
Influence dans Twitter

Définition et utilisation en recherche d'information sociale

Massinissa Ferrouk, Fatiha Boubekour, Lila Belkacemi

*Université Mouloud Mammeri de Tizi-Ouzou,
BP 17 RP, 15000 Tizi-Ouzou, Algérie*

amesnsen@yahoo.fr ; fboubekour2002@yahoo.fr ; belkacemi.lila@gmail.com

RÉSUMÉ. L'influence d'un blogueur sur Twitter véhicule une part importante de la pertinence de ses publications, elle peut donc être exploitée avantageusement en recherche d'information pour identifier les tweets pertinents. L'information sur l'influence des blogueurs est implicitement portée par la structure du réseau social de Twitter. Dans ce papier, nous exploitons cette structure pour mesurer l'influence. En particulier, nous proposons une approche adaptée de l'algorithme PageRank, pour l'analyse des liens du réseau social de rediffusion et le calcul de l'influence. Dans cette approche, l'influence du blogueur sur le réseau social est définie comme le rapport entre sa capacité à influencer d'autres blogueurs (influence imposée) et sa disposition à être influencé par d'autres blogueurs du réseau (influence subie). Nous évaluons notre score d'influence dans un contexte de RI sociale. Nous proposons alors un modèle de recherche d'information qui estime la pertinence d'un tweet comme une combinaison de sa pertinence thématique et de l'influence de son auteur. Ce modèle de recherche a été évalué sur la collection de tests TREC-Microblog 2011. Les résultats obtenus montrent tout son intérêt.

ABSTRACT. The blogger's influence on Twitter conveys an important part of his publications relevance, it can therefore be advantageously exploited in information retrieval (IR), to identify relevant tweets. The information about blogger's influence is implicitly driven by Twitter's social network structure. In this paper, we exploit this structure to measure influence. In particular, we propose an adapted approach of the PageRank algorithm for analyzing retweet-based social network links and calculating influence. In this approach, the blogger's influence in the social network is defined as the ratio of his ability to influence other bloggers (imposed influence) to his predisposition to be influenced by other bloggers in the network (suffered influence). We evaluate our ratio of influence in a context of social IR. We thus propose a social retrieval model which estimates the tweet relevance as a combination of its topic relevance and its author's influence. This retrieval model has been evaluated on TREC-microblog 2011 test collection. The obtained results show all its interest.

MOTS-CLÉS : influence, réseau social de Twitter, recherche d'information sociale, PageRank.

KEYWORDS: influence, Twitter social network, social information retrieval, PageRank.

DOI:10.3166/ISI.22.6.9-36 © 2017 Lavoisier

1. Introduction

L'évolution d'internet, et en particulier l'avènement du web 2.0, a permis l'émergence fulgurante d'une multitude de réseaux sociaux à vocations diverses (forums, wikis, blogs, microblogs, etc.). Parmi ces réseaux sociaux, la plateforme de microblogging Twitter constitue sans doute le réseau social le plus populaire sur le web avec ses quelques 313 millions d'utilisateurs (ou blogueurs) actifs par mois, et 500 millions de tweets¹ publiés par jour dans plus de 40 langues². Twitter constitue une véritable source d'informations fournies et partagées par des millions d'utilisateurs à travers le monde. La grande quantité d'informations non contrôlées et librement publiées sur Twitter, pose le problème de la qualité des messages publiés (ou tweets). La qualité d'un tweet est fortement corrélée à la *qualité* de son contenu et à l'*importance* de son auteur (*ie.* le blogueur qui l'a publié). Dans un contexte de RI, la qualité du contenu d'un tweet se réfère à sa pertinence pour un sujet donné formellement exprimé par une requête de recherche (Romero *et al.*, 2011; Anger *et al.*, 2011). Tandis que l'importance d'un blogueur se réfère généralement à son influence sur le réseau social. Cependant, il n'existe aucun moyen objectif pour dire quel blogueur est influent (Bogdan *et al.*, 2011), et les différentes approches existantes proposent d'utiliser divers critères du réseau social de Twitter pour estimer cette influence.

Dans cet article, nous proposons une nouvelle approche pour mesurer l'influence d'un utilisateur sur Twitter. A l'instar de nombreuses approches existantes, notre approche exploite les propriétés structurelles du réseau social de Twitter basé sur la relation de *retweet* pour estimer cette influence. La relation de *retweet* est en effet connue pour être un bon indicateur d'influence dans Twitter (Cha *et al.*, 2010 ; Anger et Kittl., 2011 ; Azaza *et al.*, 2015). Cependant, alors que les approches existantes considèrent que l'influence d'un blogueur est liée à ses tweets qui ont été retweetés, notre proposition se base autant sur les tweets que le blogueur a retweeté, que sur ses propres tweets qui ont été retweetés par d'autres blogueurs. L'idée est qu'un blogueur est d'autant plus influent qu'il est retweeté par beaucoup d'autres utilisateurs (il *impose* donc son *influence* à ces blogueurs) sans que lui-même ne retweete beaucoup d'autres (il *ne subit* donc *pas l'influence* des autres blogueurs). Nous mesurons alors l'influence d'un blogueur par son *ratio d'influence*, calculé comme le rapport de son *influence imposée* à son *influence subie*. L'*influence subie* définit une sorte de *passivité* du blogueur au sens de (Romero *et al.*, 2011). De plus, notre approche, à l'instar d'autres approches existantes, tient compte de la propagation d'influence sur le réseau social de Twitter. Le modèle de propagation utilisé est inspiré de l'algorithme *PageRank* (Page *et al.*, 1999), un algorithme bien connu, utilisé pour mesurer à la fois la pertinence et la présence des sites web sur internet (Riquelme et González-Cantergiani, 2016). Pour évaluer l'impact de l'influence sur la recherche d'information, nous proposons un modèle de recherche

1. Messages publiés sur Twitter.

2. Statistiques au 15 mars 2017, Source : Wikipédia.

sociale qui combine *ratio d'influence* et pertinence thématique, pour estimer la pertinence d'un tweet relativement à une requête de recherche donnée.

Le reste de ce papier est organisé comme suit. La section 2 introduit une brève description des concepts de Twitter. Dans la section 3 nous présentons quelques travaux de recherche sur l'influence dans Twitter. La section 4 détaille notre contribution. La section 5 décrit les expérimentations effectuées pour évaluer notre approche et présente les résultats obtenus. La section 6 conclut le papier et présente quelques perspectives.

2. Une brève introduction à Twitter

Twitter est la plateforme de microblogging la plus populaire sur le web avec ses quelques 313 millions d'utilisateurs actifs par mois, et 500 millions de tweets publiés par jour dans plus de 40 langues. Les utilisateurs de Twitter (ou blogueurs) partagent, à travers la plateforme, des informations (librement accessibles) sous forme de messages courts, ne dépassent pas 140 caractères, appelés *tweets*. Un tweet peut contenir du texte libre, des images ou vidéos, des *liens* (généralement compressés) vers des pages web, des *hashtags* et/ou des *mentions*. Un *hashtag* est un mot ou une suite de mots accolés précédés par un croisillon (#), servant à marquer dans le tweet courant, un événement, une date, ou une thématique bien précise. *#ElectionPrésidentielle* est un exemple de hashtag, et un tweet contenant ce hashtag est lié à l'élection présidentielle. Une *mention* dans un tweet est une référence à un utilisateur de Twitter. La mention est constituée d'une arobase @ suivie par le nom (identifiant unique délivré par Twitter) de l'utilisateur référencé. Une mention est placée soit en début d'un tweet, pour indiquer une réponse à un tweet publié par l'utilisateur référencé, soit au milieu ou en fin de tweet, pour indiquer que l'utilisateur référencé est concerné par le sujet du tweet courant.

Outre la publication et le partage de ses propres tweets, un utilisateur de Twitter peut autrement interagir avec les autres utilisateurs de la plateforme à travers une multitude de pratiques sociales existantes. Ainsi, un utilisateur peut aimer (*like*) – ou marquer comme *favoris* – un tweet publié par un autre utilisateur. Il peut aussi y répondre (*reply*). Un *reply* est un tweet envoyé en réponse directe à un autre tweet. Un utilisateur peut aussi s'abonner à d'autres utilisateurs de la plateforme afin de suivre (*to follow*) en temps réel leurs publications. L'utilisateur abonné est dit *follower*. Les utilisateurs auxquels il est abonné sont ses *amis* ou *followees* (ou *followings*). La relation d'abonnement confère au *follower* la possibilité de recevoir des messages directs (privés) de ses *followees*. Par ailleurs, un utilisateur peut rediffuser (ou *retweeter*) un tweet initialement publié par un autre utilisateur. Le *retweet* (i.e. le message *retweeté*) est envoyé à ses propres *followers*. La figure 1 résume ces interactions entre utilisateurs dans Twitter³.

3. <https://about.twitter.com/fr/company/brand-assets>

L'ensemble des utilisateurs de Twitter et de leurs multiples interactions forme le réseau social de Twitter. Ce réseau social peut être représenté par un multi-graphe $G = (U, E)$ où U est l'ensemble des blogueurs et E est l'ensemble de toutes les interactions (via les relations d'abonnement, mentions, retweets, et like) entre blogueurs. Cependant, de nombreuses approches d'analyse du réseau social de Twitter utilisent des instances de ce réseau basées généralement sur une seule relation d'interaction à la fois (à l'exemple de la relation de *retweet*, ou de la relation d'abonnement), représentant ainsi le réseau social de Twitter par un graphe simple.

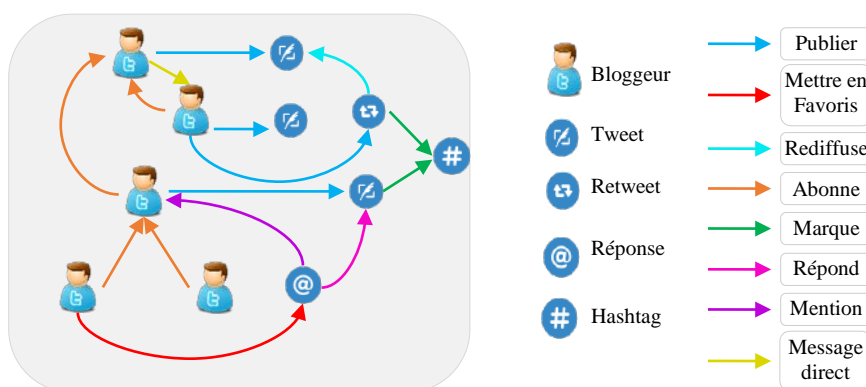


Figure 1. Les interactions dans Twitter

3. Etat de l'art

L'influence peut être définie comme la capacité d'une personne à s'imposer aux autres par la simple force de ses idées, et/ou de sa popularité, et/ou de sa notoriété. Un utilisateur d'un réseau social est influent si ses actions dans le réseau sont capables d'affecter les actions de nombreux autres utilisateurs du réseau (Riquelme et Gonzalez-Cantergiani, 2016). Autrement dit, les actions de l'utilisateur influent font réagir de nombreux autres utilisateurs du réseau. La réaction de l'utilisateur influencé consistera soit à suivre l'utilisateur influent en s'y abonnant et/ou en le mentionnant, soit à lire ses messages (i.e. ses tweets) et à y répondre et/ou à les rediffuser (ou retweeter). L'abonnement, les mentions, les réponses, les retweets, etc. sont autant d'indicateurs de réaction qui ont été largement utilisés dans de nombreux travaux de recherche en vue d'estimer l'influence d'un blogueur dans Twitter.

Ainsi, en se basant sur l'abonnement (*follow*), le *retweet* et la *mention*, les auteurs dans (Cha *et al.*, 2010) ont défini trois mesures de l'influence d'un utilisateur dans le réseau social de Twitter : *indegree*, *retweet* et *mention*. *Indegree* désigne le nombre d'abonnés d'un blogueur donné u sur le réseau. *Retweet* désigne le nombre de fois que les autres utilisateurs ont rediffusé les tweets publiés par u . Et

mention désigne le nombre de fois où les autres utilisateurs ont mentionné l'utilisateur u . Les auteurs ont montré d'une part que les facteurs de *retweet* et de *mention* sont fortement corrélés à l'influence (et sont donc de bons indicateurs de l'influence), et d'autre part que le nombre d'abonnés (à travers *indegree*) n'était pas un bon indicateur de l'influence.

Les auteurs dans (Anger et Kittl, 2011) se sont également intéressés aux relations d'abonnement, de *retweet* et de *mention*, à travers les mesures d'influence suivantes issues de l'état de l'art: le *ratio de followers* $r_f = (\#abonnés/\#abonnement)$, le *ratio de retweets* $r_{RT} = (\#retweets + \#mentions/\#tweets-publiés)$, et le *ratio d'interaction* $r_i = (\#utilisateurs-qui-retweetent-ou-mentionnent/\#followers)$. Les auteurs ont expérimenté ces approches pour retrouver et ordonner les blogueurs autrichiens les plus influents. Les résultats expérimentaux obtenus pour chacune de ces mesures ont été comparés entre eux, puis avec ceux de Klout⁴, un service en ligne populaire, pour la mesure de l'influence sur Twitter. Ces résultats, tous différents, montrent (comme le soulignent les auteurs) qu'il n'y a pas de consensus sur les indicateurs d'influence sur Twitter. Cependant, les auteurs suggèrent que le *ratio de followers* (ou d'abonnés) utilisé seul, n'est pas un bon indicateur d'influence. En effet, de nombreux blogueurs s'abonnent à d'autres blogueurs pour gagner soit un abonnement réciproque, soit une contrepartie financière (cas des organisations qui paient pour augmenter leurs audiences sur Twitter). Partant de l'hypothèse que c'est principalement les interactions positives et les réactions qui déterminent l'influence d'un utilisateur, les auteurs proposent une nouvelle mesure de l'influence, dite *potentiel de réseautage social (Social Networking Potential – SNP)* calculée comme la moyenne des *ratios de retweets* et d'*interaction* respectivement r_{RT} et r_i . Les auteurs ont utilisé leur score SNP pour retrouver et ordonner les blogueurs autrichiens les plus influents. Les résultats obtenus sont proches de ceux de Klout.

Les auteurs dans (Azaza *et al.*, 2015) combinent sept marqueurs d'influence (*Mentions, Retweets, Hashtags, URL, Réponses, Followers* et *Favoris*) dans une nouvelle mesure d'influence multicritère. Les auteurs montrent que les utilisateurs les plus influents selon les critères du nombre de *followers* ne sont pas nécessairement les candidats les plus influents en termes de *retweets* ou de *mentions*.

Pour mesurer l'influence d'un utilisateur dans le réseau de Twitter, les auteurs dans (Zhang *et al.*, 2011) proposent une approche basée sur les actions (*Retweets* et *Réponses*). L'idée est que, si un utilisateur influent retweete ou répond à un autre utilisateur, alors ce dernier est probablement influent. En outre, plus un utilisateur retweete ou répond à d'autres utilisateurs, plus son influence se propage. Les auteurs modélisent la propagation d'influence sur les réseaux de retweets (*retweet network*) et de réponses (*Reply network*) de Twitter respectivement, en utilisant l'algorithme *PageRank*.

4. <http://klout.com/>

La propagation d'influence à travers le réseau social de Twitter a été abordée dans de nombreuses autres approches qui ont adapté l'algorithme *Pagerank* (Tunklang, 2009 ; Weng *et al.*, 2010 ; Benjabeur *et al.*, 2011 ; Gayo-Avello, 2013 ; Sung *et al.*, 2013 ; Cataldi et Aaufaure, 2014) ou l'algorithme *HITS* (Romero *et al.*, 2011) pour le calcul de l'influence propagée.

Dans (Tunklang, 2009), l'auteur propose une mesure de l'influence d'un blogueur calculée sur la base du nombre de ses abonnés (*followers*) et du nombre de ses amis (*following*). Partant de l'idée qu'un utilisateur est d'autant plus influent qu'il est suivi (*followed*) par d'autres utilisateurs influents, et que lui-même ne suit pas beaucoup d'autres utilisateurs du réseau, l'auteur propose d'estimer l'influence d'un blogueur comme le rapport de ses *followers* à ses *followings* (i.e. $\frac{\#followers}{\#followings}$). L'influence est propagée sur le réseau social basé sur la relation d'abonnement. L'algorithme de propagation proposé, *TunkRank*, est une adaptation de l'algorithme *PageRank*.

Les auteurs dans (Weng *et al.*, 2010) ont noté qu'il existait une forte réciprocité entre blogueurs sur le réseau social d'abonnement. Cette réciprocité est due à l'*homophilie*. Des utilisateurs *homophiles* sont susceptibles d'avoir des sujets d'intérêt similaires. Sur la base de cette hypothèse, les auteurs proposent une extension de l'algorithme *PageRank*, appelée *TwitterRank*, qui exploite l'*homophilie* sur le réseau social d'abonnement pour mesurer l'influence propagée d'un blogueur. Cependant, les auteurs dans (Cha *et al.*, 2010) ont contredit ce travail en arguant que la quasi-totalité des données de Twitter présente une faible réciprocité.

Après étude du réseau social de Twitter basé sur la relation de *retweet*, les auteurs dans (Romero *et al.*, 2011) ont relevé que la majorité des utilisateurs sont *passifs*. Un utilisateur *passif* est un utilisateur qui ne *retweete* pas (ou peu) les messages des autres. La passivité d'un utilisateur mesure sa résistance vis-à-vis de l'influence des autres utilisateurs (Romero *et al.*, 2011). Partant de l'hypothèse qu'un utilisateur est fortement influent si 1) il a une grande influence sur de nombreux utilisateurs passifs, et 2) il est lui-même passif pour les autres utilisateurs influents, les auteurs ont proposé un algorithme inspiré de *HITS* (Kleinberg, 1999), dit *Influence-Passivity algorithm*, qui exploite la structure du réseau social de *retweet*, pour estimer simultanément la propagation d'influence sur le réseau et la passivité des utilisateurs du réseau.

Dans (Benjabeur *et al.*, 2011; 2012), les auteurs exploitent également la structure du réseau social de *retweet* pour calculer l'influence des blogueurs. Le réseau social de *retweets* dit *réseau social d'influence*, est représenté par un graphe $G = (U, E)$ où U est l'ensemble des utilisateurs et $E = U \times U$ représente l'ensemble des relations d'influence entre eux. Une relation d'influence $e(u_i, u_j)$ entre deux utilisateurs u_i et u_j (dans ce sens) traduit le fait que u_i a rediffusé au moins un tweet publié par u_j . Les relations d'influence sont pondérées. Le poids $w(u_i, u_j)$ est défini comme le rapport du nombre de tweets rediffusés par u_j et qui sont publiés par u_i , au nombre total de tweets rediffusés par u_i . Pour calculer le score d'influence, les auteurs utilisent

l'algorithme *PageRank* sur le réseau social d'influence. Le score d'influence est itérativement calculé selon la formule suivante:

$$Inf_G(u_i) = d \times \frac{1}{|U|} + (1-d) \times \sum_{u_i: e(u_i, u_j) \in E} w(u_i, u_j) \frac{Inf_G(u_j)}{O(u_j)} \quad (1)$$

Où $O(u_j)$ représente le nombre de relations d'influence à partir de l'utilisateur u_j , et $d \in [0, 1]$ un paramètre de configuration de l'algorithme *PageRank*. Les auteurs ont ensuite défini une mesure de l'expertise d'un blogueur en s'appuyant sur son profil Twitter. En particulier, l'ensemble des tweets d'un blogueur u est évalué (du point de vue de la RI), relativement à un sujet donné exprimé sous forme de requête Q . Le score de pertinence du profil représente l'expertise, $Exp(Q, u)$, du *blogueur*. Finalement, les auteurs ont défini un modèle de recherche d'information sociale dans Twitter. Dans ce modèle, la pertinence d'un tweet t pour un sujet donné formellement exprimé dans une requête de recherche Q , est estimée comme une combinaison de sa pertinence thématique d'une part, et la pertinence sociale de son auteur u_i d'autre part. La pertinence thématique d'un tweet mesure son adéquation au sujet de la requête, tandis que la pertinence sociale d'un blogueur est définie comme une combinaison de son expertise et de son influence.

Dans (Ding, 2013), l'influence d'un utilisateur de Twitter est mesurée à travers sa *propagabilité* (*spreadability*), soit sa capacité à propager l'information sur le réseau social. L'algorithme proposé, *SpreadRank* mesure cette propagabilité en se basant sur le réseau de rediffusion des tweets. Dans leur approche, les auteurs ont d'abord extrait de Twitter, un ensemble de *cascades d'informations* concernant des blogueurs chinois. Une cascade d'information est un graphe de diffusion d'un tweet, dont la racine est l'auteur du tweet, et dont les autres nœuds sont les utilisateurs qui, successivement, ont retweeté ce tweet. Un lien d'un nœud A vers un nœud B du graphe de diffusion, signifie que B a diffusé le tweet (ou retweet) de A. Dans ce graphe, les liens sont pondérés par le *ratio de retweets* (i.e. $\#retweets/\#tweets$ de l'utilisateur). L'idée est de mesurer la proportion (plutôt que le nombre) de tweets qui ont été retweetés. Intuitivement, plus ce taux est levé, plus grande est la propagabilité (et donc l'influence) de l'utilisateur. Par ailleurs, plus un utilisateur apparaît dans un grand nombre de cascades d'information, plus sa propagabilité est grande, et plus la taille d'une cascade d'information est grande, plus grande est la propagabilité de l'utilisateur associé. Ces deux caractéristiques définissant la localisation d'un utilisateur u dans l'ensemble des cascades d'informations. Par ailleurs, en analysant l'impact du délai de retweets, les auteurs ont relevé que la propagabilité d'un utilisateur est d'autant plus grande que ses tweets sont retweetés très tôt après leur diffusion. Les auteurs ont alors combiné toutes les cascades d'information pour construire le graphe de rediffusion des blogueurs considérés. Ce graphe est utilisé pour calculer la propagabilité sur Twitter. L'algorithme *SpreadRank*, utilisé pour ce calcul, est une adaptation de *PageRank* qui tient compte

à la fois, de l'intervalle de temps entre les *retweets* et des localisations des utilisateurs dans les *cascades d'information*.

Dans (Nargundkar et Rao, 2016), les auteurs proposent un modèle prédictif pour mesurer l'influence des utilisateurs de Twitter: le système *InfluenceRank*. Ce système se base sur un ensemble de caractéristiques extraites des profils Twitter des utilisateurs, et de leurs tweets publiés sur une durée de deux mois. Plutôt que de considérer tout Twitter, les auteurs ont manuellement composé un échantillon représentatif, avec 180 blogueurs les plus influents – chacun sur un sujet spécifique (politique, technologie, musique, sport, etc.) –. Les auteurs ont alors récupéré, pour chaque utilisateur de l'échantillon, les caractéristiques 1) de son profil Twitter, et 2) de ses tweets (publiés sur la période considérée). Les caractéristiques du profil sont : le score d'activité (*activity score*) – représenté par le nombre de tweets collectés sur la période considérée –, le nombre d'abonnés (*followers_count*), et le nombre de listes publiques auxquelles l'utilisateur est affilié (*liste_count*). Tandis que les caractéristiques des tweets sont: le ratio de retweets (*trt_t_ratio*), le ratio de favoris (*tf_t_ratio*), le ratio abonnés-amis (*followers_friends_ratio*), la moyenne de retweets des tweets de l'utilisateur (*rt_trt_ratio*), et la moyenne de favoris (*f_tf_ratio*). Ces caractéristiques sont ensuite linéairement combinés pour former le score d'influence *InfluenceRank*, comme suit:

$$\begin{aligned} \text{InfluenceRank} = & a \times \text{trt_t_ratio} + b \times \text{rt_trt_ratio} + c \times \text{tf_t_ratio} + d \times \text{f_tf_ratio} \\ & + e \times \text{followers_friends_ratio} + f \times \text{activity_score} \\ & + g \times \text{followers_count} + h \times \text{listed_count} \end{aligned}$$

Les coefficients *a*, *b*, *c*, *d*, *e*, *f*, *g*, et *h* sont déterminés par une approche d'apprentissage automatique basée sur la régression avec *InfluenceRank* comme variable prédictive par rapport à l'ensemble des caractéristiques utilisées. Bien qu'expérimenté sur un échantillon de données réduit, le système proposé a montré une précision acceptable.

4. Contribution

Dans ce papier, nous proposons une nouvelle mesure de l'influence d'un blogueur sur Twitter : le *ratio d'influence*. Le *ratio d'influence* d'un utilisateur est défini comme le rapport entre 1) son influence sur les autres utilisateurs du réseau, et 2) l'influence des autres utilisateurs sur lui. Pour calculer ce ratio, notre approche, déploie une adaptation de l'algorithme PageRank sur le réseau social d'influence – selon l'appellation de (Benjabeur *et al.*, 2011) – construit à partir du réseau social de *retweets*. Nous évaluons l'efficacité de notre *ratio d'influence* dans un contexte de recherche d'information sociale dans Twitter. Pour cela, nous proposons un modèle de recherche qui tient compte de l'influence. En particulier, ce modèle de recherche calcule la pertinence d'un tweet comme une combinaison de son score thématique, d'une part et du *ratio d'influence* de son auteur, d'autre part. Dans ce qui suit, nous

décrivons la structure du réseau social d'influence, puis nous détaillons nos propositions.

4.1. Le réseau social d'influence

Le réseau social de Twitter est principalement formé à partir des utilisateurs et des relations (*abonnement*, *retweet*, *mention*, etc.) qui les lient. Nous nous basons sur une instance particulière de ce réseau qui est le réseau social de rediffusion (ou réseau social de *retweets*) pour construire notre *réseau social d'influence* comme suit (cf. figure 2).

Le fait qu'un blogueur u_j rediffuse un message publié par un blogueur u_i , dénote une relation d'influence de u_i sur u_j . Le réseau social d'influence est modélisé par un graphe orienté $G = (U, E)$ où U est l'ensemble des blogueurs du réseau de rediffusion, et $E = U \times U$ est l'ensemble des relations d'influence entre eux. Dans ce réseau, une relation d'influence $e(u_i, u_j)$ d'un blogueur u_i vers un blogueur u_j traduit donc le fait que le blogueur u_j a rediffusé au moins un tweet du blogueur u_i . La figure 2 suivante explicite la relation d'influence qui naît de la relation de *retweet* entre deux blogueurs.



Figure 2. Relation de rediffusion (*retweet*) vs relation d'influence

4.2. Degré d'influence

Les relations dans le réseau social d'influence sont pondérées. Le poids $w(u_i, u_j)$ d'une relation d'influence $e(u_i, u_j)$ entre deux blogueurs u_i et u_j (dans cet ordre), exprime le degré d'influence du blogueur u_i sur le blogueur u_j . Dans (Benjabeur *et al.*, 2011), ce poids est défini comme le ratio de tweets rediffusés par u_j et qui sont publiés par u_i , soit :

$$w(u_i, u_j) = \frac{\text{nombre de tweets publiés par } u_i \text{ et rediffusés par } u_j}{\text{nombre de tweets rediffusés par } u_j} \quad (2)$$

Cette approche présente un inconvénient majeur. En effet, un blogueur u_j qui rediffuse très peu de tweets de u_i et seulement ceux-là, aura un plus fort poids qu'un blogueur u_j qui rediffuse tous les tweets de u_i , en plus de quelques autres. En

particulier, un blogueur qui rediffuse un et un seul tweet de u_i et seulement celui-là est considéré comme hautement influencé par u_i , alors qu'un blogueur qui a retweeté tous les tweets de u_i en plus de quelques autres est considéré moins influencé par u_i que le précédent.

À l'instar de (Ding, 2013), nous considérons que plus un blogueur u_j rediffuse de tweets du blogueur u_i , plus il est influencé par u_i . Nous calculons alors le degré d'influence $w(u_i, u_j)$ de u_i sur u_j comme le taux de tweets de u_i qui sont rediffusés par u_j . Formellement :

$$w(u_i, u_j) = \frac{\text{nombre de tweets publiés par } u_i \text{ et rediffusés par } u_j}{\text{nombre de tweets publiés par } u_i} \quad (3)$$

REMARQUE – Un blogueur u_i qui n'a publié aucun tweet, a une influence nulle sur tous les autres blogueurs u_j , d'où $w(u_i, u_j) = 0$.

4.3. Le ratio d'influence

Partant de l'intuition qu'un blogueur est d'autant plus influent qu'il influence beaucoup d'autres utilisateurs (eux-mêmes *influentes sur* et *influencés par* d'autres utilisateurs) sans être lui-même fortement influencé, nous proposons une nouvelle mesure de l'influence d'un blogueur, dite *ratio d'influence*, définie comme le rapport entre sa capacité à influencer d'autres blogueurs (*i.e.* son *influence imposée*) et sa disposition à être influencé par d'autres blogueurs (*i.e.* son *influence subie*). Formellement :

$$\tau_{inf}(u_i) = \frac{Infl_i(u_i)}{Infl_s(u_i)} \quad (4)$$

Tel que :

- $\tau_{inf}(u_i)$: *ratio d'influence* du blogueur u_i dans le réseau social d'influence ;
- $Infl_i(u_i)$: *influence imposée* par u_i sur les autres blogueurs du réseau social d'influence ;
- $Infl_s(u_i)$: *influence subie* par le blogueur u_i sur le réseau social d'influence ;

On en déduit deux propriétés fondamentales du *ratio d'influence* :

- Si l'*influence imposée* égale l'*influence subie*, le *ratio d'influence* égal à 1.
- Si l'*influence imposée* est inférieure (resp. supérieure) à l'*influence subie*, le *ratio d'influence* est inférieur (resp. supérieur) à 1.

La formule (4) présente néanmoins des insuffisances. En effet, si l'*influence subie* d'un blogueur u_i est nulle, le *ratio d'influence* devient indéfini (division par zéro). Pour pallier cet inconvénient, nous introduisons une constante de *lissage* au dénominateur. Or, la constante de *lissage* ainsi introduite invalide les propriétés du *ratio d'influence* (en particulier, le *ratio d'influence* n'est plus égal à 1 en cas d'égalité entre l'*influence imposée* et l'*influence subie*). Ce qui nous impose

d'introduire cette même constante au numérateur. Finalement, le *ratio d'influence* sera calculé comme suit :

$$\tau_{inf}(u_i) = \frac{(inf\ l_i(u_i) + \theta)}{(inf\ l_s(u_i) + \theta)} \quad (5)$$

Où θ est une constante quelconque. Nous avons choisi de l'initialiser à 1.

Ainsi l'importance d'un blogueur est proportionnelle à l'influence qu'il *impose* aux autres et inversement proportionnelle à l'influence qu'il *subit* des autres blogueurs. Dans ce qui suit, nous définissons formellement les facteurs *d'influence imposée* et *d'influence subie*, respectivement $inf\ l_i()$ et $inf\ l_s()$.

4.3.1. L'influence imposée

En nous basant sur la seule relation de rediffusion dans Twitter, nous définissons l'influence imposée comme la capacité d'un blogueur à faire propager ses tweets à travers le réseau social de Twitter par le biais d'autres blogueurs. La propagation des tweets est matérialisée par l'action de rediffusion (ou *retweet*). En particulier, plus les tweets d'un blogueur u_i sont rediffusés par des blogueurs importants, plus il y a de chances que ces tweets soient largement relayés sur le réseau social (en particulier par les nombreux autres blogueurs qui leur sont abonnés), ce qui a pour effet d'augmenter l'influence de u_i sur le réseau.

Le score d'influence d'un blogueur u_i sur l'ensemble des blogueurs u_j qui l'ont rediffusé dans le réseau est formalisé comme suit :

$$inf\ l_i(u_i) = \sum_{u_j: e(u_i, u_j) \in E} w(u_i, u_j) * \tau_{inf}(u_j) \quad (6)$$

4.3.2. L'influence subie

Nous définissons l'influence subie par un blogueur u_i comme étant l'influence que les autres blogueurs du réseau, chacun caractérisé par son influence propre, lui imposent. Ces blogueurs sont ceux dont les messages ont été rediffusés, au moins une fois, par u_i . L'influence subie d'un blogueur u_i est ainsi mesurée à travers ses rediffusions des tweets publiés par les autres blogueurs.

L'influence subie est formellement définie comme suit :

$$inf\ l_s(u_i) = \sum_{u_j: e(u_j, u_i) \in E} w(u_j, u_i) * \tau_{inf}(u_j) \quad (7)$$

4.3.3. Algorithme de calcul du ratio d'influence

L'algorithme de calcul du *ratio d'influence* (ou algorithme CRI) permet de calculer le *ratio d'influence* d'un blogueur connaissant les ratios d'influence des autres blogueurs (qu'il influence, ou qui l'influencent), eux même influencés par

d'autres blogueurs (qu'ils retweetent) ou influençant d'autres blogueurs (qui les retweetent), et ainsi de suite... Le calcul d'influence est alors itératif. L'algorithme démarre à partir d'un ratio initial quelconque pour tous les blogueurs (nous considérons pour notre part, une valeur initiale égale à 1), puis, le *ratio d'influence* est calculé de proche en proche, itérativement. Les itérations concernent plus particulièrement le calcul de l'influence subie et de l'influence imposée. Chaque itération fait converger les résultats vers une valeur de plus en plus précise. La valeur de cette précision, notée ε , est choisie initialement. Pour un blogueur u_i , la valeur du ratio calculée à l'itération k (soit $\tau_{inf}(u_i)^{(k)}$) est utilisée pour calculer le ratio à l'itération $k+1$ (soit $\tau_{inf}(u_i)^{(k+1)}$). L'algorithme converge, si pour tout blogueur u_i , on a $|\tau_{inf}(u_i)^{(k+1)} - \tau_{inf}(u_i)^{(k)}| \leq \varepsilon$. L'algorithme CRI est présenté en table suivante (*Algorithm 1*), où U représente l'ensemble des blogueurs du réseau social d'influence, E l'ensemble des relations d'influence, et ε la précision de calcul souhaitée. La variable booléenne *Nonconvergence* permet de contrôler la convergence (ou non) de l'algorithme, et conditionne ses itérations.

Algorithm 1. Algorithme CRI

CalculRatioInfluence (U, E, ε)
{Début
1. **Pour tout** $u_i \in U$ **faire**
2. $\tau_{inf}(u_i) \leftarrow 1$ //Initialisation du ratio d'influence à 1
3. **Fin Pour**
4. *Nonconvergence* \leftarrow **Vrai**
5. **Tant que** (*Nonconvergence* = **Vrai**) **faire**
6. *Nonconvergence* \leftarrow **Faux**
7. **Pour tout** $u_i \in U$ **faire**
8. $inf\ l_i(u_i)^{(k+1)} = \sum_{u_j: e(u_i, u_j) \in E} w(u_i, u_j) \times \tau_{inf}(u_j)^{(k)}$
9. $inf\ l_s(u_i)^{(k+1)} = \sum_{u_j: e(u_j, u_i) \in E} w(u_j, u_i) \times \tau_{inf}(u_j)^{(k)}$
10. $\tau_{inf}(u_i)^{(k+1)} = \frac{inf\ l_i(u_i)^{(k+1)} + 1}{inf\ l_s(u_i)^{(k+1)} + 1}$
11. **Si** $|\tau_{inf}(u_i)^{(k+1)} - \tau_{inf}(u_i)^{(k)}| > \varepsilon$ **alors**
12. *Nonconvergence* \leftarrow **Vrai**
 FinSi
 Fin Pour
 Fin Tant que
Fin}

4.3.4. Illustration

Considérons le mini-réseau d'influence de la figure 3, composé de 5 blogueurs A, B, C, D et E, et des relations d'influence entre eux. On suppose que les nombres de tweets publiés par A, B, C, D et E sont respectivement 50, 35, 70, 45 et 60.

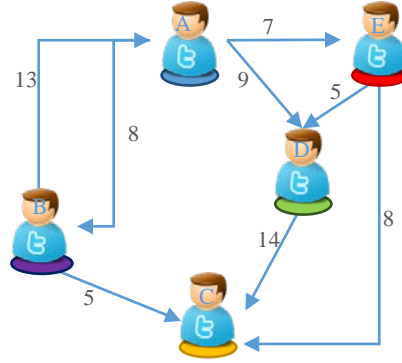


Figure 3. Un mini réseau d'influence

Dans ce réseau, chaque relation orientée ($u_i \rightarrow u_j$), est étiquetée par le nombre de tweets de u_i qui sont retweetés par u_j , et traduit l'influence de u_i sur u_j . Le poids d'influence de u_i sur u_j est calculé par la formule (3). Pour calculer le ratio d'influence de chacun des blogueurs du réseau, l'algorithme CRI est itéré jusqu'à la convergence.

Ainsi :

$$\text{A l'étape } k = 0 : \tau_{inf}(A)^{(0)} = \tau_{inf}(B)^{(0)} = \tau_{inf}(C)^{(0)} = \tau_{inf}(D)^{(0)} = \tau_{inf}(E)^{(0)} = 1$$

A l'étape $k = 1$:

$$\begin{aligned} inf_{l_i}(A)^{(1)} &= \sum_{u_j: e(A, u_j) \in E} w(A, u_j) \times \tau_{inf}(u_j)^{(0)} \\ &= w(A, B) \times \tau_{inf}(B)^{(0)} + w(A, E) \times \tau_{inf}(E)^{(0)} + w(A, D) \times \tau_{inf}(D)^{(0)} \\ &= 0.16 * 1 + 0.14 * 1 + 0.18 * 1 = 0.48 \end{aligned}$$

$$\begin{aligned} inf_{l_s}(A)^{(1)} &= \sum_{u_j: e(u_j, A) \in E} w(u_j, A) \times \tau_{inf}(u_j)^{(0)} \\ &= w(B, A) \times \tau_{inf}(B)^{(0)} \\ &= 0.3714 * 1 = 0.3714 \end{aligned}$$

$$\text{D'où : } \tau_{inf}(A)^{(1)} = \frac{inf_{l_i}(A)^{(1)} + 1}{inf_{l_s}(u_i)^{(1)} + 1} = \frac{1.48}{1.3714} = 1.07918$$

Considérant une précision de calcul $\varepsilon = 10^{-5}$, on a :

$$|\tau_{inf}(A)^{(1)} - \tau_{inf}(A)^{(0)}| = 0.07918 > \varepsilon : \text{ la convergence n'est pas encore atteinte.}$$

On poursuit le calcul du ratio d'influence pour chacun des autres blogueurs du réseau, puis on réitère l'algorithme à l'ordre 2... et ainsi de suite jusqu'à convergence de l'algorithme. Le tableau 1 présente les ratios d'influence calculés à

chaque itération. La convergence est obtenue à l'itération (7), pour la précision considérée.

Tableau 1. Calcul itératif du ratio d'influence

| Itération | A | B | C | D | E |
|-----------|---------|---------|---------|---------|---------|
| 0 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1,07917 | 1,31639 | 0,61256 | 0,93189 | 1,00717 |
| 2 | 1,02043 | 1,26069 | 0,62332 | 0,94187 | 1,01640 |
| 3 | 1,03084 | 1,26353 | 0,62148 | 0,93946 | 1,01471 |
| 4 | 1,02996 | 1,26318 | 0,62188 | 0,93978 | 1,01489 |
| 5 | 1,03007 | 1,26324 | 0,62183 | 0,93974 | 1,01487 |
| 6 | 1,03005 | 1,26323 | 0,62184 | 0,93975 | 1,01487 |
| 7 | 1,03005 | 1,26323 | 0,62184 | 0,93975 | 1,01487 |

4.4. Utilisation de l'influence en recherche d'information dans Twitter

Nous proposons d'utiliser notre score d'influence dans un contexte de recherche d'information (RI) dans Twitter. La RI a pour objectif de retrouver, parmi une collection de documents, ceux qui sont pertinents pour un besoin informationnel d'un utilisateur, formellement exprimé sous forme de requête. En RI classique, la pertinence d'un document (ou pertinence thématique) est estimée, par un *modèle de recherche*, comme le degré d'adéquation du contenu du document au contenu de la requête, généralement à travers le nombre de mots qu'ils ont en commun. L'objectif de la RI dans Twitter est de retrouver les tweets pertinents pour une requête utilisateur. La pertinence d'un tweet est estimée par un modèle de recherche social qui combine généralement des facteurs sociaux (mesure d'influence, degré d'expertise, degré de popularité de l'auteur du tweet, etc.) avec la pertinence thématique du tweet lui-même.

Nous définissons alors un modèle social pour la recherche de tweets. Dans ce modèle, un tweet est considéré comme définitivement pertinent si d'une part, il est fortement pertinent pour le sujet de la requête (*pertinence thématique*), et d'autre part son auteur est un blogueur influent (*pertinence sociale*). En effet, un blogueur influent est un blogueur que l'on a beaucoup retweeté sur le réseau, ce qui atteste de l'intérêt de ses tweets. La pertinence d'un tweet est alors calculée comme la somme pondérée de sa *pertinence thématique* et de sa *pertinence sociale*. La *pertinence thématique* d'un tweet est classiquement mesurée par le degré de similarité du tweet avec la requête. Tout modèle classique de RI adapté à la recherche de textes courts peut être utilisé pour la calculer. La *pertinence sociale* du tweet est liée à l'importance de son auteur, mesurée par son *ratio d'influence* (défini par l'équation (5)). Le score de pertinence d'un tweet t pour une requête Q est alors formellement

défini par la somme pondérée du score de pertinence thématique et du ratio d'influence, comme suit :

$$Rel(Q, t, G) = \alpha * RSV(Q, t) + (1 - \alpha) * \tau_{inf}(u_i)_{/u_i \text{ auteur de } t} \quad (8)$$

Où G est le graphe social d'influence, α un paramètre compris entre 0 et 1, et $RSV(Q, t)$ la pertinence thématique du tweet t pour la requête Q .

5. Évaluation expérimentale

Nous avons implémenté notre modèle de recherche sociale sur la plateforme de RI Lucene⁵, puis l'avons expérimenté et évalué (et avons par la même occasion évalué l'efficacité de notre *ratio d'influence*) sur une collection de test standard. Nous l'avons ensuite comparé d'une part, à une *baseline* (ie. un modèle de recherche classique basé sur un simple score de pertinence thématique), et d'autre part à un modèle de recherche sociale basé sur la mesure d'influence de (Benjabeur *et al.*, 2011) que nous avons aussi implémenté sous Lucene. Ce modèle de recherche est construit par intégration de la mesure d'influence correspondante à la *baseline*.

Dans la suite, nous désignons les modèles de recherche d'information sociale (RIS) évalués comme suit :

- *Baseline* représente un modèle de recherche classique à score thématique,
- RIS_{ratio} représente notre modèle social {*baseline* + notre score d'influence}.
- RIS_{Ben} représente le modèle social {*baseline* + score d'influence de (Benjabeur *et al.*, 2011)}. Dans ce modèle, le score de pertinence d'un tweet t pour une requête Q , étant donné le réseau d'influence G , est calculé par la formule suivante :

$$Rel(Q, t, G) = \alpha * RSV(Q, t) + (1 - \alpha) * Inf_{Ben}(u_i)_{/u_i \in G \wedge u_i \text{ auteur de } t} \quad (9)$$

Où Inf_{Ben} est le score d'influence de (Benjabeur *et al.*, 2011) tel que défini dans l'équation (1). Du modèle initial de l'auteur, nous avons délibérément omis le score d'expertise car il nous importe ici de comparer uniquement nos scores d'influence respectifs.

Dans ce qui suit, nous présentons la collection de test utilisée, ainsi que nos expérimentations et résultats.

5.1. Protocole d'évaluation

Pour évaluer notre modèle de recherche sociale, nous avons utilisé la collection de test standard TREC-Microblogs 2011. Pour ce faire, nous avons :

5. <https://lucene.apache.org/core/>

- d’abord récupéré les tweets à partir des serveurs de Twitter,
- puis construit le réseau d’influence correspondant,
- et finalement évalué notre approche sur la collection.

5.1.1. Extraction des tweets

TREC-Microblogs 2011 est constituée de quelques 16 millions de liens vers des tweets, de 49 requêtes de recherche, et des jugements de pertinence. Pour extraire ces tweets à partir des serveurs de Twitter, nous avons utilisé l’outil *twitter-tools*⁶. Nous n’avons pu récupérer qu’un sous-ensemble de 11 millions de ces tweets. Nos expérimentations ont été réalisées sur ce sous-ensemble.

5.1.2. Construction du réseau social d’influence

Pour construire le réseau social d’influence (basé sur la relation de rediffusion), nous avons identifié pour chacun des tweets récupérés, le blogueur qui l’a publié, ainsi que tous les blogueurs qui l’ont retweeté. L’ensemble de ces blogueurs et des relations de retweets qui les lient forme le *réseau social de rediffusion global* (ResG). La figure 4a représente ce réseau⁷. La figure 4b est un zoom partiel de ce réseau (où les nœuds interconnectés sont des nœuds de même couleur).

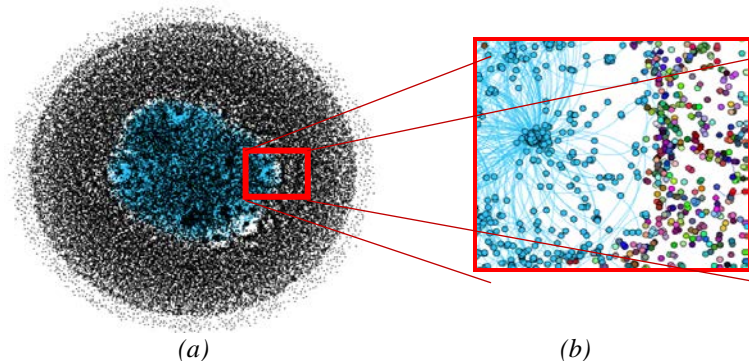


Figure 4. Réseau de rediffusion issu de TREC-Microblogs2011

Il apparaît clairement de cette figure que le réseau social ainsi formé n’est pas connexe. En effet, de nombreux blogueurs ne sont pas interconnectés via la relation de rediffusion, et forment une multitude de sous-réseaux de rediffusion (au nombre de 103 943) indépendants. Le *sous-réseau connexe le plus dense* (ResC), soit celui qui comprend le plus de blogueurs interconnectés (il apparaît en couleur bleue sur la figure 4a), compte 170 515 blogueurs. Cette situation peut s’expliquer de deux manières : soit que les blogueurs sur Twitter forment effectivement des réseaux

6. <https://github.com/lintool/twitter-tools>

7. Le graphe a été construit avec l’outil de visualisation de graphes *Gephi* (<https://gephi.org>)

communautaires indépendants, soit que notre collection (incomplète) a impacté sur la connexité du réseau de rediffusion entre blogueurs du fait que des tweets et/ou retweets (et donc des blogueurs et des relations de rediffusion) ont été omis (car n'ayant pas pu être récupérés à partir des serveurs de Twitter).

Comme il nous paraît évident que la connexité du réseau de rediffusion va impacter notre *ratio d'influence*, nous avons choisi d'étudier cet impact. Pour cela, nous construisons notre réseau social d'influence en considérant tour à tour, le réseau global de rediffusion ResG, puis son sous-réseau connexe le plus dense ResC. Le tableau 2 fournit les caractéristiques de chacun de ces réseaux, en mettant l'accent en particulier sur :

- l'ensemble B_{sub} des blogueurs qui rediffusent les tweets des autres blogueurs sans jamais être eux même retweetés. Au sens de notre mesure d'influence, ces blogueurs subissent l'influence des autres ;
- l'ensemble B_{imp} des blogueurs dont au moins un tweet est retweeté par les autres blogueurs sans qu'eux même n'aient retweeté aucun tweet. Ces blogueurs imposent donc leur influence aux autres ;
- l'ensemble $B_{sub-imp}$ des blogueurs qui ont retweeté au moins une fois et qui ont été retweetés au moins une fois. Ces blogueurs imposent leur influence et subissent aussi l'influence des autres blogueurs ;
- et l'ensemble B_{isol} des blogueurs du réseau qui ont tweeté, mais qui n'ont pas été retweetés et n'ont pas eux-mêmes retweeté d'autres blogueurs. Ces blogueurs, que l'on qualifie d'*isolés*, n'imposent pas d'influence sur les autres et ne subissent pas l'influence des autres blogueurs.

Dans ce tableau, RI/ResG désigne le réseau d'influence construit à partir du réseau de rediffusion global ResG, et RI/ResC désigne le réseau social d'influence construit à partir du réseau de rediffusion connexe ResC.

Tableau 2. Statistiques sur le réseau social d'influence

| | RI/ResG | RI/ResC |
|------------------------------|------------|---------|
| #Blogueurs | 4 019 580 | 170 515 |
| #tweets sur le réseau social | 11 280 274 | 984 478 |
| #retweets | 377 652 | 213 915 |
| #tweets par blogueur | 2,81 | 5,77 |
| #retweets par blogueur | 0,099 | 1,254 |
| $ B_{sub} $ | 192 893 | 74 657 |
| $ B_{imp} $ | 214 749 | 78 363 |
| $ B_{sub-imp} $ | 25 895 | 17 495 |
| $ B_{isol} $ | 3 586 043 | 0 |

De ce tableau, il ressort que les relations d'influence sur notre réseau social sont composées de relations d'influence imposées et de relations d'influence subies à proportions quasi-égales comme le montre la figure 5 qui fournit une représentation graphique des statistiques du tableau 3.

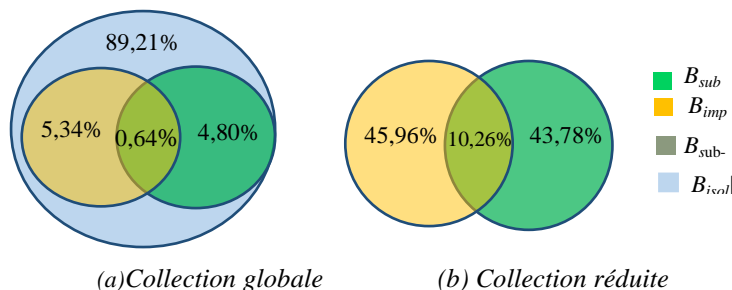


Figure 5. Blogueurs influents vs influencés dans le réseau d'influence

Le réseau d'influence choisi se prête donc bien au calcul du ratio d'influence (basé sur les relations d'influence imposée et d'influence subie). De plus, le réseau d'influence construit sur la collection globale montre un fort taux (89 %) de blogueurs isolés (ces blogueurs ont un ratio d'influence quasi nul), tandis que le réseau d'influence construit sur la collection réduite ne contient pas de blogueurs isolés puisque seuls les blogueurs interconnectés *via* la relation de rediffusion ont été retenus. On s'attend alors à ce que les performances de notre modèle d'influence – dépendant de l'interconnexion des blogueurs *via* la relation de rediffusion – soient meilleures dans le cas de la collection réduite.

5.1.3. Collection de test

Nous avons construit deux collections de test relativement aux deux réseaux d'influence considérés, comme suit :

– Dans le cas où le réseau d'influence RI/ResG est considéré : la collection de test utilisée est la collection TREC-microblogs2011 avec ses 11 millions de tweets (récupérés à partir des serveurs de Twitter), ses 49 requêtes de recherche et jugements de pertinence associés. Dans la suite, nous désignerons cette collection de test par *collection globale*.

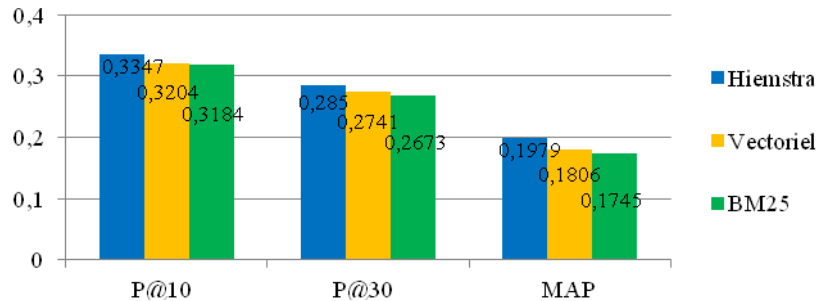
– Dans le cas où le réseau d'influence RI/ResC est considéré : la collection de test utilisée est un sous-ensemble de TREC-microblogs2011, obtenue après avoir supprimé les tweets de tous les blogueurs qui n'apparaissent pas dans le réseau RI/ResC (*i.e.* les blogueurs isolés), et en ne gardant que les requêtes relatives aux tweets publiés et/ou retweetés par les blogueurs du réseau. La collection résultante est composée de 984 478 tweets, et d'un ensemble de 39 requêtes avec les jugements de pertinence associés. Dans la suite, nous désignerons cette collection de test par *collection réduite*.

5.2. Choix de la Baseline (ou quel modèle de recherche thématique des tweets ?)

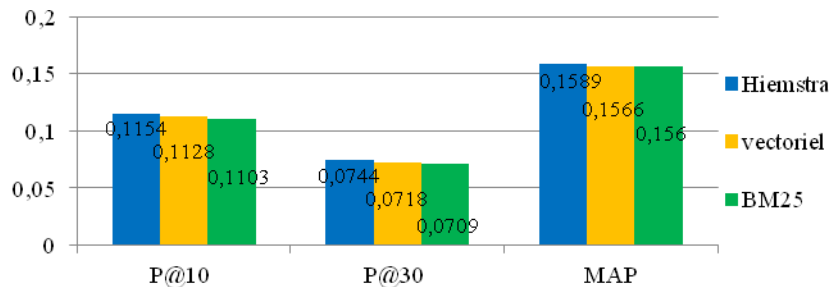
Les Tweets sont des textes très courts (limités à 140 caractères), souvent mal orthographiés et/ou utilisant un langage non conventionnel (acronymes, émoticônes, etc.). Pour cela, les systèmes de RI classiques, généralement conçus pour des textes conventionnels, ne conviennent pas particulièrement à la recherche des tweets. Comme cette problématique est hors du sujet de cet article, nous l'avons contournée en sélectionnant simplement, parmi un ensemble de systèmes de recherche d'information classiques (thématiques) existants, celui qui répond le mieux à la recherche de tweets. Nous avons pour cela, testé et évalué les trois modèles de recherche classiques suivants, qui sont implémentés dans la plateforme de RI Lucene⁸ : le modèle vectoriel (Salton *et al.*, 1975), le modèle probabiliste basé sur le score BM25 (Robertson et Jones, 1976), et le modèle de langue de Hiemstra (Hiemstra, 2001). Ces modèles de recherche ont été évalués sur la base de leurs résultats obtenus respectivement sur nos deux collections de test. Les mesures d'évaluation utilisées sont : la précision moyenne MAP (*mean average precision*), et les précisions aux rangs 10 et 30, respectivement $P@10$ et $P@30$. Les résultats de ces évaluations sont présentés en figures 6a et 6b pour les collections globale et réduite respectivement. Notons cependant qu'il importe peu dans ce cas que la collection soit construite à partir d'un réseau d'influence connexe ou pas. En effet, l'influence n'est pas prise en compte ici, seule la pertinence thématique des tweets importe.

Les résultats de l'évaluation sur la collection globale montrent des taux d'amélioration du modèle de Hiemstra par rapport au modèle vectoriel (resp. au modèle BM25) de 4,46 % (resp. 5,12%) pour $P@10$, de 3,98 % (resp. 6,62 %) pour $P@30$, et de 9,58 % (resp. 13,41 %) pour la MAP. Les résultats sur la collection réduite montrent des taux d'amélioration du modèle de Hiemstra par rapport au modèle vectoriel (resp. au modèle BM25) de 2,30 % (resp. 4,62 %) pour $P@10$, 3,62 % (resp. 4,94 %) pour $P@30$, et 1,47 % (resp. 1,86 %) pour la MAP. Le modèle de Hiemstra a présenté des taux d'amélioration significatifs dans le cas de la collection globale, et faibles dans le cas de la collection réduite. Ceci peut s'expliquer par le fait que dans la collection réduite, de nombreux tweets ont été omis (car les auteurs correspondants ne sont pas dans le réseau connexe), ce qui a impacté les résultats de la recherche et les performances des systèmes de recherche expérimentés. Il est logique alors de penser que c'est les résultats sur la collection globale qui déterminent le meilleur système de recherche thématique de tweets, soit donc le modèle de Hiemstra. C'est ce modèle que nous retiendrons comme *baseline* pour la suite de nos expérimentations.

8. http://lucene.apache.org/core/3_0_3/api/core/org/apache/lucene/search/Similarity.html



(a). Résultats sur la collection globale



(b). Résultats sur la collection réduite

Figure 6. Résultats de la recherche thématique dans les tweets

5.3. Évaluation de notre modèle de RI sur la collection de test globale

Dans cette partie, nous évaluons notre système de RI social sur la collection de test globale construite à partir du réseau d'influence RI/ResG.

5.3.1. Paramétrage des modèles de RI sociale RIS_{ratio} et RIS_{Ben}

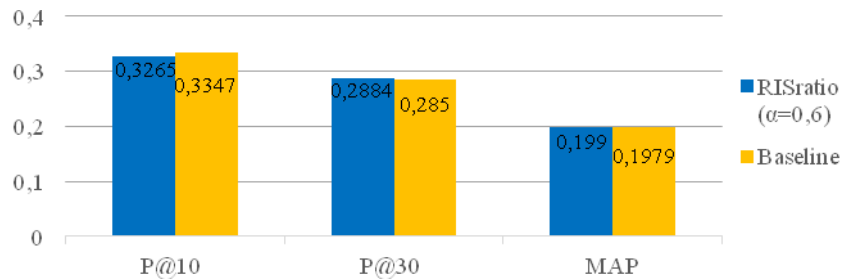
Comme le modèle de recherche RIS_{ratio} (respectivement RIS_{Ben}) dépend d'un paramètre expérimental α , nous devons au préalable déterminer la valeur optimale de celui-ci, soit donc la valeur pour laquelle le modèle de recherche considéré est le plus performant. Pour cela, nous avons évalué RIS_{ratio} (respectivement RIS_{Ben}) sur la collection de test courante, pour différentes valeurs de α . Les résultats de ces évaluations sont présentés en tableau 3. De ce tableau, il ressort que le modèle de recherche RIS_{ratio} (respectivement RIS_{Ben}) est globalement plus performant pour $\alpha = 0,6$ (respectivement 0,9). Dans la suite, les modèle RIS_{ratio} et RIS_{Ben} sont considérés respectivement avec les valeurs optimales correspondantes de α .

Tableau 3. Evaluation de RIS_{ratio} et RIS_{Ben} pour différentes valeurs de α

| | | $\alpha = 0,1$ | $\alpha = 0,2$ | $\alpha = 0,3$ | $\alpha = 0,4$ | $\alpha = 0,5$ | $\alpha = 0,6$ | $\alpha = 0,7$ | $\alpha = 0,8$ | $\alpha = 0,9$ |
|---------------|------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| RIS_{ratio} | MAP | 0,1732 | 0,1911 | 0,1961 | 0,1982 | 0,199 | 0,199 | 0,199 | 0,1989 | 0,1988 |
| | P@10 | 0,2653 | 0,3122 | 0,3224 | 0,3224 | 0,3184 | 0,3265 | 0,3265 | 0,3265 | 0,3265 |
| | P@30 | 0,2619 | 0,2796 | 0,2871 | 0,2884 | 0,2878 | 0,2884 | 0,2871 | 0,2871 | 0,2857 |
| RIS_{Ben} | MAP | 0,1962 | 0,1966 | 0,1972 | 0,1974 | 0,1975 | 0,1975 | 0,1976 | 0,1977 | 0,1979 |
| | P@10 | 0,3265 | 0,3265 | 0,3265 | 0,3286 | 0,3286 | 0,3286 | 0,3286 | 0,3306 | 0,3347 |
| | P@30 | 0,2837 | 0,2823 | 0,283 | 0,283 | 0,283 | 0,285 | 0,285 | 0,285 | 0,285 |

5.3.2. RIS_{ratio} vs la baseline

Pour évaluer l'intérêt de l'influence dans la mesure de pertinence d'un tweet, nous avons comparé les résultats de la recherche de notre modèle de RI sociale RIS_{ratio} (combinant *ratio d'influence* et score thématique) avec la *baseline* (basée uniquement sur le score thématique). Les résultats sont présentés en figure 7. Ces résultats montrent que notre système est moins performant que la *baseline* pour la précision $P@10$, et plus performant pour la précision $P@30$, avec respectivement un taux de dégradation de 2,49 % pour $P@10$, et un taux d'amélioration de 1,19 % pour $P@30$. Cependant, notre modèle est globalement (très légèrement) plus performant que la *baseline* avec un taux d'amélioration de 0,55 % pour la précision moyenne MAP. On en conclut que l'influence n'a que très légèrement impacté les performances du système.

Figure 7. RIS_{ratio} vs Baseline

Ceci peut s'expliquer par le fait que notre collection globale contient un taux élevé (89 %) de blogueurs isolés pour lesquels le *degré d'influence* est nul. Les tweets correspondants participent au score de recherche avec leur score de pertinence thématique uniquement. Parallèlement, il y a un taux très faible de blogueurs *influençants* (5,34 %) et de blogueurs *influencés* (4,8 %) dans le réseau d'influence RI/ResG considéré, dont l'influence a impacté, bien que légèrement, les résultats de la recherche. De ce fait, notre modèle de recherche (basé sur le score

thématique et le *ratio d'influence*) n'apporte pas l'amélioration escomptée sur les résultats de la recherche, mais ses performances sont néanmoins légèrement meilleurs que ceux du modèle *baseline*. On s'attend à ce que, dans un réseau plus adéquat (avec plus de relations d'influence), les résultats de notre approche soient plus probants.

5.3.3. RIS_{ratio} vs RIS_{ben}

Nous avons comparé notre modèle de recherche RIS_{ratio} basé sur le *ratio d'influence*, au modèle de recherche RIS_{ben} basé sur le *taux de retweet* de (Benjabeur et al., 2011). Les résultats sont présentés en figure 8. Cette figure montre que le modèle RIS_{ben} fournit les mêmes résultats que la *baseline*. Ce qui signifie que, dans ce modèle, seule la pertinence thématique a impacté les performances de recherche. Le *taux de retweet* n'a pas eu d'effets. Ceci peut s'expliquer d'une part, par le fait que le coefficient du score d'influence dans le modèle RIS_{ben} est très faible ($1 - \alpha = 0,1$), et d'autre part, par le fait que notre collection globale contient un taux élevé (89 %) de blogueurs isolés pour lesquels le *taux de retweets* est nul. Par ailleurs, alors que le coefficient de notre *ratio d'influence* dans le modèle RIS_{ratio} est non négligeable ($1 - \alpha = 0,4$), c'est le taux élevé de blogueurs isolés et les taux très faibles de blogueurs *influençants* (5,34 %) et de blogueurs *influencés* (4,8 %) dans le réseau d'influence RI/ResG considéré, qui ont probablement impacté négativement les performances de notre modèle. Cependant, nos résultats sont malgré tout légèrement meilleurs que ceux du modèle RIS_{ben} .

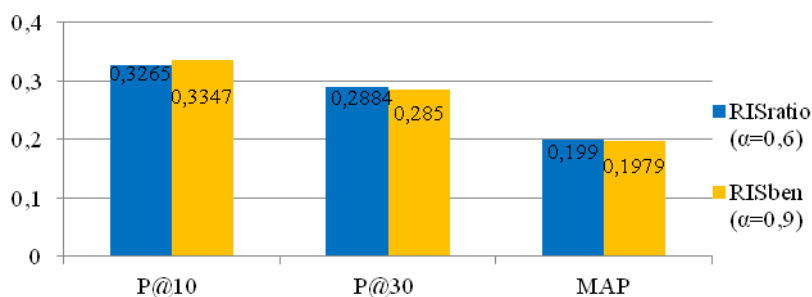


Figure 8. RIS_{ratio} vs RIS_{ben}

5.4. Évaluation de notre modèle de RI sur la collection de test réduite

Dans cette partie, nous évaluons notre système de RI social sur la collection de test réduite construite à partir du réseau d'influence connexe RI/ResC.

5.4.1. Paramétrage des modèles de RI sociale RIS_{ratio} et RIS_{ben}

Comme dans l'évaluation précédente, nous avons dans le cas de la présente collection, évalué le modèles de recherche RIS_{ratio} (respectivement RIS_{ben}) pour

différentes valeurs du paramètre expérimental α . Les résultats de ces évaluations sont présentés en tableau 4. De ce tableau, il ressort que le modèle de recherche RIS_{ratio} (respectivement RIS_{Ben}) est globalement plus performant (meilleure MAP) pour $\alpha = 0,9$ (respectivement 0,8). Dans la suite, les modèle RIS_{ratio} et RIS_{Ben} sont considérés respectivement avec la valeur optimale correspondante de α .

Tableau 4. Evaluation de RIS_{ratio} et RIS_{Ben} pour différentes valeurs de α

| | $\alpha = 0,1$ | $\alpha = 0,2$ | $\alpha = 0,3$ | $\alpha = 0,4$ | $\alpha = 0,5$ | $\alpha = 0,6$ | $\alpha = 0,7$ | $\alpha = 0,8$ | $\alpha = 0,9$ |
|---------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| RIS_{ratio} | MAP | 0,0991 | 0,1235 | 0,1452 | 0,1477 | 0,1565 | 0,1571 | 0,1587 | 0,1612 |
| | P@10 | 0,0769 | 0,1179 | 0,1179 | 0,1179 | 0,1231 | 0,1256 | 0,1256 | 0,1282 |
| | P@30 | 0,053 | 0,0667 | 0,0701 | 0,0718 | 0,0752 | 0,0761 | 0,0761 | 0,0761 |
| RIS_{Ben} | MAP | 0,1562 | 0,1561 | 0,1566 | 0,1568 | 0,1567 | 0,157 | 0,1574 | 0,158 |
| | P@10 | 0,1154 | 0,1154 | 0,1154 | 0,1154 | 0,1154 | 0,1154 | 0,1154 | 0,1103 |
| | P@30 | 0,0752 | 0,0752 | 0,0752 | 0,0752 | 0,0761 | 0,0761 | 0,0761 | 0,0761 |

5.4.2. RIS_{ratio} vs la baseline

Les résultats de l'évaluation de notre système RIS_{ratio} et de la *baseline* sur la collection courante sont présentés en figure 9.

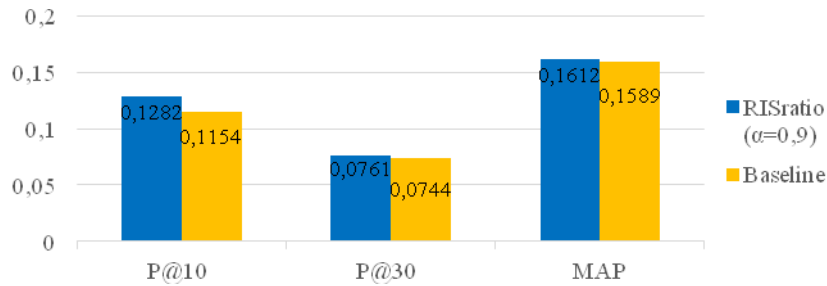


Figure 9. RIS_{ratio} vs Baseline

Ces résultats montrent une amélioration des performances de la recherche avec notre modèle basé sur la pertinence thématique et le *ratio d'influence*, par rapport au modèle de recherche thématique de la *baseline*, avec des taux d'accroissement de 11,1 %, 2,3 % et 1,4 % respectivement pour les précisions $P@10$, $P@30$, et MAP.

5.4.3. RIS_{ratio} vs RIS_{Ben}

Les résultats de l'évaluation des systèmes RIS_{ratio} et RIS_{Ben} considérés respectivement avec leurs valeurs optimales de α , sont présentés en figure 10. Ces résultats montrent que notre modèle est plus performant que le modèle RIS_{Ben} avec

des taux d'amélioration de 11,1 %, 1,2 % et 2 %, respectivement pour les précisions $P@10$, $P@30$ et MAP.

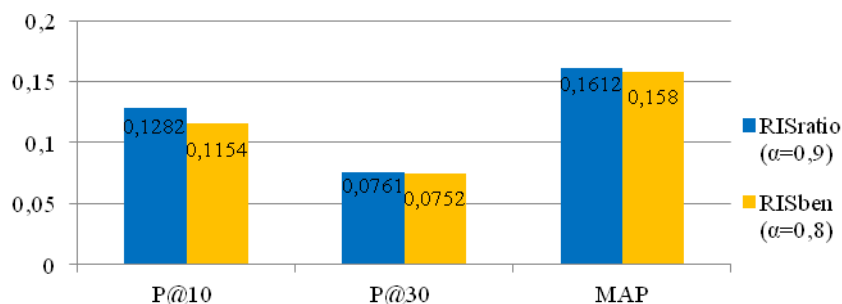


Figure 10. RIS_{ratio} vs RIS_{ben}

Nous avons ensuite comparé les systèmes RIS_{ratio} et RIS_{ben} avec les mêmes valeurs du paramètre α et à pertinence thématique égale. L'objectif étant d'isoler l'impact du *ratio d'influence* de notre système RIS_{ratio} par rapport à l'impact du *taux de retweet* de RIS_{ben} sur les résultats de la recherche. Deux valeurs de α ont été testées, respectivement 0,9 et 0,8. Ces valeurs sont les valeurs optimales retenues lors des paramétrages respectifs des modèles RIS_{ratio} et RIS_{ben} sur la collection courante. Les résultats de cette évaluation sont présentés en figure 11.

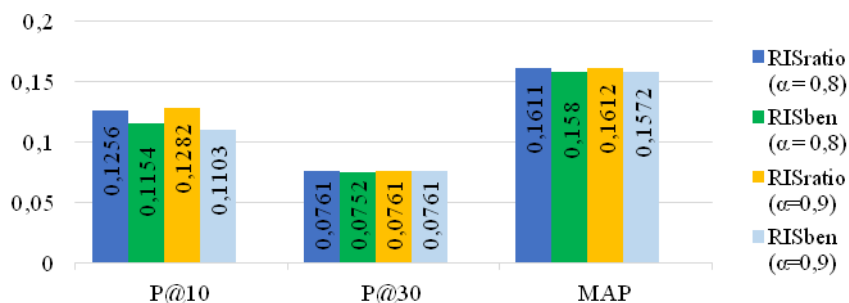


Figure 11. RIS_{ratio} vs RIS_{ben} à paramètre égal

La figure 11 montre que notre modèle RIS_{ratio} est plus performant au regard des précisions $P@10$ et $P@30$, et plus globalement aussi au regard de la MAP, et ce pour les deux valeurs de α considérées. Les taux d'amélioration enregistrés pour la valeur $\alpha = 0,9$ (soit la valeur optimale pour RIS_{ratio}) sont de 16,2 % et de 2,5 % respectivement pour les précisions $P@10$ et MAP. Et, bien que notre modèle n'ait pas enregistré d'amélioration pour la précision $P@30$, il n'y a pas eu de dégradation de performances non plus. Les taux d'amélioration apportés pour la valeur $\alpha = 0,8$,

pourtant testée comme valeur optimale pour RIS_{Ben} , sont de 8,83 %, 1,2 % et de 2 % respectivement pour les précisions P@10, P@30 et MAP.

Ces résultats montrent bien que le score d'influence basé sur le *taux de retweets* défini dans le modèle RIS_{Ben} est problématique (problème discuté en section 4.2), et que notre *ratio d'influence* pouvait pallier ce problème. En effet, nos multiples évaluations expérimentales ont montré l'intérêt de notre approche et nous confortent dans notre intuition première sur la nécessité de tenir compte des relations d'influence subie en plus de l'influence imposées, en particulier dans un contexte de forte connexité (vis-à-vis de la relation de *retweet*) entre blogueurs. Néanmoins, les performances obtenues ne sont pas probantes. Ceci pourrait s'expliquer par une mauvaise définition du ratio d'influence. Or, il est logique de penser que plus un blogueur impose son influence, plus il est influent. Et plus ce blogueur est influencé, moins il est influent. L'influence du blogueur est donc logiquement proportionnelle à son influence imposée et inversement proportionnelle à son influence subie : la définition du ratio d'influence comme rapport entre l'influence imposée à l'influence subie est correcte. Le problème pourrait alors venir de la définition de l'influence imposée et/ou de l'influence subie. Dans notre proposition, l'influence imposée d'un blogueur est proportionnelle à l'influence imposée des blogueurs qui le retweetent, et son influence subie est aussi proportionnelle à l'influence imposée des blogueurs qu'il retweete. En effet, ces mesures sont duales, de la même façon que le sont le score d'autorité et le score pivot de l'algorithme HITS (Kleinberg, 1999). Le score d'autorité d'une page web se calcule en fonction des scores pivots des pages qui la référencent. Son score pivot se calcule en fonction des scores d'autorité des pages qu'elle référence. Or, cette dualité n'est pas prise en compte dans nos mesures d'influence proposées, ce qui a probablement impacté négativement l'apport du ratio d'influence.

6. Conclusion

Nous avons proposé une mesure de l'influence d'un blogueur sur Twitter à travers un *ratio d'influence*. Le *ratio d'influence* d'un blogueur est le rapport entre sa capacité à influencer d'autres blogueurs (*influence imposée*) à sa disposition à être influencé par d'autres blogueurs (*influence subie*). Le *ratio d'influence* est calculé sur un réseau d'influence basé sur les *retweets*, par un algorithme de propagation adapté de *PageRank*. Nous avons expérimenté et évalué notre mesure d'influence dans un contexte de RI sociale. Le modèle de recherche proposé combine *ratio d'influence* et score thématique pour calculer la pertinence d'un tweet. Ce modèle de recherche a été évalué comparativement à un modèle sans influence (*baseline*) d'une part, et à un autre modèle à base d'influence (RIS_{Ben}) d'autre part. Deux séries d'évaluation ont été menées respectivement sur deux collections de tweets issues de TREC-micoblogs2011 : une collection globale, et une collection réduite. Ces collections ont permis de mettre en évidence le problème de connexité dans les réseaux d'influence associés. Nous avons alors évalué l'impact de la connexité des

blogueurs dans les réseaux d'influence considérés. Les résultats de nos différentes évaluations, ont prouvé d'une part, l'apport de l'influence en RI, et d'autre part l'intérêt de notre ratio d'influence.

Néanmoins, notre approche présente des limites et des insuffisances qui doivent être soulignées et qui ouvrent les voies de nouvelles perspectives de recherche.

– D'abord, les taux d'amélioration enregistrés restent faibles relativement à ce qu'on avait espéré. Comme nous l'avons expliqué dans notre précédente analyse, cela pourrait être dû à une définition inadéquate de la mesure d'influence imposée ou de l'influence subie. Une perspective intéressante pour pallier ce problème serait d'adapter le score d'autorité et le score pivot de l'algorithme HITS (Kleinberg, 1999), au calcul de nos mesures d'influence.

– Ensuite, le modèle d'influence proposé est statique par définition : il ne tient pas compte de la temporalité de l'influence d'un blogueur (en ce sens qu'un blogueur peut être influent sur une période de temps et pas sur une autre). Il serait alors intéressant de définir une influence relative qui caractérise un blogueur dans une période donnée. Dans un contexte de RI sociale, l'indicateur de temps peut être utilisé pour retrouver les tweets pertinents pour la requête dans une période donnée. C'est sur cette période que l'influence des blogueurs doit être considérée.

– Enfin, dans le modèle de recherche d'information social proposé, le score d'influence utilisé est générique, et ne tient pas compte du sujet de la requête. En effet, lorsqu'une requête est évaluée, les tweets pertinents pour cette requête sont retrouvés et leurs auteurs identifiés. Le ratio d'influence de chaque auteur est alors calculé sur le réseau social d'influence, indépendamment de la requête. Or, il serait logique de penser qu'un blogueur peut être influent sur un sujet et pas sur un autre. Son influence donc doit être adaptée au sujet de la requête. Pour ce faire, on pourrait construire une instance du réseau d'influence, dépendante de la requête Q (ie. dépendante des tweets t pertinents pour Q), à partir de laquelle le ratio d'influence d'un blogueur est calculé. Un tel réseau sera constitué des blogueurs u tels que u a publié t , ou a retweeté t , ou u a répondu à t . Le modèle de recherche sociale fonctionnera en deux temps : dans un premier temps, il s'agit de retrouver les tweets pertinents pour la requête de recherche. Dans un second temps, l'influence des auteurs est calculée sur l'instance du réseau d'influence dépendante de Q , puis le score du tweet est affiné avec ce score d'influence spécifique.

En perspectives, nous envisageons de développer et de mettre en œuvre nos différentes réflexions ci-haut. Nous envisageons également de comparer, autrement qu'à travers la RI sociale, notre ratio d'influence aux mesures d'influence issues d'outils professionnels tels que *Klout*. Pour cela, il nous faudra définir des mécanismes d'appariement qui permettent de comparer des mesures d'influence, sur la base des rangs (ie. classement par ordre d'influence) des blogueurs, obtenus par application de leurs algorithmes respectifs.

En outre, nous pensons également redéfinir notre score d'influence en considérant par exemple, outre la rediffusion, d'autres relations et/ou

caractéristiques du réseau social de Twitter. Le modèle de RI sociale pourrait aussi être enrichi en combinant la pertinence thématique avec d'autres facteurs sociaux en plus de l'influence.

Bibliographie

- Anger I., Kittl C. (2011). Measuring Influence on Twitter. *KNOW'11, Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*. Article n° 31.
- Azaza L., Kirgizov S., Savonnet M., Frame A. (2015). Evaluation de l'influence sur Twitter: Application au projet « Twitter aux Elections européennes 2014. *Etudier le Web politique : regards croisés*, May, Lyon.
- Benjabeur L., Tamine L., Boughanem M. (2011). Un modèle de recherche d'information sociale dans les microblogs : cas de Twitter. *Conférence sur les Modèles et l'Analyse des Réseaux: Approches Mathématiques et Informatique MARAMI'11*.
- Benjabeur L., Tamine L., Boughanem M. (2012). Active microbloggers: identifying influencers, leaders and discussers in microblogging networks. *International Symposium on String Processing and Information Retrieval*, p. 111-117.
- Bogdan G., Zygmunt A. (2015). Finding Influential Bloggers. *International Journal of Machine Learning and Computing*. vol. 5, n° 2, April 2015.
- Cano A. E., Mazumdar S., Ciravegna F. (2014). Social influence analysis in microblogging platforms – A topic-sensitive based approach. *Semantic Web*, vol. 5, n° 5, p. 357-372.
- Cataldi M., Afaure M.-A. (2014). The 10 Million follower fallacy: Audience size does not prove domain-influence on Twitter. *Knowledge and Information Systems*, p. 1-22.
- Cha, M., Haddadi H., Benvenuto F., Gummadi P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- Ding Z., Jia Y., Zhou B., Han Y., He L., Zhang J. (2013). Measuring the spreadability of users in microblogs. *Journal of Zhejiang University - Science C*, vol. 14, n° 9, p. 701-710.
- Gayo-Avello D. (2013). Nepotistic relationships in Twitter and their impact on rank prestige algorithms. *Information Processing and Management*, vol. 49, n° 6, p. 1250-1280.
- Herrera R., Cardenas J P., Alfaro R. (2014). User influence in on-line social networks. In *WACS 2014, IV Workshop on Agents and Collaborative Systems. Jornadas Chilenas de Computacion(JCC)*. November 8-14, Talca, Chile.
- Hiemstra D.(2001). *Using Language Models for Information Retrieval*. PhD thesis, Enschede. January.
- Kleinberg J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, vol. 46, n° 5, p. 604-632.
- Liu X., Shen H. Ma F., Liang W. (2014). Topical influential user analysis with relationship strength estimation in Twitter. In Zhi-Hua Zhou, Wei Wang, Ravi Kumar, Hannu Toivonen, Jian Pei, Joshua Zhexue Huang, Xindong Wu (Eds), *2014 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2014*, Shenzhen, China, December 14, p. 1012-1019.

- Nargundkar A., Rao Y. S. (2016). InfluenceRank: A machine learning approach to measure influence of Twitter users. In *International Conference on Recent Trends in information technology*.
- Page L., Brin S. Motwani R., Winograd T. (1999). *The PageRank citation ranking: Bringing order to the web*. Technical report, Stanford Digital Library.
- Räbiger S., Spiliopoulou M. (2015). A framework for validating the merit of properties that predict the influence of a twitter user. *Expert Syst. Appl.*, vol. 42, n° 5, p. 2824-2834.
- Riquelme F., González-Cantergiani P. (2016). Measuring user influence on Twitter: A survey. In *Journal of Information Processing and Management* vol. 52, n° 5, p. 949-975, April.
- Robertson S. E., Jones K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, vol. 27, n° 3, p. 129-146, mai.
- Romero D. M., Galuba W., Asur S., Huberman B. A. (2011). Influence and Passivity in Social Media. *WWW'11*, March 28-April 1, Hyderabad, India.
- Sung J., Moon S., Lee JG. (2013). The influence in Twitter: Are they really influenced? In Longbing Cao, Hiroshi Motoda, Jaideep Srivastava, Ee-Peng Lim, Irwin King, PhilipS Yu, Wolfgang Nejdl, Guandong Xu, Gang Li, Ya Zhang (Eds), *Behavior and Social Computing*, vol. 8178 of *Lecture Notes in Computer Science*, p. 95-105. Springer International Publishing.
- Tunkelang D. (2009). *A Twitter Analog to PageRank*, <http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-PageRank>.
- Weng J., Lim E.-P., Jiang J., He Q. (2010). TwitterRank: finding topic-sensitive influential Twitterers. *Proceedings of the third ACM international conference on Web search and data mining, WSDM'10*, New York, NY, USA, ACM, p. 261-270.
- Zhang J., Zhang R., Su nJ., Zhang Y., Zhang C. (2016). TrueTop: A Sybil-Resilient System for User Influence Measurement on Twitter. In *IEEE/ACM Transactions on Networking*, vol. 24, n° 5, p. 2834-2846, October.
- Zhang M., Sun C., Liu W. (2011). Identifying Influential Users Of Micro-Blogging Services: A Dynamic Action-Based Network Approach. *PACIS 2011, Proceedings.223*, <http://aisel.aisnet.org/pacis2011/223>.