



HAL
open science

La Grille de Kleinberg, l'Univers et le Reste

Céline Comte, Fabien Mathieu

► **To cite this version:**

Céline Comte, Fabien Mathieu. La Grille de Kleinberg, l'Univers et le Reste. ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2017, Quiberon, France. pp.1-4. hal-01517123

HAL Id: hal-01517123

<https://hal.science/hal-01517123v1>

Submitted on 2 May 2017

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.



Distributed under a Creative Commons Attribution 4.0 International License

La Grille de Kleinberg, l'Univers et le Reste

Céline Comte^{1,2} et Fabien Mathieu¹ †

¹Nokia Bell Labs, Route de Villejust, 91620 Nozay

²Télécom ParisTech, Université Paris-Saclay, France

Longtemps source d'anecdotes, le routage dans les petits mondes a été étudié en 2000 par Kleinberg grâce à un modèle simple qui capture l'influence de la construction de liens distants, les raccourcis, sur la navigabilité dans un graphe : la grille de Kleinberg. Tandis que le modèle initial a été enrichi par de nombreuses analyses asymptotiques et extensions, les résultats numériques sur des grilles simulées sont limités par la complexité du calcul des raccourcis. La contribution principale de l'article est l'introduction d'une nouvelle méthode de tirage par double rejet dynamique qui permet une analyse numérique détaillée. Il devient possible de choisir comme taille de la grille de Kleinberg l'univers, et le reste de l'article montre quelques applications : mise en évidence d'une certaine robustesse du routage ; proposition de nouvelles bornes asymptotiques ; observation des *six degrés de séparation*.

Une première version cet article a été présentée à *Opodis 2016* [Mat16].

Mots-clefs : Routage dans les petits mondes, grille de Kleinberg, simulation, méthode de rejet.

1 La Grille de Kleinberg

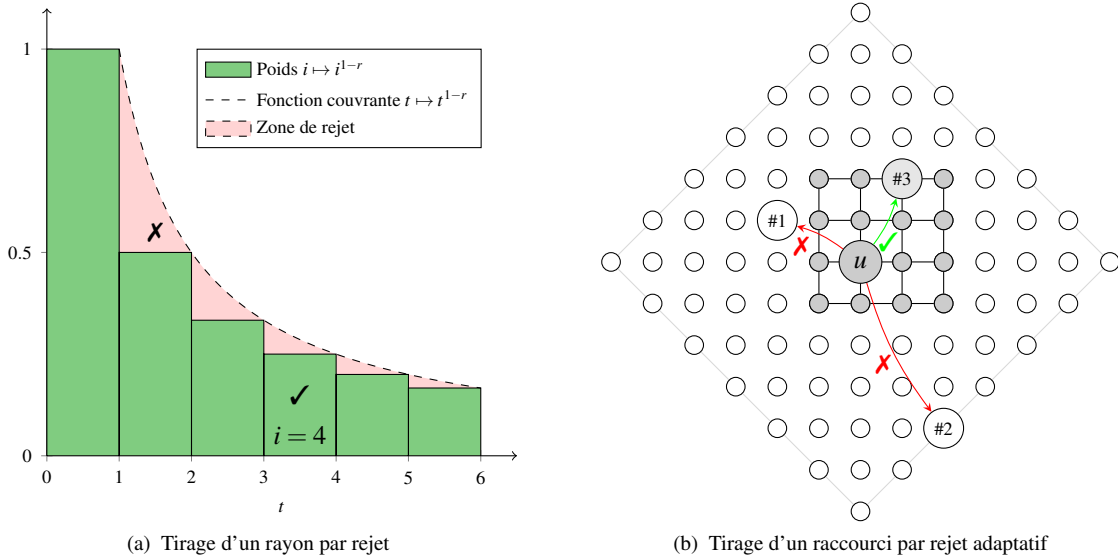
Modèle. On appelle *grille de Kleinberg* un graphe aléatoire dirigé $G(n, r, p, q)$, où n , p et q sont des entiers strictement positifs et r est un réel positif [Kle00b]. Une instance du graphe est construite à partir d'un réseau carré de $n \times n$ nœuds muni de la distance de Manhattan d qui représente une notion de proximité (géographique, sociale, ...) entre les nœuds. Un nœud u possède des voisins *locaux* qui sont les nœuds à distance au plus p de u , ainsi que q voisins *distants* ou *raccourcis* tirés de façon aléatoire et indépendante de sorte qu'un nœud $v \neq u$ soit choisi avec une probabilité proportionnelle à $(d(u, v))^{-r}$.

Les graphes $G(n, r, p, q)$ ont été introduits pour étudier le *routage décentralisé*, ou comment acheminer un message d'un nœud u à un nœud v en parcourant la grille à base de décisions locales. Un exemple central est l'*algorithme glouton* où chaque nœud choisit le prochain saut parmi ses voisins les plus proches de la destination pour la distance de Manhattan. Suivant [Kle00b], nous nous intéressons aux performances de cet algorithme, mesurées par le *temps moyen d'acheminement*, i.e. le nombre moyen de sauts nécessaires pour transmettre un message entre deux nœuds choisis aléatoirement de façon uniforme et indépendante. Sauf mention spéciale, on suppose dorénavant que $p = q = 1$ (un nœud a jusqu'à quatre voisins locaux et un raccourci) et on note $e_r(n)$ le temps moyen d'acheminement de l'algorithme glouton dans $G(n, r, 1, 1)$.

Résultats antérieurs. Les principaux résultats théoriques portant sur le modèle initial sont donnés dans les papiers originaux [Kle00b, Kle00a], où Kleinberg prouve que $e_2(n) = O(\ln^2(n))$ tandis que si $r \neq 2$, n'importe quel algorithme décentralisé se comporte en $e_r(n) = \Omega(n^\alpha)$ pour un certain $\alpha > 0$ qui dépend de r . Intuitivement, si $r < 2$, les raccourcis permettent de parcourir de grandes distances au début mais n'aident pas sur les dernières étapes ; si $r > 2$, les raccourcis ont tendance à être plus courts et ne permettent pas de parcourir de grandes distances au début. Le fait que 2 soit l'unique valeur de r donnant des distances asymptotiquement raisonnables est parfois vu comme le signe que le modèle de Kleinberg n'est pas assez robuste pour expliquer l'efficacité observée du routage glouton dans les petits mondes réels.

Au niveau de l'évaluation numérique, les résultats existants portent sur des valeurs de n relativement limitées et un petit nombre d'instances, avec d'éventuelles hypothèses simplificatrices [Kle00a, AKLP10, MNW04]. Nous allons maintenant expliquer les raisons de ces limitations, comment s'en affranchir, et les enseignements qu'une simulation numérique efficace peut apporter à l'étude de la grille de Kleinberg.

†Les auteurs sont membres du LINCS (<http://www.lincs.fr>)

FIGURE 1: Idée générale de la méthode du double rejet ($n = 4, r = 2$)

2 L'Univers (Comment le simuler ?)

Problématique. On peut estimer $e_r(n)$ en moyennant le nombre de sauts dans R réalisations indépendantes de $G(n, r, 1, 1)$, source et destination étant tirées aléatoirement de façon uniforme et indépendante à chaque réalisation. L'obstacle principal est le tirage des raccourcis. Chaque grille en possède en effet n^2 , et le coût en temps et en mémoire pour le calcul de chacun est *a priori* $\tilde{\Omega}(n^2)$ car $n^2 - 1$ candidats sont à considérer[‡]. Mutualiser les calculs semble difficile car la distribution des raccourcis dépend du nœud considéré : un coin n'a par exemple que deux voisins à distance 1 alors qu'un nœud intérieur en a quatre.

Au bout du compte, si l'on instancie R grilles, le calcul de $e_r(n)$ semble devoir requérir une complexité $\tilde{\Omega}(Rn^4)$ en temps et $\tilde{\Omega}(n^2)$ en mémoire, ce qui est inacceptable pour étudier des grilles de grande taille.

Il est toutefois inutile de calculer des grilles entières. Comme observé dans [Kle00b, MNW04, AKLP10], les raccourcis sont tirés de façon indépendante et ne sont jamais réutilisés sur un chemin donné. On peut ainsi faire appel au *principe de décision différée* [MU05] et calculer les raccourcis à la volée tandis que le chemin est construit. Ceci réduit la complexité en temps à $\tilde{\Omega}(Rn^2 e_r(n))$.

Il est possible de réduire encore la complexité en approximant la grille par un tore [Kle00b]. La distribution des coordonnées relatives des raccourcis est alors identique pour tous les nœuds, de sorte que l'on peut mutualiser leur tirage. La symétrie radiale de la distribution facilite aussi le tirage puisque l'on peut tirer successivement un *rayon* puis un *angle*.

Tirage par double rejet dynamique. Nous proposons une technique permettant d'appliquer sur la grille de Kleinberg (non torique) les mêmes optimisations que sur le tore. On peut ainsi estimer $e_r(n)$ avec un coût $\tilde{O}(Re_r(n))$ en temps et $\tilde{O}(1)$ en mémoire. L'idée centrale est de tirer le raccourci d'un nœud u donné par la méthode de *double rejet* suivante :

1. On plonge la grille initiale, de taille $n \times n$, dans une boule virtuelle B_u de rayon $N = 2(n - 1)$ centrée en u . Ce choix de N assure que la grille réelle est toujours incluse dans B_u , quel que soit le nœud u considéré, ce qui permet de mutualiser certains calculs.
2. On tire en deux étapes un nœud v à l'intérieur de B_u choisi proportionnellement à $(d(u, v))^{-r}$:
 - (Rayon) On choisit d'abord un entier i entre 1 et N , avec probabilité proportionnelle à $(4i)^{1-r}$, i.e. à i^{1-r} . Cette opération est réalisable en $\tilde{O}(1)$ grâce à une méthode de rejet, inspirée de [HD96] et illustrée en Figure 1a, qui évite le calcul explicite de la distribution des N valeurs possibles de i .

‡. La notation *tilde* signifie à un facteur poly-logarithmique près et rappelle ici le coût logarithmique de représentation d'un entier.

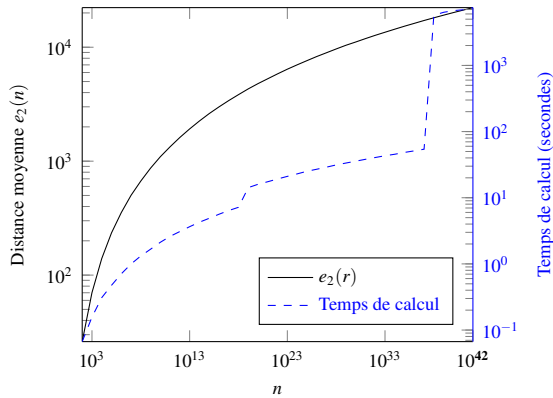


FIGURE 2: Distance et temps de calcul pour $r = 2$

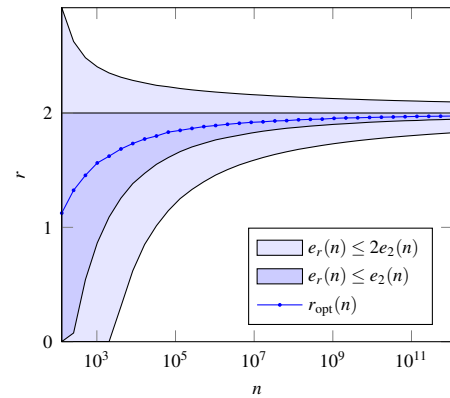


FIGURE 3: Valeurs efficaces de r

- (Angle) On tire ensuite un angle aléatoirement de façon uniforme entre 1 et $4i$, ce qui détermine de manière unique un point v de B_u choisi proportionnellement à $(d(u, v))^{-r}$.
- 3. Si v appartient à la grille initiale, alors il devient le raccourci de u , sinon on le rejette et on recommence le tirage à l'étape 2, comme illustré en Figure 1b. En moins de 8 essais en moyenne (cf. [Mat16]), on obtient par construction un nœud v appartenant à la grille initiale et tiré selon la distribution attendue.

Nous avons implémenté le double rejet et mis le code à disposition sur GitHub[§]. La Figure 2 montre le genre de résultat que l'on peut obtenir sur un PC classique (précision par défaut : $R = 10000$). Le simulateur peut monter en temps raisonnable à $n = 10^{42}$, donnant des grilles contenant 10^{84} nœuds, de l'ordre du nombre de particules contenues dans l'univers observable. Les sauts de performance vers $n = 10^{19}$ et $n = 10^{38}$ sont la conséquence visible du facteur de complexité $\log(n)$: ils correspondent aux moments où le simulateur doit modifier sa représentation des entiers.

3 Et le Reste (Applications)

Il existe une pléthore de résultats théoriques forts autour du routage décentralisé dans les petits mondes. Quel est donc l'intérêt d'un simulateur, même s'il peut simuler l'univers? Nous illustrons son utilité à travers trois études montrant qu'un simulateur permet non seulement d'obtenir des évaluations numériques précises pour des grilles de taille raisonnable mais aussi de raffiner les prédictions asymptotiques.

Valeurs efficaces de r . Dans [Kle00a], Kleinberg constate par simulation sur une approximation torique que choisir r suffisamment proche de 2 semble suffisant en pratique. Cette question de la sensibilité à l'exposant r sur des grilles de taille finie nous paraît essentielle pour décider si la grille de Kleinberg peut être acceptée comme un modèle raisonnable dans les réseaux sociaux.

Dans la Figure 3, nous testons cette hypothèse en évaluant pour n donné : la valeur r_{opt} de r qui minimise $e_r(n)$; les valeurs de r telles que $e_r(n) \leq e_2(n)$; les valeurs de r telles que $e_r(n) \leq 2e_2(n)$. r_{opt} est calculé par la méthode du nombre d'or [PTVF07] (en prenant $R = 10^6$ pour satisfaire aux besoins de précision de la méthode) et les autres valeurs sont calculées par bisection. En plus de confirmer que $r = 2$ est asymptotiquement la valeur optimale, les résultats montrent que l'intervalle de valeurs raisonnables pour des grilles de taille finie est confortable. Par exemple, pour $n \leq 2^{14}$ (environ 270 millions de nœuds), tout r entre 0,85 et 2,26 assure une distance moyenne inférieure ou égale à $2e_2(n)$.

Nouvelles bornes asymptotiques. En supposant que la distance moyenne se comporte en n^α , le simulateur nous permet d'estimer la valeur de α en fonction de n . Les résultats, présentés en Figure 4, indiquent que les bornes prouvées par Kleinberg dans [Kle00b] peuvent être améliorées et suggèrent de nouvelles bornes, consistantes avec des résultats obtenus sur l'anneau dans [BFKK01].

§. <https://github.com/balouf/Kleinberg>. À noter : la version précédente du code, publiée dans [Mat16], n'utilise pas de méthode de rejet pour le tirage du rayon. Il en résulte un coût mémoire $\tilde{O}(n)$ qui limite n à environ 10^8 sur un PC standard.

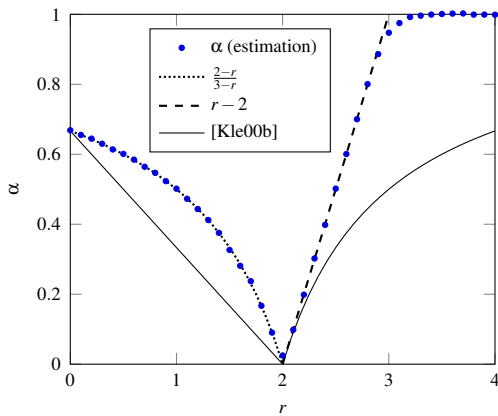


FIGURE 4: Comportement asymptotique

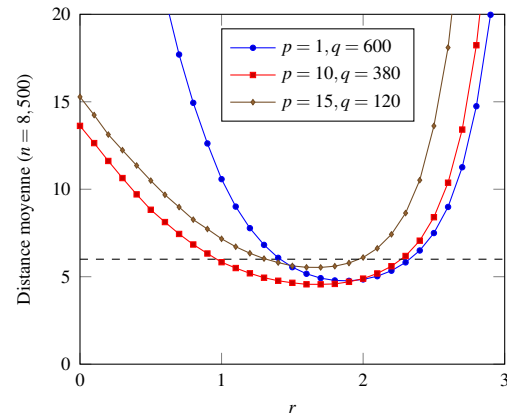


FIGURE 5: Six degrés de séparation

Six degrés de séparation. Les expériences de Milgram menées sur la population des États-Unis donnent une distance moyenne de routage proche de 6 [TM69], soit beaucoup moins qu’observé dans la Figure 2. Cela s’explique par la faible densité de la grille considérée, au plus 5 voisins par nœud pour $p = q = 1$. En choisissant des valeurs de n , p et q plus réalistes, la distance moyenne est effectivement proche de 6 pour une large plage de valeurs de r (cf. Figure 5). Le simulateur nous montre ainsi qu’en dépit de sa simplicité, le modèle de la grille de Kleinberg est suffisant pour capturer le phénomène dit des *six degrés de séparation*.

4 Conclusion

Grâce à une technique de double rejet, nous pouvons simuler le routage glouton dans la grille de Kleinberg à des niveaux jusqu’à présent inenvisageables, ce qui permet de revisiter le modèle original et de donner un éclairage nouveau sur les résultats théoriques existants. Nous espérons que notre simulateur sera utile aussi bien aux chercheurs issus des communautés théoriques qu’empiriques.

Références

- [AKLP10] S. Athanassopoulos, C. Kaklamanis, I. Laftsidis, and E. Papaioannou. An experimental study of greedy routing algorithms. In *HPCS 2010*, pages 150–156, 2010.
- [BFKK01] L. Barrière, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient routing in networks with long range contacts. In *DISC '01*, pages 270–284, London, UK, 2001.
- [HD96] W. Hörmann and G. Derflinger. Rejection-inversion to generate variates from monotone discrete distributions. *ACM Trans. Model. Comput. Simul.*, 6(3) :169–184, July 1996.
- [Kle00a] J. Kleinberg. Navigation in a small world. *Nature*, August 2000.
- [Kle00b] J. Kleinberg. The small-world phenomenon : An algorithmic perspective. In *STOC 2000*, pages 163–170, 2000.
- [Mat16] F. Mathieu. Kleinberg’s Grid Reloaded. In *OPODIS 2016*, Madrid, Spain, December 2016.
- [MNW04] G. S. Manku, M. Naor, and U. Wieder. Know thy neighbor’s neighbor : The power of lookahead in randomized P2P networks. In *STOC 2004*, pages 54–63. ACM, 2004.
- [MU05] M. Mitzenmacher and E. Upfal. *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.
- [PTVF07] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Minimization or maximization of functions. In *Numerical Recipes 3rd Edition : The Art of Scientific Computing*, chapter 10. Cambridge University Press, New York, NY, USA, 3 edition, 2007.
- [TM69] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32 :425–443, 1969.