Détection d'utilisateurs malveillants dans les réseaux sociaux

Luc-Aurélien Gauthier — Patrick Gallinari

Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
4, place Jussieu
75005 Paris

{luc-aurelien.gauthier,patrick.gallinari}@lip6.fr

RÉSUMÉ. *L'apprentissage dans les graphes s'est développé avec l'étude des grands réseaux de contenu. Parce que les techniques disponibles reposent largement sur l'utilisation des propriétés des matrices non négatives, l'apprentissage a jusqu'à présent principalement traité du cas de relations non signées. De nombreux problèmes reposent sur l'utilisation conjointe de relations positives et négatives. Nous proposons une étude empirique de quelques méthodes permettant de traiter des problèmes de classification ou d'ordonnancement dans le cas de graphes signés. Ces méthodes sont appliquées au cas de la détection d'utilisateurs malveillants dans le réseau social Slashdot.*

ABSTRACT. *Learning in graphs has become a major field in machine learning with the development and spread of social networks. The methods and applications have mainly focused until now on unsigned graphs. Many practical applications require the consideration and modeling of the simultaneous influence of positive and negative links.We propose here an experimental evaluation of classification and ranking methods in signed graphs. These methods are evaluated for the task of malicious users in the Slashdot social network.*

MOTS-CLÉS : *Réseaux sociaux, graphe signés, marches aléatoires, Slashdot*

KEYWORDS: *Social Network, signed graphs, random walk, Slashdot*

1. Introduction

L'apprentissage dans les graphes s'est principalement développé avec l'émergence des grands réseaux de contenu, ensembles d'objets à fort contenu sémantique (utilisateurs, produits, page web, etc.) partageant différents types de relations. Parmi les problématiques génériques étudiées, on peut citer la classification ou l'ordonnancement relationnels, le clustering, la prédiction de lien, la recommandation. En général, les relations correspondent à des liens positifs entre objets et modélisent des relations de préférence, de similarité, d'influence. Dans un réseau social, on aura des liens avec ses amis ou relations et on participera à des communautés d'intérêts, on n'aura pas de lien avec des individus ou sujets qui ne nous intéressent pas. Les systèmes de recommandation rechercheront des utilisateurs de profil similaire. En pratique, de nombreux réseaux exhibent des liens signés qui traduisent des influences positives et négatives. Les cas d'application sont fréquents, par exemple controverse et discussion dans les applications sociales, antagonismes, inhibition en biologie, etc. Ce n'est pourtant que très récemment que quelques travaux [KUN 09, KER 08, BRZ 08] sur la modélisation et l'utilisation des relations signées ont commencé à apparaître en apprentissage et fouille de données. D'une part, la compréhension de ces relations et des mécanismes sous-jacents est plus complexe que pour les relations de similarités. D'autre part, l'arsenal mathématique développé pour apprendre dans les données relationnelles repose principalement sur l'utilisation de matrices non négatives. C'est le cas des noyaux de graphes, des marches aléatoires, des techniques de régularisation, des méthodes de factorisation matricielle. La modélisation des réseaux signés est largement plus complexe que dans le cas non signé et est très peu développée. En particulier, l'arsenal mathématique est des plus réduits.

Nous nous intéressons ici à des problèmes de classification dans de grands réseaux signés. Nous considérons une problématique de classification qui est la détection d'individus malveillants dans des réseaux sociaux. Le cas d'étude spécifique qui a orienté nos travaux concerne le réseau Slashdot [1]. Ce site publie des documents et permet aux utilisateurs de poster des commentaires et d'indiquer si l'utilisateur apprécie ou n'apprécie pas les commentaires d'autres utilisateurs. Chaque utilisateur peut ainsi étiqueter les autres comme "ami" ou "detracteur". La tâche de classification consiste à identifier des utilisateurs malveillants, utilisateurs qui postent des commentaires malveillants ou faux, du spam, afin de détourner l'utilisation du réseau dans différents buts. La détection de ces malveillants permet de limiter leur influence.

Nous proposons une étude empirique pour faire de la classification dans ces grands graphes signés et orientés. Il s'agit d'une étude exploratoire qui a pour but d'évaluer une classe de méthodes basées sur des marches aléatoires pour cette problématique. Notre point de départ est l'étude publiée sur le corpus Slashdot par [KUN 09]. Notre contribution consiste à proposer et évaluer des améliorations et des variantes d'une série d'algorithmes.

1. http ://slashdot.org/

Nous faisons un état de l'art en section 2. Nous introduisons et discutons brièvement quelques interprétations sémantiques des relations dans les graphes signés en section 2.1 et présentons les travaux en apprentissage et mining en section 2.2. Nous introduisons les méthodes de base qui serviront de référence en section 3. Ensuite nous présentons des améliorations et variantes de ces méthodes en section 4. Les tests expérimentaux sont décrits en section 5.

2. Etat de l'art : apprentissage dans les graphes signés orientés

Les articles traitant de l'apprentissage dans les graphes signés et orientés sont encore rares. Du point de vue de la psychologie ou de la sémantique, les relations négatives apportent une information riche, mais qu'il peut être difficile d'interpréter. Du point de vue des mathématiques, les méthodes à développer ne peuvent utiliser les mêmes bases que les méthodes actuelles qui sont utilisées pour les réseaux non signés.

2.1. *Caractéristiques sémantiques des graphes signés*

Dans le cas des graphes non signés, l'interprétation des relations entre éléments d'un graphe est en général basée sur la notion de similarité ou d'amitié : l'ami de mon ami est mon ami ou qui se ressemble s'assemble. L'interprétation et l'analyse de graphes signés sont plus complexes et elles ont donné lieu à plusieurs travaux en psychologie et, en lien avec ces interprétations, en théorie des graphes. Les travaux en apprentissage ou fouille de donnée utilisent également une interprétation des liens signés, explicite ou implicite. Deux théories issues de la psychologie sont souvent mises en avant, la plus fréquente est la théorie de l'équilibre (*structural balance theory*), l'autre celle du statut. [LES 10b] aborde ces deux aspects.

La théorie de l'équilibre étudie les interactions entre liens positifs et négatifs dans les graphes. Pour cela, on définit le signe d'un cycle comme le produit du signe des arêtes qui le composent. Un graphe sera alors dit équilibré (*balanced*) si tous ses cycles sont de signe positif. Dans une triade (groupe de trois sommets) par exemple, *l'ami de mon ami est mon ami* se traduit par trois liens positifs : (i,j), (i,k) et (i,k). Ou encore *l'ennemi de mon ami est mon ennemi* par un lien positif (i,j) et deux liens négatifs (i,k) et (i,k). En terme de modélisation, cette théorie fournit des interprétations simples comme le signe de la relation entre deux sommets correspond au produit des signes des arcs qui composent le chemin entre les deux sommets. C'est cette interprétation qui est utilisée et validée par exemple dans [KUN 09]. Cette notion de théorie de l'équilibre est largement étudiée dans la littérature. C'est à Frank Harary [HAR 53] que nous devons la notion de graphe signé équilibré, dans son analyse des réseaux sociaux (Fritz Heider avait introduit la notion d'équilibre dans les graphes non signés [HEI 58]). Mais nous trouverons d'autres articles sur l'étude des réseaux sociaux, comme dans [BRZ 08], et dans des articles de théorie des graphes (appliqués ou non aux réseaux sociaux) comme [BAT 94] qui construit et étudie des semi-anneaux pour savoir si un graphe signé est équilibré. La théorie du statut est utilisée dans [LES 10b] sous la

forme de relations de dominance. Si A est meilleur que B et B meilleur que C, alors C n'est pas meilleur que A. (triade ++-) Attention toutefois, la relation de dominance n'est pas transitive. Ces statuts sont ensuite propagés sur les chemins signés.

2.2. *Apprentissage et fouille de données dans les graphes signés*

Comme mentionné il n'y a que peu de travaux sur la fouille de graphes signés. [YAN 07] étudie la découverte de communauté. [KER 08] propose une extension de PageRank au cas des graphes signés, [GUH 04] propose des mécanismes de propagation de la confiance également basés sur des méthodes itératives. [LES 10b] réalise une analyse exploratoire sur plusieurs réseaux de contenu signés et adapte la théorie du statut pour mieux modéliser les observations. [KUN 09] propose un ensemble de mesures (clustering, centralité, popularité) et les teste sur des tâches de classification et de prédiction de liens. [LES 10a] et [SYM 10] étudient la prédiction de liens signés.

3. Marches aléatoires - Méthodes de base

3.1. *Marches aléatoires*

Nous introduisons un ensemble de méthodes qui nous serviront de point de départ et de comparaison. Ce sont toutes des marches aléatoires. Ces méthodes se basent sur la construction de processus stochastique de Markov par le biais de matrices de transition. L'idée est de s'appuyer sur l'ergodicité d'une matrice pour associer au système un vecteur invariant de probabilités traduisant le score de chaque sommet. La caractéristique principale de la méthode PageRank [PAG 99] est de travailler directement sur les matrices d'adjacence de graphes orientés non fortement connexe (comme le graphe du web). Cela implique de transformer la matrice d'adjacence en une matrice ergodique.

Nous pouvons construire pour tout graphe orienté une matrice stochastique $\bar{P}$ à partir de la matrice d'adjacence A. On pose tout d'abord $P_{ij} = A_{ij} / \sum_j A_{ij}$. Si aucun sommet n'est isolé, alors la matrice P est stochastique. Puis, nous construisons une nouvelle matrice $\bar{P}$ stochastique et ergodique en ajoutant à chaque étape de la marche aléatoire une probabilité d'aller en un sommet du graphe non nécessairement relié à notre point actuel. Nous parlerons alors de saut aléatoire. Soit :

$$\bar{P} = \alpha P + (1 - \alpha) e v^T$$

$\bar{P}$ étant ergodique, la suite (P^n) converge vers une matrice strictement positive, notons cette matrice P^*. Nous avons par ailleurs, par le théorème de Perron, que la matrice $\bar{P}$ admet 1 comme valeur propre dominante (1 car la matrice est stochastique). Soit λ le vecteur propre associé, alors on a : $\lambda P = \lambda$. C'est ici que notre limite P^* va nous servir. Posons δ un vecteur quelconque, on pose alors $\delta P^* = \lambda$. Donc $\lambda P = \delta P^* P = \lambda$. Ce qui nous donne l'algorithme 1, détaillé ici car il nous sert par la suite. Cette probabilité invariante correspond aux scores des sommets.

Algorithm 1 Computing PageRank

Entrée : P la matrice stochastique d'un graphe orienté non-signé et E le vecteur uniforme unité. Sortie : R le vecteur de scores.
$R_0 \leftarrow E$
while $\delta > \varepsilon$ **do**
  $R_{i+1} \leftarrow PR_i$
  $d \leftarrow \|R_i\|_1 - \|R_{i+1}\|_1$
  $R_{i+1} \leftarrow R_{i+1} + dE$
  $\delta \leftarrow \|R_i - R_{i+1}\|_1$
end while
return R

Pour plus de détails sur l'algorithme lui-même, voir l'article original [PAG 99] ainsi que [BER 05] et [LAN 04].

3.2. *Les méthodes de base utilisées pour la prédiction d'utilisateurs malveillants*

Les méthodes utilisées pour notre détection d'utilisateurs malveillants sont introduites dans leur version de base dans [KUN 09]. Nous les décrivons brièvement ci-dessous. Nous considérerons simplement A comme notre matrice d'adjacence, décomposable en $A = A^+ - A^-$ (les deux matrices sont positives, l'une correspond aux relations positives, l'autre aux relations négatives).

3.2.1. *PageRank (PR)*

L'algorithme Pagerank peut être utilisé tel que, même s'il n'est pas adapté aux graphes signés. [KUN 09] applique l'algorithme PageRank sur la matrice d'adjacence absolue $\bar{A}$ définie par $\bar{A}_{ij} = |A_{ij}|$. Cette méthode renvoie alors une mesure de centralité, plutôt que d'impopularité. Pour obtenir un véritable classement des utilisateurs malveillants, nous avons également appliqué l'algorithme PageRank à la matrice A^- qui, rappelons le, est positive et représente la matrice d'ajacence du graphe partiel ne considérant que les relations négatives. Ici, ce sont les utilisateurs avec beaucoup de liens négatifs incidents et beaucoup de liens négatifs sortants que l'on veut caractériser comme malveillants. Ces derniers seront donc les utilisateurs de plus forts scores dans cette version de PageRank.
Que ce soit avec $\bar{A}$ ou A^-, la matrice est rendue stochastique (voir section précédente) afin d'utiliser l'implémentation de PageRank décrite ci-dessus (Algorithme 1).

3.2.2. *Signed Spectral Ranking (SR)*

Cet algorithme est une simple extension de PageRank au cas signé. On remplace la matrice d'adjacence absolue $\bar{A}$ par la matrice d'ajacence signée A dans l'algorithme 1. Même si il n'y a pas de garantie théorique, empiriquement, nous observons la convergence de la méthode dans le cas signé. Le produit AR_i renvoie un vecteur dont la

norme L1 est inférieure à celle renvoyé dans le cas non signé. Ainsi, un poids plus important est accordé au vecteur saut (par le biais du coefficient *d* dans l'algorithme). Un poids important pour *d* peut rendre ergodique la matrice signée et assurer la convergence. Cependant, cette ergodicité s'obtient au prix d'un lissage de la matrice signée initiale. Avec les graphes signés, les utilisateurs malveillants caractérisés par de nombreux liens négatifs incidents et sortants seront ceux avec les scores les plus faibles éventuellement négatifs, contrairement au cas du graphe non signé précédant. Les utilisateurs "normaux" auront des connexions positives et des scores positifs d'autant plus forts qu'ils sont populaires. La même remarque vaut pour la méthode NR décrite ci-dessous.

3.2.3. *Negative Rank (NR)*

Cette méthode proposée dans [KUN 09] revient à calculer le vecteur correspondant à la différence entre le vecteur invariant obtenu par PageRank et celui obtenu par Signed Spectral Ranking. L'idée de cette méthode est que les utilisateurs malveillants ont souvent des scores différents suivant ces deux méthodes alors que les autres ont des scores similaires. Pour modifier l'influence du vecteur PageRank sur le vecteur Signed Spectral Ranking, nous y ajoutons un paramètre β. Soit : $NR = SR - \beta PR$. Tout comme SR, ce sont les utilisateurs de plus faible score qui sont considérés comme malveillants.

4. Améliorations

Les méthodes appliquées sur notre corpus sont des méthodes générales et l'algorithme utilisé pour la marche aléatoire est un algorithme prévu au départ pour le cas non signé. En considérant la nature de notre problème, il est possible d'apporter des améliorations notables pour la détection d'utilisateurs malveillants. Nous proposons quatre types d'améliorations basés sur le critère d'incidence des noeuds, la normalisation du vecteur de score, le critère d'arrêt de l'algorithme itératif et la modification du vecteur saut.

4.1. *Le critère d'incidence des noeuds*

Nous supposons qu'un utilisateur sera considéré comme malveillant s'il donne un nombre suffisant d'avis négatifs et/ou s'il en reçoit un nombre suffisant. Nous proposons de ne garder pour notre classification que les utilisateurs ayant un degré incident supérieur à un certain seuil fixé empiriquement.

4.2. *Normalisation du vecteur de score*

Dans le but de faire varier l'importance du vecteur saut uniforme (notre d dans l'algorithme 1), nous proposons d'introduire un coefficient de normalisation dépendant de la norme du vecteur, et de tester plusieurs normes, L_1, L_2 et L_∞. Du point de vue de notre algorithme 1, nous ajoutons après l'étape 4 :

$$R_{i+1} = \frac{R_{i+1}}{\|R_{i+1}\|_X}$$

où X peut correspondre à la norme L_1, L_2 ou L_∞.

L'ajout de la normalisation peut induire des problèmes de divergence pour la méthode SR. Afin de pallier le problème, nous avons tout d'abord considéré que l'algorithme devait s'arrêter dès que le δ de l'algorithme PageRank augmentait. Cela a donné de bons résultats. Mais, nous proposons un autre critère d'arrêt, décrit ci-dessous.

4.3. *Critère d'arrêt*

Comme nous avons pu le voir dans la section sur la sémantique des graphes signés, sous l'interprétation de la sémantique d'équilibre, le signe de la relation entre deux noeuds dépend du signe de tous les chemins entre les deux points (de façon inversement proportionnelle à leur taille). Mais, dans un grand graphe comme celui étudié ci-après, représentatif du réseau social Slashdot, considérer tous les chemins entre deux sommets reviendrait à considérer la suite (P^n), qui n'est certainement pas convergente. Nous pourrions avoir, eventuellement, des sous-suites convergentes. Cependant, entre deux élèments consécutifs de la suite, les signes changent souvent. Un réseau social équilibré est tout simplement inconcevable avec un nombre aussi conséquent d'intéractions. En cherchant trop loin dans les relations d'amitié, il est facile de trouver des incohérences. C'est pourquoi nous proposons de limiter la longueur maximale des chemins entre deux sommets à considérer pour évaluer les scores. Autrement dit, nous limitons notre algorithme à un nombre restreint d'itérations.

4.4. *Vecteur de sauts non-uniforme*

Un des points sur lequel nous pouvons facilement jouer pour améliorer les résultats est de modifier le vecteur saut utilisé lors de la marche aléatoire. Nous changeons ce vecteur saut uniquement dans le cas signé, car l'algorithme SR bénéficie de la meilleure marge de progression - PR n'est pas vraiment approprié car son cadre d'application (matrice positive) le limite fortement.

Dans l'algorithme de base, ce vecteur saut est uniforme. Nous proposons de modifier ce vecteur saut pour qu'il prenne davantage en compte l'influence et l'importance des sommets. L'intuition veut que l'on favorise les visites sur les utilisateurs influents qui ont un comportement suspect.

Du point de vue de notre algorithme, certains changements sont nécessaires. Nous gardons les améliorations précédentes : critère d'incidence, normalisation et critère d'arrêt. Nous y ajoutons une nouvelle étape (voir algorithme 2 ligne 7) permettant de moduler le poids donné au vecteur saut. Ce poids noté μ, est compris entre 0 et 1. Nous utilisons également un vecteur de saut non uniforme qui est simplement le score PageRank classique (ligne 9).

Algorithm 2 Computing SR$(L_X)i+$

Entrée : P la matrice stochastique d'un graphe orienté signé, $\bar{P}$ sa version non-signée et E le vecteur uniforme unité, I la nombre d'itérations à effectuer et μ le poids accordé au vecteur saut. Sortie : R le vecteur de scores.
$R_0 \leftarrow E$
$V \leftarrow PageRank(\bar{P})$
while (#iter < I) **do**
  $R_{i+1} \leftarrow PR_i$
  $R_{i+1} \leftarrow \frac{R_{i+1}}{\|R_{i+1}\|_X}$
  $R_{i+1} \leftarrow ((1+\mu)\|R_i\|) \times R_{i+1}$
  $d \leftarrow \|R_i\|_1 - \|R_{i+1}\|_1$
  $R_{i+1} \leftarrow R_{i+1} + dV$
end while
return R

5. Expérimentations

5.1. *Les données Slashdot*

Slashdot est un site d'actualité sous forme de réseau social sur les sujets de l'informatique, des jeux vidéos ou encore de la science. Chaque utilisateur peut y proposer des articles et noter les autres utilisateurs (positivement ou négativement). Nous avons donc affaire à un réseau orienté (un utilisateur i donne son avis sur l'utilisateur j) et signé (avis positif ou négatif). Nous avons utilisé le corpus Slashdot Zoo (étudié dans [KUN 09] et [LES 10b]) qui nous a été fourni par le laboratoire *Distributed Artificial Intelligence Laboratory* de Berlin. Il correspond à un sous-ensemble du réseau Slashdot : 77 985 noeuds (utilisateurs) et 510 157 arcs. Notre but étant la prédiction d'utilisateurs malveillants, nous utilisons une liste fournie avec le corpus et comprenant 96 utilisateurs malveillants connus.

5.2. *Les résultats des algorithmes de base*

Pour évaluer les performances des différents algorithmes, nous avons choisi de calculer la précision et le rappel au rang 50, 100 et 200 et la *Précision Moyenne* au rang 100 (AP@100). Nous avons choisi ce seuil de 100 pour la précision moyenne, car d'une part il permet pour un système idéal de caractériser les 96 utilisateurs malveillants et d'autre part, il permet de comparer de façon claire les différents algorithmes.

Les résultats pour les algorithmes de base sont présentés dans le Tableau 1. La méthode NR a des résultats nettement meilleurs que les deux autres pour cette tâche. Elle reconnait davantage d'utilisateurs malveillants. Les résultats présentés dans les tableaux pour l'algorithme PageRank sont ceux obtenus en considérant la matrice

	PR	SR	NR
Précision@50	0,14	0,24	0,42
Précision@100	0,17	0,24	0,36
Précision@200	0,18	0,16	0,26
Rappel@50	0,08	0,12	0,22
Rappel@100	0,17	0,20	0,37
Rappel@200	0,38	0,34	0,54
AP@100	0,03	0,05	0,14

Tableau 1. *Résultats pour les algorithmes originaux. PR PageRank, SR Signed Ranking, NR Negative Rank*

	PRi	SRi	NRi
Précision@50	0,20 (+42%)	0,26 (+8%)	0,46 (+9%)
Précision@100	0,18 (+6%)	0,26 (+8%)	0,40 (+11%)
Précision@200	0,18 (+0%)	0,18 (+12%)	0,27 (+4%)
Rappel@50	0,11 (+37%)	0,14 (+16%)	0,24 (+9%)
Rappel@100	0,19 (+12%)	0,27 (+35%)	0,42 (+14%)
Rappel@200	0,37 (-3%)	0,37 (+9%)	0,55 (+2%)
AP@100	0,04 (+33%)	0,07 (+40%)	0,18 (+28%)

Tableau 2. *Résultats pour les algorithmes avec le critère d'incidence. Entre parenthèses, le pourcentage d'amélioration par rapport aux méthodes de base.*

d'ajacence absolue $\bar{A}$. En fait, en considérant uniquement la matrice A^-, les résultats sont très légèrement meilleurs pour notre reconnaissance. Ignorer les signes des arcs donne des performances plus faibles, ce qui est naturel. Cependant, nous utiliserons $\bar{A}$ par la suite. Les méthodes NR et SR+ (avec saut non uniforme) obtiennent de meilleurs résultats avec $\bar{A}$ comme matrice d'ajacence pour PR. Cela s'explique par le fait que ces deux méthodes utilisent PR pour pénaliser davantage les comportements malveillants des utilisateurs influents. Ce qui s'obtient avec $\bar{A}$, car PR donne une mesure de centralité, plutôt que d'impopularité (avec A^-).

5.3. *Améliorations*

Nous présentons dans la suite les scores obtenus avec chacune des améliorations proposées dans la section précédente. Nous pourrons visualiser les différents gains en reconnaissance dans les tableaux 2 et 3 ainsi que sur le graphique final.

5.3.1. *Le critère d'incidence*

Nous avons intégré le critère d'incidence aux trois méthodes de bases. Les scores obtenus sont indiqués sur le tableau 2. Le gain en qualité de reconnaissance est important ; pour chacun des trois algorithmes, le score MAP - notre score de reconnaissance - augmente de 30 à 40%, la précision de 8 à 12% en moyenne et le rappel de 10 à 25%, en moyenne également.

5.3.2. *Critère d'arrêt*

Nous avons choisi de fixer le critère d'arrêt à 5 (les résultats varient peu entre 1 et 8 itérations avant de se dégrader). Autrement dit, nous nous intéressons uniquement à un voisinage limité à une distance de 5 arêtes. Ce critère n'est nécéssaire qu'avec l'ajout de la normalisation dans la méthode Signed Spectral Ranking, car la normalisation provoque une divergence de la méthode. Nous présentons les résultats dans les sous-sections suivantes.

5.3.3. *Normalisation*

Nous avons testé les performances des algorithmes avec chacune des trois normes usuelles : L_1, L_2 et L_∞. Cette normalisation profite seulement à Signed Spectral Ranking qui travaille avec des vecteurs score non nécessairement stochastiques, contrairement à PR. Les performances pour les différentes normes sont représentées sur la figure 1 quand on utilise le critère d'incidence + la normalisation. La norme L1 donne les meilleurs résultats. Le tableau 3 donne la précision et le rappel des méthodes PR, SR et NR avec normalisation par norme L1 et critère d'incidence (colonnes SR(L_1)i et NR(L_1)i). Les gains obtenus varient suivant les méthodes. Par rapport à l'utilisation du critère d'incidence seul, ils sont importants pour l'algorithme SR, plus faibles pour l'algorithme NR.

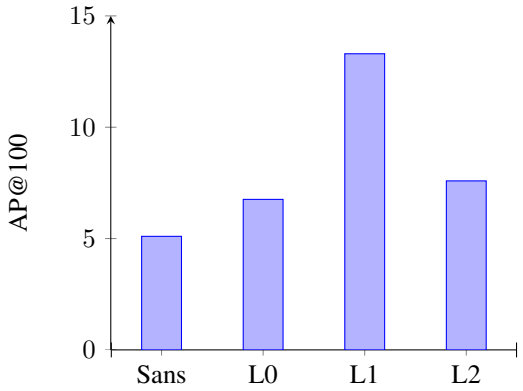

Figure 1. *Scores (AP@100) obtenu avec Signed Spectral Rank en fonction de la normalisation choisie.*

5.3.4. *Vecteur de saut non-uniforme*

Nous avons utilisé l'algorithme 2 défini dans la section précédente. Nous avons fixé expérimentalement μ à 0.3. Les gains apportés par cette méthode sont très intéressants, cela nous permet d'atteindre un score AP@100 de 0.33. À comparer aux 0.15 de la méthode Négative Rank, considérée comme la meilleure avant les améliorations. Nous donnons dans le tableau 3, les différents scores obtenus avec cette méthode. Nous remarquons cependant que si cette méthode permet d'augmenter substantiellement les scores pour Signed Spectral Ranking, la méthode Négative Rank n'est pas améliorée.

	SR(L_1)i	NR(L_1)i	SR(L_1)i+
Précision@50	0,46 (+91%)	0,50 (+19%)	0,64 (+166%)
Précision@100	0,38 (+58%)	0,38 (+5%)	0,48 (+100%)
Précision@200	0,23 (+44%)	0,29 (+11%)	0,29 (+81%)
Rappel@50	0,24 (+100%)	0,26 (+18%)	0,33 (+175%)
Rappel@100	0,40 (+100%)	0,40 (+8%)	0,50 (+150%)
Rappel@200	0,48 (+41%)	0,60 (+11%)	0,59 (+73%)
AP@100	0,20 (+300%)	0,21 (+50%)	0,33 (+560%)

Tableau 3. *Résultats pour l'algorithme avec sauts non-uniforme noté (+), comparés aux algorithmes avec normalisation et critère d'incidence noté (L_1)i. Entre parenthèses, le pourcentage d'amélioration par rapport aux méthodes de base.*

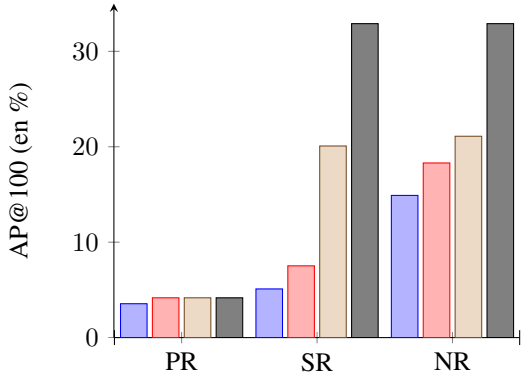

Figure 2. *Récapitulatif des scores AP@100 pour la prédiction d'utilisateur malveillants. Pour chaque méthode, les scores des algorithmes originaux (bleu clair), avec critère d'incidence (rouge) avec normalisation (beige) et saut non-uniforme (gris)*

6. Conclusion

La difficulté de l'apprentissage dans les graphes signés orientés provient de la sémantique des relations négatives et des outils mathématiques qui ne sont que partiellement étudiés dans la littérature. Si la prise en compte de ces liens négatifs est importante dans beaucoup de domaines, l'arsenal dont nous disposons ne semble pas à la hauteur. À travers notre étude exploratoire, nous avons tenté d'expérimenter quelques pistes. Les résultats que nous avons obtenus sont intéressants, ce qui ouvre quelques perspectives de recherche. La connaissance de certaines caractéristiques du graphe permet d'appliquer des algorithmes efficaces. Il reste cependant de nombreux problèmes à régler pour développer des méthodes formelles solides.

Les futurs recherches porteront sur la construction d'algorithmes convergents, semblable à PageRank, ne necessitant pas de critère d'arrêt. Cette convergence ne peut pas être obtenue pour le moment, compte tenu de la nature déséquilibrée des réseaux

construits. Une première approche pourrait ainsi être de rendre la graphe équilibré en élaborant une méthode plus complexe de saut aléatoire. Il pourrait également être intéressant d'étudier les sous-suites de (P^n). Peut-être convergent-elles vers des limites pouvant traduire une certaine séparation de notre espace de sommets.

7. Bibliographie

[BAT 94] BATAGELJ V., « Semirings for social networks analysis », *The Journal of Mathematical Sociology*, vol. 19, n° 1, 1994, p. 53-68.

[BER 05] BERKHIN P., « A Survey on PageRank Computing », *Internet Mathematics*, vol. 2, 2005, p. 73–120.

[BRZ 08] BRZOZOWSKI M. J., HOGG T., SZABO G., « Friends and foes : ideological social networking », *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, New York, NY, USA, 2008, ACM, p. 817–820.

[GUH 04] GUHA R., KUMAR R., RAGHAVAN P., TOMKINS A., « Propagation of trust and distrust », *Proceedings of the 13th conference on World Wide Web - WWW '04*, , 2004, page 403, ACM Press.

[HAR 53] HARARY F., « On the notion of balance of a signed graph », *Michigan Math J*, vol. 2, 1953, p. 143–146.

[HEI 58] HEIDER F., *The Psychology of Interpersonal Relations*, Wiley, New York, 1958.

[KER 08] KERCHOVE C. D., DOOREN P. V., « The PageTrust algorithm : How to rank web pages when negative links are allowed ? », *the SIAM International Conference on Data Mining*, , 2008, p. 346–352.

[KUN 09] KUNEGIS J., LOMMATZSCH A., BAUCKHAGE C., « The Slashdot Zoo : Mining a Social Network with Negative Edges », *Proceedings of the 18th international conference on World wide web*, 2009.

[LAN 04] LANGVILLE A. N., MEYER C. D., « Deeper Inside Pagerank », *Internet Mathematics*, vol. 1, 2004, page 2004.

[LES 10a] LESKOVEC J., HUTTENLOCHER D., KLEINBERG J., « Predicting positive and negative links in online social networks », *Proceedings of the 19th international conference on World wide web*, WWW '10, New York, NY, USA, 2010, ACM, p. 641–650.

[LES 10b] LESKOVEC J., HUTTENLOCHER D., KLEINBERG J., « Signed networks in social media », *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, , 2010, page 1361, ACM Press.

[PAG 99] PAGE L., BRIN S., MOTWANI R., WINOGRAD T., « The PageRank Citation Ranking : Bringing Order to the Web. », Technical Report n° 1999-66, November 1999, Stanford InfoLab, Previous number = SIDL-WP-1999-0120.

[SYM 10] SYMEONIDIS P., TIAKAS E., MANOLOPOULOS Y., « Transitive node similarity for link prediction in social networks with positive and negative links », *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, New York, NY, USA, 2010, ACM, p. 183–190.

[YAN 07] YANG B., CHEUNG W., LIU J., « Community Mining from Signed Social Networks », *IEEE Trans. on Knowl. and Data Eng.*, vol. 19, 2007, p. 1333–1348, IEEE Educational Activities Department.