Academic Press, London, 1985.

[Bou85] J. Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 1985.

[CGS02a] D. Coppersmith, D. Gamarnik, and M. Sviridenko. The diameter of a long range percolation graph. *Random Structures and Algorithms*, 21 :1–13, 2002.

[CGS02b] D. Coppersmith, D. Gamarnik, and M. Sviridenko. The diameter of a long range percolation graph. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 329–337, 2002.

[CM03] A. Clauset and C. Moore. How do networks become navigable ? Preprint oai :arXiv.org :cond-mat/0309415, 2003.

[CSWH00] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet : A distributed anonymous information storage and retrieval system. In *Proceedings of Designing Privacy Enhancing Technologies : Workshop on Design Issues in Anonymity and Unobservability*, pages 46–66, July 2000.

[DMW03] P.S. Dodds, R. Muhamad, and D.J. Watts. An experimental study of searching in global social networks. *Science*, 301 :827–829, 2003.

[dSP76] D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27 :292–306, 1976.

[dSPK78] I. de S. Pool and M. Kochen. Contacts and influence. *Social networks*, 1 :1–48, 1978.

[EP23] F. Eggenberger and G. Polya. Uber die statistik verketteter. *Vorgange. Zeitschrift Agnew. Math. Mech.*, 3 :279–289, 1923.

[ER59] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicas 6*, 290(7), 1959.

[FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Computer Communications Rev.*, 29 :251–262, 1999.

[FG03] P. Fraigniaud and P. Gauron. Brief announcement : an overview of the content-addressable network d2b. In *Proceedings of the ACM 22st Symposium on Principles of Distributed Computing (PODC)*, page 151, 2003.

[FGP04] 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.

[FMNP05] M. Flammini, L. Moscardelli, A. Navarra, and S. Perennes. Asymptotically optimal solutions for small world graphs. In *LNCS Proceedings of the19th International Symposium on Distributed Computing (DISC)*, pages 414–428, 2005.

[Fra05] P. Fraigniaud. Greedy routing in tree-decomposed graphs : a new perspective on the small-world phenomenon. In *Proceedings of the 13th Annual European Symposium on Algorithms (ESA)*, 2005. To appear.

[GKL03] A. Gupta, R. Krauthgamer, and J.R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003.

[Gla98] M. Gladwell. Six degrees of lois weisberg. *The New Yorker*, 74(41) :52–64, 1998.

[GP03] Cyril Gavoille and David Peleg. Compact and localized distributed data structures. *Journal of Distributed Computing*, 16 :111–120, May 2003. PODC 20-Year Special Issue.

[GPPR01] C. Gavoille, D. Peleg, S. Pérennes, and R. Raz. Distance labeling in graphs. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 210–219, 2001.

[Gri90] R.I. Grigorchuk. On growth in group theory. In *Proceedings of the International Congress of Mathematicians, Kyoto*, pages 325–338, 1990.

[Hei01] J. Heinonen. *Lectures on analysis on metric spaces*. Springer-Verlag, New York, 2001.

[HKK04] K. Hildrum, R. Krauthgamer, and J. Kubiatowicz. Object location in realistic networks. In *Proceedings of the 16th annual ACM symposium on Parallelism in algorithms and architectures (SPAA)*, pages 25–35, 2004.

[HM03] K. Horowitz and D. Malkhi. Estimating network size from local information. *Inf. Process. Lett.*, 88(5) :237–243, 2003.

[HM05] S. Har-Paled and M. Mendel. Fast construction of nets in low dimensional metrics and their applications. In *Proceedings of the 21th ACM Symposium on Computational Geometry*, 2005. To appear.

[HMRe98] M. Habib, C. McDiarmid, J. Ramirez, and B. Reed (ed.). *Probabilistic methods for algorithmic discrete mathematics*. Springer-Verlag, New York, 1998.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. John Wiley & Sons, New York, 2000.

[Kai99] J. Kaiser. It's a small web after all ? *Science*, 285(1815), 1999.

[KB78] P. Killworth and H. Bernard. Reverse small world experiment. *Social networks*, 1 :159–192, 1978.

[KL03] R. Krauthgamer and J.R. Lee. The intrinsic dimensionality of graphs. In *Proceedings of the 35th ACM Symposium on Theory of Computing (STOC)*, pages 438–447, 2003.

[Kle00] J. Kleinberg. The Small-World Phenomenon : An Algorithmic Perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.

[Kle02] J. Kleinberg. Small-world phenomena and the dynamics of information. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS) 14*, Cambridge, MA, 2002. MIT Press.

[KR02] D. R. Karger and Matthias Ruhl. Finding nearest-neighbors in growth-restricted metrics. In *Proceedings of the 34th annual ACM symposium on Theory of computing (STOC)*, pages 741–750, 2002.

[LNJK+05] D. Liben-Nowell, J.Novak, R. Kumar, R. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Science of the USA*, 102(33) :11623–11628, 2005.

[Mil67] S. Milgram. The small world problem. *Psychology Today*, 61(1), 1967.

[MN04] C. Martel and V. Nguyen. Analyzing kleinberg's (and other) small-world models. In *Proceedings of the 23rd Annual ACM Symposium*

on Principles of Distributed Computing (PODC), pages 179–188, 2004.

[MN05] C. Martel and V. Nguyen. Analyzing and characterizing small-world graphs. In *Proceedings of the 16th ACM-SIAM symposium on Discrete Algorithms (SODA)*, 2005. To appear.

[MNR02] D. Malkhi, M. Naor, and D. Ratajczak. Viceroy : a scalable and dynamic emulation of the butterfly. In *Proceedings of the ACM 21st Symposium on Principles of Distributed Computing (PODC)*, pages 183–192, 2002.

[MNW04] G. S. Manku, M. Naor, and U. Wieder. Know thy neighbor's neighbor : the power of lookahead in randomized p2p networks. In *Proceedings of the 36th ACM Symposium on Theory of Computing (STOC)*, pages 54–63, 2004.

[New01] M.E.J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98 :404–409, 2001.

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2) :167–256, 2003.

[NW99] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263 :341–346, 1999.

[PRR99] C. G. Plaxton, R. Rajaraman, and A. W. Richa. Accessing nearby copies of replicated objects in a distributed environnement. *Theory Comput. Syst.*, 32(3) :241–280, 1999.

[PSV00] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letter*, 86(14) :3200–3203, 2000.

[RD01] A. Rowstron and P. Druschel. Pastry : Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *IFIP/ACM Int. Conf. on Distr. Syst. Platf.*, pages 329–350, 2001.

[RFH+01] S. Ratsanamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content-addressable network. In *Proceedings of the ACM annual conference of the Special Interest Group on Data Communication (SIGCOMM)*, pages 161–170, 2001.

[Sim55] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4) :425–440, 1955.

[Sli05] A. Slivkins. Distance estimation and object location via rings of neighbors. In *Proceedings of the 24th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, 2005. To appear.

[SMK+01] I. Stoica, R. Morris, D. Karger, M. Frans Kaashoek, and H. Balakrishnan. Chord : a scalable peer-to-peer lookup service for internet applications. *Proceedings of the ACM annual conference of the Special Interest Group on Data Communication (SIGCOMM)*, 2001.

[TM69] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32 :425–443, 1969.

[TZ05] M. Thorup and U. Zwick. Approximate distance oracles. *Journal of the ACM*, 52(1) :1–24, 2005.

[WDN02] D.J. Watts, P.S. Dodds, and M. Newman. Identity and search in social networks. *Science*, 296 :1302–1305, 2002.

[WS98] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(440–442), 1998.

[Yul24] G. U. Yule. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis. *Philos. Trans. Roy. Soc. London, Ser. B*, 213 :21–87, 1924.

[ZGG02] H. Zhang, A. Goel, and R. Govindan. Using the small-world model to improve Freenet performance. In *Proceedings of IEEE INFOCOM*, 2002.

[Zho04] S. Zhou. A channel assignment problem for optical networks modelled by Cayley graphs. *Theoretical Computer Science*, 310(1-3) :501–511, 2004.

[ZHS+04] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. Kubiatowicz. Tapestry : A resilient global-scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, 22(1) :41–53, 2004.

[Zip49] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA, 1949.

Bibliographie personnelle

[1] E. Lebhar et N. Schabanel. Almost optimal decentralized routing in long-range contact networks. Actes LNCS de la conférence *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 894–905, 2004.

[2] E. Lebhar et N. Schabanel. Close to optimal decentralized routing in long-range contact networks. *Theoretical Computer Science*, vol. 348, issue 2-3, pages 294–310, 2005.
Note : version longue de [1] contenant la généralisation aux modèles de Kleinberg de dimension quelconque.

[3] P. Duchon, N. Hanusse, E. Lebhar, et N. Schabanel. Could any graph be turned into a small world ? À paraître dans le numéro spécial de *Theoretical Computer Science* sur les Réseaux Complexes, 2005.

[4] P. Duchon, N. Hanusse, E. Lebhar, et N. Schabanel. Could any graph be turned into a small world ? *Brief announcement*, actes LNCS de la conférence *DISC 2005*, volume 3724, pages 511–513, 2005.
Note : version courte de [3].

[5] M. Habib, D. Kelly, E. Lebhar, et C. Paul. Can transitive orientation make sandwich problems easier ? À paraître dans le numéro spécial de *Discrete Mathematics* pour *Eurocomb'03*, 2005.

[6] M. Habib, E. Lebhar, et C. Paul. A note on finding all homogeneous set sandwiches. *Information Processing Letters*, 87(3), pages 147–151, 2003.

[7] P. Duchon, N. Hanusse, E. Lebhar, et N. Schabanel. A fully distributed scheme to turn a network into a small world. *Article soumis*.

Abstract : The purpose of this thesis is to study the algorithmic aspects of the small world phenomenon in large interaction networks. Experimental observations showed that large interactions networks (e.g. social or computer networks) share common global properties. One of them is the small world phenomenon which consists in the existence of very short paths between any pair of nodes, that can be discovered using only a partial knowledge of the network. We are interested in this algorithmic characteristic of the small world effect, its application to decentralized routing, and its emergence in real networks.

We propose a new decentralized routing algorithm on the Kleinberg random graph model of small world, which computes paths of length $O(\log n.(\log\log n)^2)$, asymptotically shorter than those computed by existing algorithms ($O((\log n)^2)$). This algorithm could be used in peer-peer networks. We develop this study by comparing the edge load induced by the various algorithms proposed on this model.

By trying to exhibit the minimal characteristics of a graph required to turn it into a small world by adding random shortcuts, we propose a new small world model which generalizes Kleinberg's one. It consists in adding a distribution of links depending on the ball sizes of the underlying graph metric. Furthermore, this model can be simply extended to produce any degree distribution, including in particular the famous power-law. Finally, we propose the first distributed scheme which turns a network into a small world by adding one link per node ; this is a first step towards the understanding of the emergence of the small world phenomenon in real networks.
Keywords : small world graphs, routing algorithms, random models, distributed algorithms.

Résumé : L'objet de cette thèse est l'étude des aspects algorithmiques de l'effet petit monde dans les grands réseaux d'interaction. Les observations expérimentales ont montré que les grands réseaux d'interactions (sociales, informatiques, biologiques), présentaient des propriétés macroscopiques communes. Une d'elles est l'effet petit monde qui consiste en l'existence de chemins très courts entre toutes les paires de noeuds qui peuvent être découverts en n'utilisant qu'une vue locale du réseau. Nous nous intéressons à cette caractéristique algorithmique de l'effet petit monde, à son application au routage informatique décentralisé, et à son émergence dans les réseaux réels.

Nous proposons un nouvel algorithme de routage décentralisé sur le modèle aléatoire de petit monde de Kleinberg, qui calcule des chemins de longueur $O(\log n.(\log\log n)^2)$, asymptotiquement plus courts que ceux des algorithmes existants (en $O((\log n)^2)$). Cet algorithme pourrait également s'appliquer aux réseaux pair-à-pair. Nous précisons cette étude en comparant les charges induites pas les différents algorithmes proposés sur ce modèle.

En tentant d'exhiber les caractéristiques minimales d'un graphe qui permettent de l'augmenter en un petit monde par l'ajout de raccourcis aléatoires, nous proposons un nouveau modèle de petit monde qui généralise celui de Kleinberg. Il s'agit d'ajouter une distribution de liens dépendant de la taille des boules de la métrique des distance sous-jacente. Ce modèle peut par ailleurs être étendu simplement pour produire toute distribution des degrés, dont en particulier la fameuse loi de puissance. Enfin, nous proposons le premier schéma distribué qui permette de transformer un réseau de diamètre quelconque en petit monde en ajoutant un seul nouveau lien par noeud, il s'agit d'un premier pas vers la compréhension de l'émergence naturelle du phénomène dans les réseaux réels.
Mots-clés : graphes petits mondes, algorithmes de routage, modèles aléatoires, algorithmique distribuée.