

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira –Bejaia
Faculté des Sciences Exactes
Département d'Informatique
Option Administration et sécurité des réseaux



Mémoire de fin de cycle

En vue de l'obtention du Diplôme de Master en Informatique

Thème

Détection des rumeurs dans les réseaux sociaux

Réalisé par:

M^{elle} BENSALÉM Khadidja

M^{elle} SOUALMI Samiha

Devant le jury composé de :

Président du jury: M^r SAADI Mustapha

Examineur: M^r BOUDRIES Abdelmalek

Promoteur: M^r AMAD Mourad

Co-Promoteur: M^r BAADACHE Abderrahmane

Année universitaire 2016/2017

Remerciements

On dit souvent que le trajet est aussi important que la destination. Cinq années nous avons permis de comprendre la signification de cette phrase toute simple. Ce parcours ne s'est pas réalisé sans défis et sans soulever de nombreuses questions pour lesquelles les réponses ont suscité de longues heures de travail.

Nous souhaitons vivement que nos parents qui nous ont épaulés moralement, puissent trouver dans ces lignes l'expression de notre profonde gratitude.

*Nos remerciements s'étendent à notre promoteur **M^r AMAD Mourad** pour sa patience, sa disponibilité et surtout pour ses judicieux conseils, qui ont contribué à alimenter nos réflexions et Nos vifs remerciements également à **M^r BAADACHE Abderrahmane**.*

Nous tenons à remercier chacun des membres du jury pour nous avoir fait l'honneur d'examiner et d'évaluer notre travail et de l'enrichir par leurs propositions.

Nous ne saurons épuiser ces remerciements sans gratifier nos enseignants qui ont su nous donner une formation appréciable durant tout notre cursus.

Dédicaces

La vie n'est qu'un éclair, et le jour de réussite est un jour très cher :

A mes très chers parents

A ma sœur Radia, son mari Abd rezzak et ses enfants Ala et Mohamed anes

A ma sœur Asma et son mari Reda

A mes sœurs Saida, Sara et Fatima

A toute ma famille sans exceptions

A mes amis et camarades de la promotion ASR, et tous ceux qui m'ont aidé.

Khadija

Dédicaces

A mes chers parents

A mes frères et sœurs

A toutes ma famille sans exception

A tous mes amis et camarades de la promotion ASR.

Samia

Table des matières

Table des figures

Liste des tableaux

Liste des algorithmes

Liste des abréviations

Introduction générale.....	1
1 Généralités sur les réseaux sociaux.....	2
1.1 Introduction.....	2
1.2 Bref historique	3
1.3 Définition	4
1.4 Types des réseaux sociaux.....	5
1.4.1 Typologisation des OSNs selon l'évolution et l'apparition.....	5
1.4.2 Typologisation des réseaux sociaux numériques selon la fonctionnalité	8
1.4.3 Typologisation selon wikipédia	8
1.4.4 Typologisation selon le point de vue des chercheurs	8
1.5 Intérêts des réseaux sociaux.....	9
1.6 Principaux des réseaux sociaux	12
1.6.1 Exemple de réseaux sociaux grands publics	12
1.6.2 Exemple de réseaux sociaux professionnels	13
1.7 Cartographie des réseaux sociaux numériques	15
1.8 Fonctions et fonctionnalités.....	15
1.9 Enjeux des réseaux sociaux numériques.....	17
1.10 Problèmes ouvert des réseaux sociaux	18
1.10.1 Problèmes de détection des communautés.....	18
1.10.2 Problème de sécurité.....	18
1.10.3 Problème de détection des rumeurs	19
1.11 Objectif de ce mémoire.....	19
1.12 Conclusion	20

2 Etat de l'art sur les méthodes de détection des rumeurs	21
2.1 Introduction.....	21
2.2 Contexte et objectif de ce mémoire	22
2.3 Définitions	23
2.4 Aspects liées aux rumeurs	23
2.5 Rumeur via les réseaux sociaux.....	24
2.6 Représentation des réseaux sociaux.....	25
2.7 Méthodes de détection des rumeurs.....	26
2.7.1 Greedy source set size (GSSS)	26
2.7.2 Modèles de propagation d'influence dans un réseau social	26
2.7.3 Détection d'influenceurs dans un réseau social.....	27
2.7.4 Approches de minimisation de l'influence négative dans un réseau social.....	28
2.8 Critiques sur les méthodes de détection des rumeurs	29
2.9 Synthèse et Analyse.....	29
2.10 Conclusion	34
3 Proposition d'une nouvelle méthode de détection des rumeurs	35
3.1 Introduction	35
3.2 Première étape : Identification des sources de la rumeur	36
3.2.1 Identification de notre méthode (première étape).....	37
3.2.2 Algorithme de la méthode proposée	37
3.3 Deuxième étape : Comment détecter les rumeurs et non rumeurs	38
3.2.1 Identification de notre méthode (deuxième étape).....	38
3.2.2 Algorithme de différencie la rumeur et non rumeur	39
3.4 Conclusion	41
4 Evaluation des performances de la méthode proposée	42
4.1 Introduction	42
4.2 Implémentation	43
4.2.1 Gephi	43

4.3 Etude de cas d'un twitter	45
4.3.1 Modélisation et évaluation d'une cartographie Twitter avec Gephi	45
4.3 .2 Statistiques de type centralité eigenvector	46
4.3. 3 Statistiques de type modularité	47
4.3 .4 Statistiques de type degré	48
4.4 Conclusion.....	49
Conclusion et Perspectives.....	50

1.1 Représentation d'un réseau social	4
1.2 Panorama des médias sociaux	10
1.3 Les différents types des médias sociaux	12
1.4 Carte mondiale des réseaux sociaux	15
2.1 Pyramide des besoins d'information	22
2.2 Matrice réordonnable de J. Bertin, extraite de la Sémiologie graphique	26
4.1 Représentation d'une cartographie Twitter	45
4.2 Statistiques de type centralité eigenvector.	46
4.3 Statistiques de type modularité	47
4.4 Statistiques de type degré	48

1.1 Les avantages et les inconvénients des réseaux sociaux.....	14
2.1 Synthèse des méthodes de minimisation de l'influence négative dans les réseaux sociaux	30

3.1 Algorithme de l'identification de la source des rumeurs	38
3.2 Algorithme calcule le nombre de sources	40

COOPOL	Cooperative Politique
GSSS	Greedy Sources Set Size
ICM	Independent Cascade Model
LTM	Linear Threshold Model
OSN	Online Social Network
P2P	Peer to Peer
PS	Partie Socialiste
SMS	Short Message Service
UMP	Union Mouvement Populaire

Introduction générale

Au cours des dernières années, les réseaux sociaux en ligne sont devenus de plus en plus populaires. Ces types d'applications ont revendiqué avec succès leurs places parmi les services les plus célèbres sur Internet (*Au 31 mars 2015, le leader Facebook l'un des plus célèbres plates-formes de réseaux sociaux en ligne compte 1,44 milliards d'utilisateurs actifs par mois*). Simultanément avec cette formidable croissance des plates-formes de réseaux sociaux en ligne, des inquiétudes croissantes des utilisateurs sur leur vie privée et la protection de leurs données sont augmentées. Comme la gestion et le contrôle des données de l'utilisateur sont généralement centralisées, les fournisseurs d'accès aux OSNs ont aujourd'hui un privilège sans précédent pour accéder aux données privées de chaque utilisateur, les opinions et les relations sociales, ce qui rend la fuite des informations de la vie privée à grande échelle et l'utilisation abusive des données extrêmement possible, par exemple, LinkedIn a fui des millions de mots de passe de ses utilisateurs [1] et Facebook a passé des informations commerciales sensibles des utilisateurs au public sans leurs consentements [2].

Le but de ce mémoire est de proposer une méthode pour déterminer si une information est une rumeur ou non et la source de la rumeur.

Ce mémoire est structuré en quatre chapitres comme suit :

Le premier chapitre intitulé « **Généralités sur les réseaux sociaux** » est consacré à la présentation de certains concepts fondamentaux relatifs aux réseaux sociaux numériques.

Le second chapitre nommé « **Etat de l'art sur les méthodes de détection des rumeurs** » définit la rumeur via les réseaux sociaux réseaux, citant ainsi les trois modèles de propagation d'influence, les deux métriques les plus référencés dans les travaux de minimisation d'influence. En outre, les approches de minimisation de l'influence négative et résumant ces méthodes dans un tableau de synthèse et analyse.

Le troisième chapitre « **Proposition d'une nouvelle méthode de détection des rumeurs** » nous avons proposé une méthode pour déterminer si une information est une rumeur ou non et la source de la rumeur. Notre méthode utilise une très faible quantité d'informations de source, à savoir, lequel d'un ensemble de nœuds de contrôle a entendu la pièce d'information à portée de main. Pour trouver la source de la rumeur, notre algorithme évalue la probabilité que chaque nœud soit la source, calculé à partir de la connectivité des nœuds et du chemin le plus court distances.

Le dernier chapitre, nous avons présenté l'étude de cas sur le réseau social Twitter et évaluer les performances de l'algorithme proposé.

Enfin, nous terminerons ce mémoire par une conclusion générale et quelques perspectives qu'on souhaite accomplir prochainement.

Chapitre 1

Généralités sur les réseaux sociaux

1.1 Introduction

Les réseaux sociaux sont omniprésents depuis l'apparition d'Internet. Ils permettent aux différents utilisateurs d'interagir en communauté et de se regrouper selon des critères qui leur sont importants. Ces réseaux sociaux sont de différents types. Certains sont connus de tous (*ex. Facebook¹, Twitter², LinkedIn³*) et comptent des millions de membres. D'autres exploitent des niches moins connues et peuvent passer relativement inaperçus ou rester confidentiels, tels les réseaux d'entreprise.

Dans ce chapitre, nous allons présenter les Online Social Network (OSNs), leur définition, leur intérêt et leurs enjeux.

¹ <http://www.facebook.com>

² <http://www.twitter.com>

³ <http://www.linkedin.com>

1.2 Bref historique

Les réseaux sociaux en ligne sont apparus en 2002 avec le site américain Friendster [3], premier à utiliser le principe du cercle d'amis en ligne. Puis vint Myspace [4], crée en 2003, qui dépassa Friendster. A l'origine, l'objectif de Myspace était de permettre à des musiciens de proposer certains de leurs morceaux de musique à l'écoute et de construire leur réseau en devenant amis avec d'autres membres. Ce site est rapidement devenu populaire auprès des jeunes qui l'adoptèrent pour rester en contact avec leurs amis et s'en faire de nouveaux, l'argument musical étant relégué au second plan.

Facebook est né en 2004 [8]. Cette plateforme créée par Mark Zuckerberg était à l'origine destinée aux étudiants d'Harvard⁴ souhaitant communiquer entre eux. Face à son succès fulgurant, le site est devenu grand public en 2006. Le cas de Facebook marque un tournant dans la démocratisation des réseaux sociaux sur Internet. Pour beaucoup, ce fut une porte d'entrée vers l'univers du Web 2.0⁵ et des réseaux sociaux. Premier réseau social au monde en 2008, il comptait 350 millions de membre en 2009 et en compte 500 millions en Juillet 2010.

Twitter est apparu en 2006 [8]. Cette plateforme repose sur le principe du microblogging⁶ : les messages postés par les utilisateurs sont limités à 140 caractères. Très populaire aux Etats-Unis où de nombreuses personnalités y ont un compte... le site connaît néanmoins une très forte croissance, supérieure à celle de Facebook. D'après comScore⁷ (*société américaine spécialisée dans l'analyse Web*), le réseau aurait enregistré une croissance annuelle de 109% en termes de visiteurs uniques entre Juin 2009 et Juin 2010 [8]. De nombreux internautes s'en servent comme une véritable source d'information, notamment pour faire de la veille.

Au départ c'était un outil de communication, les réseaux sociaux sont devenus des outils de diffusion de l'information qui, comme nous allons le voir, ont changé notre rapport à l'information et la façon dont nous la découvrons et la partageons. Tous les utilisateurs des réseaux sociaux sont désormais de potentiels producteurs et diffuseurs de l'information.

⁴ Harvard university : <http://www.harvard.edu>

⁵ Le **Web 2.0** est l'évolution du web vers plus de simplicité (ne nécessitant pas de connaissance techniques ni informatique pour les utilisateurs) et d'interactivité (permettant à chacun, de façon individuelle ou collective, de contribuer, d'échanger et de collaborer sous différentes formes). L'expression « **Web 2.0** » désigne l'ensemble des techniques, des fonctionnalités et des usages du World Wide Web qui ont suivi la forme originelle du web, en particulier les interfaces permettant aux internautes contribuent à l'échange d'information et peuvent interagir (partager, échanger, etc.) de façon simple, à la fois avec le contenu et la structure des pages, mais aussi entre eux, créant ainsi notamment le Web social (**Wikipédia**) .

⁶ Le microblogging désigne l'activité de création de contenus courts sur des réseaux sociaux de type Twitter.

⁷ <https://www.comscore.com>

1.3 Définition

Dans la littérature, il existe plusieurs définitions de réseaux sociaux, nous présentons quelques-unes :

Un réseau social est constitué à la fois par un ensemble de personnes liées entre elles et par la force de ces liens. On peut aussi dire qu'un réseau social est un ensemble d'individus liés entre eux par des liens caractérisés par un degré de familiarité variable qui va de simple connaissance aux liens familiaux les plus étroits [5]. La figure représente un exemple de réseau social.

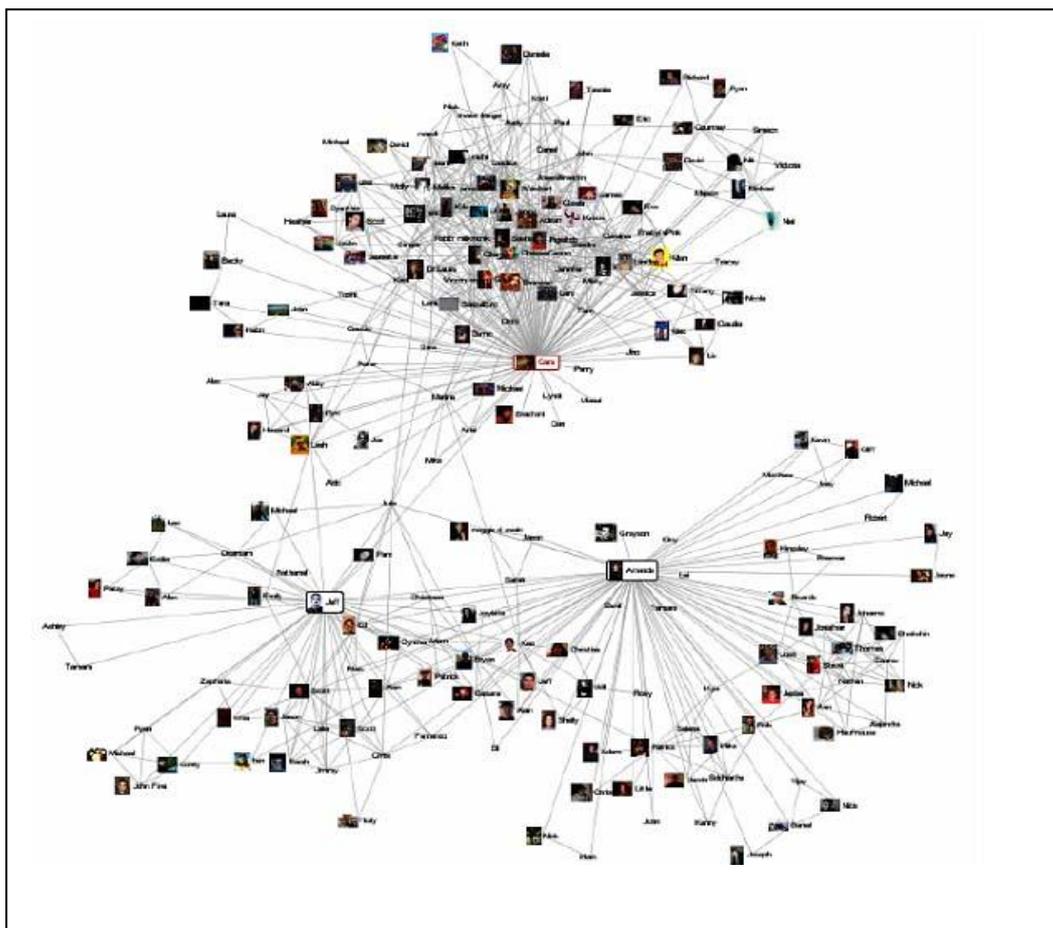


Figure 1.1 : Représentation de réseau social [6]

Pour Yahoo (*acteur dans le domaine des réseaux sociaux avec son Yahoo!360°*), un réseau social est « un terme assez large qui désigne des sites Internet qui aident leurs utilisateurs à créer leur propre profil Internet et partager une partie de leurs contenus préférés, y compris des photos et de la musique » [7].

Par réseau social, nous entendons donc toute plateforme en ligne dont la finalité est de mettre en relation des membres, et sur laquelle un individu peut s'inscrire librement,

construire son propre réseau, produire du contenu, le partager et interagir avec les membres de son réseau. Un membre peut y créer un profil public visible par tous ou privé, visible par ses contacts uniquement. C'est ce profil qui servira de carte de visite à l'internaute sur le réseau social dont il est membre. L'intérêt de telles plateformes est notamment de pouvoir suivre l'actualité des membres de son réseau et d'éventuellement la commenter [8].

Une autre définition dans [9] : Le réseautage social (*distinct du concept de réseau social en sociologie*) se rapporte à une catégorie des applications d'Internet pour aider à relier des amis, des associées ou d'autres individus employant ensemble une variété d'outils. Ces applications, connues sous le nom de « service de réseautage social en ligne » (*en anglais Social Networking*) deviennent de plus en plus populaires. Elles peuvent aussi permettre une meilleure distribution artistique, en favorisant la formation de contacts, et en invitant des artistes à assurer une visibilité de leur travail (*ex. musique, vidéo, photographie*).

Selon la définition proposée par Boyd et Ellison (2007) [10], les réseaux sociaux sont des espaces d'échange sur Internet qui permettent aux individus de construire des profils publics ou semi publics, associés à une liste de contacts inscrits sur le même site [11].

Actuellement, Les OSNs intéressent différents domaines de recherche tels que : sociologie, psychologie, communication, informatique, marketing, ...

1.4 Types de réseaux sociaux

Dans le monde des réseaux sociaux, on trouve plusieurs classifications suivant plusieurs critères, parmi ces typologies :

1.4.1 Typologisation des OSNs selon l'évolution et l'apparition

Nous retrouvons également une tentative de typologisation des réseaux sociaux numériques, selon laquelle les auteurs dressent l'évolution et l'apparition des réseaux. Pour eux, il existe huit types de réseaux : les réseaux généralistes, les réseaux politiques, les réseaux hyper locaux, les réseaux d'universités, d'entreprises, les réseaux associatifs, ceux des communautés d'intérêts et les réseaux de passionnés. Voici leurs typologies :

- **Réseaux généralistes**

Sont plutôt centrés autour de l'individu, et non plus de l'activité professionnelle. Les questions sont plus orientées sur les goûts culturels des membres. Ces sites permettent indirectement de nouer des affinités personnelles, sans pour autant avoir vocation unique d'être des sites de rencontre [12]. Ces sites permettent de créer et d'agrandir son cercle d'amis, le plus connu étant Facebook, le plus intime étant Meetic⁸. Ils sont organisés sous la forme de groupes spécifiques à des thématiques [13].

⁸ <http://www.meetic.fr/>

- **Réseaux politiques**

Du fait que le monde politique s'interroge beaucoup sur les réseaux sociaux et notamment après avoir mis pas mal de temps à apprivoiser les blogs, ils entrent de plein pied dans la réflexion des politiques avec deux expériences: la campagne d'Obama⁹ et celle de Ségolène Royale¹⁰. Ainsi nous trouvons en France « réseaux sociaux numériques politiques » par exemple : Créateurs du possible¹¹ de l'UMP et le coopol¹² (**Coopérative politique**) rattaché au PS (*partie socialiste*).

- **Réseaux hyperlocal**

Dans le but de renforcer les liens au niveau local, ce type de réseau a vu le jour. L'idée sous-jacente est de mieux connaître ses voisins de promouvoir et d'encourager les solidarités [3]. Par exemple : Voisineo¹³ et Peuplade¹⁴, la Ruche à Rennes¹⁵. De plus, l'hyperlocal peut aussi s'exprimer par la géolocalisation lié à la mobilité permise par les téléphones portables comme les iPhones (*ex : Foursquare*¹⁶).

- **Réseaux d'universités**

Comme Zeeya¹⁷, Réseau campus¹⁸, etnoka¹⁹.

⁹<http://bababillgates.free.fr/index.php/comment-obama-a-utilise-le-webmarketing-pour-remporter-lelection-americaine/>

¹⁰ Voir la Ségosphère, une cartographie réalisé par Linkfluence sur le site :<http://www.observatoire-presidentielle.fr/?pageid=12>

¹¹ <http://www.lescreateursdepossibles.com/>

¹² <https://www.lacoopol.fr/>

¹³ <http://www.voisineo.com/>

¹⁴ <http://www.peuplade.fr/home/nHome.php>

¹⁵ <http://www.ruche.org>

¹⁶ <http://www.foursquare.com>

¹⁷ <http://zeeya.net/>

¹⁸ <http://reseaucampus.com/>

¹⁹ <http://etnoka.fr/> <http://fr.beboomer.com/>

- **Réseaux d'entreprises**

Via les OSNs, ces entreprises peuvent évaluer leur présence en termes d'opportunité et de risque.

- **Réseaux associatifs**

Les auteurs donnent deux exemples : l'Association Française de Sociologie et les réseaux des créatifs culturels.

- **Réseaux sociaux de communautés d'intérêts**

Ce sont des réseaux sociaux qui proposent des types de relations beaucoup plus spécifiques et qui pour certains s'apparentent à des communautés d'intérêts. D'un côté, il existe des réseaux liés aux âges de la vie comme Beboomer²⁰ qui s'adresse aux seniors actifs de la génération baby-boom²¹. De l'autre côté, il y a des réseaux sociaux réservés aux enfants comme Globe2child²². Des réseaux encore plus spécifiques existent comme Memoree²³ dont l'objectif est de conserver les mémoires des siens, vivants ou décédés en partageant des souvenirs.

- **Réseaux de passionnés**

Les auteurs terminent leur typologie en notant que des réseaux comme Culture visuelle²⁴ qui traite de l'image sous toutes ses formes est déjà plus qu'une communauté d'intérêts et que Knowtex²⁵. Est le réseau des passionnés de la science.

²⁰ <http://fr.beboomer.com/>

²¹ Le baby-boom ou « pic de natalité » est une augmentation importante du taux de natalité dans certains pays, juste après la fin de la seconde guerre mondiale.

²² <http://www.globe2child.org/xwiki/bin/view/main/webhome>.

²³ <http://www.memoree.fr/>

²⁴ <http://culturevisuelle.org/>

²⁵ <http://www.knowtex.com/>

1.4.2 Typologisation des réseaux sociaux numérique selon la fonctionnalité

D'autres chercheurs ont tenté de classer et de catégoriser les réseaux sociaux. Pascal Faucompré [14] a essayé dans un ticket intitulé : *Ras le bol des réseaux sociaux ?*, d'élaborer un classement suivant la fonctionnalité :

- **Networkings** : qui permettent les échanges entre les professionnels.
- **Bloglikes** : ils ressemblent vaguement aux blogs et sont souvent le refuge d'ados en mal de reconnaissance.
- **Spécialisés** : ils regroupent des communautés autour d'un thème bien précis.
- **Micro-blogging** : chat publique instantané.
- **Fourres-tout** : ce sont les inclassables qui se servent du collaboratif ou du participatif pour alimenter leur service.
- **Open-sources** : plateformes qui permettent aux utilisateurs de créer leurs propres réseaux.

1.4.3 Typologisation selon wikipédia

Sur Wikipédia, nous trouvons un classement de réseaux sociaux selon trois catégories : Réseaux ouverts, Réseaux sur invitation et Services en ligne de réseautage professionnels qui favorisent les rencontres professionnelles, les offres de poste et la recherche de profils.

1.4.4 Typologisation selon le point de vue des chercheurs

Thelwall [14] catégorise les OSNs selon leurs trois objectifs qu'il nomme respectivement : socialisation, réseautage et navigation (*sociale*) :

- **Réseaux sociaux de socialisation**

Cette catégorie se caractérise par son aspect récréatif et conçue pour les loisirs de communication sociale entre les membres. Les connexions sont souvent utilisées pour trouver et afficher des listes d' « amis » existants d'ores et déjà. De plus, les connexions sont souvent utilisées pour trouver d' « amis » existants hors ligne, comme par exemple : MySpace et Facebook et Cyworld²⁶ (*un monde visuel coréen lancé en 2001*).

- **Réseaux sociaux de réseautage**

Utilisés davantage pour trouver de nouveaux contacts et peuvent servir pour trouver de nouveaux contacts et entrer en connexions avec des personnes inconnues auparavant comme c'est le cas de LinkedIn ou Viadeo, site de réseautage à caractère professionnel.

²⁶ <http://us.cyworld.com>

- **Réseaux sociaux de navigation**

Comme Digg²⁷ ou Del.icio.us²⁸, qui sont des sites de partage de liens Internet (*connu sous le social bookmarking*). Ce type de réseaux est un moyen pour aider les utilisateurs à trouver une information ou des ressources. Autrement dit, nous trouvons des listes de contacts, listes permettant l'accès à l'information et aux ressources associés à ceux-ci. Les membres peuvent soit lire les propositions mises en avant en page d'accueil, soit utiliser la navigation sociale en lisant les informations postées ou recommandées par leur amis, ou bien pour certains, recourir plusieurs objectifs [5].

1.5 Intérêts des réseaux sociaux

L'écosystème des médias sociaux illustré dans la figure 1.2 s'organise autour de quatre grands usages : la publication, le partage, la discussion et le réseautage. À chacun de ces usages correspondent des services, certains étant dédiés à une fonction bien particulier (*ex. Instagram qui ne sert qu'à publier des photos depuis son Smartphone*), tandis que d'autres sont plus versatiles (*ex. Tumblr est une plateforme de blogs créée en 2007 à New York qui est difficile à caser*).

Les quatre grands usages sont complémentaires : les utilisateurs publient des contenus, en partageant d'autres, cela génère des conversations qui leur permettent de développer leur réseau de contacts. Ceci étant dit, les services sont classés selon les quatre grands usages suivant :

1. **La publication** avec les plateformes d'hébergement de blog (*ex. WordPress, Blogger, Live Journal, TypePad, Over-Blog, SquareSpace, Medium, ...*), la nouvelle génération de services de publication minimalistes (*ex. Svbtle, Ghost, Sett*), les wikis (*ex. Wikipedia, Wikia, Mahalo...*) et les services intermédiaires de publication / partage comme Tumblr.
2. **Les services de partage** de photos (*ex. Flickr, Imgur, 500px Pinterest, ...*), de vidéos (*ex. YouTube, Vimeo, Dailymotion...*), de musique (*ex. Spotify, Deezer, SoundCloud, MySpace...*), de liens (*Delicious, Scoop.if*), de lieux (*ex. Foursquare, Swarm*), les applications mobiles (*ex. Instagram, Slingshot, Riff, Vine...*), les communautés d'acheteurs (*ex. TheFancy, Polyvore, Shopstyle, Bezar, Lyst, Yeay...*) ainsi que des communautés verticales comme ces trois-là dédiées aux créateurs (*ex. Behance, Dribbble, DeviantArt*).
3. **La discussion** avec les plateformes conversationnelles (*ex. Quora, Reddit, Github, Tieba, Baidu, Disqus, Muut*), les outils de communication grand public (*ex. Skype, Sina Weibo, Tencent Weibo*), les applications mobiles de communication (*ex. Facebook Groups, BlackBerry Messenger, MessageMe, Telegram, Pheed, Hike, Wire, Bleep...*), et les outils de communication professionnels (*ex. Slack, Yammer, Chatter, Jive Chime, Caliber...*).

²⁷<http://digg.com>

²⁸<http://delicious.com>

4. Le réseautage avec les réseaux sociaux grand public (ex. *Tagged*, *Nextdoor*, *Notabli*, *Ello*...), et leurs équivalents asiatiques et russes (ex. *Qzone*, *Vkontakte*, *RenRen*, *Mixi*, *StudiVZ*...), les services de rencontre (ex. *Badoo*, *OKCupid*...), les applications mobiles de rencontre (ex. *Tinder*, *Skout*), et les réseaux sociaux BtoB (ex. *LinkedIn*, *Viadeo*, *Xing*).



Figure 1.2 : Panorama des médias sociaux 2016 [15]

Selon le panorama dressé ci-dessus, il est opportun de constater quatre grands domaines d'application qui représente l'intérêt des réseaux sociaux [5] :

a - Outils d'expression

Qui permettent aux utilisateurs de prendre la parole, de discuter et d'agréger leurs productions, ils comportent :

- **Outils de publication** : dans lesquels nous trouvons les blogs (ex. *Blogger*, *Typepad*, *WordPress*), les plateformes de wiki (ex. *Wikipédia*, *Wetpaint*, *Wikia*), les plateformes de microblogging tels que *Twitter*, les portails de news et de «journalisme citoyen» (ex. *Digg*, *Wikio*, *Le Post*) ainsi que les outils de livecast (ex. *JustinTV*, *Ustream*, *BlogTV*).
- **Outils de discussion** : il s'agit des plateformes de forum (ex. *Phorum*) et de forum vidéo (Seismic), des logiciels et services de messagerie instantanée (ex. *LiveMessenger*, *Meebo*,

eBuddy et Y! Messenger) des services de gestion de commentaires comme Cocomment, Backtype, IntenseDebate et Disqus.

- **Services d'agrégation**, par exemple : Friendfeed, Profilactic, LifeSteam et autres.

b – Services de partage

Qui permettent la publication et le partage de contenu. Ils comportent :

- **Outils de partage de contenu** : de vidéos (*ex. YouTube, Dailymotion, Vimeo*), de photos (*ex. Flickr, SmugMug, Picasa et Fotolog*), de musique (*ex. Last.fm, Deezer, ...*), de liens (*ex. Delicious, Reddit, ...*), de documents (*ex. Slideshare, Scrib, Slideo*).
- **Partage de produits** : offerts par les services de recommandations (*ex. Crowdstorm, ThisNext, StyleHive*), de suggestions d'évolution (*ex. User Voice, GetSatisfaction*) et d'échange (*ex. LibraryThing, Shelfari, SwapTree*).
- **Partage de lieux** : en fonction des adresses (*ex. BrightKite, Loopt, Whrrl, Moximity*), des évènements (*ex. Upcoming, Zvents, EventFul, Socializr*), des voyages (*ex. TripWolf, TripSay, Driftr, Dopplr*).

c – Services de réseautage

Leur but est la mise en relation des individus.

- Les réseaux de recherche d'anciens camarades (*ex. CopainsDavant, Trombi, MyYearBook*), de personnes (*ex. MyLife*) ou de « conjoints » (*ex. Badoo*).
- Les réseaux de niche : PatientsLikeMe, Dogster, etc.
- Les réseaux B to B : LinkedIn, Viadeo, Xing, etc.
- Les réseaux mobiles : Groovr, MocoSpace, etc.
- Les outils de création/gestion de réseaux : Ning, KickApps, CrowdVine, ...

d – Services de jeux en ligne

- Les portails de « casual games »: Zynga, SGN480, ThreeRings, PlayFish, CasualCafe, ChallengeGames.
- Les « MMORPG »: jeux de rôle massivement multi-joueurs comme world of Warcraft, EverQuest, Lord of the Rings Online, EVE Online, Lineage, Dofus, Runespace.
- Les « MOG » : jeux massivement multijoueurs, par exemple : Drift City, Maple Story, Combat Arms, Quake Live.
- Les « casual MMOG » qui selon l'auteur, se positionnent à mi-chemin entre les deux catégories précédentes (MMORPG ET MOG) : Puzzle Pirates, Club Penguin, Ncopcts, Gaia Online, SmallWorlds, OurWorld.

La figure 1. 3 illustre bien la richesse et la diversité des médias sociaux.

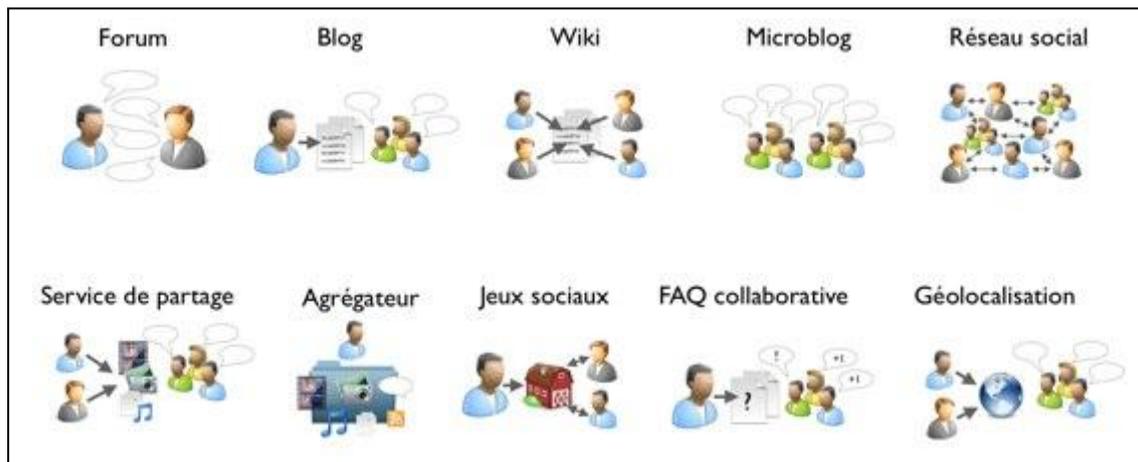


Figure 1 .3 : Les différents types de médias sociaux [15]

1.6 Principaux réseaux sociaux

1.6.1 Exemple de réseaux sociaux grands publics

- **Facebook [8]:**

c'est le réseau social le plus connu : proche d'un milliard de membres d'inscrits (à autre moment). Le principe est d'échanger avec sa communauté d'amis sur tout et n'importe quoi. L'inscription est obligatoire pour l'utiliser. Pour être amis sur Facebook avec une personne, il faut lui envoyer une demande et que cette dernière l'accepte. Facebook permet également de réagir sur les commentaires et news postés par ses amis via le « Like » ou J'aime. C'est un moyen pour dire que l'on a trouvé un commentaire ou un post à son goût. Il est devenu fréquent d'entendre le verbe « Liker » dans une conversation. Facebook permet beaucoup d'autres chose : discussion instantanée, envoi de message direct, identifier des amis sur une photo ...

- **Twitter [8]:**

c'est le « petit » qui monte en flèche. Il s'agit d'une plateforme de microblogging. Comme Facebook, Twitter permet de partager les tweets avec d'autres. Le fonctionnement est toutefois différent de Facebook : une limitation à 140 caractères par message, la possibilité de suivre d'autres comptes, pas de demande d'invitation, le partage de photo, de vidéo ou d'article se fait par l'utilisation de lien. En gros, on peut publier des SMS avec des liens. Sur Twitter, contrairement à Facebook, nous n'avons pas un mur mais une timeline ou fil. Alors que Facebook est destiné à échanger avec ses amis, Twitter, lui, possède un côté plus pro et relationnel. La notion d'influence y est très présente de par sa logique de suivi d'informations données par d'autres. Avec Twitter, nous lisons les tweets de personnes que nous suivons les Retweeter pour les partager avec vos followers.

- **Youtube [8]**

il peut aussi être classé dans les réseaux sociaux puisqu'il permet de partager ses vidéos et de commenter les vidéos postées. Youtube appartient à Google. Il n'est pas nécessaire d'être inscrit pour regarder les vidéos postées. Mais, pour y déposer vidéos et commentaires, une inscription à Youtube est obligatoire. La logique d'amis avec les autres internautes inscrits existe et Youtube permet également de suivre des thèmes pour être informé de leurs actualités. La grande majorité du contenu présent provient de particuliers [5].

- **Klickr [8]**

il appartient également à Google. Il est distribué par Yahoo et se prononce Flickère. Flickr met à notre disposition un espace pour poster des photos et des vidéos. Il existe également une logique d'amis et groupe d'amis comme Facebook ce qui nous permet de choisir le niveau de partage souhaité : tout le monde, quelque amis, notre famille... nous pouvons également géolocaliser le lieu où a été prise la photo. Nos amis peuvent commenter nos photos.

1.6.2 Exemple de réseaux sociaux professionnels

- **Linkedin [8]**

un réseau professionnel international permet la mise en relation entre des professionnels. Il offre un espace de présentation de ses compétences et expériences qui peuvent être consultable par le public. Très utile pour le recrutement. En Mars 2011, le site revendique plus de 130 millions de membres issus de 170 secteurs d'activités dans plus de 200 pays et territoires.

- **Viadeo [8]**

il est le pendant français du réseau social LinkedIn. Il permet lui aussi de construire et de gérer son réseau professionnel. Viadeo est plus populaire et plus connu en France que LinkedIn. Il offre à peu près les mêmes possibilités que LinkedIn.

- **Xing [8]**

c'est une plateforme allemande qui permet de construire et d'agréger son réseau professionnel. Il possède 3,5 millions d'utilisateurs répartis sur plus de 190 pays [16].

Le tableau suivant décrit les avantages et les inconvénients des réseaux sociaux :

Nom	Facebook	Twitter	Linkedin
Avantages	- Créer une page Facebook simple et la plateforme offre de plus en plus d'outils qui permettent notamment aux entreprises et marques de suivre la progression de leur nombre d'adaptes ainsi que leur données démographiques.	- On peut s'informer en temps réel et faire la même chose pour diffuser une information importante et pertinente. Il n'y a aucun problème à publier fréquemment puisque les tweets se suivent rapidement.	- C'est un réseau social professionnel par excellence, qui permet de bâtir un réseau à partir de ses expériences de travail mais également autour de ses intérêts et compétences.
Inconvénients	- Facebook demande une interaction avec les adeptes. On ne peut pas diffuser une information sur Facebook et ignorer la réaction des adeptes.	- Sur Twitter, on ne peut pas envoyer un message qui contient plus de 140 caractères. Twitter oblige les entreprises à être concises et claires.	- L'interaction sur LinkedIn, à part dans certains groupes, est vraiment limitée. Une interface un peu moins accessible ne rend pas les choses faciles.

Tableau 1.1 : Tableau comparatif entre les réseaux sociaux

1.7 Cartographie des réseaux sociaux numériques

Une cartographie des réseaux sociaux édité en Janvier 2016, montrant les sites de réseautage social les plus populaires par pays, selon Alexa²⁹.

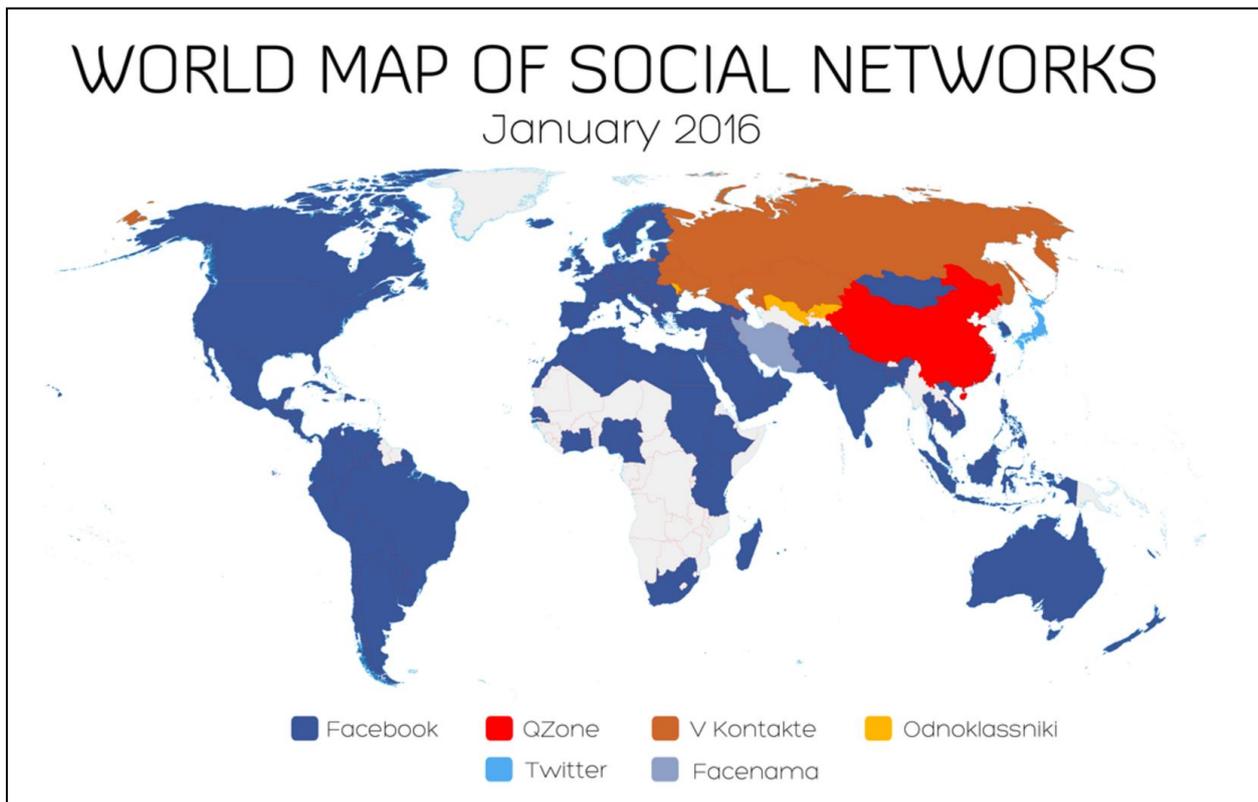


Figure 1.4 : Carte mondiale des réseaux sociaux [17]

Nous pouvons noter la très forte implantation de Facebook en Europe Occidentale, sur le continent américain, en Afrique du Nord, ainsi qu'en Australie.

1.8 Fonctions et fonctionnalités

Le fonctionnement de ces sites est toujours le même. Pour créer sa page d'accueil, on demande au nouvel arrivant de définir un profil qui constituera la représentation qu'il a de lui-même sous la forme de textes, de photos, de vidéos, de musiques et de liens. Des questionnaires lui imposent parfois de préciser ses goûts, ses affinités, ses opinions, etc. Facebook permet aux utilisateurs d'enrichir leur profil avec toutes sortes d'applications. Les listes d'« amis » font le lien entre les profils. Chaque participant au site peut demander à

²⁹ <http://www.alexa.com>

d'autre participants au même site de faire partie de ses « amis » ; si sa proposition est acceptée, le nom et une photo (*ou un avatar*) de l'« amis » sont affichés sur le site de l'autre. Par « contagion » les participants du site étendent ainsi continuellement leur liste « amis ». Mais qui dit « amis » sur Facebook ne dit pas forcément camarade dans la vraie vie [18].

La liste d'« amis » ne se résume pas à l'exposition des liens forts d'amitié qui ont une réalité dans la « vraie vie ». Elle permet aussi de mesurer l'audience que chaque participant a auprès d'autres participants, souvent inconnus dans un premier temps, qui partagent les mêmes activités, goûts, opinions, etc. Ces liens sont dits faibles. Deux pratiques différentes tournent autour de cette distinction : l'accès d'un profil peut être restreint aux « amis », liens forts ou faibles, ou au contraire ouvert à tous, pour faciliter la croissance de la liste.

La recherche à tout prix de nouveaux « amis » peut conduire à accepter des inconnus dans la liste, sans vérifier leur identité, ni s'ils sont véritablement des « amis » d'« amis ». C'est un des moyens couramment utilisés pour atteindre les données protégées d'un compte.

La troisième par ordre d'importance est la fonction de revue publique ; « **Témoignages** », « **Commentaires** », « **Panneau d'affichages** [19]. Cette fonctionnalité essentielle des réseaux sociaux permet à tout visiteur de déposer des **commentaires** sur une page d'accueil (*La sienne ou celle de quelqu'un d'autre*). Ces commentaires sont lisibles par toutes les personnes autorisées à accéder au profil (*Mur sur Facebook*). Toutes ces possibilités permettent une médiatisation de soi via un profil détaillé et des contenus diversifiés (*statuts, photos, liens*), ainsi qu'une socialisation active avec les membres de son entourage au moyen des commentaires.

La plupart des sites sociaux propose l'échange de messages privés entre participants. Ils peuvent aussi offrir des services de partage de photos, de vidéos, et offrir des interactions avec les téléphones mobiles. Ces trois fonctionnalités - profils, listes d'amis, commentaires – constituent la structure de base des réseaux sociaux numériques, même si certains d'entre eux proposent des fonctions complémentaires. Les réseaux sociaux numériques permettent à leurs visiteurs de naviguer d'un « Ami » à l'autre et de correspondre avec quiconque possède un profil visible [19].

Les réseaux sociaux numériques permettent la navigation d'un profil à l'autre et rendent accessible la communication avec tout participant. Le mode d'accès le plus répandu est cependant basé sur l'existence de groupes d'« amis » préexistants, ayant rejoint le site ensemble, pour conserver ensuite le contact entre eux (*élèves d'une classe, d'une école, etc.*) et pouvoir communiquer deux à deux tout en faisant partie d'un groupe [18].

1.9 Enjeux des réseaux sociaux numériques

Les enjeux dans le domaine des réseaux sociaux sont multiples tant pour les utilisateurs que pour les acteurs. Du côté des utilisateurs, les réseaux sociaux regroupent toutes sortes d'individus provenant de tous secteurs. Ainsi, sur le réseau LinkedIn, les utilisateurs du service viennent de tous les horizons : plus de 120 secteurs d'activités sont représentés et aucun secteur ne représente plus de 11% de la base des inscrits. D'après Scott Allen³² que 11,8% des utilisateurs de LinkedIn sont PDG, 10,2% sont vice-présidents ou directeurs généraux et 1,3% sont membres d'un conseil d'administration. Cela ne vaut pas que pour les réseaux sociaux professionnels. Sur MySpace, par exemple, il est possible de trouver la plupart des grands groupes de musiques qui ont créé leur page MySpace officielle.

De ce fait, il est possible de retrouver quasiment n'importe quelle personne quel que soit sa situation géographique ou son poste au sein de la société [11]. En plus de pouvoir visualiser des informations sur les personnes, les réseaux sociaux offrent l'opportunité d'entrer en contact avec toutes ces personnes. Les liens entre tous les membres d'un réseau sont les profils personnalisés, ce que l'on peut appeler la carte d'identité numérique. Ainsi, les réseaux sociaux permettent de gérer son identité numérique ainsi que sa réputation en ligne.

Lorsque l'utilisateur remplit sa fiche, il a le choix d'y intégrer les informations qu'il souhaite et de cacher celles qu'il estime privées. Les réseaux sociaux lui offrent donc de la visibilité et lui permettent de contrôler son « extimité ». Plus l'internaute arrive à se mettre en avant et se rendre visibles seront les informations choisies. De même selon les informations entrées sur la fiche d'identité, l'utilisateur va pouvoir se mettre en avant devant telles ou telles personnes et dans un cadre bien précis (*recherche d'emploi, contact pour développer une entreprise...*)

Et en raison de la popularité croissante des réseaux sociaux en ligne (OSN) et énorme quantité de données partagées sensibles, la préservation de la vie privée devient un enjeu majeur pour les utilisateurs OSN. Les services de réseaux sociaux existants sont centralisés et les entreprises qui fournissent les services ont l'autorité exclusive de contrôler toutes les données des utilisateurs. Les utilisateurs ont également peu de contrôle sur comment et ce que les informations les concernant sont présentés à leur amis en ligne.

La présentation de leur information dépend en grande partie sur la conception du service de réseau social que les utilisateurs utilisent plusieurs architectures décentralisées ont récemment été proposées pour des OSN décentralisée basée sur les réseaux P2P comme une alternative des architectures OSN centralisées.

³⁰ Editeur d'onlineBusinessnetworks.com qui fournit une étude sur l'utilisation de LinkedIn

Il s'agit essentiellement là d'enjeux économiques mais aussi de visibilité. Ainsi les milliers d'utilisateurs inscrits offrent, indirectement, une source de revenu importante. MySpace compte entre 80 et 100 millions de profils créés dont un million rempli en détail.

Il est ainsi possible de les cibler très précisément pour leur proposer du contenu publicitaire en adéquation avec leurs passions et leurs centres d'intérêts. Outre la publicité directe, les réseaux sociaux, grâce à les nombres importants d'utilisateurs, offrent aux grands groupes (audiovisuels, musicaux, informationnels ...) un beau support de diffusion avec un large public qu'ils peuvent toucher de manières très pertinentes. Ainsi, des chaînes comme celles du groupe Fox ou encore des maisons de disques profitent des réseaux sociaux pour diffuser des contenus adaptés au profil des membres de communautés. De plus, il est très facile d'infiltrer la communauté (*création de page personnelle pour un utilisateur fictif, mise en ligne de vidéo marketing...*) afin de mettre en place une opération de marketing viral.

Ainsi, il est essentiel pour les acteurs du monde des réseaux sociaux d'accroître le nombre d'utilisateurs qui est au final leur vrai fonds de commerce.

1.10 Problèmes ouvert des réseaux sociaux

1.10.1 Problème de détection des communautés

Le problème de la détection de communauté dans les réseaux est un sujet relativement récent, mais qui a très rapidement conduit à une grande quantité de travaux.

Lorsque l'on étudie des réseaux de terrain, de grande taille et/ou représentent des données complexes tel que les réseaux sociaux, le nombre de groupes que l'on cherche à obtenir ne peut être connu à l'avance. Ce qui a à nous intéresser à un autre problème plus complexe, celui de la détection de communautés.

On peut définir le problème de la manière suivante : pour un réseau donné, comment le décomposer en un nombre inconnu de groupes de nœuds de manière à ce que ces groupes de nœuds satisfasse efficacement le problème de la minimisation des liens inter-communautés, et la maximisation des liens intra-communautés.[20]

1.10.2 Problème de sécurité

En plus d'être l'origine de fuites d'informations nuisibles, les sites de réseaux sociaux peuvent être utilisés comme plateforme d'attaque contre votre système ou pour mener des escroqueries. Voici quelques étapes à suivre afin de vous protéger.

- **Ouverture de session** : Protégez votre compte de réseau social avec un mot de passe fort.
- **Chiffrement** : De nombreux sites comme Facebook, Google+ ou Twitter vous permettent de forcer le chiffrement de toutes les communications (appelé HTTPS). Si cela est possible, activez cette option.

- **Applications :** Certains sites de réseaux sociaux vous permettent d'ajouter ou d'installer des applications d'une tierce partie comme des jeux. Gardez à l'esprit que très peu (voir aucun) contrôles de qualité sont effectués sur ces applications et qu'elles peuvent avoir un accès complet à votre compte ainsi qu'aux données que vous partagez. Des applications malveillantes peuvent utiliser cet accès dans le but d'interagir avec vos amis en votre nom et ainsi voler ou utiliser des données personnelles. Soyez attentifs et installez uniquement les applications provenant de sites connus, assurez-vous également qu'après installation elles soient régulièrement mises à jour. Lorsque vous n'utilisez plus l'application, supprimez-la [21].

1.10.3 Problème de détection des rumeurs

Puisque le problème d'Internet et des media sociaux n'est pas la quantité, mais la qualité des informations : quel sens donner à une information ? Mais également quelle valeur donner à une information reçue ou lue sur les media sociaux ?

1.11 Objectif de ce mémoire

Le but de notre mémoire est de détecter si une information partagé sur les réseaux sociaux est une rumeur ou pas.

1.12 Conclusion

De nos jours, les réseaux sociaux numériques sont devenus des outils de communication incontournables. Dans ce chapitre, nous avons présenté les réseaux sociaux numériques, nous avons commencé par donner une définition, ensuite nous avons cité quelques typologies des OSNs, leurs intérêts, les enjeux.

Parmi les enjeux des réseaux sociaux est que la plupart reposent sur une architecture centralisée qui présentent deux problèmes. Tout d'abord, les informations sur un seul site OSN est inutilisable dans les autres. Deuxièmement tels sites ne permet pas aux utilisateurs plus contrôle sur la façon dont leurs informations personnels sont diffusés, ce qui entraîne des problèmes de confidentialité potentiels.

L'adoption d'une architecture décentralisée est une meilleure alternative pour pallier ces enjeux.

Plusieurs architectures décentralisées ont récemment été proposées pour OSN décentralisé.

Dans le chapitre suivant, nous allons présenter les différentes méthodes de détection ou éventuellement la prévention des rumeurs

Chapitre 2

Etat de l'art sur les méthodes de détection des rumeurs

2.1 Introduction

Les utilisateurs des réseaux sociaux sont envahis par le flux d'information diffusé continuellement et ils sont généralement influencés par ce dernier. Ainsi, ils rediffusent l'information, commentent des publications, achètent des produits ou des services et parfois organisent des évènements, signent des pétitions, etc.

La rumeur est une influence sociale car elle peut prendre son origine dans un groupe ou dans une institution pour agir sur les comportements, les croyances et les opinions d'un individu. Mais elle peut aussi partir d'un fait et avoir une influence sur toute une société. Il s'agit souvent d'une influence inconsciente car on ne se rend pas compte de la réalité de la rumeur et encore moins de l'impact de son message sur notre vie quotidienne.

Dans ce chapitre, nous allons présenter les différentes méthodes de détection des rumeurs et quelques algorithmes et approches.

2.2 Contexte et objectif de ce mémoire :

- **La fiabilité des informations trouvées :**

Dans le besoin d'information, il y a une hiérarchie à l'image de la pyramide des besoins de MASLOW³¹. Cette hiérarchie est définie dans Figure 1: Pyramide des besoins d'information. Dans ce besoin, nous abordons donc la question de la fiabilité de l'information. Un besoin stratégique pour la prise de décision. Alors que le besoin de communiquer est déjà satisfait par les moyens technologiques ou non, récents ou non.

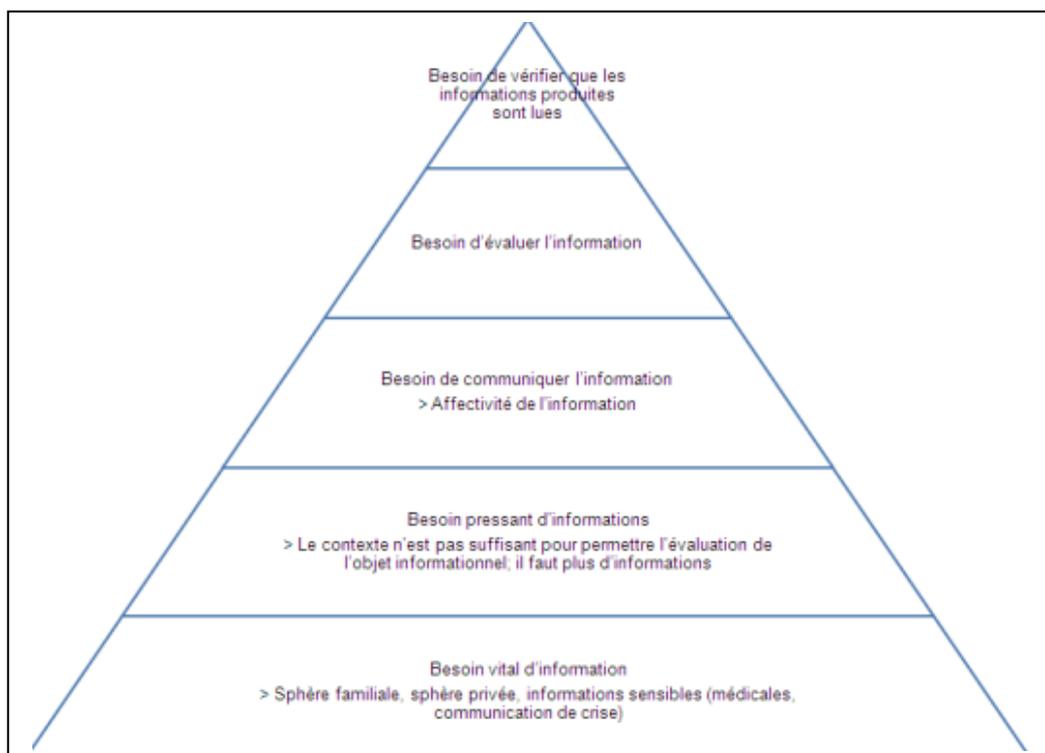


Figure 2.1: Pyramide des besoins d'information [22]

³¹ Abraham Maslow est un **célèbre psychologue américain du XXème siècle**. Beaucoup le considère d'ailleurs comme le père de l'approche humaniste. Il s'est notamment fait connaître à partir des années 1940 avec *A Theory of Human Motivation*, parue pour la première fois en 1943.

- **Informations trouvées sur les media sociaux**

Les media sociaux sont donc des supports d'échanges de messages et d'informations. Ces messages et informations ont un contenu. Ce contenu a de la valeur pour l'expéditeur et le destinataire. Cette valeur est plus ou moins importante en fonction de la problématique de chacun.

Mais la valeur de l'information envoyée par un autre membre du même media peut être perçue comme plus importante uniquement sur la base de la confiance faite. Sur les media sociaux, il n'y a pas entre les membres de hiérarchie ni d'autorité supérieur. Tous les membres ont la même quantité de connaissances et d'intérêts à participer à l'intelligence collective, donc à répondre à un besoin d'information exprimé.

Les échanges d'informations avant prise de décision ont donc aussi lieu, en partie, sur les media sociaux. Pour chacune de ces informations, l'origine de l'information, sa source, est évaluée différemment que s'il s'agissait d'une communication de masse.

2.3 Définitions

- **Information**

Pour la différencier entre autres de la « rumeur », une information est définie par Yves-François LE COADIC comme : « *une connaissance inscrite (enregistrée) sous forme écrite (imprimé ou numérisée), orale ou audiovisuelle sur un support spatiotemporel* [22].

- **Opinion**

Une opinion est un jugement de valeur individuel sur un objet par une personne dans un instant limité et identifiable. Une opinion n'est donc pas une information car elle est particulièrement applicable et valable pour la personne qui l'exprime. Elle est subjective [22].

- **Rumeur**

D'après Laurent Gaildraud, une rumeur est avant tout un bruit informel. Il existe, il persiste, il s'évapore. Il repart aussi vite qu'il est arrivé. Il s'agit d'une ou de plusieurs informations qui circulent et qui sont transmises par un vecteur (*individus, internet...*). Il n'y a pas forcément de source déterminée [23].

2.4 Aspects liés aux rumeurs

Tout d'abord, on peut décrire la rumeur selon sept caractéristiques réparties en trois classes (Cette classification a été formalisée par M.L.Rouquette³²) [24].

³² Michel-Louis Rouquette, Ses travaux ont marqué la psychologie sociale francophone, notamment dans les champs de l'étude de la pensée sociale (à travers les concepts de rumeurs, de représentations sociales et de nexus) et de la psychologie politique.

- **Situation**

La rumeur apparaît dans une situation de crise, mais elle n'est pas toujours le signe d'un dysfonctionnement social.

Les canaux formels de communication ne véhiculent qu'une information réduite sur certains événements ou aspects de cette situation. C'est-à-dire que devant la privation d'information les individus vont créer des rumeurs.

- **Processus de transmission**

La rumeur se transmet oralement de personne à personne, par la bouche-à-oreille, mais aussi par les médias (*presse, radio, internet*). Les canaux sont donc formels (*médias*) et informels (*bouche-à-oreille*).

Cette communication a lieu entre des individus également impliqués dans la situation.

- **Contenu**

Le contenu de la rumeur connaît différentes distorsions au cours de son processus de transmission.

Ce contenu traduit la pensée de désir de la population, elle témoigne de l'exercice d'une pensée sociale, la rumeur devient une sorte d'écran projectif où se déchiffre une dynamique socio affective.

Il entretient un rapport avec l'actualité.

2.5 Rumeur via les réseaux sociaux

Une rumeur via les réseaux sociaux est une nouvelle qui est diffusée via un site de réseautage. Dans un laps de temps très court, elle va prendre une ampleur considérable de manière virale. Sur internet, et dans les réseaux sociaux en particulier, plusieurs caractéristiques sont ajoutées à celles reconnues habituellement à la rumeur. Instantanéité, décentralisation, visibilité et internationalité sont les plus conséquentes [25]:

- **Instantanéité:** connectés en permanence, grâce à des outils technologiques de plus en plus abordables, les internautes ont très facilement accès à leur messagerie instantanée et aux réseaux sociaux. Cela favorise évidemment la diffusion rapide d'une information.

- **Décentralisation :** sur internet, la diffusion de l'information n'est pas réglementée par les principes déontologiques du journalisme professionnel comme ce l'est dans les médias d'information traditionnels. Tout un chacun a accès aux outils qui lui permettent d'éditer du contenu en ligne. Nous sommes donc tous des informateurs potentiels. L'information surgit

sur nos profils Facebook, nos chronologies Twitter, dans le flux de nos mails ou sur le blog d'une personnalité influente. Nous sommes tous susceptibles d'être les « relayeurs » d'une rumeur.

- **Visibilité** : les réseaux sociaux sont basés sur le principe d'une liste de contacts qui s'échangent des informations. Relayer les propos « d'amis » dans sa propre sphère est donc le premier pas vers une visibilité accrue par les relais successifs.

- **Internationalité** : les réseaux « d'amis » ne connaissant pas de frontières, les nouvelles, et donc les rumeurs, font rapidement le tour du globe de mail en mail ou de profil en profil. La confirmation ou l'infirmité d'une rumeur ne parviendra donc pas forcément aux oreilles « électroniques » de chaque internaute. Internet et les réseaux sociaux laissent donc libre cours aux croyances et offrent ainsi une plus longue vie aux rumeurs que lorsque celles-ci se propagent par la bouche à oreille humain.

2.6 Représentation des réseaux sociaux

La représentation la plus courante des graphes ou des réseaux est la représentation nœud-lien. Cette représentation a donc l'avantage d'être familière à la majorité des scientifiques. Cependant, elle souffre de problèmes de lisibilité lorsque les réseaux représentés sont soit grands (*beaucoup de nœuds*), soit denses (*beaucoup de liens*). Ces problèmes de passage à l'échelle sont d'autant plus importants que les données à analyser sont de plus en plus nombreuses ; ils sont devenus rédhibitoires avec l'apparition des réseaux sociaux en ligne qui sont à la fois grands et denses.

- **Représentation Matricielle**

Les graphes ont deux représentations canoniques : les diagrammes nœud-lien et les matrices d'adjacence. Une matrice d'adjacence représente chaque sommet d'un réseau à la fois comme une ligne et comme une colonne. Si deux sommets sont connectés, la case correspondant à l'intersection de la ligne et de la colonne est marquée. Traditionnellement, on utilise une valeur numérique (*0* marquant l'absence de connexion, *1* marquant la présence), soit par une marque graphique comme dans la Figure 2.2.

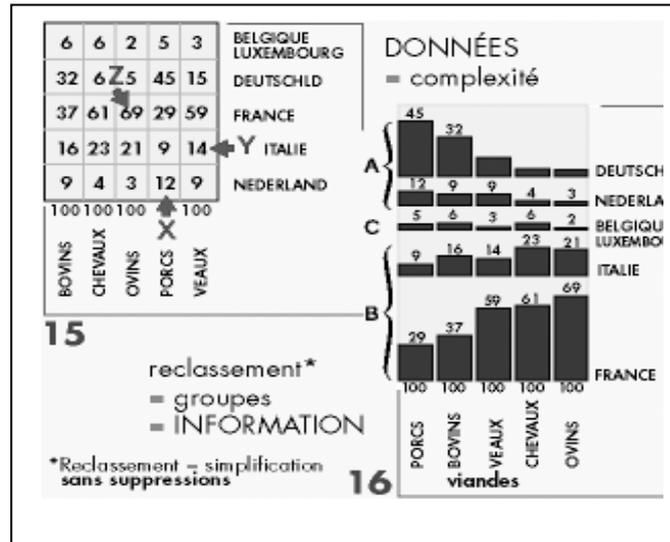


Figure 2.2 : Matrice réordonnable de J. Bertin, extraite de la Sémiologie graphique [26].

2.7 Méthodes de détection des rumeurs :

2.7.1 Greedy Sources Set Size (GSSS)

Si une rumeur est initiée intentionnellement par une personne, elle n'est pas corroborée par d'autres personnes. Par conséquent, en absence de collusion, il n'y a qu'une seule source de la rumeur dans le réseau.

Si une rumeur est initiée par un petit groupe de personnes en collusion, le nombre de sources indépendantes est juste de la taille du groupe.

A l'inverse, si une information n'est pas une rumeur, il peut y avoir de nombreuses sources indépendantes de l'information. Il est donc important d'estimer correctement le nombre de sources indépendantes.

2.7.2 Modèles de propagation d'influence dans un réseau social

Les modèles de propagation d'influence dans un réseau social les plus utilisés sont: *Independent Cascade model (ICM)*, *Linear Threshold Model (LTM)* et *Modèles basés sur le sujet* [27]. Dans ces modèles, le réseau est représenté par un graphe orienté. A l'instant $t=0$, un ensemble de nœuds initiateurs de la diffusion d'une nouvelle idée sont actifs. A un instant t , si un nœud adopte la nouvelle idée, il devient actif, sinon il est inactif. On suppose qu'un nœud inactif peut passer à l'état actif mais un nœud actif le restera tout au long du processus de propagation. Un nœud actif tente d'activer ces voisins. Le processus se poursuit jusqu'à ce qu'il n'ait plus d'activations possibles.

a. Independent Cascade Model

Dans ce modèle, une probabilité $p_{u,v}$ est associé à chaque lien (u,v) où u et v sont deux nœuds du réseau. $p_{u,v}$ est la probabilité que u réussit à activer v (u influence v).

Cette probabilité peut correspondre au taux de communication entre les deux nœuds, à la proximité géographique ou se basant sur un historique de processus de propagation antécédent (*par apprentissage*) [28].

b. Linear Threshold Model

Dans ce modèle, un poids $w_{u,v}$ est affecté à chaque lien (u, v) tel que la somme des poids des liens entrant à v est inférieur à 1. Chaque nœud v est doté d'un seuil θ_v .

A un instant t , les nœuds parents de v qui sont actifs tentent de l'activer. v ne sera actif que si la somme des $w_{u,v}$ (u est un parent de v actif) est supérieur au seuil θ_v . Ce modèle correspond à dire qu'un utilisateur du réseau n'adopte une idée que si une proportion de ses relations l'ont déjà adopté.

c. Modèles basés sur le sujet

Dans ce modèle, l'analyse de l'influence porte sur un sujet spécifique. Par exemple, un utilisateur peut influencer un autre utilisateur sur le choix d'un produit, mais pas ou moins facilement sur une opinion politique. Une extension des deux modèles précédents (*ICM et LTM*) en prenant en compte le sujet de l'influence a été proposée par Barbieri, Bonchi et Manco [29]. Dans leur nouvelle version, la probabilité $p^z_{u,v}$ associé à chaque lien (u, v) représentent la force de l'influence de u sur v par rapport à un sujet z .

2.7.3 Détection d'influenceurs dans un réseau social

Les influenceurs dans un réseau social sont importants à détecter puisqu'ils constituent l'ensemble de départ dans un processus de diffusion d'influence. Si on cherche à maximiser l'influence dans le réseau, ces influenceurs seront privilégiés et seront les premiers porteurs d'information. Par contre, si on cherche à minimiser l'influence, ils seront à protéger, à bloquer ou à solliciter pour une contre campagne par exemple.

Les deux métriques les plus référencés dans les travaux de minimisation d'influence dans les réseaux sociaux sont données comme suit [28] :

a. Degré d'un nœud :

Le nombre de voisins d'un nœud. On parle aussi d'*in-degree* ou *out-degree* pour désigner respectivement, le nombre de voisins avec une relation entrante ou sortante.

b. Betweenness d'un nœud :

Le nombre des plus courts chemins entre chaque nœud vers tous les autres nœuds traversant ce nœud.

2.7.4 Approches de minimisation de l'influence négative dans un réseau social

Les travaux de recherche sur la minimisation de l'influence négative dans les réseaux sociaux peuvent être classés selon leur approche en trois classes : approche basée sur le blocage de nœuds, approche basée sur le blocage de lien et approche basée sur l'influence compétitive [28].

a. Approche basée sur le blocage de nœuds

Le principe est de bloquer un ensemble minimal de nœuds pour minimiser la propagation d'influence négative dans le réseau. Les nœuds sont généralement sélectionnés parmi les nœuds les plus influencés.

Plus formellement, le problème est défini comme suit : soit un réseau représenté par un graphe orienté $G = (V, E)$. V est l'ensemble de nœuds, $E \subset V \times V$ est l'ensemble des liens. Nous supposons qu'une information négative est propagée dans le réseau et un ensemble initial de nœuds I est influencé. Le but est de minimiser le nombre de nœuds influencés en bloquant un ensemble S de k nœuds, $S \subseteq \{V\}$ et k est une constante donnée. La fonction objective est :

$$\text{Minimiser } \sigma \{I|V \setminus S\} \dots \dots \dots (1)$$

Où $\sigma \{I|V \setminus S\}$ représente le nombre de nœuds influencés par I quand les nœuds de S sont bloqués. Dans cette approche, on suppose qu'une information négative est propagée dans le réseau social et un ensemble de nœuds est infecté (*influencé*). Le but est de minimiser le nombre de nœuds infectés en bloquant k nœuds parmi les non infectés. Un algorithme glouton est utilisé où à chaque itération on sélectionne un nœud parmi les nœuds non infectés qui maximise la décrémentation du nombre de nœuds infectés. L'algorithme proposé donne de meilleurs résultats que si on sélectionne les top- k nœuds non infectés selon leur out degree ou betweenness [28].

b. Approche basée sur le blocage de liens

Le principe de cette approche est de bloquer un ensemble minimal de liens pour minimiser la propagation de l'influence négative dans le réseau.

Plus formellement, le problème est défini comme suit : soit le réseau $G = (V, E)$.

Nous supposons qu'une information négative est propagée dans le réseau et un ensemble initial de nœuds I est influencé. Le but est de minimiser le nombre de nœuds influencés en bloquant un ensemble S de k liens, $S \subseteq \{V\}$ et k est une constante donnée.

La fonction objective est :

$$\text{Minimiser } \sigma \{I|V \setminus S\} \dots \dots \dots (2)$$

Où $\sigma \{I|V \setminus S\}$ représente le nombre de nœuds influencés par I quand l'ensemble de liens de S sont bloqués [28].

c. Approche basée sur l'influence compétitive

Le principe consiste à sélectionner un ensemble minimal de nœuds qui adopteront une contre campagne afin de minimiser l'effet de l'influence négative.

Plus formellement, le problème est défini comme suit : Soient le réseau $G = (V, E)$ un ensemble S_A de nœuds qui adoptent une contre campagne, un délai de détection de la contre campagne d . Le but est de trouver un ensemble S_C de k nœuds qui diffuseront une contre information pour sauver le maximum de nœuds. La fonction objective est :

$$\text{Maximiser } \sigma \{S_C ; S_A, d\} \dots \dots \dots (3)$$

Où $\sigma \{S_C; S_A, d\}$ représente le nombre de nœuds influencés par S

2.8 Critiques sur les méthodes de détection des rumeurs :

- Les solutions proposées se sont fixé comme seul objectif, la minimisation de l'influence négative en négligeant le temps d'exécution, un paramètre très important dans les réseaux sociaux. Les algorithmes proposés sont des algorithmes gloutons qui donnent des résultats qui approchent la valeur optimale mais qui sont longs en termes de temps d'exécution.
- L'évaluation de l'influence négative a été exprimée, dans la plus part des travaux, par le nombre de nœuds infectés. Ce nombre peut ne pas être significatif si ces nœuds, par exemple, ont très peu de relations ou tout simplement n'ont rien à perdre s'ils sont infectés.

2.9 Synthèse et Analyse

Le tableau 2.1, résume les travaux cités précédemment.

Approche	Contribution	Modèle diffusion		Algorithmes					Données d'expérimentation			Résultats	
		ICM	LTM	Utilisés		Testés avec			Réseau	Nombre de nœuds	Nombre de liens	Minimisation de l'influence	Temps d'exécution
				Glouton	autres	Degré	Betweenness	autres					
Bloquer des nœuds	Fixer k nœuds parmi les nœuds non infectés		X	X	X	X	X	X	Enron email communication network	36692	367662		+
	prend en compte le sujet de l'influence.	X			X	X	X		Sina microblog	2000	14426	+	
									Facebook	4039	88234		
Bloquer des liens	Fixer k lien pour minimiser l'influence négative (initiateur de l'approche).	X		X		X	X	X	blog network	12047	79920	+	
	Adaptation du travail précédent sous le modèle de diffusion Linear Threshold		X	X		X	X		Japanese Wikipedia network	9481	245044		
									blog network	12047	79920		
									Japanese Wikipedia network	9481	245044		

	Formalisation et preuve théorique que la fonction objective est supermodulaire		X	X		X	X	X	FORESTFIRE network	500	1691	+	
									Memetracker network				
	Evaluation du temps d'exécution	X		X		X	X		Facebook data set	4039	88234	+	-
									Digger data set	8193	56440		
Influence compétitive	La contre campagne est lancée après un délai, délai de détection d'une campagne de désinformation	X		X	X	X		X	2009 Santa Barbara network	26455	53132	+	
									2008 Santa Barbara network	12814	184482		
									2008 the Monterey Bay network	6117	62750		
	Protection de noeuds : Décontaminer les noeuds les plus influents	X	X	X		X		X	NetHEPT network	15233	31398	+	
									Facebook	63000	1.5 million		

	Sauver le maximum de nœuds avant un délai donné. Prend en compte le taux de transmission entre deux nœuds.			X		X		X	Facebooklike social network	1899	20296	+	
	Minimiser l'infection des communautés voisines Les nœuds protecteurs sont définis parmi les nœuds qui ont des relations avec les communautés voisines.			X		X		X	Enroll Email Communication network	36692	367662	+	
									e-print arXiv Collaboration Network 15233 58891 En	15233	58891		

Tableau 2.1 : Synthèse des méthodes de minimisation de l'influence négative dans les réseaux sociaux. [28]

En analysant le tableau 2.1, nous tirons les conclusions suivantes :

– La problématique de minimisation d'influence négative dans un réseau social est assez récente. Les premiers travaux publiés datent de 2008. Cependant, la problématique n'a pas attiré beaucoup l'attention de la communauté de recherche. Les travaux s'orientent plutôt vers l'influence positive et sa maximisation dans les réseaux sociaux.

– Parmi les trois approches, les deux approches consistant à bloquer des nœuds ou des liens ne peuvent être appliquées que par un administrateur du réseau social qui possède ce privilège de blocage d'utilisateurs ou de relations. Ceci limitera l'application de ces approches par un responsable marketing, par exemple, qui veut limiter la diffusion d'une contre publicité.

– Dans les trois approches (*blocage de nœuds, blocage de lien, influence compétitive*), le problème est modélisé comme un problème d'optimisation. Il s'agit dans tous les cas, de trouver un minimum de nœuds/ liens à bloquer ou un minimum de nœuds qui lanceront une campagne compétitive afin de minimiser le nombre de nœuds influencé négativement. Il est prouvé que la fonction objective est sous/super modulaire. Ceci garantit qu'une solution avec un algorithme glouton donne des résultats qui approchent la valeur optimale de $(1-1/e)$, soit 63%. La plupart des travaux ont donc opté pour une solution avec un algorithme glouton. Cependant, ces algorithmes sont très longs en termes de temps d'exécution.

– La majorité des travaux comparent les résultats obtenus à des heuristiques basées sur des mesures très utilisées telles que le degré et le betweenness. L'algorithme glouton donne toujours de meilleurs résultats en termes de minimisation du nombre de nœuds influencés négativement. Peu de travaux ont fait une évaluation en termes de temps d'exécution.

Ces derniers montrent que les algorithmes gloutons ne sont pas performants en comparaison avec les heuristiques citées précédemment.

– Les travaux proposés ne comparent pas leurs résultats aux travaux qui les précèdent. On n'a donc pas une idée sur la meilleure solution proposée jusqu'au là, ni en terme de minimisation d'influence ni en terme de temps d'exécution.

– Les données d'expérimentation sont généralement des données issues de réseaux sociaux réels. Seulement, la taille des réseaux testés reste limitée.

Elle varie entre 500 et 63 000 nœuds tandis que les réseaux actuels comprennent des centaines de millions de nœuds.

2.10 Conclusion

L'influence dans les réseaux sociaux est un sujet en plein essor. La problématique de minimisation de l'influence négative, par ses applications et ses avantages apportés aux réseaux sociaux, suscite l'intérêt de la communauté de recherche. Dans ce chapitre nous avons présenté la rumeur et les différentes approches de détection et de minimisation, ensuite nous avons cités les modèles de propagation d'influence dans un réseau social et quelques approches de minimisation de l'influence négative, en particulier leurs principes de fonctionnement.

Dans le chapitre suivant, nous allons proposer une nouvelle méthode de détection des rumeurs (*minimisation des effets négatifs*).

Chapitre 3

Proposition d'une nouvelle méthode de détection des rumeurs

3.1 Introduction

Les réseaux sociaux sont des médias populaires pour le partage d'informations. Dans ce chapitre, nous allons proposer une nouvelle méthode de détection des rumeurs. En effet, notre méthode est basée sur deux étapes. Premièrement, nous étudions la question de l'identification des sources de rumeurs en l'absence d'informations de provenance complètes sur la propagation des rumeurs. Deuxièmement, nous étudions comment les rumeurs (*fausses déclarations*) et les non-rumeurs (*information réelle*) peuvent être différenciés.

3.2 Première étape : Identification des sources de la rumeur

La première question que nous allons étudier dans ce cas est la suivante :

Si une rumeur est initiée par une seule source dans un réseau social, comment la source peut-elle être identifiée ?

Un réseau social est modélisé sous la forme d'un graphe $G = (V, E)$ où V est l'ensemble de toutes les personnes et E est l'ensemble de arrêtes où chaque arrête représente l'information entre deux individus. Nous supposons qu'un ensemble de k nœuds présélectionnés nommé M tel que ($M \subseteq V$) sont nos nœuds de contrôle (*moniteurs*). A des fins de rumeur, compte tenu d'une information spécifique, un moniteur (*nœud de contrôle*) indique s'il a reçu la rumeur ou non. Nous désignons l'ensemble des nœuds de contrôle qui ont reçu la rumeur par M^+ et l'ensemble des nœuds de contrôle qui ne l'ont pas reçu par M^- (où $M^+, M^- \subseteq M$). Nous appelons les premiers (M^+) moniteurs positifs et les derniers (M^-) moniteurs négatifs.

Pour identifier la source d'une rumeur, nous utilisons l'intuition selon laquelle la source doit être proche des moniteurs positifs mais loin des moniteurs négatifs. Par conséquent, pour chaque nœud x , notre algorithme calcule les quatre paramètres suivants :

1 Accessibilité à tous les moniteurs positifs : Nous calculons d'abord combien de moniteurs positifs sont accessibles depuis chaque nœud.

2 Distance aux moniteurs positifs : Parmi les nœuds qui peuvent atteindre tous les moniteurs positifs, les nœuds qui sont plus proches, en moyenne, sont préférées. En d'autres termes, pour chaque nœud x , nous calculons la distance totale

$$D_t = \sum_{m \in M^+ \text{ et } m \text{ est atteint par } x} d(x, m) \quad (1)$$

Où $d(x, m)$ est la distance de x à m , et trier les sources présumées en augmentant la distance totale des moniteurs positifs.

3 Réactivité aux moniteurs négatifs : Parmi les nœuds qui peuvent atteindre tous les nœuds dans M^+ et avoir la même distance totale à ces moniteurs positifs, nous utilisons l'accessibilité aux moniteurs négatifs en tant que troisième métrique. Pour chaque nœud x , nous comptons des moniteurs dans M^- qui ne sont pas accessibles de x et préfèrent des comptes plus importants.

4 Distance aux moniteurs négatifs : En dernière mesure, nous utilisons également la distance aux moniteurs négatifs. Pour chaque nœud x , nous calculons la distance totale

$$D_t = \sum_{m \in M \text{ et } m \text{ est atteint par } x} d(x, m) \quad (2)$$

Il est plus naturel que les moniteurs négatifs soient loin de la source des rumeurs, de sorte que les nœuds avec de grandes valeurs de distance totale sont préférés.

3.2.1 Identification de notre méthode (1^{ère} étape)

On propose que tous les moniteurs soient au moins **d** sauts nécessairement. Pour ce faire, il s'agit au hasard de la liste de tous les nœuds. Ensuite, à partir du premier nœud, il vérifie s'il s'agit d'au moins **d** sauts de tous les moniteurs déjà sélectionnés. Si c'est le cas, le nœud est sélectionné comme moniteur et le nœud suivant est coché. Notez que le premier nœud est toujours sélectionné comme moniteur. Ceci est répété jusqu'à ce que **k** moniteurs soient sélectionnés où il est impossible de sélectionner d'autres moniteurs. Notre méthode de sélection est appelée **Dist** et qu'est la plus grande **d** qui peut choisir **k** moniteurs. Pour ce faire, il commence par une grande valeur de **d** et diminue à chaque fois qu'il ne parvient pas à choisir **k** moniteurs, en commençant par le plus petit **d** jusqu'à ce qu'il puisse trouver **k** moniteurs.

3.2.2 Algorithme de la méthode proposée

D'après la proposition de notre méthode d'identification de la source des rumeurs et les paramètres ci-dessus qui nous calculons dans notre algorithme. Ce dernier va identifier les sources des rumeurs. L'algorithme est donné comme suit :

Algorithme d'identification de la source

Entrée : $G(V, E)$: graphe

M : ensemble de nœud de contrôle

$M^+ : \{m_1, m_2, \dots, m_n\}$

Sortie : X : ensemble de sources

$D[i]$: vecteur de distance

Debut

Pour $i \leftarrow 1$ à N faire

$D[i] \leftarrow 0$;

 Pour $m_j \in M^+$ faire

$D[i] \leftarrow D[i] + \text{distance}(m_i, m_j)$;

 Fin pour ;

Fin pour ;

$i_{\min} \leftarrow 0$;

$\min \leftarrow D[0]$;

Pour $i \leftarrow 1$ à N faire

 Si $D[i] < \min$ faire

$\min \leftarrow D[i]$;

$i_{\min} \leftarrow i$

 Finsi ;

$X \leftarrow i_{\min}$;

Fin pour ;

Algorithme 3.1 : Algorithme de l'identification de la source des rumeurs.

3.3 Deuxième étape : Comment détecter les rumeurs et non- rumeurs

Dans cette étape, nous allons expliquer comment faire pour détecter si l'information est une rumeur ou non- rumeur.

On propose que :

1 L'information est une rumeur

Si une rumeur est initiée intentionnellement par une personne et elle n'est pas répétée par d'autres personnes, ce qui signifie que il n'y a qu'une seule source de la rumeur dans le réseau.

Par contre si une rumeur est initiée par un petit groupe de personnes en collusion, le nombre de sources indépendantes est égal à la taille du groupe.

2 L'information n'est pas une rumeur

A l'inverse, si une information n'est pas une rumeur, il peut y avoir de nombreuses sources indépendantes de l'information. Il est donc important d'estimer correctement le nombre de sources indépendantes.

3.3.1 Identification de notre méthode (2^{ème} étape)

Initialement, il faut calculer le nombre de sources, En effet pour chaque source x , il faut trouver ou calculer l'ensemble des moniteurs positifs qui sont mentionné dans la première étape (*nombre des nœuds de contrôle qui ont reçu la rumeur M^+*). Si on trouve le nombre des moniteurs est petit, ça signifie que le nombre de sources est également petit par conséquent l'information est une rumeur. Par contre si le nombre des moniteurs positifs est grand, ça signifie qu'il peut y avoir de nombreuses sources indépendantes de l'information qui n'est pas une rumeur.

3.2.2 Algorithme de différencie la rumeur et non- rumeur

$C \leftarrow \{ \}$

Pour chaque $m \in V$, $P_m \leftarrow \{ \}$ **Faire**

Pour chaque $m \in M^+$, S_m l'ensemble de nœuds qui ont un chemin vers m **Faire**

Tant que $M^+ \neq \{ \}$ **Faire**

Soit x l'un des éléments les plus fréquents dans toutes les S_m où $m \in M^+$.

Ajouté x au C .

Pour chaque $m \in M^+$ **Faire**

Ajouté m au P_x et **supprimer** m de M^+ if $d(x, m) \leftarrow \infty$

Fin Tant que ;

Algorithme 3.2 : Algorithme calcule le nombre de sources

3.4 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle méthode pour déterminer si une information est une rumeur ou non. Pour trouver la source de la rumeur, notre algorithme calcule quatre paramètres qui sont utilisés pour trouver le nombre de moniteurs positifs et négatifs accessibles depuis chaque nœud et calculer la distance totale entre les moniteurs positifs ou négatifs ainsi que le nœud qui est la source à retrouver. Pour détecter si l'information est une rumeur ou non, notre algorithme calcule le nombre de sources et pour chaque source, on calcule le nombre de moniteurs positifs tel que le nombre de sources dépend de nombre de moniteurs positifs. S'il est petit, le nombre de sources est petit ce qui signifie que à l'information est une rumeur. Sinon, s'il est grand, le nombre de sources l'est aussi, ce qui signifie que l'information n'est pas une rumeur.

Chapitre 4

Evaluation des performances de la méthode proposée

4.1 Introduction

Après avoir vu le fonctionnement des algorithmes proposés dans le chapitre précédent, dans ce chapitre nous allons voir tout ce qui concerne l'évaluation des algorithmes proposés. Ce chapitre est organisé comme suit : premièrement, nous décrivons le logiciel Gephi Ensuite, nous présentons l'étude de cas d'un Twitter. Enfin nous allons évaluer les performances de l'algorithme proposé.

4.2 Implémentation

Pour l'implémentation de notre proposition, nous utilisons logiciel gephi pour analyser la cartographie Twitter.

4.2.1 Gephi

Est un logiciel pour visualiser, analyser et explorer en temps réel les graphes (aussi appelés réseaux ou données relationnelles) de tout type. L'outil permet de représenter, ordonner, agencer les structures, formes et couleurs pour révéler les propriétés cachées d'un réseau via des saillances visuelles.

Les travaux produits sont exportables dans plusieurs formats, notamment le PDF qui permet ainsi d'être visionné par un ensemble large de destinataires. Gephi peut également exporter au format .csv les données importées.

Les formats supportés par Gephi : GEXF14; GDF; DOT (language); GraphML; Graph Modelling Language.

Que représenter avec Gephi ?

Sur les réseaux sociaux :

- Les conversations autour d'une thématique (#hashtag ; combinaison de requêtes)
- L'écosystème d'un compte Twitter
- L'activité et l'écosystème d'une page Facebook

Sur le web :

- Dynamique de diffusion et de relations entre des sites web et blogs

Autres possibilités :

- Cartographie textuelle d'un corps de texte (nécessite un travail de filtre avec Automap)

Introduction à la représentation cartographique

Pour produire un réseau, deux informations sont nécessaires: une liste des acteurs composant le réseau et une liste des relations entre ces acteurs. On appellera les acteurs « nœuds » ou « node » et les relations « arêtes » ou « arc ». Le label correspond au nom du nœud donc de l'acteur.

L'arc est orienté, cela veut dire que la relation va du compte 1, au compte 2. Ce type de relation est par exemple utilisé pour montrer que le compte 1 a envoyé un tweet ou suit le compte 2 en fonction des données collectées.

Dans le cas de Twitter, on obtient par exemple deux types de liens si on se centre sur un nœud :

- Les liens entrants : Le compte Twitter est suivi / mentionné par une personne identifiée
- Les liens sortants : Le compte Twitter suit / mentionne une personne identifiée

Dans un graphe, un cycle est une chaîne simple dont les extrémités coïncident. On ne rencontre pas deux fois le même sommet, sauf celui choisi comme sommet de départ et d'arrivée.

On parle de « graphe connexe » si chaque node du graph possède au moins un lien de liaison avec tous les autres points. Deux sommets sont adjacents s'ils sont reliés par une arête. Un arbre constitue un graphe connexe ne présentant aucun cycle.

Un graphe est dit complet si toutes les paires de sommets sont adjacentes.

4.3.2 Statistiques de type centralité eigenvector

C'est une mesure statistique, permet d'identifier des personnes importantes au sein du réseau.

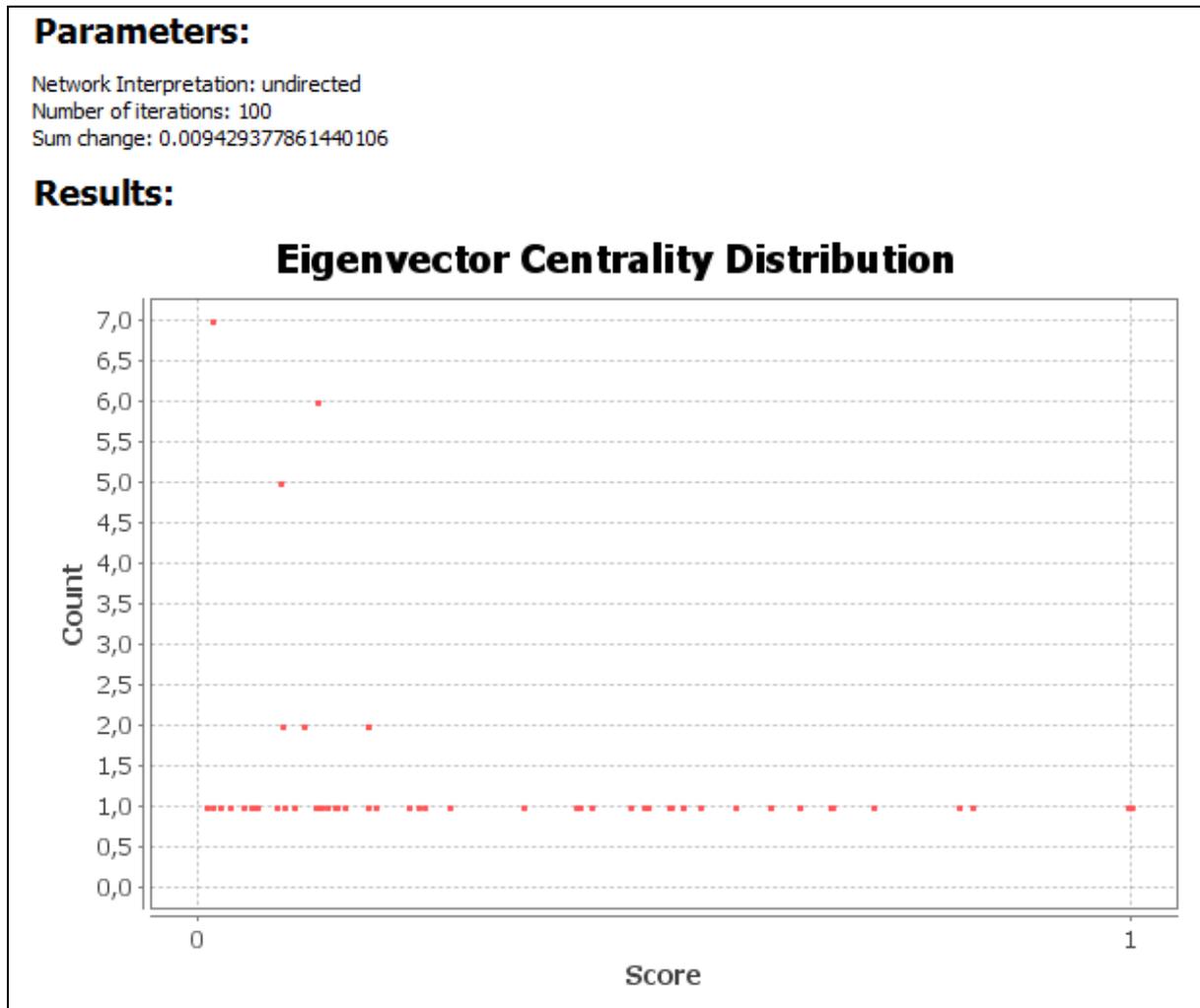


Figure 4.2 : Statistiques de type centralité eigenvector.

4.3.3 Statistiques de type modularité

Permet d'identifier les roues de personnes existantes.

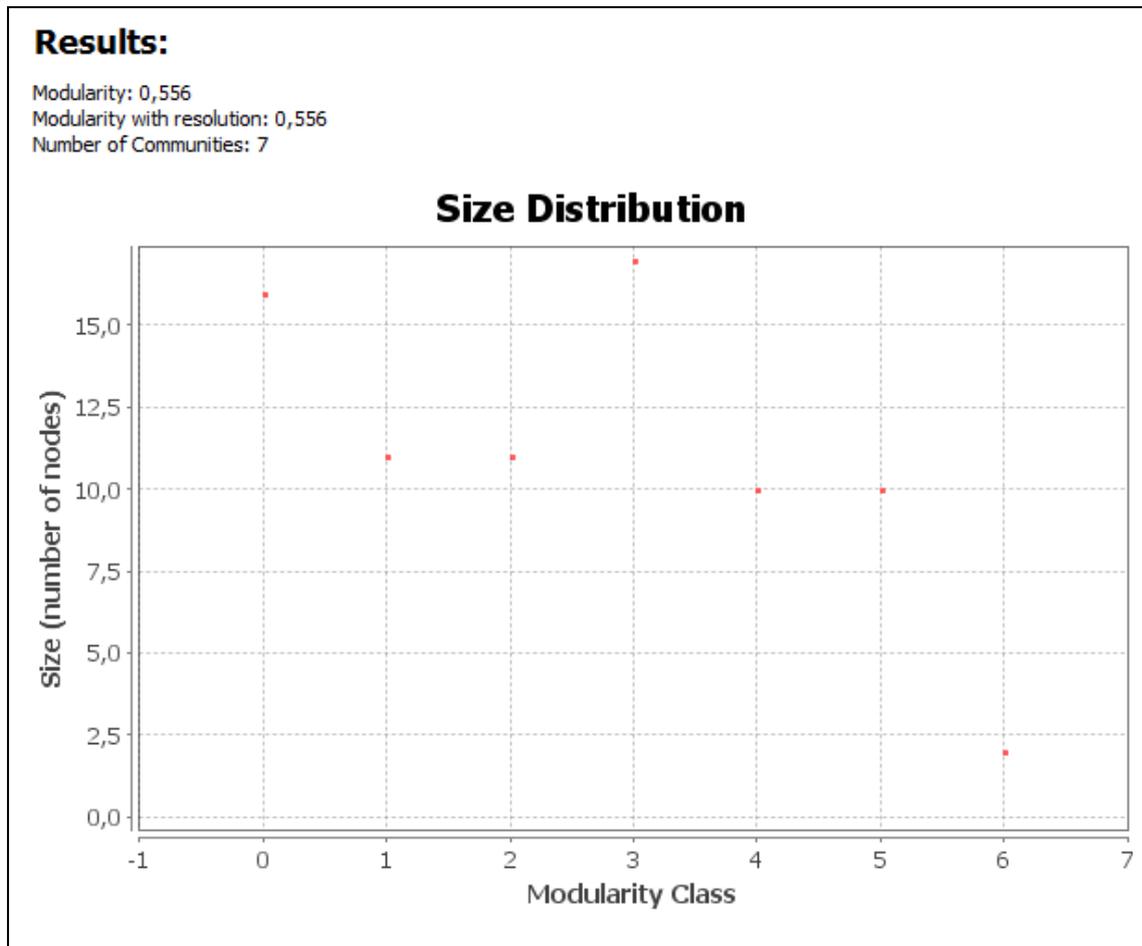


Figure 4.3 : Statistiques de type modularité.

4.3.4 Statistiques de type degré

Permet de calculer le degré d'une relation au sein du réseau.

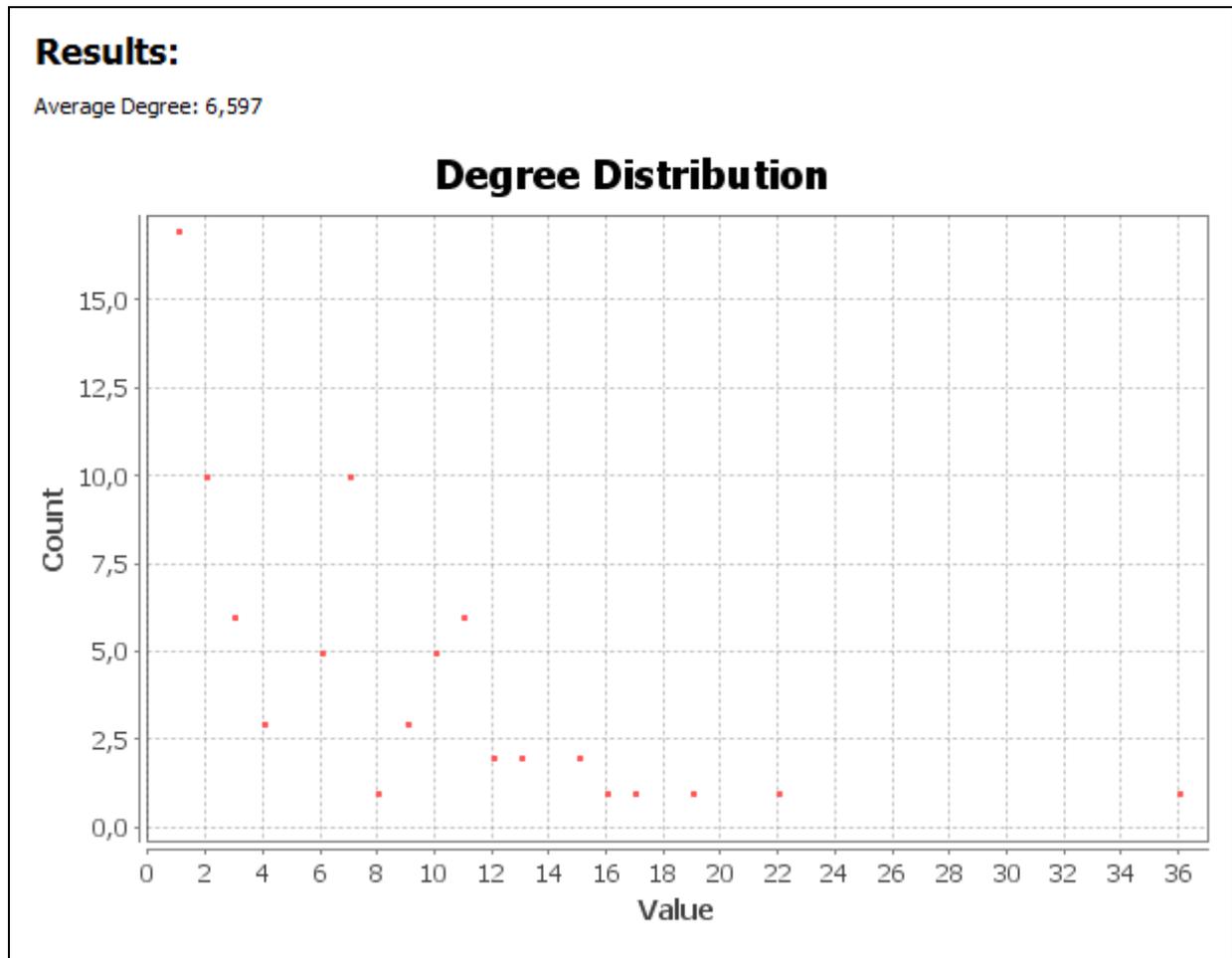


Figure 4.4 : Statistiques de type degré.

4.4 Conclusion

Dans ce chapitre, nous avons d'abord analysé une cartographie Twitter en utilisant le logiciel Gephi, puis dans l'évaluation, nous avons présenté quatre résultats d'analyse, la première c'est statistiques de type centralité eigenvector, la deuxième c'est de type modularité et la dernière de type degré.

Conclusion et Perspectives

Les réseaux sociaux permettent à n'importe qui de partager de l'information avec un grand nombre de personnes, cette information pouvant aller du simple texte (documents, annotations, commentaires, ...) aux images, vidéos, ou tout autre contenu. Chaque utilisateur sur un réseau possède un profil qui lui est propre et qui guide ses choix sur le réseau. De plus, il est relié à un certain nombre d'autres par choix personnel.

Le travail a débuté par une étude axée sur les généralités des réseaux sociaux, Ensuite un deuxième chapitre a été consacré sur état de l'art des méthodes de détection des rumeurs.

Le troisième chapitre a été entamé par une proposition d'une nouvelle méthode de détection des rumeurs. Enfin les résultats de simulations ont été présentés dans le dernier chapitre.

En guise de perspectives, nous envisageons de faire les points suivants :

La première perspective, découle du fait que la méthode proposée traite les graphes orientés, nous souhaitons étendre cette méthode pour traiter d'autres types de graphes.

La deuxième perspective, qui semble être, une continuité logique de ce travail, est l'adaptation de la méthode globale proposée pour détecter des rumeurs dans les réseaux sociaux.

Références Bibliographies

- [1]: KEEPER SECURITY [En ligne]. <http://blog.keepersecurity.com/2012/06/07/linkedin-passwords-leaked-6-5-million-accounts-compromised/>
- [2]: DWYER Catherine, "Privacy in the Age of Google and Facebook," *Technology and Society Magazine, IEEE*, vol. 30, no. 3, pp. 58-63, 2011.
- [3]: FRIENDSTER. [En ligne].<http://www.friendster.com/>
- [4]: MYSPACE. [En ligne].<http://fr.myspace.com/>
- [5]: ZEMMAR Nisrine, "Réseaux Sociaux numériques : essai de catégorisation et cartographie des controverses," Université Rennes 2, Thèse de doctorat 2012.
- [6]: HEER Jeffrey, BOYD Danah, "Vizster: Visualizing online social networks," *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on. IEEE*, pp. 32-39, 2005.
- [7]: LES Z'ED. [En ligne]. <http://les-zed.com/qu'est-ce-que-les-reseaux-sociaux/>
- [8]: CREFF Marie, "Réseaux sociaux : quelles opportunités pour les services ? Le cas de l'assistance en ligne d'Orange," Institut national des techniques de la documentation, Mémoire de fin d'étude INTD 2010.
- [9]: Bibliothèque municipale de Lyon. [En ligne]. http://www.pointsdactu.org/article.php3?Id_article=1293
- [10]: ELLISON Nicole, et al, "Social network sites : definition, history, and scholarship," *Journal of Computer- Mediated Communication*, vol. 13, no.1 , pp. 210-230,2007.
- [11]:TORLOTING Philippe, "Enjeux et perspectives des réseaux sociaux," Institut Supérieur du Commerce, Paris, mémoire de fin d'étude 2006.
- [12]: LIEN OPTIONNEL. [En ligne]. <http://www.lien-optionnel.com/reseaux-sociaux.html>
- [13]:LESPETITSDEBROUILLARDS.[Enligne].http://lespetitsdebrouillardspc.org/IMG/pdf/reseaux_sociaux.pdf
- [14]:THELWALL Mike, "Social network sites: Users and uses," M.Zelkowitz (Ed.) *Advances Incomputers Elsevier*, vol. 76, pp. 19-73, 2009.
- [15]: FREDERIC cavazza. [En ligne]. <http://www.fredcavazza.net/2015/05/29/panorama-des-medias-sociaux-2015/>
- [16]:CANALBLOG. [En ligne]. [Http://reseauxlapie.canalbog.com](http://reseauxlapie.canalbog.com)

- [17]: SOCIALTIMES. [En ligne]. <http://www.adweek.com/socialtomes/vincos-blog-world-map-of-social-networks-january-2016/626833>
- [18]: FILLIETTAZ Francois, GREGORI Macro, "Comprendre les réseaux sociaux numériques," Direction des systèmes d'information et service écoles- médias, 2011.
- [19]: BOYD danah, "Social Network Sites: Public, Private, or What? " Knowledge Tree May 2007.
- [20] : NEDIOUI MOHAMED ABDELHAMID, Fouille et apprentissage automatique dans les réseaux sociaux dynamiques, 2014/2015.
- [21] : OUCH ! Sécurité des réseaux sociaux, Septembre 2011.
- [22] : Christine DU, M. Olivier POMMERET, thèse professionnelle en intelligence économique et management des connaissances, promotion 2012.
- [23]: FRANÇOIS-BERNARD HUYGHE, Faux, rumeurs et désinformation dans le cyberspace, Edition L'IRIS, 2013.
- [24]: Adeline Michel, A.C.Sordet et E. Moraillon, les rumeurs en tant que phénomène d'influence sociale, Ecole de psychologues praticiens de Lyon, 2004.
- [25]: Jean-Noël Kapferer : Rumeurs. Le plus vieux média du monde, Le Seuil, Coll. Points actuels, 1990.
- [26] : Nathalie Henry, Jean-Daniel Fekete , Représentations visuelles alternatives pour les réseaux sociaux ,11 janvier 2013.
- [27]: Shakarian, P., Bhatnagar, A., Aleali, A., Shaabani, E. and Guo, R. (2015). The Independent Cascade and Linear Threshold Models. *SpringerBriefs in Computer Science*, pp.35-48.
- [28] : Zakia Challal, Kamel Boukhalfa, Minimisation de l'influence négative dans les réseaux sociaux, 2016.
- [29]: Barbieri, N., Bonchi, F. and Manco, G. (2012). Topic-Aware Social Influence Propagation Models. *2012 IEEE 12th International Conference on Data Mining*.

[27]: Ivan Graham, with some revisions by Nick Britton, Mathematical Sciences, University of Bath, MATLAB manual and introductory tutorials, February 9, 2005

Webographie

[28]: http://fr.wikipedia.org/wiki/Utilisation_du_logiciel_Gephi_pour_l'analyse_cartographique – Master Intelligence Economique et Stratégies Compétitives.

Résumé

L'information qui se propage dans les réseaux sociaux peut entraîner de nombreuses fausses affirmations. Par exemple, les rumeurs sur certains sujets peuvent se propager rapidement menant à un grand nombre de nœuds signalant les mêmes observations. Dans ce mémoire, nous décrivons une approche pour trouver la source des rumeurs et évaluer la probabilité qu'une information soit en fait une rumeur, en l'absence d'informations sur la provenance des données. Nous modélisons le réseau social sous la forme d'un graphe orienté, où les sommets représentent des individus et des arêtes dirigées représentent le flux d'information (Par exemple, qui est sur Twitter). Un certain nombre de nœuds de contrôle sont injectés dans le réseau dont le travail consiste à signaler les données qu'ils reçoivent. Notre algorithme identifie les rumeurs et leurs sources en observant lequel des nœuds de contrôle a reçu l'information donnée et qui n'a pas reçu. Nous montrons que, avec un nombre suffisant de nœuds de contrôle, il est possible de reconnaître la plupart des rumeurs et leurs sources avec une grande précision.

Mots-clés: Réseaux sociaux, propagation de rumeurs.

Abstract

Information that propagates through social networks can carry a lot of false claims. For example, rumors on certain topics can propagate rapidly leading to a large number of nodes reporting the same observations. In this paper, we describe an approach for finding the rumor source and assessing the likelihood that a piece of information is in fact a rumor, in the absence of data provenance information. We model the social network as a directed graph, where vertices represent individuals and directed edges represent information flow (e.g., who follows whom on Twitter). A number of monitor nodes are injected into the network whose job is to report data they receive. Our algorithm identifies rumors and their sources by observing which of the monitors received the given piece of information and which did not. We show that, with a sufficient number of monitor nodes, it is possible to recognize most rumors and their sources with high accuracy.

Keywords: Social Networks, Rumor Spreading.