Un modèle gravitationnel du Web

Toufik Bennouas, Mohamed Bouklit et Fabien de Montgolfier

LIRMM, 161 rue Ada, 34392 Montpellier Cedex 5, France
{bouklit, bennouas, montgolfier}@lirmm.fr

Cet article fournit un nouveau modèle du Web, permettant de détecter les cybercommunautés, de visualiser l'ensemble des pages hypertextes, et d'avoir une mesure d'audience. Il s'inspire du modèle PageRank, les pages étant modélisées comme des particules massives et les liens hypertextes comme des forces gravitationnelles.

Keywords: graphe du Web, graphes petits mondes, modélisation du Web, cybercommunautés, représentation du Web, mesure d'audience.

1 Introduction

Le Web (ensemble des pages hypertextes disponibles sur Internet) est devenu une partie intégrante de la vie quotidienne de millions de gens. La nature même des médias électroniques, ainsi que la volonté de ses inventeurs [BLCL+94] lui ont donné une nature **hypertexte** : les documents sont structurés en **pages**, qui se *pointent* les unes vers les autres, par un système de références.

La croissance exponentielle du Web rend problématique l'appréhension de sa structure globale. Pourtant, une connaissance du contenu et de la structure du Web est indispensable pour réaliser de nombreuses tâches essentielles à la vie de l'internaute, telles que la **recherche d'information** (où trouver une page sur tel sujet ?) ou la **mesure d'audience** (ma page est-elle populaire ?).

Ces problèmes ont conduit les chercheurs à élaborer des modèles et outils divers. Un outil simple, et directement inspiré de la structure hypertexte, consiste à modéliser le Web comme un graphe orienté formé par les pages Web et les liens hypertextes qui les relient. L'analyse de ce **graphe du Web** a permis d'améliorer la performance des moteurs de recherche. 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** [PBMW98]. Le classement des pages est fait en utilisant un indice numérique, le *rang*, calculé pour chaque page [BJM02].

Cette étude propose un outil contribuant à la solution de trois problèmes :

- identifier les **cybercommunautés**, groupes de pages partageant le même centre d'intérêt. Il existe en effet des définitions concurrentes et plus ou moins empiriques, basées sur la sémantique, la co-citation ou des sous-graphes particuliers [Kle98, GKR98, ERC+00].
- fournir un outil de **visualisation** de la structure du Web. Appréhender un aussi vaste objet est une gageure !
- offrir une mesure d'**audience** des pages Web.

Pour ce faire, nous proposons un modèle **particulaire** : les pages Web deviennent des *particules* évoluant dans un espace euclidien tridimensionnel. Les liens hypertextes se traduisent en **forces** gravitationelles s'éxerçant sur ces pages ; ainsi le mouvement de l'ensemble est induit par sa structure hypertexte. Enfin, l'audience d'une page donne un **poids** à la particule.

Nous nous sommes inspirés du modèle cosmologique du Big Bang [Haw88], qui décrit comment la matière, uniformément répartie dans l'univers à son commencement, a été façonnée en galaxie par deux actions : la **gravitation** et l'**expansion**. La première tend à **regrouper** les particules qu'elle lie, tandis que la seconde, dilatation de l'espace (qui *diminue* à mesure que l'univers viellit) tend à **écarter** les particules sans relation. Notre modèle adapte ces deux phénomènes au Web. Ils agissent au cours du temps et isolent les ensembles densément liés de pages, qui conservent la somme de leurs masses et se regroupent en globules. Ils nous permettent de proposer une nouvelle définition *par émergence* des cybercommunautés.

Nous nous sommes inspirés des lois physiques, et en particulier cosmologiques, qui produisent de bonnes métaphores pour décrire le monde du Web. Ainsi, notre modèle fait apparaître une tendance des pages à se regrouper dans l'espace, au gré des forces gravitationnelles subies, en *galaxies*. Elles contiennent des pages de même sujet, spatialement proches les unes des autres : ce sont des cybercommunautés.

L'autorité des particules fournit une autre analogie avec la masse : une page de référence se comportera comme un soleil, immobile autour d'un nuage de planètes ayant trait au même sujet. À notre connaissance, il n'existe pas de travaux antérieurs présentant le problème sous cet angle.

2 Le modèle PageRank

Lancé en 1998, le moteur de recherche Google est devenu un des plus utilisés. Une des clefs de son efficacité est le facteur *PageRank* [PBMW98], un indice numérique (le «rang») qui est attribué à chaque page et reflète sa popularité. Mais comment connaître l'audience d'une page sans avoir de mesure d'accès réelle (comme des compteurs) à sa disposition ? On peut tirer parti de la *structure hypertexte* du Web. Un lien hypertexte vers une page est interprété comme un **vote positif** en faveur de la page pointée. Cette sémantique est vraie pour une grande majorité des liens hypertextes inter-sites. Parmi plusieurs pages traitant d'un même sujet, celle ayant le plus de *liens entrants* est donc supposée être le choix des internautes rédacteurs, et *a fortiori* des internautes surfeurs. On peut ordonner les résultats d'un moteur de recherche selon le degré entrant décroissant.

Mais un tel indice est faible. Il est par exemple facile de rendre une page Web intéressante en créant plusieurs pages fictives qui pointent vers elle. Dans [PBMW98], les auteurs proposent un modèle de *conservation du rang* et un algorithme permettant son calcul : PageRank. Il modélise le comportement d'un surfeur aléatoire, passant d'une page à l'autre au gré des liens hypertextes. Tous les liens sortants d'une page sont supposés équiprobables (cet axiome est discutable et nous ne l'utilisons pas). Le Web devient alors une chaîne de Markov, dont le vecteur stationnaire est la probabilité de présence de l'internaute probabiliste sur une page donnée. Cette mesure est assimilée à la *popularité* de la page ; elle est en tous cas une bonne mesure de son *accessibilité*. Elle est *robuste* aux changements temporels locaux du Web et aux tentatives de *spamming*.

Soit A la matrice telle que $A[p,q] = \frac{1}{d^+(p)}$ s'il existe un lien hypertexte dans la page p vers la page q, et 0 sinon. A est une matrice sous-stochastique que l'on nommera *matrice du Web*. Soit n la dimension de A (nombre de pages Web) et $\overrightarrow{E}$ le vecteur $(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n})^t$. Le vecteur PageRank $\overrightarrow{R}$ est la solution (unique) de l'équation $\overrightarrow{R} = d\, A^t\ \overrightarrow{R} + (1-d)\ \overrightarrow{E}$.

Le terme $(1-d)\ \overrightarrow{E}$, appelé ici *facteur zappe* (car il représente la probabilité, pour notre surfeur aléatoire, de «zapper» vers une page équiprobablement choisie du Web), est ajouté pour assurer l'existence d'une solution (que les pages sans successeurs menaceraient sinon). Il est pris égal à 0.85 par [PBMW98] afin d'accélérer la convergence de l'algorithme. On peut critiquer le fait que le *zappe* suive une loi de distribution uniforme [BJM02]. PageRank peut être vu comme une *distribution* à chaque étape du rang d'une page à toutes les pages qu'elle pointe (cf. figure 1).

3 Cybercommunautés

Une *cybercommunauté* est un ensemble de pages web partageant le même sujet. Cette définition est bien sûr inutilisable, car faisant appel à la subjectivité de chacun. Parfois la volonté explicite des auteurs de pages structure la communauté (webrings, etc) mais souvent ils n'ont pas eux-même clairement conscience de leur cybercommunauté d'appartenance. De nombreuses définitions concurrentes de cette notion ont été proposées, parmi lesquelles on peut citer :

– [GKR98], qui applique la méthode *HITS* de [Kle98] à la définition et à la détection des cybercommunautés. L'ensemble des pages recherché a une structure bipartie : un ensemble de *hubs* (catalogues de pages) pointant un ensemble d'*autorités*, pages de référence du domaine (voir figure 2).

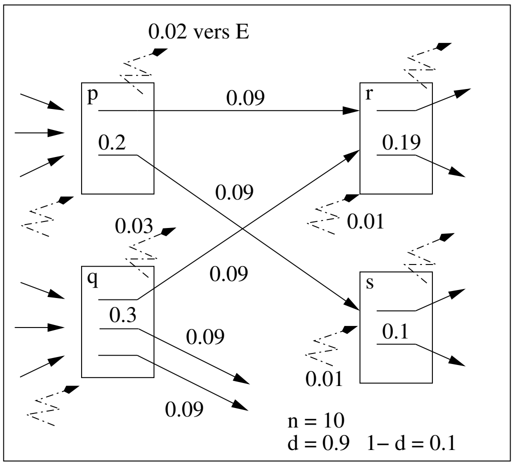


FIG. 1 - Propagation de rang

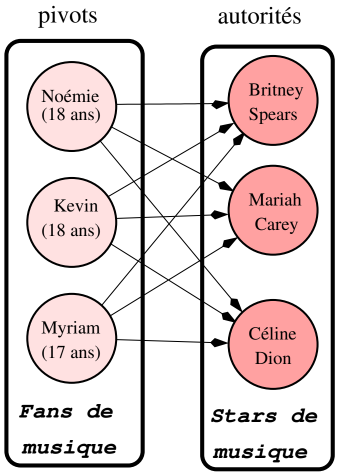


FIG. 2 - Exemple de communauté

– [KRRT99] définit aussi les communautés à partir d'une structure bipartie, le *core*. La *cocitation* est la base de ces deux méthodes.
– [FLG00] définit une cybercommunauté comme un ensemble de pages contenant plus de liens internes que de liens externes (le rapport) étant paramétrable ; un algorithme de coupe minimale permet de les trouver.
– [WS98] définit les graphes *small world* (petits mondes) comme possédant une distance moyenne entre deux sommets courte et une *coefficient de clustérisation* élevé, c'est-à-dire que le voisinage d'un sommet donné contient beaucoup de liens. [Ada99] a mesuré pour des crawls une distance moyenne de 3.1 liens seulement entre deux pages et un coefficient de clustérisation de 0.1, montrant l'adéquation de ce modèle au Web. Ils proposent une définition des communautés par *centres* (pages d'excentricité minimale) et *attracteurs* (pages très liées) dans l'ensemble des résultats d'un moteur de recherche.
– [ERC$^+$00] contient un répertoire de diverses défininfitions.

Dans notre modèle, les communautés sont définies comme étant des pages proches *géographiquement* dans un certain espace (voir 4.4).

4 Description du modèle

4.1 Modélisation de l'Univers

Notre modélisation du Web distingue deux entités. La première est le graphe du Web dont les sommets sont les pages Web. L'arc (p, p') existe si, et seulement si, il existe un lien hypertexte dans la page p pointant la page p'. Ce graphe est la donnée du problème. La deuxième entité est l'**espace** et le **temps** au sein desquels évolue le système. Pour obtenir un modèle se rapprochant le plus possible de la physique, nous avons pris l'espace euclidien $\mathcal{E} = \mathbb{R}^3$, mais les nécessités de l'algorithmique nous ont fait choisir un temps discret $\mathcal{T} = \mathbb{N}$. Les pages y sont présentes en tant que particules massives dotées de coordonnées tridimentionnelles. Augmenter le nombre de dimentions sépare plus les cybercommunautés ; nous pensons que qu'avoir trois dimensions est suffisant et naturel.

4.2 Modélisation des actions

Les forces mettent en mouvement les pages/particules. Une interaction gravitationelle n'a lieu qu'entre pages unies par un lien hypertexte. Cette interaction respecte le principe galiléen *d'action et de réaction* : la force subie par la page pointée est la même que celle subie par la page qui pointe. Le sens de l'hyperlien ne compte que pour le transfert de masse (*cf. infra*). Nous avons utilisé simplement la force de gravitation newtonienne :

$$F_{pq} = \mathcal{G}\frac{m(p).m(q)}{dist(p,q)^2}$$

où p et q sont des particules, et dist(p, q) est la distance euclidienne entre p et q dans $\mathbb{R}^3$.

L'autre action subie par les particules est l'**expansion** de l'univers, qui les sépare au commencement. Nous avons pris une définition où l'univers a un centre O. Un point P de l'espace est translaté en un instant t en suivant

$$\overrightarrow{OP}_{t+1} = (1+\lambda e^{-\alpha t}).\overrightarrow{OP}_t$$

L'expansion s'arrête asymptotiquement (assez vite, car $\sum e^{-\alpha t}$ converge). C'est ce qui nous a conduit à choisir ce modèle d'expansion parmi ceux proposés par la cosmologie.

4.3 Modélisation des transferts de masse

La masse représente l'*autorité* d'une page. Elle varie au cours du temps, ce qui viole le bon sens physique. Le transfert de masse blesse l'intuition dans la mesure où l'énergie ne se conserve pas ; il s'inspirent du modèle PageRank [PBMW98], en y ajoutant la notion de distance.

À chaque étape, une page **répartit** la totalité de sa masse entre les pages qu'elle pointe. La masse circule donc le long des liens hypertextes. Cette circulation respecte la loi des nœuds pour chaque page(si l'on considère une page Web comme un composant électronique et le transfert de masse comme un courant électrique alors le courant en entrée est égale au courant en sortie du composant) ; cela pose un problème pour les pages sans successeur. Une page transfère préférentiellement sa masse à ses proches voisines (renforcement local) ; la proportion de masse transférée est asymptotiquement nulle avec la distance. Enfin, la masse totale du système se conserve. Mais il contribue au renforcement mutuel des pages spatialement proches en cybercommunautés.

La masse d'une page p à l'étape t est définie comme suit :

$$m_t(p) = \sum_{q \text{ pointant } p} \frac{1}{dist_t(p,q)^{\delta}} \frac{m_{t-1}(q)}{S_t(q)} \quad \text{avec} \quad S_t(q) = \sum_{q \text{ pointant } r} \frac{1}{dist_t(q,r)^{\delta}}$$

Cette loi se ramène à celle de PageRank pour $\delta = 0$, la masse étant alors équitablement répartie entre les successeurs de la page, dont $S(q)$ est le degré sortant. Nous utilisons $\delta = 2$ par cohérence avec la force gravitationelle.

Pages sans successeurs

[BJM02] souligne le problème posé par les pages sans successeur, d'où la nécéssité du *facteur zappe* brassant la masse totale du système, accélerant la convergence. Le transfert de masse devient :

$$m_t(p) = d\Big(\sum_{q \text{ pointant } p} \frac{1}{dist_t(p,q)^{\delta}} \frac{m_{t-1}(q)}{S(q)}\Big) + (1-d) \sum_{r \in P} m_{t-1}(r)$$

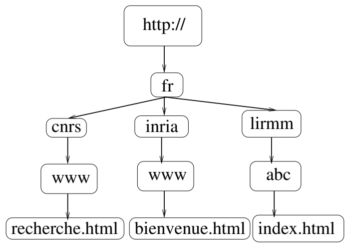
Suivant [PBMW98], nous prenons $d = 0.85$. Signalons enfin que les pages sans successeur et les erreurs d'arrondi font perdre de la masse au système. La masse est donc renormalisée après chaque itération pour que la masse totale se conserve : la masse perdue est redistribuée équitablement. La masse d'une particule ne converge pas en général vers une valeur donnée.

4.4 Cybercommunautés

Nous proposons trois définitions des communatés, basées sur la proximité géographique.

4.4.1 Communautés sphériques

Un k-regroupement des pages est un ensemble de k sphères de l'espace, de rayons variables et pouvant se chevaucher. Une page est *intérieure* au regroupement si elle est contenue dans une sphère. L'ensemble des *k-cybercommunautés* est le k-regroupement qui *maximise* la masse totale des pages intérieures, et *minimise* le volume total des sphères. Il s'agit donc du regroupement des pages qui «colle» le mieux à la structure. Cette définition impose le nombre de communautés et est très dure à calculer (optimisation non linéaire). Nous avons plutôt utilisé des définitions par *regroupement*.



Ordre suivant l'arbre lexicographique.
domaines/serveurs/pages.

FIG. 3 - Arbre des URLs

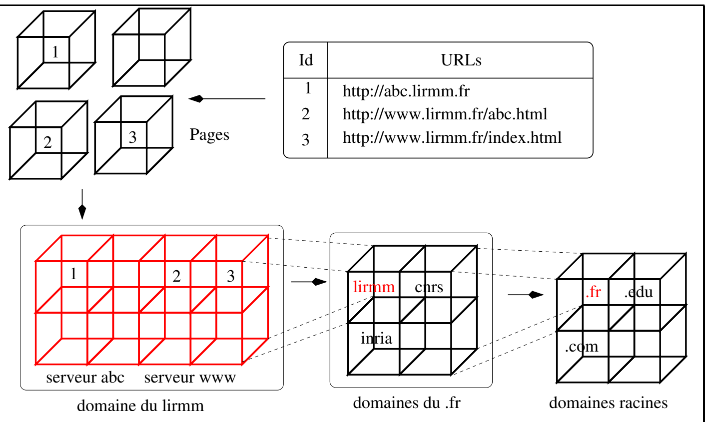


FIG. 4 - Représentation en cubes

4.4.2 Communautés dichotomiques

Soit le couplage du graphe complet des distances qui minimise la longueur totale des arêtes. L'algorithme de calcul remplace chaque arête par son barycentre (tenant compte des deux masses) et recommence avec le graphe obtenu. On forme ainsi une forêt binaire. L'algorithme s'arrete quand la densité des communautés trouvée devient inférieure à une valeur critique.

4.4.3 Communautés par attracteurs

Cette définition suppose que les communautés s'organisent autour de quelques pages centrales de grande masse se trouvant au centre de l'amas. Une valeur arbitraire de masse est fixée, définissant les attracteurs. Un rayon minimal est aussi fixé, en-deça duquel les attracteurs sont supposés être de la même communauté. Ensuite, une page est rattachée à son attracteur le plus proche (sauf si elle est vraiment trop loin de tout attracteur).

4.5 Conditions initiales

Notre modèle renforce la proximité des pages proches ; il est donc très sensible aux conditions initiales. Nous avons pris le parti de faire la répartition initiale selon les sites. Les pages Web du crawl sont d'abord regroupées en un arbre (domaines/serveurs/répertoires). Puis cet arbre est parcouru en largeur. Un nœud à f fils donne naissance à un *cube* dans l'espace, dans lequel chacun de ses fils prend place comme cube de côté $\sqrt[3]{f}$, jusqu'aux feuilles qui donnent les points (voir figures 3 et 4).

Remarquons que les pages d'un même site sont très denses en **liens navigationnels** qui les lient les unes aux autres. Nous avons estimé leur densité à 95 % lors de nos expérimentations ! Ces liens doivent être préalablement enlevés, sous peine de ne détecter que les sites et non les cybercommunautés. Paradoxalement, rien ne lie donc les sommets proches initialement : chacun est libre de migrer vers sa cybercommunauté.

5 Implémentation du modèle

5.1 Implémentation en machine

Le calcul d'une itération (passage de l'instant t à $t+1$) se fait en temps linéaire par rapport au nombre de sommets et d'arcs du graphe. Les liens n'ont pas besoin d'être en mémoire : une seule passe le long du fichier des listes d'adjacence suffit à faire les calculs. Le facteur limitant est la *mémoire vive* plus que le temps, car chaque sommet occupe 32 octets (position, vitesse et masse), limitant à quelques dizaines de millions de sommets les expérimentations. Le programme pourrait facilement être parallélisé pour vaincre cette barrière. Le choix des constantes $\mathcal{G}$, λ, α et d se fait empiriquement. α est fixée pour que l'espace décuple asymptotiquement de volume (ce choix est arbitraire).

5.2 Graphes utilisés

Nous avons utilisé deux sortes de jeux de données : tout d'abord des graphes artificiels. Nous avons en particulier testé des graphes *petits mondes* [WS98, Ada99] qui nous ont permis de vérifier que ces derniers se regroupaient bien en galaxies. Pour ce faire, nous avons utilisé un algorithme de réorientation aléatoire des arêtes proposé par Watts et Strogatz [WS98] permettant de générer un graphe intermédiaire entre un graphe régulier et un graphe aléatoire sans altérer le nombre de sommets dans le graphe. Partant d'un graphe k-régulier (tous ses sommets ont degré k) à n sommets disposé en anneau, l'algorithme réoriente chaque arête avec une probabilité p. Leur construction leur permet de génerer un graphe *petits mondes* intermédiaire entre régularité ($p = 0$) et désordre ($p = 1$).

Nous avons également utilisé des *crawls*, parcours réels d'une partie du Web par des robots [Pag], images forcément incomplètes du Web mais qui en donnent une bonne idée. Le graphe «théorique» et instantané diffère nécéssairement des différents avatars que peut en fournir un crawler ; l'existence des pages dynamiques le rend potentiellement infini.

5.3 Extraction des résultats

Un algorithme implémentant notre modèle fournit facilement une animation vidéo, donnant un résultat assez esthétique. Nous avons implémenté la définition de 4.4.2. La définition 4.4.3 se programme en $O(n^2 \log^2 n)$ ce qui est trop long pour de gros crawls, des heuristiques pour trouver le couplage sont nécéssaires.

6 Résultats

Les figures 5 à 7 présentent l'état typique des particules pour un anneau construit selon la définition des *petits mondes* de [WS98], avec une probabilté p croissante d'arêtes court-circuits. Les graphes aléatoires se dissolvent rapidement dans l'espace (figure 7) tandis les graphes réguliers se resserent en une communauté unique (début du porcessus figure 5). Les authentiques *petits mondes* se regroupent bien en communatés distinctes (figure 6). Les figures suivantes présentes différentes vues de crawls réels En quelques itérations, nous voyons se former à l'écran des cybercommunautés. Par ailleurs, nous avons constaté que 80 % des pages ont tendance à quitter leur emplacement d'origine (site) pour migrer vers une cybercommunauté.

7 Conclusion et perspectives

Partant d'une idée originale de représentation du Web, nous avons aussi trouvé une définition alternative de la notion de cybercommunauté, qui donne des résultats honorables. On a beaucoup dit que le graphe du Web possède une structure de *petits mondes* : notre modèle en fournit une preuve *visuelle*, les mondes en question se formant et tournant effectivement à l'écran !

Le Web est un objet très dynamique, et notre modèle se prête particulièrement bien à l'évolution des liens et des pages. Sur le plan algorithmique, ce n'est en revanche pas aussi simple ; il serait intéressant de travailler dans ce sens.

Un des défis actuels est le **stockage** du graphe du Web, sa structure hypertexte ayant des milliards de liens [BBH$^+$98, HN99, RSWW01]. Les codages à base d'arbre [GLV02] sont une bonne solution. Or, notre méthode de clusterisation fournit un arbre, différent de l'arbre des sites, et encore plus dense en liens : la probabilité d'un lien non-navigationnel entre deux feuilles proches est grande, tandis qu'elle est faible dans l'arbre des sites. Bien sûr, pour les liens navigationnels, c'est l'inverse. Mais une combinaison des deux représentations fournirait assurément un codage hybride d'une grande puissance.

Un autre problème est la construction de **crawlers intelligents** qui parcourent le Web en économisant la bande passante, donc en sélectionnant des pages *a priori* meilleures [CGMP98, NW01]. Toute mesure d'audience permet de faire un choix : par définition, les successeurs d'une page au fort PageRank auront un fort PageRank ; ils doivent être retrouvés prioritairement. Notre modèle propose un autre critère : on peut chercher à obtenir une image d'une *région* du Web, en se concentrant sur les pages proches dans l'espace. À cause de la densité de liens, on obtiendra ainsi rapidement des pages pertinentes.

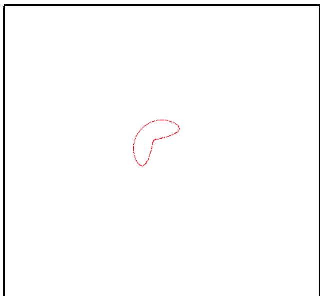


FIG. 5 - $p = 0$

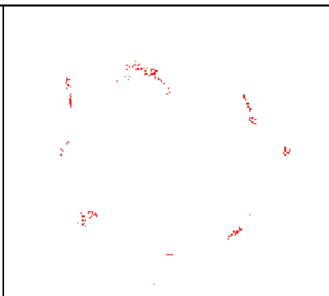


FIG. 6 - $p = 0.25$

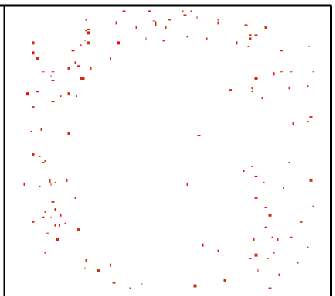


FIG. 7 - $p = 1$

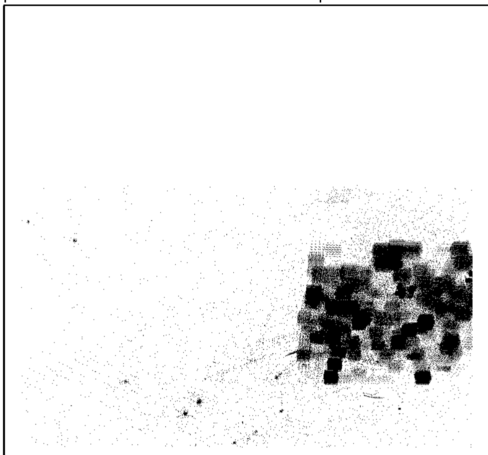


FIG. 8 - crawl de 300000 pages

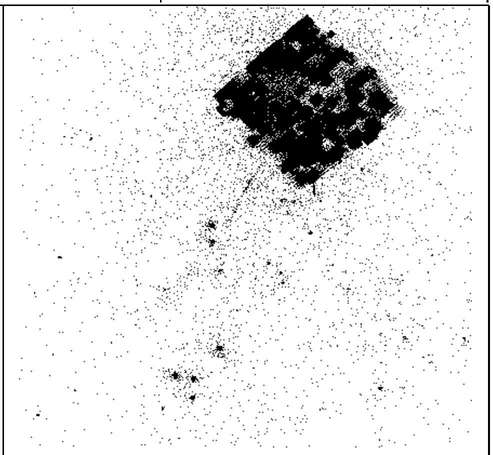


FIG. 9 - autre vue du même instant

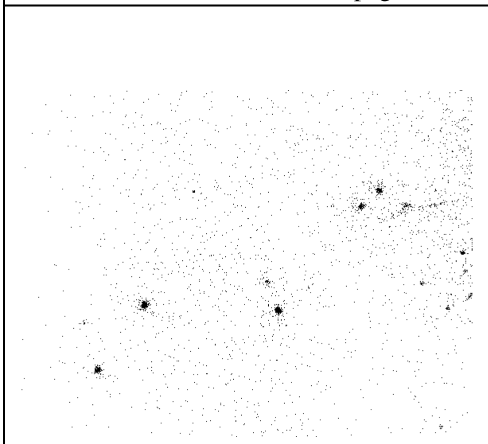


FIG. 10 - Quelques «galaxies»

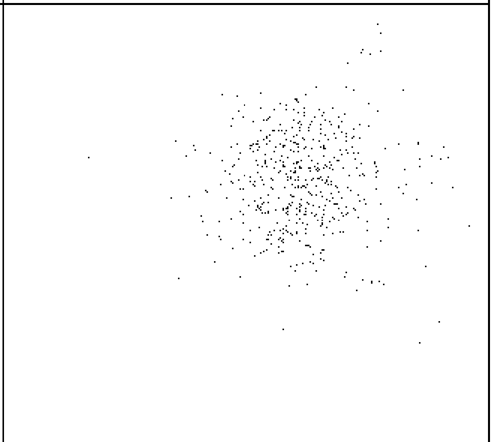


FIG. 11 - Zoom sur un «sytème solaire»

Références

[Ada99] L. Adamic. The small world web. In S. Abiteboul and A.-M. Vercoustre, editors, *Proc. 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL*, number 1696, pages 443–452. Springer-Verlag, 1999.

[BBH+98] K. Bharat, A. Broder, M. Henzinger, P. Kumar, and S. Venkatasubramanian. The connectivity server : Fast access to linkage information on the web. In *Proceedings of the 7th International World Wide Web Conference(WWW7)*, Brisbane, Australia, 1998.

[BJM02] M. Bouklit and A. Jean-Marie. Une analyse de pagerank, une mesure de popularité des pages web. In *Proceedings ALGOTEL'02*, Mèze, France, 2002.

[BLCL+94] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret. The world wide web. *Communications of ACM*, 37(8) :76–82, 1994.

[CGMP98] J. Cho, H. García-Molina, and L. Page. Efficient crawling through URL ordering. *Computer Networks and ISDN Systems*, 30(1–7) :161–172, 1998.

[ERC+00] K. Efe, V. Raghavan, C. Henry Chu, A. Broadwater, L. Bolelli, and S. Ertekin. The shape of the Web and its implications for searching the Web, 2000.

[FLG00] G. Flake, S. Lawrence, and C. Lee Giles. Efficient identification of web communities. In *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 150–160, Boston, MA, 2000.

[GKR98] D. Gibson, J. Kleinberg, and P. Raghavan. Inferring web communities from link topology. In *UK Conference on Hypertext*, pages 225–234, 1998.

[GLV02] J. Guillaume, M. Latapy, and L. Viennot. Efficient and simple encodings for the web graph. In *Proceedings of the 11-th international conference on the World Wide Web*, 2002.

[Haw88] S. W. Hawking. *A Brief History of Time*. Bantam, NY, 1988.

[HN99] A. Heydon and M. Najork. Mercator : A scalable, extensible web crawler. *World Wide Web*, 2(4) :219–229, 1999.

[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, 1998.

[KRRT99] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the Web for emerging cyber-communities. *Computer Networks (Amsterdam, Netherlands : 1999)*, 31(11–16) :1481–1493, 1999.

[NW01] M. Najork and J. Wiener. Breadth-First Crawling Yields High-Quality Pages. In *Proceedings of the 10th International World Wide Web Conference*, pages 114–118, Hong Kong, 2001. Elsevier Science.

[Pag] Larbin Home Page. http ://larbin.sourceforge.net/.

[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.

[RSWW01] K. Randall, R. Stata, R. Wickremesinghe, and J. Wiener. The link database : Fast access to graphs of the web, 2001.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(1–7) :440–442, 1998.