

Une analyse de PageRank, une mesure de popularité des pages Web

Mohamed Bouklit and Alain Jean-Marie

LIRMM

161, rue Ada, 34392 Montpellier Cedex 5 France.

Web: <http://www.lirmm.fr/~bouklit> <http://www.lirmm.fr/~ajm>

L'analyse du graphe formé par les pages Web et les liens hypertextes qui les relient, communément appelé graphe du Web, a permis d'améliorer la performance des moteurs de recherche actuels. Ainsi, lancé en 1998, le moteur de recherche Google classe les pages grâce à la combinaison de plusieurs facteurs dont le principal porte le nom de PageRank. Ce dernier est un indice numérique qui utilise le nombre de liens pointant sur les pages.

Dans cet article, nous analysons une variante du modèle PageRank proposée brièvement par [Kle98] que nous justifions à l'aide d'un modèle probabiliste de l'internaute. Nous montrons ensuite en quoi les travaux antérieurs sur PageRank[PBMW98, BP98, Kle98, Hav99] conduisent naturellement à un tel modèle. Nous analysons l'algorithme PageRank[PBMW98] de calcul du rang et en proposons un nouveau adapté à notre variante. Pour conclure, nous nous intéressons à la convergence de ces deux algorithmes.

Keywords: graphe du Web, moteur de recherche, Google, PageRank, modèle de l'internaute, chaînes de Markov

1 Introduction

Les moteurs de recherche ont développé des méthodes de tri automatique des résultats. Leur but est d'afficher dans les dix à vingt premières réponses les documents répondant le mieux à la question. Dans la pratique, aucune méthode de tri n'est parfaite mais leur variété offre à l'utilisateur la possibilité de traquer l'information de différentes manières : cette variété augmente donc ses chances d'améliorer ses recherches.

La méthode de tri la plus ancienne et la plus utilisée est la méthode de tri par pertinence : on la trouve dans les sites Voila, Lycos, AltaVista, Excite, InfoSeek, ... Elle est basée sur le nombre d'occurrences des termes de la recherche dans les pages, de leur proximité, de leur place dans le texte. Malheureusement, cette méthode présente l'inconvénient d'être facile à détourner par des auteurs désireux de placer leurs pages en tête de liste : pour cela, il suffit de répéter les mots importants soit dans l'en-tête, soit dans le texte en utilisant des techniques de *spamming* (écrire le texte en blanc sur fond blanc par exemple) pour modifier à son avantage le classement. Les limites du tri par pertinence ont alors conduit à rechercher, à partir de principes tout à fait différents, d'autres méthodes indépendantes du contenu des documents. C'est dans ce contexte que sont apparues des méthodes de tri basées sur une notion de popularité (co-citation, mesure d'audience). Ainsi, lancé en 1998, le moteur de recherche Google[Goo98] classe les pages grâce à la combinaison de plusieurs facteurs dont le principal porte le nom de *PageRank*[PBMW98]. Plus précisément, le classement des pages est fait en utilisant un indice numérique (le «rang») calculé pour chaque page.

En amont du processus, il y a tout d'abord les robots qui chالutent continuellement le Web dans l'intention de découvrir de nouvelles pages et à défaut de mettre à jour les anciennes. Ces pages sont stockées dans un entrepôt de données. Viennent ensuite les hyperliens qui sont stockées séparément pour former un sous graphe du Web. Ce graphe est

alors utilisé pour le calcul des rangs de page. Le rang d'une page permettra en particulier d'ordonner les résultats d'une requête d'un usager. Dans [PBMW98], les auteurs proposent un modèle de *conservation du rang* et un algorithme permettant son calcul : l'algorithme PageRank.

Cet article a pour objets : de synthétiser les différentes versions de la définition du rang qui ont été proposées, de préciser leur relation avec la théorie des chaînes de Markov, et d'apporter des éléments théoriques validant les algorithmes qui sont proposés pour le calcul. En effet, la littérature contient différentes propositions de lois pour la conservation du rang. D'autre part, si l'article originel fait allusion aux chaînes de Markov, l'analyse n'est pas poussée jusqu'à son terme, ce qui entraîne un certain manque de clarté dans le modèle et l'algorithme qui en est déduit. Enfin, plusieurs algorithmes ont été décrits, en relation avec les différents modèles, mais peu d'arguments autres qu'empiriques permettent d'affirmer qu'ils convergent bien, et la valeur du rang qu'ils calculent n'est pas toujours claire.

La suite de cet article est organisée comme suit. La section 2 introduit tout d'abord une variante du modèle PageRank proposée brièvement par [Kle98] que nous justifions à l'aide d'un modèle probabiliste de l'internaute. Nous rappelons ensuite une série de travaux antérieurs autour de PageRank en mettant en évidence leur contribution au modèle choisi [PBMW98, BP98, Kle98, Hav99]. La section 3 décrit l'algorithme PageRank et en propose un nouveau adapté à notre variante. La dernière section s'intéresse à la convergence de ces deux algorithmes.

2 Modélisation

Dans cette section, nous présentons ce modèle sous sa forme fonctionnelle d'abord, puis vectorielle. Nous lui apportons ensuite une interprétation probabiliste. Nous montrons enfin en quoi les travaux antérieurs sur PageRank conduisent de toute évidence à un tel modèle.

2.1 Le modèle fonctionnel

Soit $G = (S, U)$ le graphe orienté formé par les pages Web : S et les liens hypertexte qui les relient : U . G est obtenu par une succession de parcours du Web.

Le rang R d'une page p respecte la loi de *conservation* suivante :

$$R(p) = c \left[d \times \left(\sum_{\forall q/q \text{ pointe sur } p} \frac{R(q)}{d^+(q)} \right) + (1-d) \times E(p) \right] \quad (1)$$

où d est un facteur d'amortissement choisi dans $[0, 1]$, $d^+(q)$ le degré externe de q et c une constante de renormalisation permettant de garantir que $\sum_{p \in S} R(p) = 1$ (voir l'analyse en section 4). De plus, on suppose que : $\forall p \in S, R(p) > 0, E(p) > 0$ et $\sum_{p \in S} E(p) = 1$.

E représente donc une loi de distribution sur l'ensemble des pages de S . Généralement, E est une loi de distribution uniforme : $\forall p \in S, E(p) = \frac{1}{n}$ où $n = |S|$. Mais il a été proposé que cette distribution puisse être «personnalisée» [BMPW98].

On peut se demander ce qui motive donc le choix d'un tel modèle. Ce dernier cache en fait une idée intuitive de propagation de rang assez intéressante. Ainsi, une page reçoit un rang élevé si la somme des rangs des pages pointant sur elle est élevée.

La figure 1 illustre une propagation de rang d'une paire de pages à l'autre. On y suppose $d = 0.9$ et $n = 10$. En observant la page p sur cette figure, nous remarquons que :

- $d = 90\%$ de son rang (soit 0.18) est redistribué équitablement sur ses liens sortants (soit $\frac{0.18}{2} = 0.09$) affectant ainsi le rang des pages pointés par p .
- $1 - d = 10\%$ de son rang (soit 0.02) est dissipée au profit d'une répartition globale sur l'ensemble du graphe contribuant ainsi à alimenter chaque page d'un rang égal à $\frac{1-d}{n} = \frac{0.1}{10} = 0.01$.

Nous pouvons vérifier par exemple que le rang de la page r est bien 0.19 :

$$\begin{aligned} R(r) &= d \times \frac{R(p)}{d^+(p)} + d \times \frac{R(q)}{d^+(q)} + \frac{1-d}{n} = 0.9 \times \frac{0.2}{2} + 0.9 \times \frac{0.3}{3} + \frac{0.1}{10} \\ &= 0.19. \end{aligned}$$

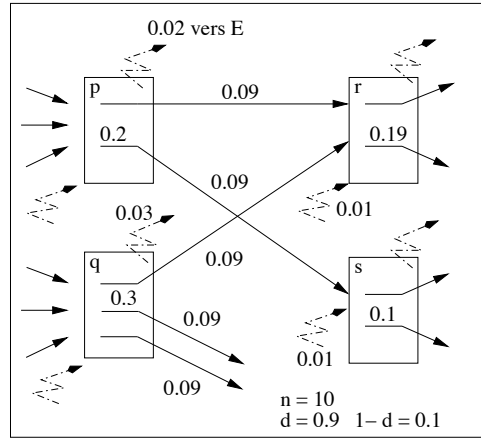


FIG. 1: propagation de rang

2.2 Le modèle vectoriel

Le modèle fonctionnel précédent peut être reformulé sous forme vectorielle. Soit E et R les vecteurs colonne formés respectivement des $E(p)$ et des $R(p)$. Soit A la matrice telle que l'entrée $A[p, q] = 0$ s'il n'existe pas de liens sortants de la page p vers la page q . Sinon, $A[p, q] = \frac{1}{d^+(p)}$. A est donc une matrice sous-stochastique.

L'équation (1) s'écrit :

$$R = c (d A^t R + (1-d) E) \quad (2)$$

avec c maximal et $\sum_{p \in X} R(p) = 1$. A^t désigne la matrice transposée de A .

En effet, le terme sommation de l'équation (1) peut être représenté par la multiplication matricielle $A^t \times R$.

Écrire que $\sum_{p \in S} R(p) = 1$ équivaut matriciellement à écrire que la norme ℓ_1 du vecteur R vaut 1, $\|R\|_1 = \sum_{i=1}^n |R_i| = 1$. Comme $\forall i \in \{1, 2, \dots, n-1, n\}$, R_i est positif. On obtient donc : $\sum_{i=1}^n R_i = 1$. L'équation (2) peut être reformulée comme suit : $R = c(d A^t + (1-d) E \times \mathbf{1})R$ où $\mathbf{1}$ désigne le vecteur ligne ne contenant que des 1. En effet, comme $\sum_{i=1}^n R_i = 1$, on obtient alors : $\mathbf{1} \times R = 1$. Il en découle que : $E = E \times \mathbf{1} \times R$ et par conséquent R est un vecteur propre de la matrice $d A^t + (1-d) E \times \mathbf{1}$ pour la valeur propre $1/c$. Cette propriété sera exploitée au paragraphe 4.1.

2.3 PageRank, un modèle probabiliste de l'internaute

La définition de l'équation (2) évoque la distribution stationnaire d'une marche aléatoire du graphe G . Intuitivement, elle simule donc le comportement d'un internaute surfant sur le Web.

À chaque page, notre internaute λ a la possibilité

- soit de cliquer sur l'un des liens sortants ($A^t R$) avec la probabilité d .
- soit de zapper, avec la probabilité $1-d$ cette fois, sur une page choisie aléatoirement selon la distribution de E .

[PBMW98] utilise l'intuition du comportement aléatoire d'un internaute pour interpréter l'introduction du vecteur E dans leur modèle de rang. En effet, pour échapper aux circuits

sans issue, il est nécessaire «de temps en temps» de sauter aléatoirement vers une page quelconque du Web. Cependant, les auteurs n'exploitent pas plus loin cette interprétation. En fait, leur modèle de rang est alors celui de l'équation (4), qui n'admet pas d'interprétation probabiliste directe. Nous reviendrons sur cette interprétation au paragraphe 4.1.

2.4 Travaux antérieurs

Avant d'aboutir au modèle (1), les concepteurs de PageRank et d'autres auteurs ont exploré plusieurs possibilités.

[PBMW98] introduit le modèle PageRank en proposant une première fonction de rang pondéré simple :

$$R(p) = c \sum_{\forall q/q \text{ pointe sur } p} \frac{R(q)}{d^+(q)} \quad (3)$$

Le réel c est une constante de renormalisation de manière à ce que $\sum_{p \in S} R(p) = 1$. Hélas, cette fonction n'est valide que dans le cas de graphes fortement connexes.

En effet, l'existence de *puits de rang* dans le graphe pousse [PBMW98] à généraliser en dotant chaque page d'un rang initial (source de rang). Ainsi chaque sommet p se voit attribuer un rang de $E(p) > 0$. (3) devient alors :

$$R(p) = c \left(\sum_{\forall q/q \text{ pointe sur } p} \frac{R(q)}{d^+(q)} + E(p) \right) \quad (4)$$

avec c maximal et $\sum_{p \in S} R(p) = 1$.

[BP98] propose ensuite une variante empirique du modèle en introduisant un facteur de pondération $d \in [0, 1]$, ce qui donne :

$$R(p) = d \times \left(\sum_{\forall q/q \text{ pointe sur } p} \frac{R(q)}{d^+(q)} \right) + 1 - d \quad (5)$$

avec pour condition que $\sum_{p \in S} R(p) = 1$.

Kleinberg, dans [Kle98], tente d'unifier les équations (4) et (5). Il considère un graphe orienté $G = (S, U)$ de matrice d'adjacence M . La probabilité d'une transition de la page i vers la page j dans [BP98] peut être perçue, selon Kleinberg, comme égale à $B_{ij} = \frac{d}{d^+(i)} M_{ij} + \frac{1-d}{n}$. Soit la matrice B , la matrice dont les entrées sont les B_{ij} . Le vecteur rang R est alors une solution non nulle et positive de l'équation $B^t R = R$, et par conséquent il correspond au vecteur propre dominant de la matrice B^t .

Ce modèle est repris dans [Hav99]. Selon cet auteur, l'introduction du paramètre d'amortissement est destiné à améliorer la «qualité» du PageRank en garantissant la convergence vers un unique vecteur rang. La matrice A est explicitement supposée stochastique en éliminant itérativement les pages sans liens. Au regard des travaux antérieurs, nous avons donc fait le choix de suivre l'interprétation (1) de PageRank qui nous semble être bien mieux fondée que l'originelle.

3 Algorithmes

3.1 L'algorithme PageRank

[PBMW98] décrit l'algorithme *PageRank* de calcul du rang, basé sur la récurrence $R_{n+1} = A^t R_n$.

Remarquons que le facteur μ augmente la stabilité numérique de l'algorithme en forçant la condition $\|R_{n+1}\|_1 = 1$. Une normalisation alternative (et plus robuste) consisterait à multiplier R_{n+1} par un facteur approprié. D'après [PBMW98], l'utilisation de μ peut avoir un léger impact sur l'influence de E . Les auteurs n'apportent aucune preuve de la convergence de cet algorithme.

Algorithme 1: PageRank

Données :

- une matrice sous stochastique A d'un sous graphe du Web $G = (S, U)$
- un vecteur E positif tel que $\|E\|_1 = 1$
- un réel ε

Résultat : le vecteur propre dominant R de la matrice A^t

début

$R_0 = E;$

répéter

$R_{n+1} = A^t R_n;$

$\mu = \|\|R_n\|_1 - \|R_{n+1}\|_1;$

$R_{n+1} = R_{n+1} + \mu E;$

$\delta = \|\|R_{n+1} - R_n\|_1;$

jusqu'à $\delta < \varepsilon;$

fin

Les résultats empiriques rapportés dans [PBMW98] indiquent une convergence rapide de l'algorithme en pratique : en quelques dizaines d'itérations, une approximation raisonnable de R est atteinte sur un graphe de 322 millions de liens. Les auteurs suggèrent que l'explication pourrait provenir d'une propriété d'*expansion* du graphe du Web, et font référence à [MR95]. En effet, pour un graphe expansif, on sait donner une borne supérieure pour les valeurs propres de A^t différentes de la valeur propre principale (voir l'analyse en section 4). En fait, ces résultats s'appliquent à des graphes non orientés de degré constant, donc pas directement au problème des pages du Web. Mais il est certain que les relations entre la topologie du graphe et la vitesse de convergence de l'algorithme restent à explorer. Il existe des outils analytiques et géométriques permettant d'attaquer le problème en dehors du cadre des graphes expansifs [SC96].

3.2 Proposition d'algorithme

La forme vectorielle a permis de voir que le calcul du rang revient à la résolution de systèmes linéaires. D'un autre côté, la formulation fonctionnelle de ce calcul suggère de transformer les équations de *conservation* du rang telles que (1) (ou sa forme vectorielle (2)) en équation de *propagation* du rang de la forme :

$$R_{n+1} = c_n (d A^t R_n + (1-d) E), \quad c_n \text{ t.q. } \|R_{n+1}\|_1 = 1, \quad (6)$$

dont (2) constitue l'équation au point fixe associée. Étant donnée la taille des systèmes en question, les méthodes itératives de résolution sont de toutes façons les plus appropriées.

Nous proposons maintenant l'algorithme *PageRank2*, un algorithme itératif de calcul du rang, basé sur la récurrence (6).

Observer que le vecteur initial R_0 est mis à la distribution E «par défaut» mais peut être choisi autrement. Par exemple, prendre le résultat d'un calcul précédent peut souvent accélérer la convergence. Cette remarque s'applique à l'algorithme précédent, même si E a l'interprétation de «rang initial», voir (4).

4 Analyse

Du point de vue théorique, les premières questions qui se posent sont : a) existe-t-il un vecteur de rangs R , et une constante c solutions de l'équation de conservation (2)? b) cette solution est-elle bien définie (unique)? c) l'algorithme itératif (6) converge-t-il vers cette solution? La réponse à ces trois questions est apportée par l'analyse des valeurs propres de la matrice. Comme il l'a été observé au paragraphe 2.2, il est possible de reformuler (2) en :

$$(d A^t + (1-d) E \times \mathbf{1}) R = \frac{1}{c} R. \quad (7)$$

Algorithme 2: PageRank2**Données :**

- une matrice sous-stochastique A d'un sous graphe du Web $G = (S, U)$
- un vecteur E positif tel que $\|E\|_1 = 1$
- un réel ε
- un réel d choisi dans $[0, 1]$

Résultat : le vecteur propre dominant R de la matrice $d A^t + (1 - d) E \times \mathbf{1}$ **début**

$$R_0 = E;$$

$$B^t = dA^t + (1 - d)E \times \mathbf{1};$$
répéter

$$R_{n+1} = B^t R_n;$$

$$\mu = \|R_{n+1}\|_1;$$

$$R_{n+1} = \frac{R_{n+1}}{\mu};$$

$$\delta = \|R_{n+1} - R_n\|_1;$$
jusqu'à $\delta < \varepsilon$;**fin**

Le problème est donc : existe-t-il un vecteur positif R qui soit le vecteur propre de la matrice $d A^t + (1 - d) E \times \mathbf{1}$, pour une valeur propre $1/c > 0$?

Le Théorème de Perron-Frobenius permet de répondre. Nous incluons ci-après une version de ce théorème (voir par exemple [HJ85]). On rappelle qu'une matrice est *irréductible* si son graphe est fortement connexe, et *apériodique* si le p.g.c.d. des longueurs des circuits est 1.

Théorème 1 (Perron-Frobenius) *Soit A une matrice positive, irréductible et apériodique.*

Alors il existe une valeur propre r de A telle que :

- a) $r > 0$;
- b) r est associée à un vecteur propre gauche $x > 0$ et un vecteur propre droit $y > 0$;
- c) pour toute autre valeur propre λ de A , $|\lambda| < r$;
- d) soit la matrice $L = yx/xy$: alors $(A/r)^n \rightarrow L$ quand $n \rightarrow \infty$, et pour tout $\phi > \frac{1}{r} \max\{|\lambda|, \lambda \text{ valeur propre de } A \neq r\}$, il existe C tel que pour tout n ,

$$\|(A/r)^n - L\|_1 \leq C \phi^n . \quad (8)$$

La valeur propre r est dite valeur propre principale (ou : de Perron-Frobenius) de A .

4.1 Existence de solutions

Le Théorème de Perron-Frobenius s'applique à la matrice $B^t = d A^t + (1 - d) E \times \mathbf{1}$ quand $0 \leq d < 1$. En effet, nous avons supposé que $E > 0$, donc $B_{pq} = dA_{pq} + (1 - d)E(p) > 0$. Ceci implique l'irréductibilité et l'apériodicité de B . Il existe donc une solution $R > 0$ au système (7), avec pour constante c l'inverse de la valeur propre principale de B .

Dans le cas où $d = 1$, nous avons $B = A$ et le théorème ne s'applique que si A est fortement connexe (et également apériodique). Dans le cas qui nous occupe, ceci revient à demander que le graphe du Web soit fortement connexe. Ce n'est pas le cas, et c'est ce qui a conduit à l'abandon du modèle simple (3) de conservation du rang. En effet, dans ce cas, on montre que seules les pages situées dans une composante fortement connexe terminale peuvent avoir un rang strictement positif.

Un cas bien connu de la littérature est celui où la matrice A est *stochastique* (c'est-à-dire dont les lignes ont une somme qui vaut 1). En effet, dans ce cas la matrice $B = dA + (1 - d)\mathbf{1}^t \times E^t$ l'est aussi pour tout $d \in [0, 1]$. Or, la valeur propre principale d'une matrice stochastique (et de sa transposée) est $r = 1$. Le système (7) a alors pour solution $c = 1$ et R s'interprète comme le vecteur de *probabilité stationnaires* de la chaîne de Markov

dont la matrice de transition est B (voir par exemple [KS60]). Cette chaîne de Markov est précisément celle parcourue par un internaute ayant le comportement décrit au paragraphe 2.3. Les résultats classiques permettent de savoir quand le problème (3) admet une solution unique, et ce qui se produit dans les autres cas.

Mais en général, la matrice A n'est pas stochastique car il existe des pages dont aucun lien ne permet de sortir. Telle que la matrice a été construite (voir au paragraphe 2.2), la ligne correspondant à une telle page est nulle. On obtient alors que A et B sont *sous-stochastiques*, et que leur valeur propre principale est $r < 1$ (donc $c > 1$). Le modèle probabiliste que nous avons imaginé n'est donc pas valide dans ce cas. Ce qui lui manque est la spécification du comportement quand l'internaute atteint une page-impasse *mais* qu'il ne «zappe» pas. On pourrait imaginer qu'il revient sur la (ou une) page précédente, ce qui revient à compléter les pages sans liens avec un lien «par défaut» vers leur page mère. La matrice A redevient alors stochastique, et l'interprétation probabiliste tient.

Enfin, notons que le Théorème 1 permet de prouver aussi l'existence de solutions aux modèles de rang (4) et (5), même si pour eux l'interprétation probabiliste n'est pas pertinente.

4.2 Convergence des algorithmes

Nous nous intéressons à présent à la convergence de l'algorithme itératif *PageRank2* de calcul du rang, basé sur la récurrence (6), c'est-à-dire $R_{n+1} = c_n B^t R_n$.

Soit $c = 1/r$, r la valeur propre principale de B . La récurrence se résout en :

$$R_n = (c_{n-1}/c)(cB^t)R_{n-1} = \prod_{m=0}^{n-1} (c_m/c)(cB^t)^n R_0 = \gamma_n (cB^t)^n R_0,$$

où γ_n est la constante telle que $\|R_n\|_1 = 1$. Appliquons la partie d) du Théorème 1. Nous choisissons comme vecteur propre droit de B^t le vecteur R tel que $\|R\|_1 = 1$. Il lui correspond un vecteur propre gauche S tel que $SR = 1$. Nous avons alors : $(cB^t)^n = RS + O(\phi^n)$. Par conséquent,

$$R_n = \gamma_n (RS + O(\phi^n)) R_0 = \gamma_n R (SR_0) + O(\phi^n) = \delta_n R + O(\phi^n).$$

Comme $\|R_n\|_1 = 1$, il vient : $\delta_n = 1 + O(\phi^n)$. Finalement, nous avons :

$$R_n = R + O(\phi^n). \quad (9)$$

L'algorithme *PageRank2* converge donc vers R , et sa vitesse de convergence est géométrique. Le facteur de convergence dépend du rapport des deux plus grandes valeurs propres de B .

De même, si A est stochastique, irréductible et aperiodique, l'algorithme *PageRank* converge vers un vecteur R et la convergence est géométrique. On en déduit que le nombre d'itérations nécessaires pour approcher R à une distance de ϵ est de l'ordre de $\log \epsilon / \log \phi$.

5 Conclusions

Nous avons décrit les principes de la méthode PageRank, et montré ses liens avec la théorie des chaînes de Markov. Nous avons également décrit des algorithmes pour calculer le rang des pages, et expliqué leur convergence. De nombreuses questions restent en suspens.

D'un point de vue pratique, la méthode fait intervenir deux paramètres *a priori* : le facteur de pondération d et la distribution E . Or, il est clair que la valeur du rang obtenue dépend de ces deux paramètres. Les résultats empiriques s'accordent sur une valeur $d = 0.85$, en utilisant pour E la distribution uniforme [BP98]. Peut-on améliorer ce choix ? Le modèle probabiliste de l'internaute que nous décrivons comme support (implicite) au modèle de rang prédit que le nombre de «clics» consécutifs suit une distribution géométrique de moyenne $d/(1-d)$. Cela pourrait donner une façon empirique de trouver d , mais cette loi est contredite par les résultats rapportés dans [HPPL98] et les études qui en ont découlé.

D'un point de vue théorique, il reste à expliquer la convergence rapide des algorithmes, c'est-à-dire la petitesse de la valeur ϕ dans (9). Parmi les outils abordant cette question,

nous avons mentionné les graphes expansifs [MR95], et les approches analytiques ou géométriques du calcul du «trou spectral» [SC96]. Un autre sujet intéressant serait de trouver comment exprimer le rang comme fonction plus explicite du graphe du Web. Par exemple, certains travaux font apparaître que les distributions du rang et du degré des pages suivent toutes deux une loi de puissance[PRU].

Du point de vue algorithmique, la question est d'augmenter encore la vitesse de calcul du rang, sachant que le graphe est énorme et dynamique. Une idée serait d'utiliser les résultats précédemment calculés (incrémentalité), mais aussi de décomposer le calcul en «blocs» selon les vitesses d'évolution respectives (parallélisme, asynchronisme). Pour avancer dans cette direction, il est sans doute nécessaire d'en savoir plus sur la structure du graphe lui-même, et sur sa dynamique.

Enfin, il nous semble intéressant d'étudier l'utilisation de cette notion de rang et de popularité dans d'autres domaines que celui de la recherche d'informations sur le Web : par exemple pour la visualisation de graphes complexes, il est utile de savoir mettre en valeur les sommets les plus pertinents.

Références

- [BMPW98] S. Brin, R. Motwani, L. Page, and T. Winograd. What can you do with a Web in your Pocket? *Data Engineering Bulletin*, 21(2) :37–47, 1998.
- [BP98] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7) :107–117, 1998.
- [Goo98] Google. <http://www.google.com/>, 1998.
- [Hav99] T. Haveliwala. Efficient computation of PageRank. Technical report, Computer Science Department, Stanford University, 1999.
- [HJ85] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [HPPL98] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, and R.M. Lukose. Strong regularities in World Wide Web surfing. *Science*, 280 :95–97, april 1998.
- [Kle98] J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, San Francisco, California, 25–27 January 1998.
- [KS60] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer Verlag, 1960.
- [MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [PBMW98] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking : Bringing Order to the Web. Technical report, Computer Science Department, Stanford University, 1998.
- [PRU] G. Pandurangan, P. Raghavan, and E. Upfal. Using PageRank to Characterize Web Structure. *Manuscrit*.
- [SC96] L. Saloff-Coste. Lectures on finite Markov chains. In G.R. Grimmet E. Giné and L. Saloff-Coste, editors, *Lecture Notes on Probability Theory and Statistics*, number 1665 in LNM, pages 301–413. Springer Verlag, 1996.