Université de Montréal

Plateforme pour se protéger tant de soi-même que de ses "amis" sur Facebook

par

Charles Hélou

Département d'informatique et de recherche opérationnelle

Faculté des arts et des sciences

Thèse présentée à la Faculté des arts et des sciences

en vue de l'obtention du grade de Doctorat en Informatique

Août, 2012

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée :

Plateforme pour se protéger tant de soi-même que de ses "amis" sur Facebook

Présenté par :
Charles Hélou

a été évaluée par un jury composé des personnes suivantes :

Guy Lapalme, président-rapporteur
Gilles Brassard, directeur de recherche
Esma Aïmeur, co-directrice
Jean Meunier, membre du jury
Mikhail Atallah (Purdue University), examinateur externe
William Skene, représentant du doyen de la FES

Résumé

Les réseaux sociaux accueillent chaque jour des millions d'utilisateurs. Les usagers de ces réseaux, qu'ils soient des particuliers ou des entreprises, sont directement affectés par leur fulgurante expansion. Certains ont même développé une certaine dépendance à l'usage des réseaux sociaux allant même jusqu'à transformer leurs habitudes de vie de tous les jours. Cependant, cet engouement pour les réseaux sociaux n'est pas sans danger. Il va de soi que leur expansion favorise et sert également l'expansion des attaques en ligne. Les réseaux sociaux constituent une opportunité idéale pour les délinquants et les fraudeurs de porter préjudice aux usagers. Ils ont accès à des millions de victimes potentielles.

Les menaces qui proviennent des amis et auxquelles font face les utilisateurs de réseaux sociaux sont nombreuses. On peut citer, à titre d'exemple, la cyberintimidation, les fraudes, le harcèlement criminel, la menace, l'incitation au suicide, la diffusion de contenu compromettant, la promotion de la haine, l'atteinte morale et physique, etc. Il y a aussi un « ami très proche » qui peut être très menaçant sur les réseaux sociaux : soi-même. Lorsqu'un utilisateur divulgue trop d'informations sur lui-même, il contribue sans le vouloir à attirer vers lui les arnaqueurs qui sont à la recherche continue d'une proie.

On présente dans cette thèse une nouvelle approche pour protéger les utilisateurs de *Facebook*. On a créé une plateforme basée sur deux systèmes : *Protect_U* et *Protect_UFF*. Le premier système permet de protéger les utilisateurs d'eux-mêmes en analysant le contenu de leurs profils et en leur proposant un ensemble de recommandations dans le but de leur faire réduire la publication d'informations privées. Le second système vise à protéger les utilisateurs de leurs « amis » dont les profils présentent des symptômes alarmants (psychopathes, fraudeurs, criminels, etc.) en tenant compte essentiellement de trois paramètres principaux : le narcissisme, le manque d'émotions et le comportement agressif.

Mots-clés: Réseaux sociaux, Facebook, Protect_U, Protect_UFF, Vie privée, Sécurité, Profil psychopathique, Fraudeur

Abstract

Social networks deal every day with millions of users (individuals or companies). They are directly affected by their rapid expansion. Some have developed a certain dependence on the use of social networks and even transform their everyday lifestyle. However, this craze for social networking is not always secure. It is obvious that their expansion promotes and serves the increase of online attacks. Social networks are an ideal opportunity for criminals and fraudsters to take advantage of users. They give access to millions of potential victims.

Threats coming from “friends” on social networks are numerous: cyberintimidation, fraud, criminal harassment, moral and physical threats, incitement to suicide, circulation of compromising contents, hatred promotions, etc. There is also a “very close friend” who could cause us problems with his behavior on social networks: ourselves. When a user discloses too much information about himself, it contributes unwittingly to attracting scammers who are continually looking for preys.

This thesis presents a new approach to protect Facebook users. We created a platform based on two systems: *Protect_U* et *Protect_UFF*. The first system tries to protect users from themselves by analysing the content of their profiles and by suggesting a list of recommendations in order to reduce the publication of private information. The second system aims to protect users from their “friends” who have profiles presenting alarming symptoms (psychopaths, fraudsters, criminals, etc.) taking into account essentially three main parameters: narcissism, lack of emotions and aggressive behaviour.

Keywords: Social Network, Facebook, Protect_U, Protect_UFF, Privacy, Security, Psychopathic profile, Fraudster

Tables des matières

Liste des tableaux

Liste des figures

À ma famille

Remerciements

Je voudrais remercier avec une très grande gratitude mon directeur de recherche Monsieur Gilles Brassard qui n'a épargné aucun effort pour m'offrir toujours son soutien académique, financier et moral. Il était toujours présent pour moi pour m'encourager et me supporter durant les bonnes et les moins bonnes périodes que j'ai dû traverser tout au long de mes études doctorales.

Je tiens aussi à vivement remercier ma co-directrice de recherche Madame Esma Aïmeur d'avoir cru en moi et de m'avoir donné la possibilité de travailler sur un sujet de recherche très passionnant. Son dévouement et ses nombreux conseils m'ont guidé tout au long de ce travail et m'ont permis d'avancer.

Je souhaite aussi exprimer toute ma reconnaissance envers les membres de jury d'avoir accepté de m'accorder de leur temps pour évaluer mon travail.

Mes remerciements vont également à toutes les personnes qui m'ont aidé de près ou de loin pour l'accomplissement de cet ouvrage, notamment Monsieur Pascal Vincent, Monsieur Benoit Dupont et Monsieur Max Mignotte pour leurs judicieux conseils.

Je voudrais aussi remercier tous les professeurs et personnels du Département d'informatique et de recherche opérationnelle de l'Université de Montréal de nous avoir assuré un environnement de recherche agréable, distingué, respectueux et civilisé.

Finalement, mes remerciements iront pour ma famille qui, malgré la distance qui nous sépare, a toujours été à mes côtés.

Chapitre 1 : Introduction

« Mon Dieu, gardez-moi de mes amis. Quant à mes ennemis, je m'en charge . » (Voltaire)

« Les amis se prétendent sincères, or ce sont les ennemis qui le sont » (Schopenhauer)

Ces proverbes révèlent bien la vulnérabilité de l'être humain vis-à-vis de la défaillance de ses amis. Il est normal de se protéger de ses ennemis, mais quoi faire pour se protéger de ses amis? On connaît très souvent la position à prendre contre ses ennemis mais très rarement celle qu'on doit prendre contre ses amis. Par définition, un ami est une personne de confiance. C'est quelqu'un qui est proche de nous, qui nous connaît bien, qui connaît bien nos faiblesses et nos forces. Or, une confiance mal placée mène toujours à de sérieux problèmes et à de regrettables résultats. Ceci est vrai dans la vraie vie mais aussi, et surtout, dans les réseaux sociaux en ligne.

Les réseaux sociaux accueillent chaque jour des millions d'utilisateurs. Les usagers de ces réseaux, qu'ils soient des particuliers ou des entreprises, sont directement affectés par leur fulgurante expansion. Certains ont même développé une certaine dépendance à l'usage des réseaux sociaux allant jusqu'à transformer leurs habitudes de vie de tous les jours. Ils cherchent à rester connectés avec leurs amis tout au long de la journée pour ne rien manquer : en restant chez soi, en travaillant, en se déplaçant (grâce au téléphone mobile intelligent), en mangeant, en regardant la télévision, etc.

Cependant, cet engouement pour les réseaux sociaux n'est pas sans danger. Il va de soi que leur expansion favorise et sert également l'expansion des attaques en ligne. Les réseaux sociaux constituent une opportunité idéale pour les délinquants et les fraudeurs de porter préjudice aux usagers. Ils ont accès à des millions de victimes potentielles. De plus, l'achalandage fait en sorte que les traces des actes criminels et frauduleux commis soient difficilement repérables. Il n'est pas toujours facile de détecter les gens mal intentionnés. Dans la majorité des cas les fournisseurs du service ne procurent à leurs utilisateurs qu'une protection minimale (ce n'est pas une priorité pour eux).

Les menaces qui proviennent des amis et auxquelles font face les utilisateurs de réseaux sociaux sont nombreuses. On peut citer à titre d'exemple, la cyber-intimidation, les fraudes, le harcèlement criminel, la menace, l'incitation au suicide, la diffusion de contenu compromettant, la promotion de la haine, l'atteinte morale et physique, etc. Comment s'en protéger?

Il y a aussi un « ami très proche » qui peut être très menaçant sur les réseaux sociaux : soi-même. Lorsqu'un utilisateur divulgue trop d'informations sur lui-même, il contribue sans le vouloir à attirer vers lui les arnaqueurs qui sont à la recherche continue d'une proie. Afficher, à titre d'exemple, sa date de naissance, son numéro de téléphone, l'institution bancaire avec laquelle on fait affaire, son lieu de travail, son emplacement géographique, offre gratuitement aux fraudeurs des informations personnelles et confidentielles qui pourront être exploitées contre soi-même. Renforcer les paramètres d'un compte (du profil) et savoir gérer les paramètres de confidentialité est primordial pour contrer les attaques. Les comptes et profils sur les réseaux sociaux ne sont jamais trop sécurisés.

D'un autre côté, on ne peut pas toujours être certains de l'authenticité de la personne avec qui nous communiquons sur un réseau social. Celle-ci a peut-être volé une identité ou incarné une identité fictive. Il n'est d'ailleurs pas invraisemblable que cette même personne cherche à nous retirer de l'information à des fins personnelles. De plus, accepter un ami sur un réseau social lui donne accès à la liste de tous nos amis. C'est pourquoi, ne pas accepter les demandes d'amis de personnes que l'on ne connaît pas représente une bonne mesure préventive.

Les émotions des utilisateurs de réseaux sociaux contribuent parfois à leurs malheurs. En effet, en suscitant leur intérêt et leur curiosité (en leur promettant par exemple de l'argent, un cadeau, un voyage, etc.), un fraudeur a de forte chance à ce que sa victime tombe sous la tentation et clique par exemple sur un lien malveillant. La compassion peut aussi jouer son rôle. Elle peut inciter parfois les utilisateurs à envoyer de l'argent à un ami présumé qui prétend être victime de vol. Recevoir des courriels ou des messages suspects

avec des pièces jointes provenant de personnes qu'on ne connaît pas devrait mettre la puce à l'oreille.

Comment empêcher un utilisateur de se nuire à lui-même?

Comme *Facebook* est actuellement le réseau social le plus utilisé au monde, on l'a adopté dans cette thèse comme base sur laquelle on a créé une plateforme qui se charge de protéger un utilisateur de soi-même et de ses amis. Notre plateforme est constituée de deux modules : *Protect_U* et *Protect_UFF*. Le premier permet d'analyser le niveau de risque du profil de l'utilisateur et le classe sur une échelle de quatre niveaux : *peu risqué*, *moyennement risqué*, *risqué* et *critique*. Il propose ensuite à l'utilisateur un ensemble de recommandations lui permettant de rendre son profil plus sécuritaire. Pour cela, il applique entre autres un filtre communautaire impliquant les « amis proches » de l'utilisateur. Ce dernier a le dernier mot : il peut appliquer ces recommandations comme il peut les ignorer. *Protect_UFF*, quant à lui, a comme mission d'analyser le texte et les images publiques que les amis ont partagé avec autrui dans le but de détecter les profils qui présentent des symptômes alarmants (profil psychopathique, fraudeur, criminel, etc.). Pour cela, *Protect_UFF* évalue pour chaque profil le niveau de trois paramètres fondamentaux : le narcissisme, le manque d'émotions, et le comportement antisocial et agressif.

Ce travail est réparti comme suit: le chapitre 2 introduit les principaux réseaux sociaux qui sont actuellement en vogue et met l'emphase sur *Facebook* et ses fonctionnalités; le chapitre 3 précise les principaux dangers et risques auxquels font face en général les utilisateurs de réseaux sociaux; le chapitre 4 présente les principaux traits de personnalité d'un psychopathe, d'un fraudeur à col blanc et d'un criminel; le chapitre 5 introduit la notion de catégorisation de textes et les techniques de comparaison d'images et de fouilles d'objets et met en évidence celles qui ont été appliquées dans notre étude; le chapitre 6 expose les principaux travaux existants qui sont en lien avec notre thèse et présente les solutions que nous avons adoptées; le chapitre 7 explique l'architecture, le fonctionnement et la validation de *Protect_U*; le chapitre 8 présente l'architecture et le

principe de fonctionnement de *Protect_UFF*; le chapitre 9 conclut en évoquant les travaux qui pourront être appliqués dans le futur pour améliorer et enrichir notre plateforme.

Chapitre 2 : *Facebook* et les réseaux sociaux

À la fin des années 1990, les réseaux sociaux sont apparus sur Internet réunissant des personnes via des services d'échanges personnalisés. Ainsi, un réseau social peut être considéré comme un ensemble de personnes réunies par un lien social. L'office québécois de la langue française (Office québécois de la langue française 2012) définit le réseau social comme étant une « communauté d'internautes reliés entre eux par des liens, amicaux ou professionnels, regroupés ou non par secteurs d'activités, qui favorisent l'interaction sociale, la création et le partage d'informations ».

Fogel et Nehmad considèrent les réseaux sociaux comme un espace social sur Internet permettant le partage d'informations, la communication et la collaboration entre les divers acteurs (Fogel et Nehmad 1999). Selon eux, le principe des réseaux sociaux repose sur trois composantes clés, soit le profil de l'utilisateur (publication d'informations personnelles), l'établissement d'un cercle d'amis (création de listes et sous-listes selon des intérêts communs) et l'interaction entre utilisateurs (échange d'informations).

En général, un utilisateur de réseau social bien informé peut bloquer ou restreindre pour autrui l'accès aux informations sensibles de son profil selon le niveau de confiance qui le lie avec ses amis. Malheureusement, beaucoup d'utilisateurs de réseaux sociaux donnent facilement accès au grand public à une grande quantité d'informations privées, ce qui va menacer dans bien des cas leurs vies privées et leur sécurité.

Les raisons pour lesquelles les individus utilisent les réseaux sociaux diffèrent d'une personne à une autre. Selon l'agence *Up2Social,* les raisons qui ont été fréquemment évoquées sont : retrouver des camarades de classe ou d'anciens amis, faire des relations et des réseaux professionnels, partager des passions et des hobbies, faire des rencontres et des nouveaux amis et organiser des événements (Up2social 2011).

2.1. Évolution des réseaux sociaux

Les réseaux sociaux connaissent leur début avec la mise en place du réseau *SixDegrees.com* (1997). Ce réseau a permis à l'utilisateur d'établir un profil, de générer un cercle d'amis avec lequel il pouvait communiquer et de donner accès aux profils des « amis des amis ». Toutefois, le réseau rencontra de la difficulté à se développer et à maintenir ses services, causant ainsi sa fermeture en 2000. Plusieurs réseaux sociaux ont vu le jour entre les années 1997 et 2001. Nous citons à titre d'exemple, *BlackPlanet, LiveJournal, LunarStorm* et *AsianAvenue*.

Le concept de réseau social professionnel fait réellement surface à partir de 2001, avec *Ryse.com*. Ce dernier fait ses débuts en tant que réseau social, avant de susciter l'intérêt d'investisseurs potentiels, responsables de plusieurs réseaux dérivés, notamment *Tribe.net*, *Friendster* et *LinkedIn*. Cependant, la concurrence entre ces réseaux sociaux professionnels a fait ses effets : *Ryse.com* n'a pas pu continuer parce qu'il n'a pas réussi à attirer assez d'utilisateurs, *Friendster* a pris un virage plus ludique et seul *LinkedIn* a pu garder une vocation professionnelle. Par la suite, plusieurs réseaux sociaux, destinés aux professionnels, ont vu le jour tels que *Xing*, *Plaxo* et *Viadeo*. Contrairement à *LinkedIn*, qui fut pendant de longues années uniquement disponible en anglais, *Xing* propose son site depuis le début dans plusieurs langues, entre autres allemand, anglais, espagnol, français, italien, portugais, néerlandais, suédois, finlandais, chinois, japonais, coréen, russe, polonais, hongrois et turc. *Viadeo* est surtout utilisé en France et en Espagne tandis que Plaxo aux États-Unis.

À partir de 2003, nous assistons à la naissance de beaucoup de réseaux sociaux se dirigeant vers le même concept élaboré par *Friendster* tout en ciblant des communautés bien spécifiques selon leurs emplacements démographiques, religions ou intérêts. Ainsi, à titre d'exemple, *Google Orkut* est devenu le réseau social préféré des Brésiliens et des Indiens, *Mixi* des Japonais, *Hi5* des résidents de l'Amérique Latine, *MyChurch* de certains chrétiens, *Care2* des associations caritatives, *Dogster* des amateurs de chiens, etc.

L'introduction de *Facebook* en 2004 marque le début d'une période de réseautage importante. Depuis, nous assistons sans cesse à la naissance de nouveaux réseaux innovateurs, dont le plus récent est Google+.

2.2. Distribution des utilisateurs des réseaux sociaux aux États-Unis et au Canada

Selon l'étude menée par Experian Hitwise[1] en janvier 2012 sur la part des dix plus grands médias sociaux du trafic internet aux États-Unis, *Facebook* tient la plus grande part du marché avec 63,8% de nombre d'utilisateurs. *Youtube* vient loin derrière avec 19.7% d'utilisateurs. La distribution des utilisateurs sur les dix premiers réseaux sociaux aux États-Unis est donnée par la figure suivante.

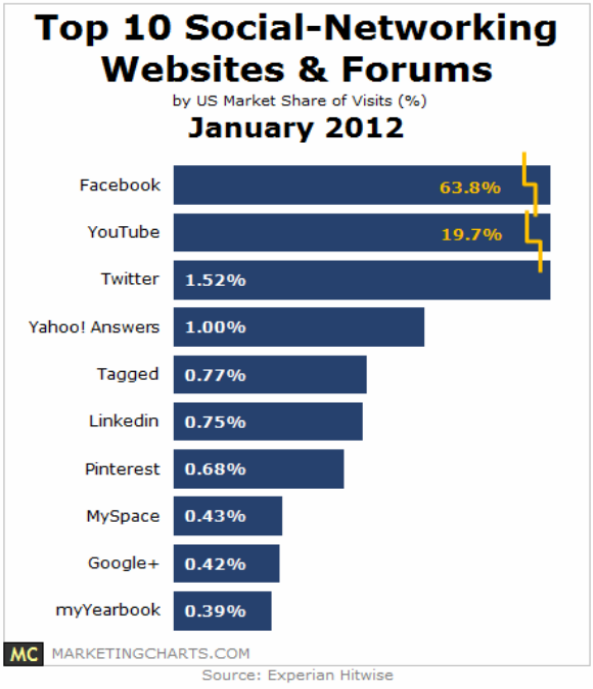
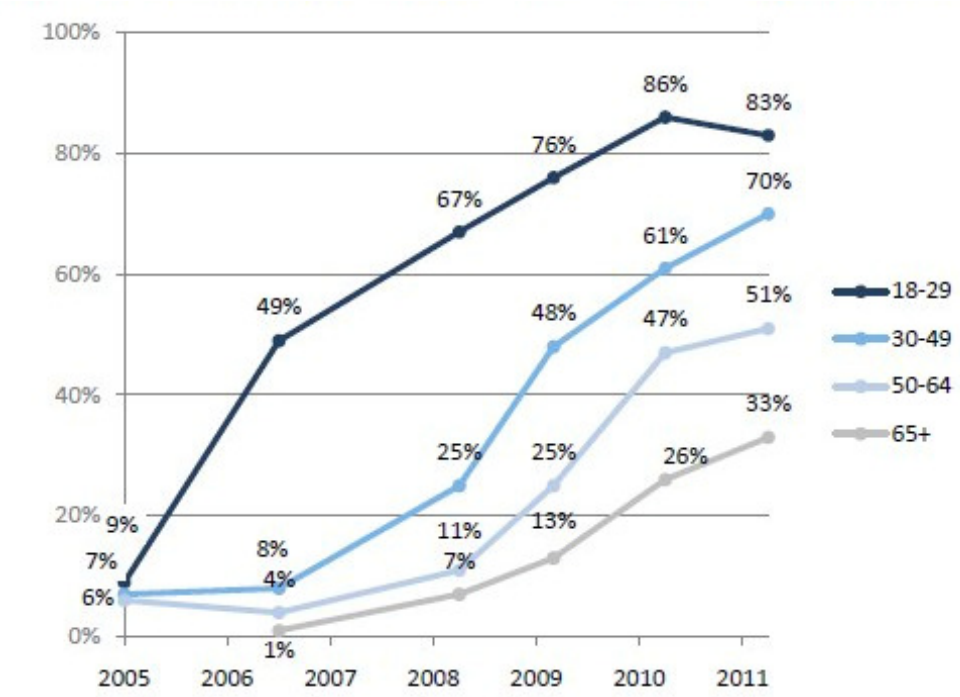


Figure 1. Les dix plus populaires réseaux sociaux aux États-Unis en janvier 2012

[1] http://weblogs.hitwise.com/james-murray/2012/04/instagram_snaps_into_top_10_so.html

Les utilisateurs de réseaux sociaux n'ont cessé d'augmenter depuis 2005, et ceci pour presque toutes les tranches d'âges. L'étude menée par *PewInternet*[2] en 2011 montre bien que les réseaux sociaux sont très populaires auprès de la population âgée entre 18 et 29 ans. Le taux d'utilisateurs de cette tranche d'âge passe d'un maigre 7% en 2005, à 49% en 2006, pour atteindre ensuite 86% en 2010 avant de marquer une légère régression en 2011 (83%). Cette augmentation est également repérable chez les 30-49, 50-64 et 65+, cependant à plus faible ampleur (voir figure ci-dessous).



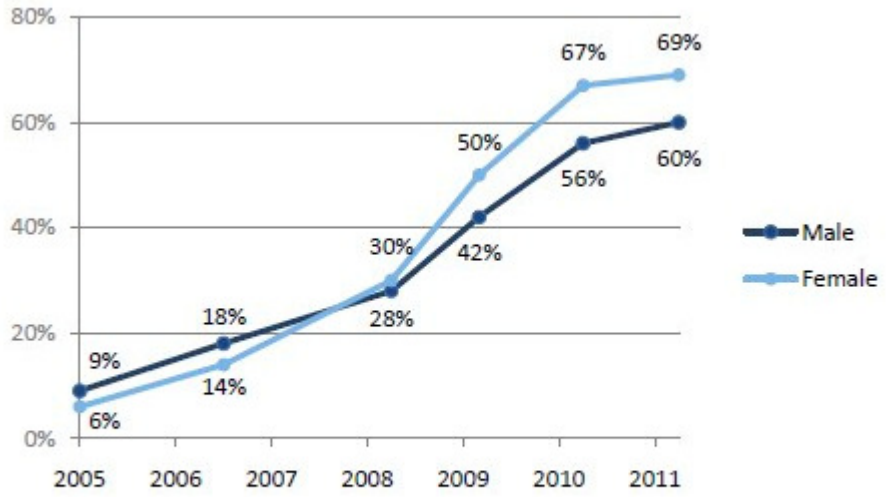
Note: Total n for internet users age 65+ in 2005 was < 100, and so results for that group are not included.

Source: Pew Research Center's Internet & American Life Project surveys: February 2005, August 2006, May 2008, April 2009, May 2010, and May 2011.

Figure 2. Utilisation des réseaux sociaux aux États-Unis par tranches d'âge entre 2005 et 2011

[2] http://www.pewinternet.org/Reports/2011/Social-Networking-Sites/Report.aspx?view=all

Cette même étude révèle qu'à partir de l'année 2008 le pourcentage des femmes internautes utilisant un réseau social aux États-Unis a bondi de 30% à 69%. Cette augmentation est cependant moins marquée pour les hommes : elle passe de 28% à 60%. Cette évolution est illustrée dans la figure suivante :



Source: Pew Research Center's Internet & American Life Project surveys: February 2005, August 2006, May 2008, April 2009, May 2010, and May 2011.

Figure 3. Utilisation des réseaux sociaux parmi les internautes des États-Unis par genre entre 2005 et 2011

Selon l'agence *Web Blakkat*[3], les médias sociaux qui remportent le plus d'adhésion au Canada sont *Facebook*, *Twitter* et *LinkedIn*. En 2011, 86% des internautes utilisaient *Facebook*, 19% *Twitter* et 14% *LinkedIn*. Ces chiffres tendent à confirmer l'utilisation grandissante des médias sociaux que ce soit au niveau personnel ou professionnel des Canadiens. Les statistiques révèlent aussi que 60% des Canadiens utilisent les réseaux sociaux, toute catégorie d'âge confondues. Les 18-34 ans sont les plus représentés (Voir figure 4).

[3] http://www.blakkat.net/blog/2011/infographie-utilisation-reseaux-sociaux-canada/

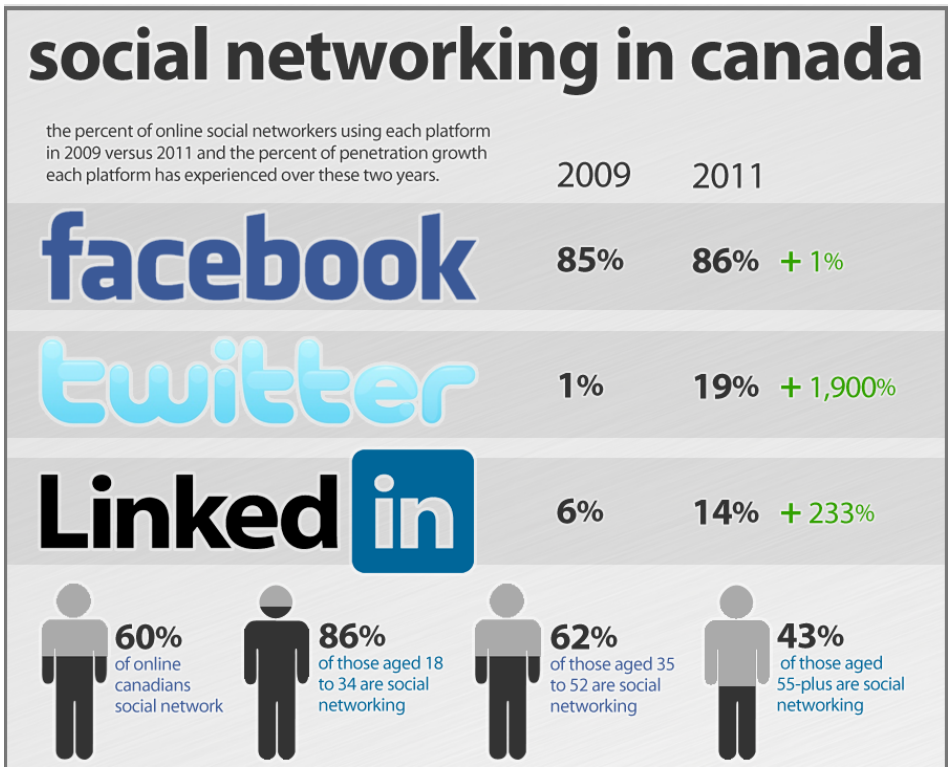


Figure 4. Distribution de l'utilisation des réseaux sociaux au Canada

Au Canada, les femmes internautes utilisent davantage les réseaux sociaux que les hommes. Selon un sondage réalisé par *Ipsos Canada*[4] en 2011, 37% des internautes Canadiennes ont affirmé qu'elles visitent un réseau social au moins une fois par jour contre uniquement 24% des internautes Canadiens.

2.3. Aperçu de réseaux sociaux populaires

Les réseaux sociaux étant en compétition constante entre eux, plusieurs ont vu rapidement le jour pour disparaître peu de temps après. Dans cette section, nous donnons un bref aperçu de six réseaux sociaux populaires qui ont réussi à surmonter la compétition. Ils sont présentés selon l'ordre chronologique de leur apparition.

[4] http://www.webfuel.ca/canada-social-media-statistics-2011/

LinkedIn

Le réseau social *LinkedIn* fut fondé en 2002 par Reid Coffman, Allen Blue, Konstantin Guericke, Eric Ly et Jean-Luc Vaillant. Il fut officiellement lancé en mai 2003. Il se démarque par une orientation purement professionnelle. Le 2 août 2012, il comptait plus de 175 millions d'utilisateurs distribués à travers 200 pays[5]. Il affirme sur son site Web que durant la première moitié de 2012 deux nouveaux membres s'ajoutaient au réseau chaque seconde. Il est présentement disponible en 18 langues : Allemand, Anglais, Coréen, Espagnol, Français, Indonésien, Italien, Japonais, Malais, Néerlandais, Norvégien, Polonais, Portugais, Roumain, Russe, Suédois, Tchèque et Turc.

LinkedIn est basé sur le principe de la connexion (pour entrer en contact avec un professionnel, il faut le connaître auparavant ou qu'une de nos connexions intervienne) et du réseautage (mise en relation professionnelle). Ainsi, il existe 3 niveaux de connexions :

- le premier niveau, ou nos contacts directs
- le deuxième niveau, ou les contacts de nos contacts
- le troisième niveau, ou les contacts de nos contacts de deuxième niveau.

Il existe 2 profils sur *LinkedIn* dont un public accessible par tous (abonnés ou non). Lors de la création d'un profil, l'utilisateur peut dévoiler plusieurs informations professionnelles telles que son éducation, son expertise et son expérience de travail. Il peut même afficher son curriculum vitae. Le réseau permet également aux compagnies de créer des profils pour afficher par exemple des offres d'emploi.

Sur son site Web, Philippe Buschini[6] a publié le mode public et le mode privé de son profil. Avec le mode public, un utilisateur a accès à moins d'informations que le mode privé. Voici un exemple :

---

[5] http://press.linkedin.com/about
[6] http://www.buschini.com/livre-augmente-les-supports-pour-etre-visible/

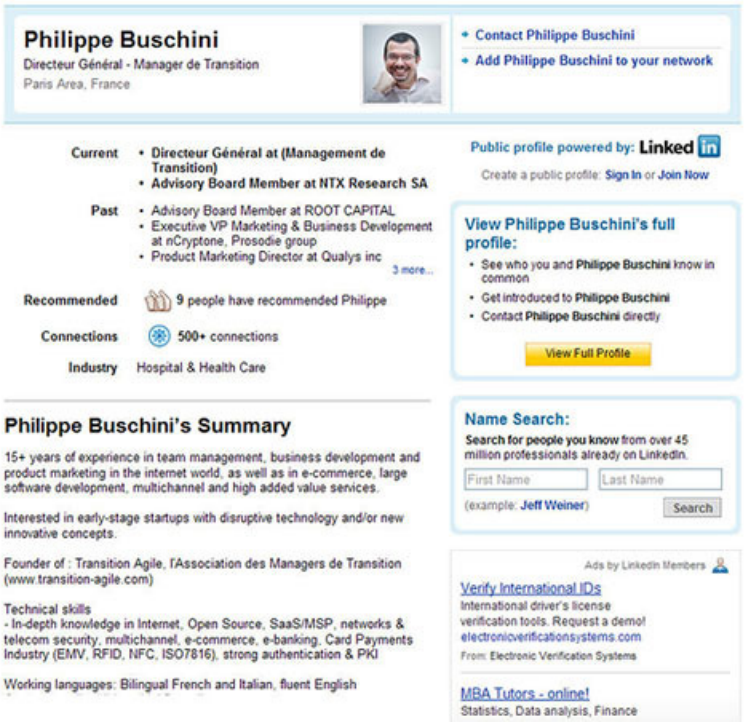


Figure 5. Mode public de *LinkedIn*

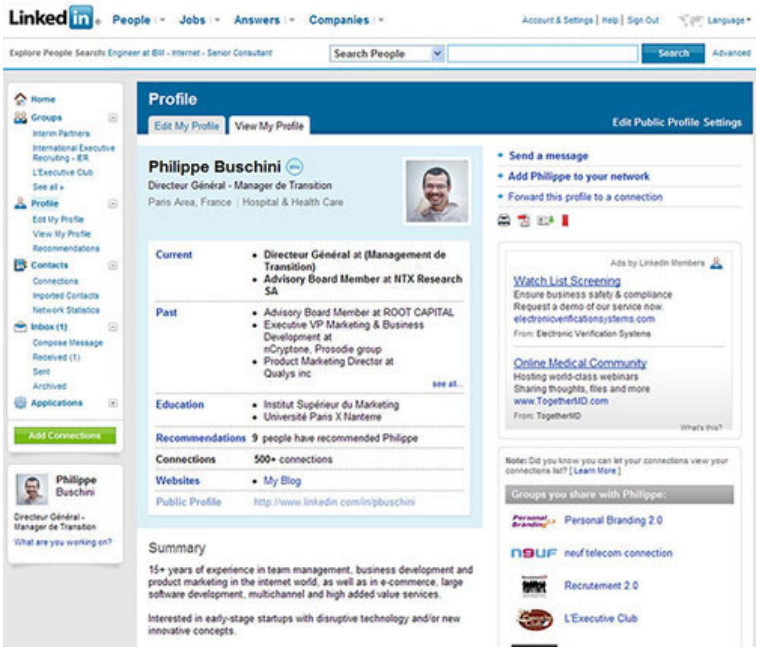


Figure 6. Mode privé de *LinkedIn*

MySpace

Le réseau *MySpace* fait son entrée sur le marché en 2003. Selon le fondateur du réseau, Tom Anderson, *MySpace* cherchait à attirer, dans un premier temps, les déserteurs du réseau *Friendster*. À l'origine, il ne s'était pas adressé à une communauté particulière, dans un souci de demeurer ouvert pour tout le monde. Cependant, avec le temps, de plus en plus d'artistes musiciens s'inscrivaient sur le réseau, lui procurant une sorte de vocation musicale. Certains groupes de musique avaient été préalablement bannis du réseau *Friendster* suite à des infractions aux règles d'inscription, ils se sont alors dirigés vers *MySpace*, entrainant avec eux tous leurs fans. Ces derniers ont également joint le réseau afin de rester en contact avec leurs groupes favoris. Depuis, Il est réputé pour héberger de nombreuses pages internet de groupes de musique et de DJs qui y entreposent et présentent leurs compositions musicales.

Le déclin de *MySpace* commença en 2008, face à l'expansion fulgurante de *Facebook*. Afin de pouvoir rivaliser avec ce dernier, il procède au remodelage du design de son interface, la rendant assez similaire à celle de son concurrent. Malheureusement, ce remodelage ne l'a pas beaucoup aidé : il a perdu 10 millions d'utilisateurs durant uniquement les mois de janvier et février 2011. Le nombre de ses utilisateurs passe de 73 millions à 63 millions en l'espace de quelques semaines (The Telegraph, mars 2011[7]). Depuis, le réseau a préféré adopter une vocation plutôt artistique en se concentrant davantage sur les nouvelles des artistes et des musiciens.

Vous trouverez à la page suivante un extrait de la page d'accueil de *MySpace* :

[7] http://www.telegraph.co.uk/technology/myspace/8404510/MySpace-loses-10-million-users-in-a-month.html

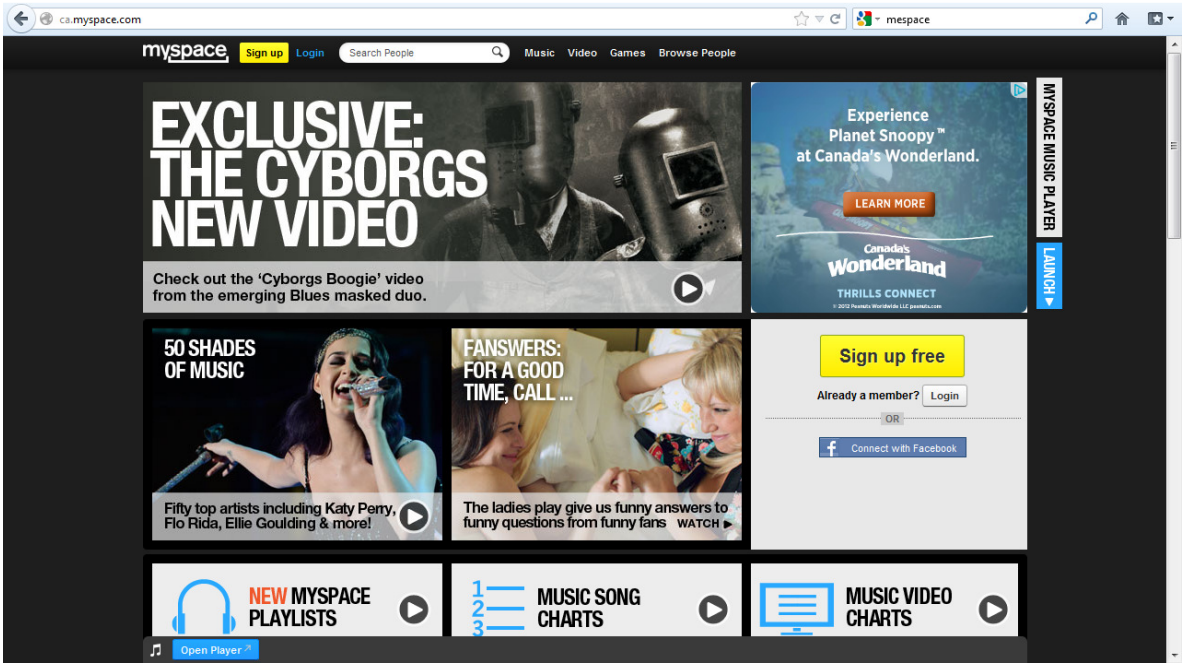


Figure 7. Page d'accueil *MySpace*

Facebook

Facebook connait ses débuts le 4 février 2004. Il a été fondé par Mark Zuckerberg, qui était étudiant à l'Université Harvard, dans le but de retrouver des amis avec qui il a perdu le contact. Ses collègues Eduardo Saverin, Andrew McCollum, Dustin Moskovitz et Chris Hughes se sont joint à lui par la suite pour l'aider à promouvoir le site. L'inscription a d'abord été limitée aux étudiants de l'Université Harvard. En effet, afin de pouvoir s'inscrire en tant qu'utilisateur sur *Facebook.com*, une adresse de courriel *harvard.edu* était nécessaire. En l'an 2005, *Facebook* s'ouvre peu à peu à d'autres réseaux universitaires, pour enfin devenir accessible à toute personne détenant une adresse électronique.

Il est défini comme étant un « réseau social sur Internet permettant à toute personne possédant un compte de créer son profil et d'y publier des informations, dont elle peut contrôler la visibilité par les autres personnes, possédant ou non un compte. L'usage de ce réseau s'étend du simple partage d'informations d'ordre privé (par le biais de photographies,

liens, textes, etc.) à la constitution de pages et de groupes visant à faire connaitre des institutions, des entreprises ou des causes variées. L'intégralité des informations publiées sur ces deux supports, à l'inverse du profil, peut être consultée par n'importe quel internaute sans qu'il soit nécessaire d'ouvrir un compte …»[8].

Fin juin 2012, *Facebook* comprenait 955 millions d'utilisateurs actifs dont 552 millions réguliers (qui l'utilisaient tous les jours)[9]. Environ 81% des utilisateurs actifs se trouvent à l'extérieur des États-Unis et du Canada. Certaines études indiquent cependant que le nombre d'utilisateurs est surestimé[10]. Selon Alexa Internet[11], il est en 2012 le site le plus visité du monde devant *Google*.

La structure et les fonctionnalités de *Facebook* sont présentées en détails un peu plus loin dans ce chapitre. La page principale d'un compte *Facebook* a la forme suivante :

Figure 8. Profil *Facebook*

[8] http://fr.wikipedia.org/wiki/Facebook
[9] http://newsroom.fb.com/content/default.aspx?NewsAreaId=22
[10] http://www.futura-sciences.com/fr/news/t/internet/d/le-nombre-dadeptes-de-facebook-est-il-surestime_24839/
[11] http://www.alexa.com/topsites

YouTube

YouTube a été créé en février 2005 par trois anciens employés de *PayPal* : Chad Hurley, Steve Chen et Jawed Karim. Il permet d'héberger des vidéos que les utilisateurs peuvent envoyer, visualiser et partager entre eux. L'identification de ce site de partage de vidéos dans la catégorie des réseaux sociaux n'est pas évidente. La notion d'amis sur ce site diffère de celle des autres réseaux sociaux : demander d'être ami avec quelqu'un veut souvent simplement dire vouloir s'abonner au flux des vidéos postées par cette personne.

Dans l'article intitulé « *Les sciences sociales et le Web 2.0 : YouTube est-il un réseau social?*[12] » l'auteure précise que *YouTube* n'est pas un simple système de diffusion de vidéo qui pourrait être comparé, comme c'est souvent le cas, à un média traditionnel pour lequel la conquête de l'audience la plus nombreuse serait le but unique et privilégié de tous. Elle utilise la notion de "media circuit" pour désigner la manière dont l'échange de vidéo peut s'organiser différemment en fonction du réseau social. Elle montre comment les fonctionnalités très particulières de *YouTube* permettent aux utilisateurs de jouer différemment sur les circuits de diffusion de leurs productions. Sur *YouTube* les utilisateurs produisent certes les vidéos, mais ils produisent aussi leur audience – et manifestement il existe plusieurs manières de construire cette audience.

Selon les dernières statistiques (août 2012) fournis par *YouTube* lui-même[13] on note que :

- plus de 800 millions d'utilisateurs consultent *YouTube* chaque mois,
- plus que 4 milliards de vidéos sont visionnés chaque jour,
- 72 heures de vidéo sont téléchargées par minute,
- 70% du trafic *YouTube* provient de l'extérieur des États-Unis,
- En 2011, il comptait plus de 1000 milliards de vues.

---

[12] http://www.internetactu.net/2008/02/11/les-sciences-sociales-et-le-web-20-youtube-est-il-un-reseau-social-47/

[13] http://www.youtube.com/t/press_statistics

- Il est localisé dans 39 pays et 54 langues.

La figure suivante présente un aperçu d'un profil *YouTube* :

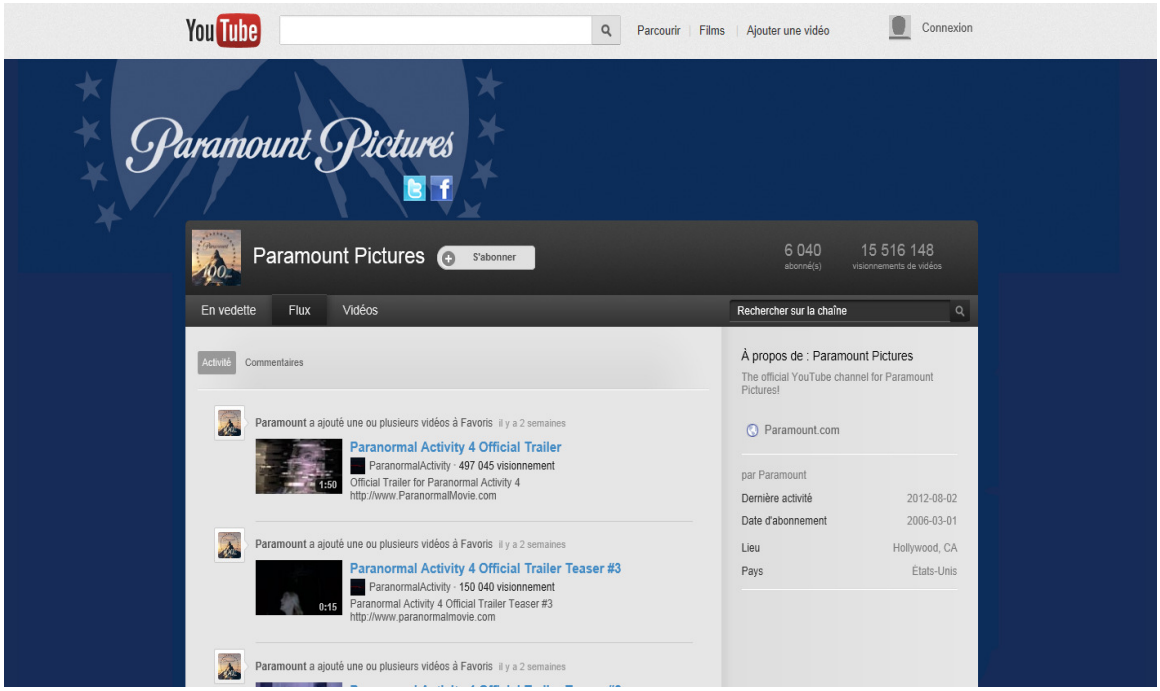


Figure 9. Profil *YouTube*

Twitter

Twitter est un réseau d'informations en temps réel qui permet à son utilisateur d'envoyer de brefs messages appelés « *tweets* ». Il a été fondé en 2006 par Jack Dorsey. L'idée de départ lancée par son fondateur était de permettre aux utilisateurs de pouvoir partager facilement leurs petits moments de vie avec leurs amis[14]. *Twitter* a comme rôle maintenant de garder ses utilisateurs connectés avec les dernières histoires, idées, opinions et actualités sur des événements qui les intéressent.

Il est considéré comme un *microblog* qui permet de publier de courts messages, plus court que dans les blogs classiques, dans le but de diffuser plus souvent des informations en se limitant au minimum utile. Il se trouve ainsi à mi-chemin entre messageries

[14] http://www.youtube.com/watch?v=uRXsgEBFy6A

instantanées et blogs. Les messages normalement sont de type texte mais peuvent contenir des images ou des vidéos embarqués. La taille maximale d'un *tweet* est de 140 caractères, obligeant ainsi les utilisateurs à être concis dans leur rédaction. Outre cette concision imposée, la principale différence entre *Twitter* et un blog traditionnel réside dans le fait qu'il n'invite pas les lecteurs à commenter les messages postés. Il se différencie aussi des autres réseaux sociaux par sa simplicité d'utilisation, ce qui en fait un des principaux facteurs de son succès. Début 2012, *Twitter* possédait quelque 465 millions de comptes ouverts[15] et il y a échange plus que 177 millions de *tweets* en moyenne par jour[16].

Pour pouvoir utiliser *Twitter*, un utilisateur doit créer un profil et choisir un « nom d'utilisateur » unique qui lui sera propre. Le nom d'utilisateur ne peut cependant être repris par plusieurs utilisateurs (contrairement à d'autres réseaux sociaux tel que *Facebook*). « Suivre » un utilisateur de *Twitter* veut dire être intéressé à recevoir et lire tous les messages qu'il émet. Dans la version française de l'interface, un *follower* était appelé initialement « suiveur », ce qui fut remplacé par « abonné ». *Twitter* est un réseau social asymétrique, c'est-à-dire n'engageant pas de réciprocité. Ainsi, il est possible pour un utilisateur de restreindre la lecture de ses messages en gardant privé l'accès à son compte, en évitant de le rendre public. Il est cependant bon de noter que l'accès privé n'est pas le mode par défaut de *Twitter*. Les comptes qui ne sont pas privés sont bien publics, ce qui implique qu'un *tweet* peut être repris et cité ailleurs (dans les médias ou en justice par exemple[17]).

La figure suivante donne un aperçu d'un profil *Twitter* actif :

---

[15] http://www.journaldunet.com/ebusiness/le-net/nombre-d-utilisateurs-twitter-0112.shtml
[16] http://blog.twitter.com/2011/03/numbers.html
[17] http://www.ozap.com/actu/arnaud-dassier-mis-en-examen-pour-un-tweet/440414

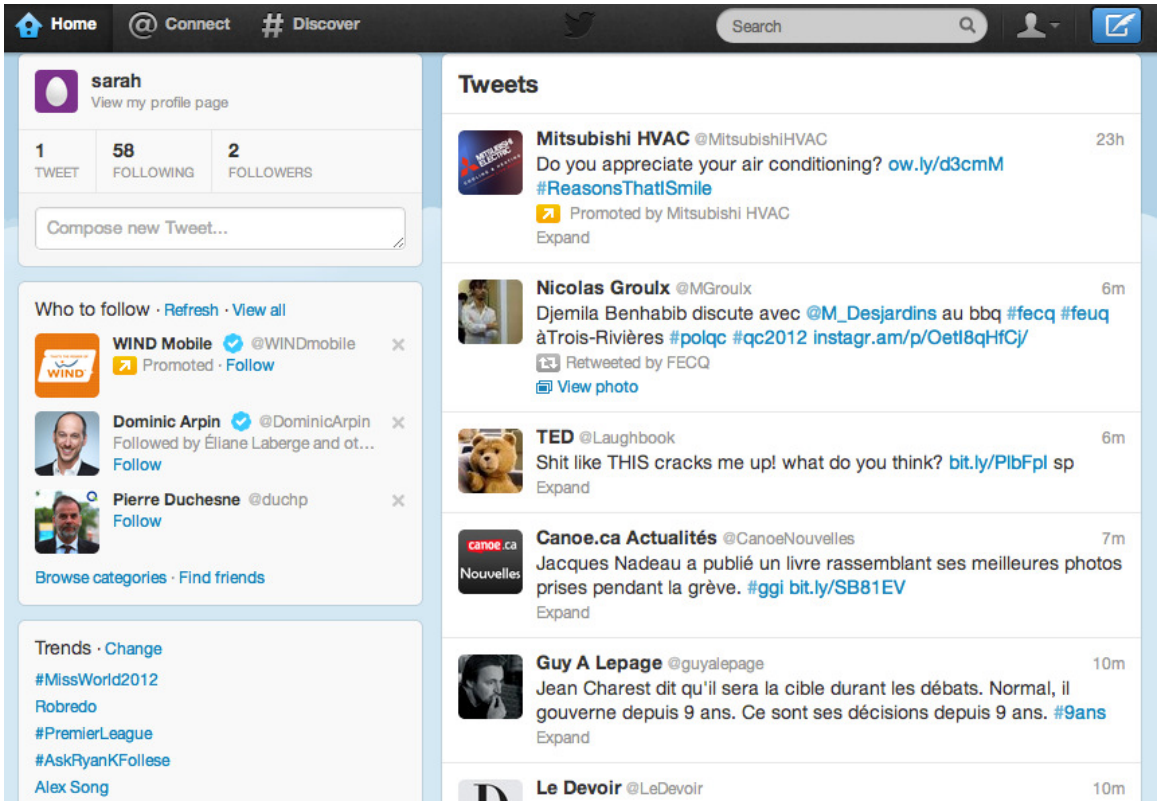


Figure 10. Profil *Twitter*

Google+

Google+ est l'application de réseautage social de *Google*. Elle était accessible au départ, à partir du 28 juin 2011, sur invitation uniquement. On l'a rendu accessible au grand public le 20 septembre 2011[18]. Elle est présentée par nombre de média comme un produit destiné à concurrencer Facebook[19,20]. *Google+* n'est pas un premier coup d'essai en termes de réseautage social pour *Google* : avec les lancements successifs d'*Orkut* en janvier 2004, de *Wave* en septembre 2009 et de *Buzz* en février 2010, le groupe avait déjà fait plusieurs efforts dans cette direction mais sans jamais rencontrer le succès escompté[21].

---

[18] http://fr.wikipedia.org/wiki/Google%2B#cite_note-21
[19] http://www.nytimes.com/2011/06/29/technology/29google.html?_r=2&ref=technology
[20] http://obsession.nouvelobs.com/high-tech/
[21] http://www.wired.com/business/2011/06/googles-schmidt-social/

Google+ offre à ses utilisateurs l'accès à des applications qui lui sont propres telles que *Picasa* (traitement et organisation d'images développé par *Google* en 2002), *Gmail* (messagerie Google) et *Google Maps* (service de localisation géographique). De plus, il met en avant trois nouveaux services[22] :

- **les cercles** (« circles ») : Il s'agit d'une catégorisation des contacts. L'utilisateur peut créer des groupes de contacts différents en fonction des informations qu'il souhaite partager avec eux. La figure ci-dessous donne un exemple d'utilisation des cercles. Ce système remplace la « liste d'amis » typique d'autres sites comme *Facebook*.

Figure 11. Utilisation des cercles dans *Google+*

[22] http://www.pcinpact.com/news/64359-reseau-social-google-facebook-cercles.htm

- **les bulles** (« hangouts ») : Il s'agit en fait d'un système de chat vidéo, proche de celui de *Skype*, pouvant réunir entre deux et dix personnes maximum. Afin de ne pas être dérangé constamment, l'utilisateur peut inviter un de ses contacts à discuter via vidéo. Si une conversation est déjà entamée avec une personne d'un cercle donné, les membres dudit cercle seront mis au courant immédiatement.
- **les déclics** (« sparks ») : c'est une forme de suggestion et de partage de contenu (vidéos, livres, articles, etc.) par thème avec certains contacts.

Google+ met aussi à la disposition de ses utilisateurs un certain nombre de jeux.

Il est bon de noter que malgré la variété des réseaux actuellement en compétition, certains d'entre eux sont complémentaires. À titre d'exemple, la majorité des contenus du web 2.0 (images, textes, etc.) peut être partagé sur *Facebook* et *Twitter* en même temps à l'aide de boutons interactifs. Le bouton « *Tweet* » permet de partager une information sur le réseau tout en se conformant à la limite de 140 caractères, tandis que le bouton « *Share* » permet de divulguer cette information sur *Facebook*. La figure ci-dessous donne quelques exemples de boutons interactifs :

Figure 12. Boutons interactifs

2.4. Principaux Termes de *Facebook*

Facebook possède un glossaire[23] de termes lui étant propre. Nous exposons dans cette partie certains de ces termes jugés importants dans la compréhension de son fonctionnement.

Amis (*Friends*)

Ils représentent les personnes avec qui un utilisateur peut entrer en contact et partager des informations (images, textes, vidéos, etc.). Tout le principe de *Facebook* repose sur cette notion de partage d'informations entre les réseaux d'amis.

Lorsqu'un utilisateur crée un profil sur *Facebook*, il peut procéder à la recherche d'amis en utilisant la barre de recherche située en haut de chaque page *Facebook*. Une fois la personne recherchée trouvée, l'utilisateur doit cliquer sur le bouton « Ajouter à mes amis ». À ce moment, une demande d'ajout à la liste d'amis sera envoyée à cette personne. Une fois la confirmation obtenue, cette personne devient alors ami(e) avec l'utilisateur et apparaît dans sa liste d'amis *Facebook*. Il est à noter que les paramètres de confidentialité peuvent limiter la recherche de certains utilisateurs.

À tout moment un utilisateur peut retirer un ami de sa liste d'amis. Il peut aussi décider de « bloquer » un ami. Dans ce cas là, l'ami bloqué ne sera plus en mesure de retrouver l'utilisateur sur le réseau. Il ne pourra plus également voir ou accéder aux activités et aux informations postées par l'utilisateur.

Paramètres du compte (*Account settings*)

Les paramètres du compte d'un utilisateur permettent de gérer les préférences de base pour son compte. Il peut entre autres modifier son nom, son adresse électronique de connexion, son mot de passe, ses préférences de notification ou ses fonctions de sécurité supplémentaires.

---

[23] http://www.facebook.com/help/glossary

Si l'utilisateur décide de désactiver son compte, il disparaîtra du service Facebook, c'est-à-dire personne ne pourra le trouver par l'intermédiaire d'une recherche, bien que certaines informations comme les messages qu'il avait envoyés puissent continuer à être visibles pour d'autres. Cependant, *Facebook* garde une copie du contenu de son compte (les amis, les photos, les centres d'intérêts, etc.).

Paramètres de confidentialité (*Privacy settings*)

Les paramètres de confidentialité permettent à l'utilisateur de gérer les options de confidentialité de son compte Facebook. Il peut, par exemple, indiquer qui pourra lui envoyer des demandes d'ajout d'amis et des messages. Pour toute autre information qu'il partage sur Facebook, il a la possibilité de choisir les personnes qui recevront chacune de ses publications. Normalement, l'utilisateur a le choix entre quatre options :

- Public
- Amis (et les amis des personnes identifiées).
- Moi uniquement
- Personnaliser (comprend des groupes spécifiques, des listes d'amis et des personnes que l'utilisateur a choisi d'inclure ou d'exclure)

Journal (*Timeline*)

Le journal d'un utilisateur (anciennement appelé profil) est le recueil qui englobe l'ensemble des photos, interactions, publications, expériences et activités de l'utilisateur. Il affiche les évènements par ordre chronologique. On y retrouve sa photo, son nom et prénom, sa date de naissance, sa ville de naissance, sa ville actuelle, son niveau d'éducation, le nom de son employeur actuel, son orientation politique, sa religion, ses intérêts musicaux, ses citations favorites et bien plus. L'utilisateur peut choisir de n'afficher qu'une partie de ces renseignements, et ce, grâce aux paramètres de sécurité du compte. Il peut également presque tout cacher ou tout afficher.

Mur (*Wall*)

Le mur est un espace du journal où l'utilisateur peut publier et échanger du contenu avec ses amis. Cet échange peut être sous forme de messages textes, d'images, de vidéos ou de liens vers du contenu sur Internet.

Toute activité exécutée par l'utilisateur et ses amis sera annoncée sur son mur. À titre d'exemple, si l'utilisateur écrit un commentaire, crée un nouvel album photo, ou encore ajoute un nouvel ami, une annonce sera affichée sur son mur pour le souligner. Cependant, *Facebook* donne la possibilité à l'utilisateur de restreindre l'accès à son mur et de limiter l'accès à seulement quelques amis. En tout temps, l'utilisateur peut choisir de masquer une actualité visible sur son mur, et ce, en appuyant sur le bouton « X » apparaissant à côté de l'actualité. Cette dernière ne sera alors plus visible.

Messages (*Messages*)

Les messages jouent un rôle majeur dans l'échange des messages privés, des discussions instantanées, des messages électroniques et des textos avec des amis.

Les utilisateurs de *Facebook* peuvent par exemple échanger des messages à travers la messagerie instantanée (Chat). Pour lancer une discussion, il suffit de cliquer sur le nom de l'ami avec qui un usager souhaite clavarder pour que *Facebook* ouvre une fenêtre de discussion.

Facebook met également à la disposition de ses « échangeurs de messages » un ensemble d'icônes leur permettant de révéler leur humeur lors d'une discussion. La figure de la page suivante donne un aperçu de ces icônes :

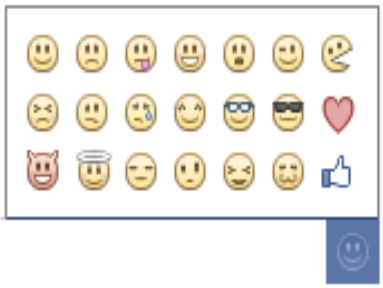


Figure 13. Émotions sur Facebook

Statut (*Status*)

C'est une fonction qui permet à l'utilisateur de faire un commentaire ou d'exprimer un avis. Semblable à un *tweet*, un statut est généralement court. Il exprime un point de vue sans entrer trop dans les détails.

Photos (*Photos*)

C'est une fonction qui permet de partager des images et de marquer (*tag*) les personnes qui y figurent. Un utilisateur *Facebook* peut insérer une photo de profil, télécharger des photos, publier une photo sur un mur, ajouter des photos dans les messages et les groupes, créer et gérer un album de photos, etc.

L'album de photos est défini par un titre et une description. Son créateur peut préciser qui peut le voir. Chaque album comporte une limite maximale de 1000 photos. Il n'existe cependant pas de limite au nombre d'albums autorisés. Toute photo de l'album peut être « aimée », « commentée » et « partagée » par les amis.

La figure de la page suivante donne un exemple des albums photos :

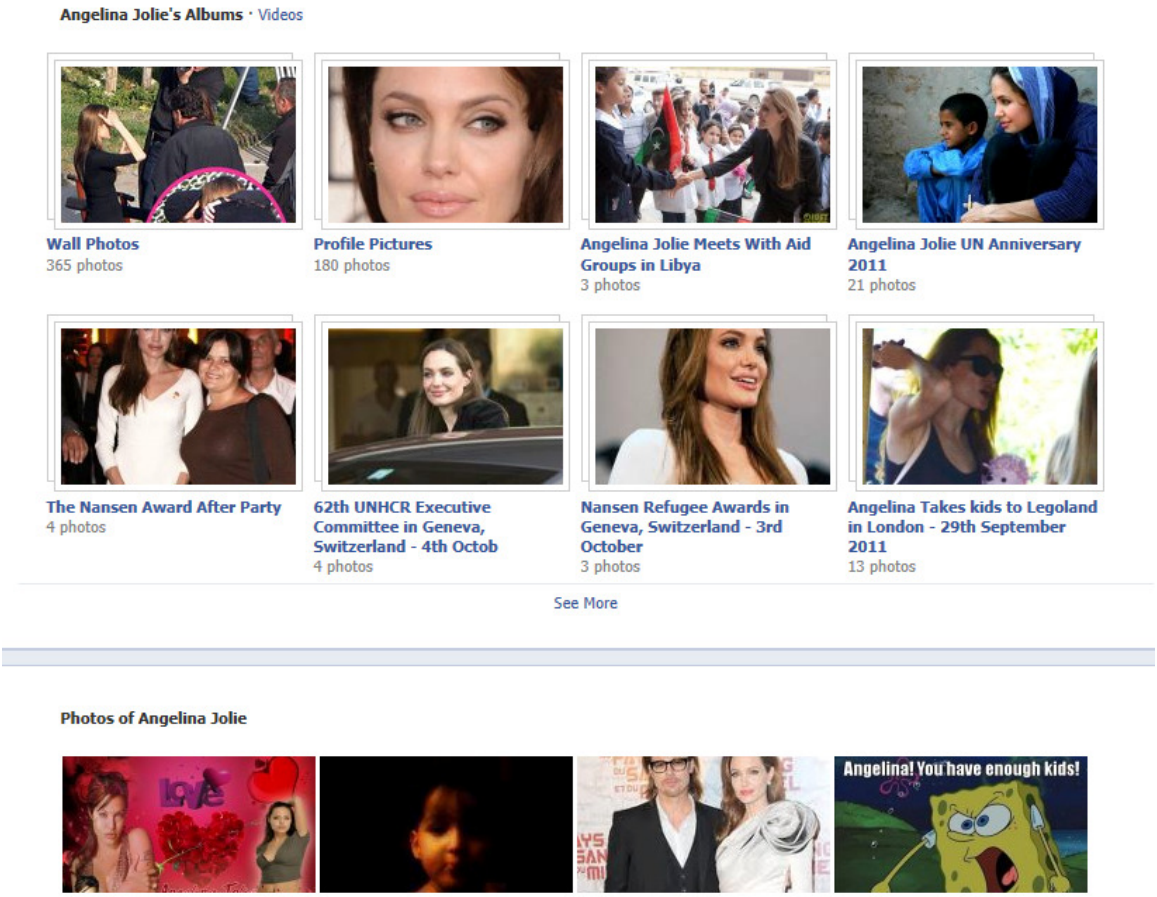


Figure 14. Albums photo *Facebook* d'Angelina Jolie

Appel vidéo (*Vidéo Chat*)

L'appel vidéo est une fonctionnalité de *Facebook* qui permet de communiquer avec un ami avec l'image et le son. C'est une option disponible dans les fenêtres de clavardage du réseau social. Si l'ami ne répond pas à l'appel vidéo, il est alors possible de lui laisser un message vidéo enregistré. La date et l'heure de l'appel seront enregistrées dans l'historique, mais contrairement à la discussion instantanée, l'appel vidéo ne sera pas sauvegardé par le réseau. Seuls les amis de l'utilisateur peuvent l'appeler par vidéo.

Groupe (*Group*)

Les utilisateurs de *Facebook* peuvent se regrouper autour de sujets d'intérêts communs sous formes de *groupes* (ex : fans de cinéma, anciens étudiant d'un lycée, etc.).

Tout utilisateur peut créer un groupe sur *Facebook*. Pour en créer un, l'utilisateur doit se rendre à sa page d'accueil, puis cliquer sur « Créer un groupe ». Il sera alors possible à l'utilisateur de déterminer le nom du groupe, de spécifier qui peut voir le groupe et son contenu, et de déterminer les paramètres de confidentialités. Les membres du groupe peuvent inviter uniquement leurs amis à se joindre à eux.

Il existe trois options de confidentialité :

- **Ouvert** : tout utilisateur de *Facebook* peut voir le groupe et le rejoindre. Les groupes ouverts apparaissent dans les résultats de recherche et tout le monde peut en voir tout le contenu (photos, vidéos et fichiers).
- **Fermé :** tous les utilisateurs de *Facebook* peuvent voir le nom et les membres d'un groupe ainsi que les personnes invitées à rejoindre ce groupe, mais seuls les membres peuvent accéder aux publications correspondantes. Pour rejoindre un groupe fermé, l'intéressé doit recevoir une invitation d'un membre de ce groupe.
- **Secret :** ces groupes n'apparaissent pas dans les résultats de recherche et les personnes qui n'en sont pas membres ne peuvent rien voir, ni le contenu, ni la liste des membres. Le nom du groupe ne s'affichera pas sur le profil (journal) de ses différents membres. Pour rejoindre un groupe secret, il faut être ajouté par un membre existant de ce groupe.

Fil d'actualité (*News feed*)

C'est la liste permanente des mises à jour sur la page d'accueil de l'utilisateur. Il affiche les publications des amis et les actualités des pages qu'il suit. Ce fil d'actualité peut

contenir aussi les identifications de photos, les demandes d'ajout à une liste d'amis, les mises à jour d'événements, les inscriptions à des groupes ainsi que d'autres activités.

Page (*Page*)

Les pages permettent aux entreprises, marques et personnalités de communiquer avec des personnes sur *Facebook*. Les administrateurs de ces pages peuvent publier des informations et mettre à jour leur fil d'actualité à l'attention des personnes qui les « suivent ». Les pages sont destinées à un usage professionnel et officiel : elles permettent aux organisations, entreprises, célébrités ou groupes de musique d'être présents sur *Facebook*. Elles permettent à leurs propriétaires de partager leurs histoires et de communiquer avec les usagers. Elles peuvent être personnalisées : on peut ajouter des applications, publier des actualités, héberger des évènements, etc. Le but est d'intéresser et d'élargir le public grâce à des publications régulières. Par conséquent, les personnes qui aiment ces pages recevront les mises à jour dans leur fil d'actualité.

Seul le représentant officiel d'un organisme, d'une entreprise, d'un artiste ou d'un groupe musical est autorisé à créer une page. On ne peut créer une page que si on représente officiellement une véritable organisation. Les pages sont gérées par des administrateurs qui disposent de profils *Facebook* personnels. Les pages ne sont pas des comptes *Facebook* distincts, elles partagent les mêmes informations de connexion qu'un profil personnel. Il s'agit simplement d'entités différentes tout comme les groupes et les évènements. Une fois qu'un administrateur a configuré une page au sein de son profil, il peut ajouter des administrateurs supplémentaires pour l'aider à la gérer.

La figure de la page suivante est un exemple de page *Facebook* :

Figure 15. Page *Facebook*

Article (*Note*)

Un article permet à un utilisateur de s'exprimer dans un format enrichi. Il peut contenir du texte, des images, des vidéos et des liens vers d'autres pages Web. Les articles sont normalement publics. Tout usager peut les consulter et les commenter.

Identification (*Tagging*)

C'est une action très populaire sur *Facebook*. Elle permet d'associer un ami à un emplacement, un statut, une photo ou un vidéo. Toute publication dans laquelle l'utilisateur

identifie un ami sera automatiquement visible sur le journal de l'utilisateur et l'ami recevra une notification en ce sens. Si l'ami décide de retirer l'identification, l'utilisateur ne pourra plus l'identifier de nouveau dans ce contenu.

Si l'utilisateur identifie une personne dans sa publication et que le partage est défini sur « Amis » ou sur un partage plus vaste, la publication peut être vue par le public que l'utilisateur a sélectionné, ainsi que par les amis de la personne identifiée. Cependant, il est possible d'activer la « vérification du profil » pour vérifier et approuver toutes les publications marquées avant qu'elles n'apparaissent sur le journal de l'utilisateur ou pour empêcher certaines personnes d'y accéder lorsqu'elles consultent ce journal.

J'aime (*Like*)

Le bouton « J'aime » de *Facebook* permet aux internautes de donner un avis positif sur le contenu d'une publication et de s'associer à des sites Web favoris. Lorsqu'un utilisateur clique sur le bouton « J'aime » sous un contenu publié par lui ou par l'un de ses amis, ça permet aux autres utilisateurs de savoir ce qu'il apprécie sans avoir à laisser de commentaires. Avec ce bouton intégré sur de nombreux sites Web, *Facebook* dispose d'une cartographie des centres d'intérêt de ses usagers. Il peut tracer chaque clic sur un site Web, combien de temps un utilisateur y passe et quels sont ses préférences.

Il est possible aussi pour un utilisateur d'« aimer » une page, ce qui signifie établir un lien avec celle-ci. À ce moment, la page va apparaître sur le journal de l'utilisateur et ce dernier va apparaître dans la liste des personnes qui aiment cette page.

S'abonner (*Subscribe*)

Le bouton « s'abonner » permet à un utilisateur d'entrer en contact avec des personnes qui l'intéressent (journalistes, célébrités, politiciens, etc.), même s'il n'est pas ami avec elles. Il permet également de voir les mises à jour publiques de ces personnes, sans pour autant les ajouter à la liste d'amis. Uniquement les utilisateurs âgés de 18 ans et plus peuvent autoriser les abonnés.

Évènement (*Event*)

Facebook permet à l'utilisateur d'organiser des évènements et d'y inviter ses amis. Pour cela, il doit déterminer la date, l'heure, l'emplacement et la nature de l'évènement. Il peut s'agir par exemple d'une réunion entre collègues, d'un anniversaire, d'un souper, d'un mariage ou d'une manifestation. Les amis invités à l'évènement peuvent accepter l'invitation, la refuser ou bien manifester une hésitation (*peut-être*). Les invités peuvent en tout temps changer d'avis.

Lieu (*Check in*)

Facebook met à la disposition de l'utilisateur une option lui permettant d'annoncer à sa liste *d'amis* l'endroit où il se trouve. Cette option permet également à l'utilisateur de repérer l'ensemble de ses *amis* se trouvant à proximité de son emplacement. Cependant, il ne sera possible de repérer que les *amis a*yant déjà procédés à un *check in*.

Poke (*Poke*)

Lorsqu'un utilisateur « poke » un de ses amis, il cherche à attirer son attention ou à le saluer. Un poke peut être envoyé aux amis confirmés, aux personnes d'un même réseau ou aux amis des amis.

Questions Facebook (*Facebook questions*)

Cette fonctionnalité permet d'obtenir des recommandations, de poser des questions et d'apprendre des amis et d'autres personnes sur *Facebook*. Elle a été conçue pour obtenir des réponses ou des votes courts et rapides. Toutes les personnes qui voient la question peuvent répondre. Les réponses sont filtrées pour afficher celles des amis en premier.

Lorsqu'une question est posée dans un groupe, seuls les membres du groupe peuvent consulter et répondre à cette question.

Notification (*Notification*)

C'est un message électronique à propos des mises à jour de l'activité sur *Facebook*.

2.5. Conclusion

La popularité des réseaux sociaux est liée à plusieurs facteurs. Ils ont grandement facilité la communication entre les gens à travers le monde entier. Ils ont brisé la barrière de la distance. Ils ont permis de rencontrer en ligne de nouvelles personnes avec des intérêts similaires ou différents et de retrouver d'anciennes connaissances. Ils ont aidé les gens à mieux s'exprimer à travers divers moyens (textes, images, vidéos, musiques, etc.) et de mieux se faire connaître. Ils ont favorisé l'implication des usagers dans plusieurs aspects de la vie de tous les jours (social, politique, artistique, etc.) en leur permettant de proposer des actions concrètes (levée de fonds pour une cause humanitaire par exemple) et d'organiser des évènements. Les réseaux sociaux ont permis aussi aux professionnels d'aller chercher directement les partenaires et les clients. Ils ont facilité le recrutement d'employés. Ils ont aidé à mieux comprendre les besoins des clients et à mieux les cibler en organisant par exemple des sondages et des enquêtes. Cette diversité enrichissante a grandement contribué à l'expansion des réseaux sociaux.

Cependant, le mauvais usage des réseaux sociaux peut avoir de sérieux inconvénients. Ils peuvent être source de distraction et d'énorme perte de temps pour les jeunes et les adultes. De plus, c'est un terrain très fertile pour les prédateurs de tous genres qui guettent continuellement le comportement des utilisateurs. Ils constituent une menace permanente à leur vie privée, tranquillité et intégrité physique.

Ce chapitre a présenté un bref aperçu historique des réseaux sociaux et a mis l'accent sur les principaux termes de *Facebook* qui permettent d'aider le lecteur à mieux comprendre son fonctionnement. Cette étape est primordiale puisque *Facebook* est utilisé comme base par notre plateforme qui doit récolter et analyser plusieurs informations sur les utilisateurs tels que leurs paramètres de comptes, leurs paramètres de confidentialité, leurs photos identifiées, leurs status, leurs commentaires et leurs articles pour qu'elle soit en mesure de les aider.

Dans le chapitre suivant, nous exposons en détails les principaux dangers auxquels font face les utilisateurs de réseaux sociaux.

Chapitre 3 : Principaux dangers dus aux réseaux sociaux

L'évolution du web 2.0 au cours des dernières années a facilité les moyens de communication entre les usagers, favorisant du même coup l'expansion de la criminalité, notamment la fraude. Les médias sociaux se classent parmi les interfaces les plus populaires sur le web. Ils permettent d'établir et de maintenir les relations avec un cercle d'amis et d'échanger entre eux. L'augmentation en flèche de leur popularité, ainsi que leur croissance constante au cours des dernières années fait en sorte que les fraudeurs ont accès à un bassin de victimes toujours plus grand. Ainsi, les réseaux sociaux, tels que Facebook, permettent de faciliter toutes les dérives qui portent atteinte à la vie privée et à l'intégrité physique des utilisateurs.

Pour mieux se protéger des dangers auxquels un utilisateur de réseau social peut faire face, il est important de comprendre les différents crimes et fraudes existants. Ce chapitre met en évidence les principales menaces rencontrées sur le web 2.0 en général et sur les réseaux sociaux plus spécifiquement notamment la fraude élaborée, le vol d'identité, l'atteinte à la personne, les attaques informatiques et les actions problématiques.

3.1. La fraude élaborée

La *fraude élaborée* représente un acte de tromperie commis dans le but de réaliser un gain potentiel sans pour autant procéder à un vol d'identité. Les auteurs du rapport intitulée *La Fraude via les médias sociaux* (Ryan, et al. 2011) ont remarqué que :

« La fraude élaborée se commet principalement dans des sites d'annonces classées parce qu'elle vise généralement le gain personnel et que les sites de réseautage social sont (pour l'instant) davantage utilisés pour communiquer avec des amis que pour réaliser des transactions commerciales.»

Elle peut se présenter sous différentes formes :

Fraude par abus de confiance

Dans le cas d'une fraude par abus de confiance, le fraudeur abuse de la confiance d'un acheteur dans un contexte de transaction réalisée en ligne, en ne livrant pas le bien ou le service demandé après avoir été payé. Les fraudes par abus de confiance sont exécutées généralement selon l'un des deux scénarios suivants :

- Un individu, agissant à terme de vendeur d'un produit ou d'un service en ligne, exige une avance de fond pour ensuite ne pas livrer le produit ou le service voulu après avoir encaissé le montant transféré.
- Un vendeur en ligne peut ne pas envoyer un bien qu'un acheteur a déjà payé. Parfois, le bien envoyé peut ne pas correspondre au produit initialement offert ou est une contrefaçon. Une fois la transaction complétée et l'argent envoyé, la victime réalise souvent trop tard le tort causé. L'argent est généralement irrécupérable.

Fraude de location immobilière

La fraude de location immobilière consiste tout simplement à obtenir de l'argent pour un service que le fraudeur n'a pas l'intention d'offrir. Elle est similaire à la fraude par abus de confiance sauf qu'elle touche le domaine de l'immobilier spécifiquement et est généralement plus sophistiquée. Le fraudeur affiche sur Internet une annonce indiquant par exemple que sa maison est à louer. Lorsqu'un locataire potentiel entre en contact avec lui pour lui annoncer son intérêt envers la propriété à louer, le fraudeur use du prétexte qu'il réside temporairement dans un autre pays pour ne pas la lui faire visiter. Puisque le loyer demandé est généralement bon marché le « futur » locataire ne s'objecte pas et accepte d'envoyer l'argent par la poste. À ce moment, le fraudeur encaisse l'argent et ne redonne plus jamais signe de vie.

Fraude par usage de faux

La fraude par usage de faux diffère de ce qui précède par le fait que le fraudeur agit en terme d'acheteur et non pas en terme de vendeur. Cette fraude est commise par l'utilisation des modes de paiement contrefaits (donner un faux chèque). Le but évidemment est de se procurer un bien gratuitement. Il arrive parfois que le fraudeur envoie plus d'argent au vendeur et lui demande de le rembourser.

L'offre de service sans permis

L'offre de service sans permis est la création d'une annonce en ligne offrant un service pour lequel l'annonceur n'a pas les qualifications ou les autorisations légales nécessaires (par exemple, annonce de déménageurs opérant sans les assurances et le permis requis par la juridiction commerciale dans laquelle ils offrent leurs services). Cependant, elle ne constitue pas un vol d'identité en soi car le fraudeur ne prend pas l'identité d'une personne spécifique.

D'un autre côté, le marketing en masse, contrairement au marketing personnalisé, est une technique de marketing fondée sur le lancement d'un nouveau produit en s'adressant au plus grand nombre de personnes possibles. En 2011, comme l'illustre bien la figure 16, la fraude par marketing de masse (FMM) au Canada a touché 2594 victimes et a engendré une perte de 11290031,16$ (Centre antifraude du Canada 2011).

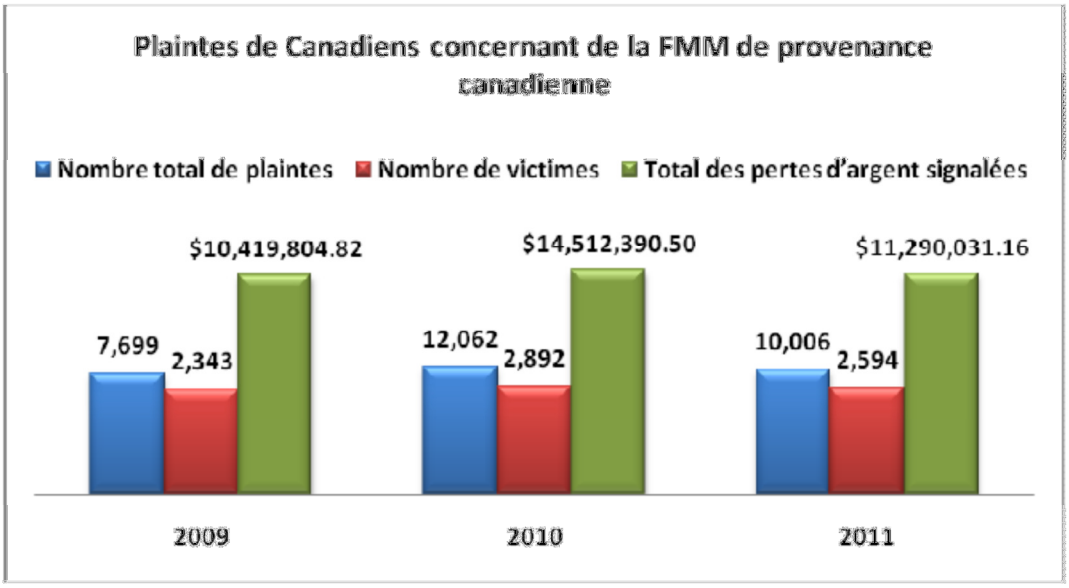


Figure 16. L'évolution des plaintes de Canadiens concernant des activités de FMM de provenance canadienne en terme de nombre total de plaintes, de nombre de victimes et du total des pertes d'argent signalées

3.2.Le vol d'identité

Le *vol d'identité* représente la collecte et l'utilisation de renseignements personnels d'un individu, sans son consentement. L'usage de ces renseignements sert généralement à des fins criminelles.

Le vol d'identité touche entre autres tout individu qui « obtient ou a en sa possession des renseignements identificateurs sur une autre personne dans des circonstances qui permettent de conclure raisonnablement qu'ils seront utilisés dans l'intention de commettre un acte criminel dont l'un des éléments constitutifs est la fraude, la supercherie ou le mensonge » (article 402.2 (1) du code criminel canadien), mais également toute personne qui « transmet, rend accessible, distribue, vend ou offre en vente, ou a en sa possession à une telle fin, des renseignements identificateurs sur une autre personne sachant qu'ils seront utilisés pour commettre un acte criminel dont l'un des

éléments constitutifs est la fraude, la supercherie ou le mensonge ou ne se souciant pas de savoir si tel sera le cas » (article 402.2 (2) du code criminel canadien). Ainsi, le vol d'identité concerne non seulement l'action de s'approprier les informations personnelles d'une personne et par conséquent son identité, mais également tout acte de trafic d'informations propres à un autre individu dans le but de commettre des actes illégaux.

Plusieurs scénarios de vol d'identité ont été recensés (Ryan, et al. 2011) :

- Le premier scénario est de se faire passer pour une autre personne (personnification) dans le but de l'humilier ou de lui transférer la responsabilité d'un acte frauduleux.
- Le second scénario consiste à personnifier une célébrité en créant un profil Internet sous cette identité pour se faire passer pour une personne connue. Compte tenu de la popularité d'une célébrité, son compte attire un grand nombre d'internautes, ce qui permet au voleur d'identité de tirer profit de cette notoriété. Le but est de porter atteinte à la réputation de la célébrité, de s'immerger dans son quotidien, d'obtenir des gains personnels ou tout simplement de s'amuser.
- Le troisième scénario consiste à personnifier un individu dans le but d'obtenir des gains financiers. En piratant son profil Facebook, par exemple, un fraudeur peut facilement s'adresser à ses proches en leur envoyant un message de détresse leur demandant de lui envoyer de l'argent. Vu que le message provient d'un compte contenant des informations personnelles, les chances de succès de la fraude augmentent. Ainsi, le fraudeur se sert généralement d'un appel ou d'un courriel d'une première victime après avoir volé son identité, puis prétend se faire passer pour cet ami ou ce membre de la famille, qui a des ennuis. Les ennuis employés varient d'une arnaque à une autre. Toutefois, les plus populaires sont les arrestations, accident de la route et problème empêchant de retourner au pays. Ces ennuis nécessitent

l'envoi d'argent immédiatement pour diverses raisons, comme payer une caution, des frais médicaux, ou un billet d'avion pour rentrer au pays. Ce stratagème est également connu sous l'appellation « l'arnaque des grands-parents ».

- Le quatrième scénario repose sur l'abus de confiance. Le fraudeur prend le rôle d'un entrepreneur à la recherche d'employés et demande aux intéressés de lui fournir des informations personnelles ou même un rapport de crédit. Il se sert ensuite de ces informations pour émettre frauduleusement des cartes de crédit au nom de ces individus.

En 2011, le nombre de victimes de vol d'identité répertorié au *Centre antifraude du Canada* (Centre antifraude du Canada 2011) s'élevait à 17002 victimes pour une perte totale de 13204091,97$ (voir figure 17). Malgré la diminution du nombre de victimes de fraude par rapport à l'année 2010 (18284 victimes), les pertes monétaires avaient considérablement augmenté. Ces chiffres peuvent être attribués au fait que les fraudeurs arrivent à mieux cibler leurs besoins et leurs proies, diminuant ainsi les tentatives non fructueuses.

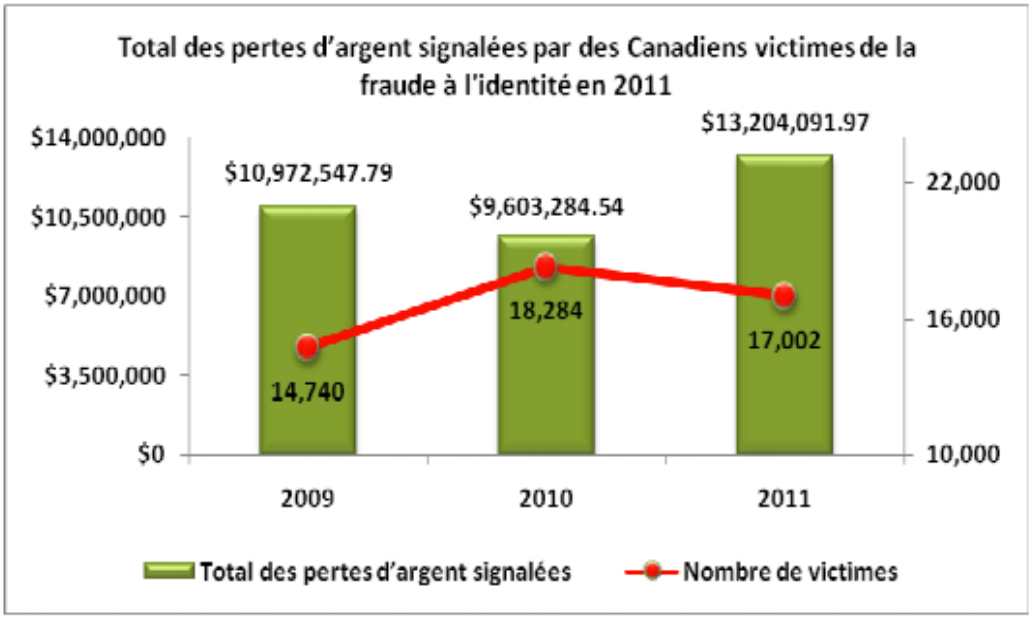


Figure 17. Totale des pertes d'argent signalées par des Canadiens victimes de la fraude à l'identité en 2011

3.3.L'atteinte à la personne

Les atteintes à la personne sont divisées généralement en deux catégories : les atteintes physiques et les atteintes à la considération de la personne (vie privée, liberté, dignité, etc.).

L'atteinte physique à la personne se manifeste généralement par des agressions, des actes de meurtres ou tentatives de meurtres, des voies de faits ou de vol avec violence. Ces attaques arrivent généralement suite à un discours haineux ou une propagande malveillante accompagnée de propos violents sur les réseaux sociaux.

L'atteinte morale à la personne est plus fréquente sur les réseaux sociaux et prend généralement l'une des formes suivantes :

Le harcèlement criminel

Il s'agit de « répétition d'actes, pendant un certain temps, qui amènent la victime à craindre raisonnablement pour sa sécurité » (ministère de la Justice Canada, 2004). Parmi

les exemples de harcèlement criminel on trouve « le fait de suivre une personne ou de communiquer avec elle de façon répétée; de surveiller la maison ou le milieu de travail d'une personne de façon répétée; ou de menacer directement une personne ou un membre de sa famille de manière à lui faire craindre pour sa sécurité ou pour celle d'une de ses connaissances » (Milligan 2009). De même, tout comportement envers une personne précise qui lui cause une détresse émotive intense et tout geste, acte ou parole visant à importuner, accabler et tourmenter verbalement une autre personne sont considérés aussi comme du harcèlement criminel.

Comme le montre bien la figure ci-dessous, le taux de harcèlement criminel a affiché une hausse graduelle entre les années 2000 et 2009 (information la plus à jour sur le harcèlement criminel chez Statistique Canada) (Milligan 2009).

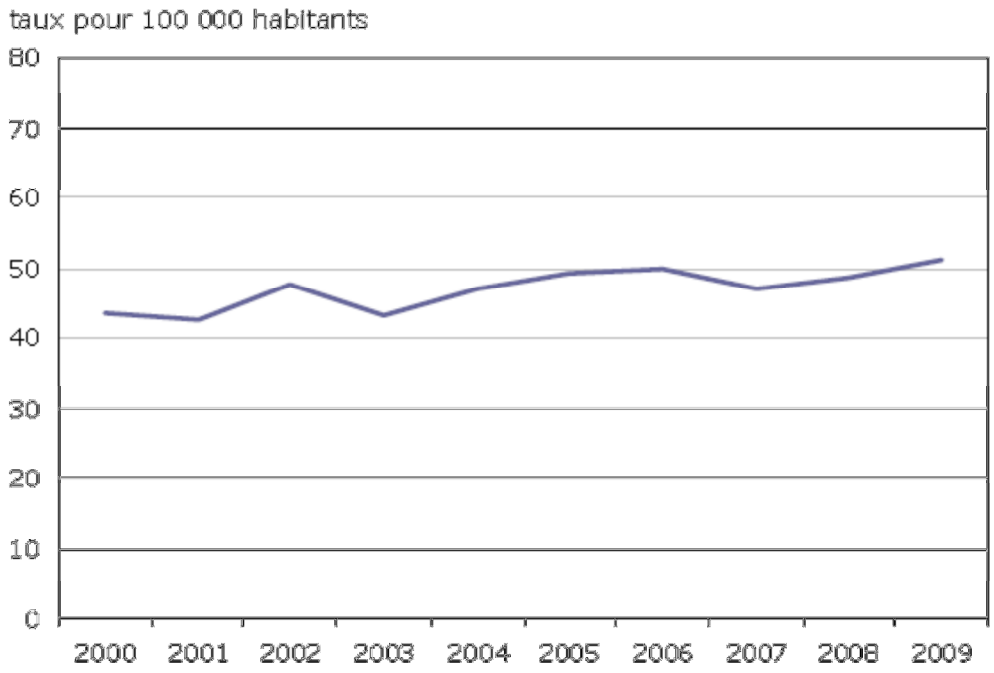


Figure 18. Taux de harcèlement criminel au Canada de 2000 à 2009

La menace

La menace sur Internet est une forme d'intimidation souvent appelée la « *cyberintimidation* » qui peut se manifester par une parole ou un geste marquant une certaine hostilité. Elle est définie comme étant « un acte agressif et intentionnel commis par un groupe ou un individu en utilisant des formes électroniques de communication, de façon répétée et sur une personne qui ne peut se défendre facilement » (Smith, et al. 2008).

La menace peut avoir des conséquences très sérieuses sur les victimes, pouvant aller jusqu'à la mort ou le suicide. Très souvent, dans le cas de menace, le rapport des forces entre l'agresseur et la victime penche clairement en faveur du premier ((Olweus 1987); (Rigby 1993); (Nansel, et al. 2001); (Vandebosch and van Cleemput 2008)).

Les voies d'expression des menaces sont nombreuses: la menace verbale, la lettre ou le message électronique (anonyme ou signé), les appels malveillants et le harcèlement criminel sont les formes les plus communes. L'agresseur peut aussi faire appel à une tierce personne pour menacer sa victime.

Incitation au suicide

Les jeunes utilisateurs des réseaux sociaux sont susceptibles de devenir la proie de certains malades mentaux qui éprouvent du plaisir à voir les autres souffrir. Ces jeunes peuvent être encouragés à s'enlever la vie devant une webcam pour tout simplement satisfaire la soif de gens à la recherche de sensations fortes. Un des cas de suicide célèbre est celui de Nadia Kajouji, étudiante à l'université Carleton d'Ottawa. William Melchert-Dinkel (États-Unis) entre en contact avec Kajouji sur le web, d'abord par le biais de forum puis à travers la messagerie instantanée (*Chat*) dans le but de la persuader de se pendre devant sa webcam, afin qu'il puisse la regarder mettre fin à ses jours. Il cache en réalité sa vraie identité derrière un personnage fictif, se faisant passer pour une infirmière dans la trentaine. Ses efforts portent fruit lorsque Kajouji semble convaincue de mettre un terme à sa dépression chronique. Elle choisira de se noyer au lieu de se pendre, dans l'espoir de

camoufler son suicide en accident de patinage. Il existe plusieurs autres cas d'incitation au suicide reliés à Melchert-Dinkel (CBC news 2011).

Les chercheurs Robin Skinner et Stephen McFaull (Skinner et McFaull 2012) de l'Agence de santé publique du Canada soulignent que :

« Des sites pro-suicide sur l'Internet peuvent augmenter le risque de suicide encore plus en donnant des détails sur les différentes façons de se suicider et en évaluant ces méthodes en ce qui concerne leur efficacité, la quantité de douleur impliquée, et la longueur de temps pour produire la mort. »

Ils mentionnent également que la prévalence et l'influence d'internet, et des médias sociaux dans la vie des jeunes canadiens doivent être prises au sérieux alors que de nombreux sites, blogues et forum de discussion font la promotion du suicide.

3.4. Les attaques informatiques

Une attaque informatique est l'exploitation d'une faille d'un système informatique (système d'exploitation, logiciel) ou de la malveillance d'un utilisateur à des fins généralement préjudiciables. Les attaques informatiques sont pour la plupart lancées automatiquement à partir de machines infectées (par des virus, chevaux de Troie, vers, etc.) à l'insu de leur propriétaire. Plus rarement, il s'agit de l'action directe de pirates informatiques.

Une attaque informatique n'est pas toujours fructueuse. La réussite d'une telle attaque dépend de la vulnérabilité du système informatique visé, ainsi que des mesures mises en place pour le bloquer (pare-feu, système de détection d'intrusions, clustering, etc.).

Selon le site de la sécurité publique du Canada (Sécurité publique Canada 2011) il existe plusieurs façons d'accéder à l'information à partir du cyberespace : « Les attaquants peuvent exploiter des vulnérabilités dans les logiciels ou le matériel. Ils peuvent exploiter

des vulnérabilités sur le plan de la sécurité en amenant par la ruse des personnes à ouvrir des courriels infectés ou à visiter des sites Web corrompus qui leur permettent d'infecter les ordinateurs avec des maliciels. Ils peuvent aussi profiter des personnes qui n'appliquent pas les pratiques de base en matière de cyber-sécurité, comme changer souvent leurs mots de passe, mettre à jour périodiquement les logiciels antivirus et utiliser seulement des réseaux sans fil protégés ».

Les motivations qui se cachent derrière les attaques informatiques sont variées. Nous citons à titre d'exemple :

- Le vol d'informations personnelles délicates (numéro de compte bancaire, mot de passe, documents secrets, etc.), de secrets industriels ou des propriétés intellectuelles;
- La perturbation du bon fonctionnement d'un système;
- L'utilisation du système comme rebond pour attaquer un autre système;
- L'utilisation des ressources du système notamment lorsqu'il possède une bande passante élevée;
- Le désir de prouver ses compétences techniques.

Parmi les attaques informatiques nous retrouvons : le détournement de clics (*Clickjaking*), l'hameçonnage (*Phishing*), le pourriel (*spamming*) et le code malveillant (*malware*).

Le détournement de clics (*Clickjacking*)

C'est une technique de piratage qui consiste à utiliser les propriétés HTML d'une page web pour manipuler son affichage, et ce en superposant un calque transparent à la page visible, afin de tromper l'internaute qui, en cliquant sur un lien ou un bouton, apparemment inoffensif, exécute des actions, sans s'en apercevoir, sur la page dissimulée.

Très souvent, le but est de rediriger l'internaute vers un site malveillant, d'augmenter artificiellement le taux de pages vues sur un site Web, ou d'activer à distance à l'insu de l'utilisateur sa webcam ou son microphone afin de l'espionner. Les attaques de détournement de clic touchent de plus en plus les réseaux sociaux, notamment *Facebook*. En effet, des cas de « *likejacking* » sont devenus de plus en plus fréquent. Le « *likejacking* » est une technique informatique frauduleuse visant les utilisateurs de *Facebook*. Il fonctionne de manière virale, grâce à des sites appâts présentant des vidéos drôles, insolites ou coquins. En cliquant sur le bouton lançant la vidéo, le visiteur clique en réalité sur un bouton « J'aime » caché, ce qui se traduit, s'il est connecté à Facebook, par la création automatique et involontaire d'un « statut » sur son « mur ». Pour cela, Facebook avise ses utilisateurs à partir de son centre d'aide[24] en leur affichant le message suivant :

« Certains sites web malveillants contiennent du code qui peut entraîner votre navigateur à agir sans que vous le sachiez ou ne le vouliez. Le fait de cliquer sur un lien sur un de ces sites web pourrait lui permettre de publier du contenu sur votre journal *Facebook*, par exemple. Ne cliquez jamais sur des liens qui pourraient vous sembler suspects, même s'ils vous ont été envoyés par un ami. Veillez également à prévenir la personne qui vous a envoyé le lien si vous remarquez quelque chose de suspect. »

D'ailleurs *Facebook* et l'état américain de Washington poursuivent en justice la société « *Adscend media* » en l'accusant de « *clickjacking* » alléguant qu'elle a récolté 1,2 millions de dollars par mois à l'aide de faux liens acheminant les utilisateurs vers d'autres sites (theguardian 2012).

Le Hameçonnage (*Phishing*)

L'hameçonnage est une technique utilisée par les fraudeurs pour obtenir des renseignements personnels des internautes dans le but de voler leurs identités. Elle consiste à faire croire à la victime qu'elle s'adresse à un tiers de confiance (banque, gouvernement,

[24] http://www.facebook.com/help/?faq=103840806372798

etc.) afin de lui soutirer des informations délicates telles qu'un mot de passe, un numéro de carte de crédit ou une date de naissance (Coronges, Dodge et al. 2012).

Les attaques d'hameçonnage ont un impact très négatif sur les institutions financières qui perdent la confiance de leurs clients. Cette perte est souvent plus dévastatrice que les pertes financières engendrées par ces attaques (Sevgi 2007). La réussite de l'attaque est assez souvent basée sur l'exploitation des émotions. En effet, en suscitant l'intérêt et la curiosité de l'internaute, le délinquant a de forte chance de réussir son coup. (Timm et Perez 2010).

L'hameçonnage est en constante croissance grâce à l'ingéniosité des fraudeurs : ils arrivent continuellement à trouver de nouvelles méthodes pour contourner les contrôles informatiques et ainsi avoir toujours plus de victimes. Selon le Centre anti-fraude du Canada (Centre antifraude du Canada 2011), 5% des personnes visées par l'hameçonnage répondent aux courriels qu'ils reçoivent. L'étendue des tentatives d'arnaques par hameçonnage est très importante. En effet, on a recensé un total approximatif de 27 200 incidents d'hameçonnage en 2011, contre seulement 17 000 en 2010. Les stratagèmes des attaques d'hameçonnage ciblent de plus en plus les institutions financières. On a recueilli un total de 6 256 signalements de plaintes d'hameçonnage entre les mois de juin et juillet 2011, soit une augmentation de 78% par rapport à la même période pour l'an 2010 (Centre antifraude du Canada 2011). Cette augmentation est également attribuable au fait que les fraudeurs sont de plus en plus compétents en matière d'usurpation d'identité. Ils reproduisent les logos et les sites web des institutions dont ils volent l'identité, et adoptent le même type de discours officiels dans leurs messages, flouant ainsi facilement leurs victimes qui croient avoir à faire avec leurs vraies institutions financières, compagnies de carte de crédit ou organismes gouvernementaux.

Les réseaux sociaux sont particulièrement vulnérables aux attaques d'hameçonnage à cause des services de messagerie électronique qu'ils offrent. De plus, les pirates peuvent exploiter les réseaux sociaux pour identifier le cercle d'amis des victimes et récolter une

grande quantité d'information sur eux. Ils seront par la suite une cible facile à flouer puisqu'ils pourront utiliser les noms et les adresses de leurs amis afin de paraître plus crédible.

Notons qu'en 2011, plus de 112472 attaques d'hameçonnages ont était signalées à travers le monde. C'est beaucoup plus que les 42624 attaques signalées en 2010, mais moins que le record de 2009 qui a atteint 126 697 attaques (Anti-Phishing Working Group 2011).

Kaspersky Lab précise que pendant le mois de juin 2012, 0,01% des courriels échangés à travers le monde étaient infectés. Les réseaux sociaux ont été les plus exposés aux attaques d'hameçonnage (25,15%), suivi par les institutions financières (23,50%) (Kaspersky Lab 2012). La figure suivante présente la distribution des principaux organismes qui ont été victimes d'hameçonnage durant le mois de juin 2012 :

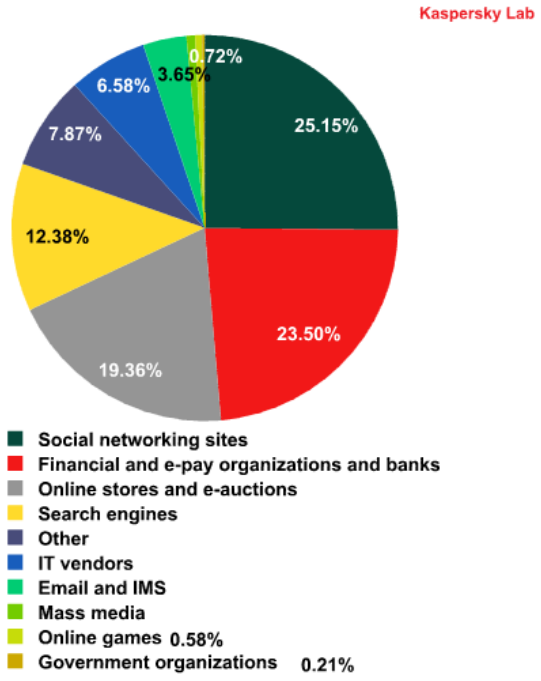


Figure 19. Distribution des principaux organismes victimes d'hameçonnage durant le mois de juin 2012

Le pourriel (*spam)*

Le pourriel est une communication électronique non sollicitée. Il s'agit normalement d'envoi massif de messages publicitaires. Il peut s'attaquer à divers médias électroniques notamment les courriels, les forums de discussion, les réseaux sociaux, les moteurs de recherches, les wikis, les messageries instantanées et les blogues (Stone-Gross et Holz 2011). Le pourriel est aussi utilisé pour diffuser d'autres menaces comme les logiciels espions, l'hameçonnage et les logiciels malveillants.

Selon *Industrie Canada* (Industrie Canada 2005), « le pourriel menace directement la viabilité d'Internet comme moyen efficace de communication et la prospérité économique, à l'efficience des services public au développement d'une cyberéconomie ».

Face à l'augmentation considérable et importante du nombre de pourriels sur le web, une loi anti-pourriel a été adoptée au Canada afin de lutter contre ce fléau. En effet, et ce en vertu de la loi C-28 visant l'élimination des pourriels sur l'ensemble des réseaux internet et sans fil, le CRTC « pourrait imposer des sanctions administrative pécuniaires (SAP) pouvant atteindre 1 million de dollars pour les particulier et 10 millions de dollars pour les entreprises fautives […] Le commissariat à la protection de la vie privée utiliserait ses outils et son cadre d'application de la loi actuels pour exécuter les dispositions de cette législation ».

En juin 2012, la Chine et l'Inde étaient les deux pays qui ont envoyé le plus de pourriels en direction de l'Europe selon *Kaspersky Lab* (voir figure 20). Le pourcentage des pourriels provenant de la Chine a atteint 36,6% du volume total des pourriels alors que le pourcentage des pourriels provenant de l'Inde a atteint 12,6% (Kaspersky Lab 2012).

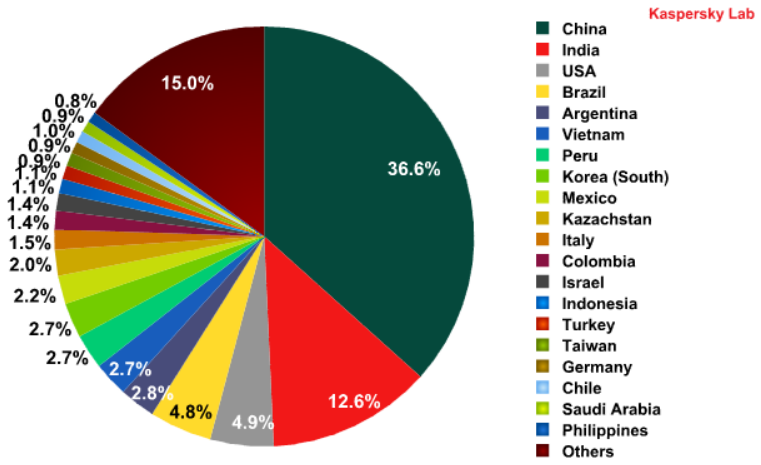


Figure 20. Principaux pays sources de pourriels envoyés en direction de l'Europe en juin 2012

Les deux pays qui ont envoyé le plus de pourriels en direction des États-Unis durant le mois de juin 2012 étaient les États-Unis eux-mêmes (39,2%) et la Chine (9,1%). La figure ci-dessous donne un aperçu sur le classement des principaux pays qui ont pris les États-Unis comme cible durant cette même période.

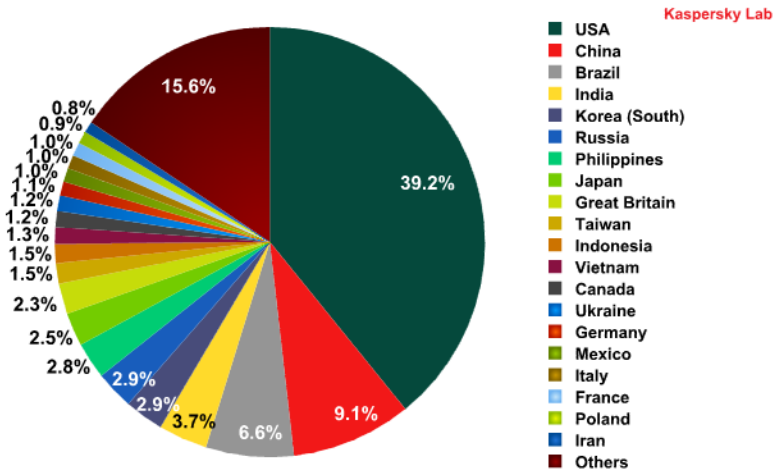


Figure 21. Principaux pays sources de pourriels envoyés en direction des États-Unis en juin 2012

Le contenu des pourriels recensés pendant le mois de juin 2012 tournent principalement au tour de cinq sujets (voir figure ci-dessous) : les finances personnelles, les médicaments et services de santé, les annonces frauduleuses d'emplois, le contenu pornographique et les jeux d'argent.

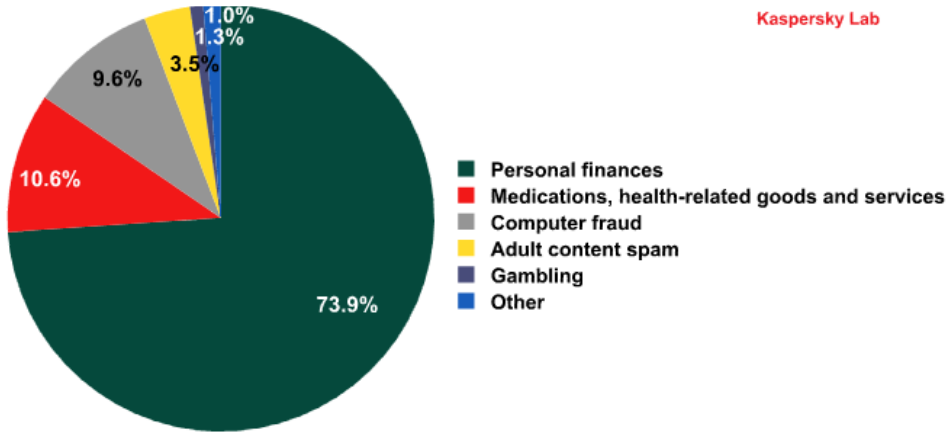


Figure 22. Catégories des pourriels en juin 2012

Le Code malveillant (*malware*)

Un code malveillant est un code qui permet de perturber le bon fonctionnement d'un ordinateur, de collecter des informations sensibles ou d'accéder et de prendre le contrôle d'un système informatique privé. C'est un terme général utilisé pour désigner un code malicieux et intrusif tels que les virus informatiques, les vers et les chevaux de Troie.

Selon les données de KSN (*Kaspersky Security Network*), au premier trimestre 2012, les logiciels de Kaspersky Lab ont détecté et neutralisé près de un milliard de programmes malveillants, soit 28% de plus que le trimestre précédent. La moitié de ces attaques sont des tentatives d'insertion de code malveillant via Internet. Ceci représente une augmentation de 10 % par rapport au trimestre précédent (Namestnikov 2012).

De plus, l'usage des machines zombies figure maintenant parmi les technologies de base utilisées par les cybercriminels. Une machine zombie est un ordinateur contrôlé à l'insu de son utilisateur par un pirate informatique, et est exploitée le plus souvent à des fins malveillantes. On a noté, au premier trimestre 2012, une nette croissance des réseaux de zombies décentralisés et administrés via des réseaux sociaux (Namestnikov 2012).

Les applications les plus souvent visées par les codes malveillants sont celles que les utilisateurs doivent mettre à jour régulièrement. Comme le montre la figure ci-dessous, 66 % de ces attaques sont imputables aux codes d'exploitation de deux applications : *Adobe Reader* et *Java* (Namestnikov 2012).

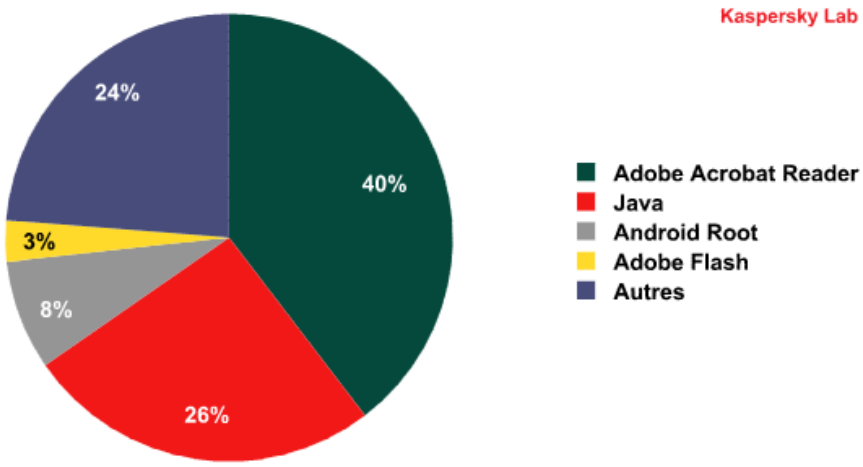


Figure 23. Applications dont les vulnérabilités ont été exploitées par des codes d'exploitation Internet au premier trimestre 2012

Les pays suivants représentaient 84 % des ressources Internet utilisées pour diffuser des programmes malveillants du premier trimestre de l'année 2012 :

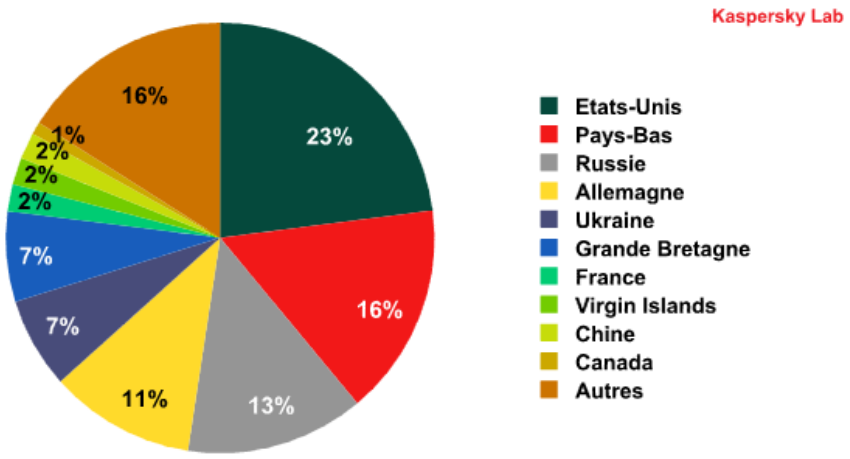


Figure 24. Répartition des sites Web abritant des programmes malveillants par pays au premier trimestre 2012

Les pays dont les internautes ont été le plus exposés au risque d'infection via Internet, durant le premier trimestre 2012, sont énumérés dans la figure suivante :

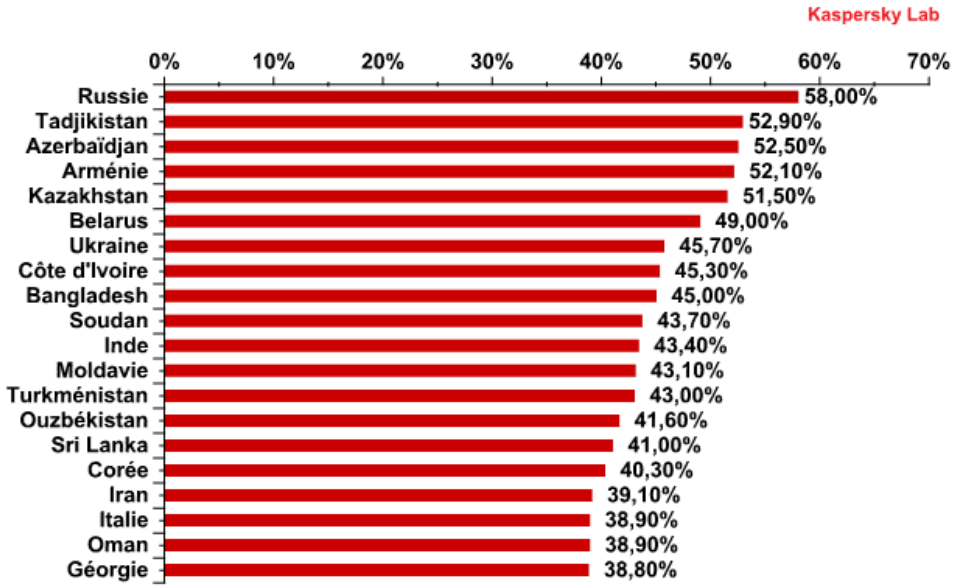


Figure 25. Top 20 des pays dont les internautes ont été le plus exposés au risque d'infection via Internet au premier trimestre 2012

3.5. Les actions problématiques

Une action problématique sur les réseaux sociaux se traduit très souvent par la déclaration de propos haineux vis-à-vis d'un groupe identifiable, ou par la diffusion de contenu (textes, images, vidéo, etc.) à caractère offensant.

Diffusion de contenu compromettant

Il s'agit essentiellement de diffusion de contenu désobligeant lié généralement au terrorisme, à la pornographie, à la pédophilie, au trafic d'organes humain ou à la violence. Tout contenu agressif, illégal, menaçant, injurieux, calomnieux, vulgaire, obscène, raciste, trompeur, impoli ou inapproprié est considéré aussi comme contenu compromettant.

Notons qu'il existe des systèmes appelés « d'étiquetage » qui permettent la classification des sites Internet en fonction de leur contenu (Heller 2000). La raison d'être de ces systèmes est la volonté de protéger les jeunes internautes du contenu compromettant qui circule sur le Web.

À titre d'exemple, le système d'étiquetage *RSACi* proposait quatre critères et cinq niveaux pour chaque critère pour déterminer si le contenu des sites Web et notamment les jeux vidéo est offensant ou pas. Les principaux critères considérés sont : (1) violence, (2) nudité, (3) sexe et (4) langage. Plus le niveau attribué est élevé, plus le contenu du site ou du jeu est inadéquat. En octobre 2010, le système *RSACi* a été interrompu[25] par le *Recreational Software Advisory Council*.

SafeSurf [26] est un système plus élaboré que le *RSACi*. Les pages Web peuvent contenir jusqu'à dix étiquettes pour décrire leur contenu. Chaque étiquette présente neuf niveaux. Les dix étiquettes proposées par *SafeSurf* sont : (1) nudité; (2) violence; (3) grossièreté; (4) thèmes hétérosexuels; (5) thèmes homosexuels; (6) sexe, violence et grossièretés; (7) intolérance vis-à-vis de la race, des croyances religieuses ou du sexe d'une

[25] http://www.icra.org/
[26] http://www.safesurf.com/

personne; (8) prônant la consommation de stupéfiants; (9) autres thèmes adultes; (10) jeux de hasard. Pour chaque classe, il existe neuf niveaux d'âge sur lesquels se décline chacune des dix catégories : (1) tous les âges; (2) enfants plus âgés; (3) adolescents; (4) adolescents plus âgés; (5) supervision parentale recommandée; (6) adultes; (7) réservés aux adultes; (8) adultes uniquement; (9) explicitement pour adultes (Heller 2000).

Néanmoins, ces systèmes sont loin d'être parfaits. En raison de leur complexité et en l'absence d'un large consensus, ce type de catalogage est encore assez peu utilisé par les éditeurs de pages Internet. Par conséquent, l'utilisateur qui opte pour ce genre de protection doit savoir qu'il ne va réellement filtrer qu'une faible part des contenus circulant sur le réseau.

Promotion de la haine

La promotion de la haine est l'utilisation de tout moyen de communication électronique, d'usage des nouvelles technologies de communication, dans le but de divulguer, partager, promouvoir des propos haineux, antisémites, racistes, extrémistes ou terroristes. La propagande de propos haineux est considérée comme criminelle par les articles 318 et 319 du Code Criminel du Canada. Ce dernier énonce que « quiconque, par la communication de déclarations en un endroit public, incite à la haine contre un groupe identifiable, lorsqu'une telle incitation est susceptible d'entrainer une violation de la paix, est coupable (…) » et « Quiconque, par la communication de déclarations autrement que dans une conversation privée, fomente volontairement la haine contre un groupe identifiable est coupable (…) ».

Au Canada, le nombre et le taux de crimes haineux déclarés par la police ont tous les deux diminué en 2010. Selon *Statistique Canada* (Statistique Canada 2012), « les crimes haineux déclarés par la police sont des affaires criminelles qui, après enquête par la police, sont déterminées comme ayant été motivées par la haine d'un groupe identifiable. L'affaire peut cibler la race, la couleur, l'origine nationale ou ethnique, la religion,

l'orientation sexuelle, la langue, le sexe, l'âge, l'incapacité mentale ou physique, ou d'autres facteurs tels que la profession et les convictions politiques ».

Il y a eu 1401 crimes haineux en 2010, soit 4,1 crimes haineux pour chaque tranche de 100000 habitants. Ce taux était de 18 % inférieur à celui noté en 2009. En 2010, trois facteurs de motivation étaient principalement à l'origine de plus de 95 % des crimes de haine : la race ou l'origine ethnique, la religion et l'orientation sexuelle. Le graphique suivant illustre le taux de chaque type de motif de crime par 100 000 habitants :

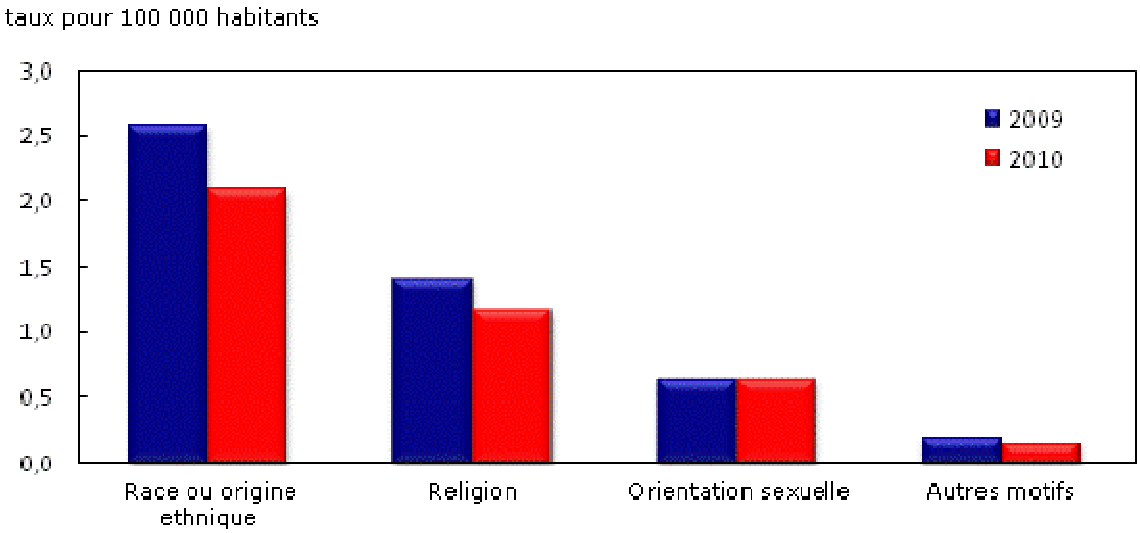


Figure 26. Crimes de haine déclarés par la police, selon le type de motif, 2009 et 2010

3.6. Quelques dangers propres aux réseaux sociaux

Les utilisateurs des réseaux sociaux peuvent être victimes de certains dangers imperceptibles à première vue. Nous nous contentons de citer les dangers suivants (Collée 2009) :

La notion trompeuse de communauté

Beaucoup d'utilisateurs de réseaux sociaux confondent les structures de communication du monde « réel » avec celles du cyberespace. Ils croient que c'est

sécuritaire de publier des données personnelles sur les plateformes des réseaux sociaux parce qu'ils se sentent entre amis comme dans la vraie vie. Si ces utilisateurs ne se rendent pas compte de la façon avec laquelle leurs informations de profils sont partagées, ils pourraient être séduits par cette notion de « communauté » et être encouragés à partager des renseignements personnels sensibles. De plus, le nom de certaines plateformes, comme « *MySpace* », crée l'illusion de l'intimité sur le Web.

Perte du contrôle des données publiées

Les données personnelles ainsi que les notes ou les commentaires (très souvent préjudiciables), une fois publiés, peuvent être conservés longtemps même lorsque l'utilisateur les aura supprimés du site original et même en cas de fermeture du réseau social. Il peut exister des copies sur des sites tiers ou chez d'autres usagers. De plus, certains fournisseurs de service ignorent parfois toute demande de suppression de données ou de profils de leurs utilisateurs. D'une certaine façon, l'utilisateur perd le contrôle sur son identité.

Collecte de données secondaires par les fournisseurs de services

Les fournisseurs des réseaux sociaux peuvent collecter beaucoup d'informations secondaires (emplacement géographique, âge, etc.) sur leurs utilisateurs dans le but de personnaliser leurs services pour le ciblage et le transfert des données aux tiers par revente. L'utilisation des données de profil dans la personnalisation de la publicité leur assure une bonne source de revenu.

Reconnaissance faciale

Les photos publiées peuvent devenir des identificateurs biométriques universels. Une fois qu'on réussit à lier un nom à une photo, il devient relativement facile de retracer la personne sur d'autres réseaux et dans d'autres photos. Cela met en danger sa vie privée et sa sécurité ainsi que celles de ses amis. De plus, les informations que les photos conservent

sous forme de métadonnées (date, heure, emplacement géographique, etc.) peuvent servir pour localiser une personne ou trouver un lieu.

Exploration et fouille des données (*Datawarehouse* et *Datamining*)

L'exploration et la fouille des données consiste à récolter et analyser les traces que les utilisateurs d'un réseau social laissent derrière eux (nombre d'amis, opinions et publications affichés, contenu des éléments multimédia utilisés, etc.), afin de cibler leurs modes de consommation et de déterminer leurs différents profils sociaux. Elles permettent au fournisseur du service de faire de la publicité dirigée en fonction des profils obtenus.

3.7. Conclusion

Ce chapitre a mis l'accent sur les différents dangers qui guettent les utilisateurs des réseaux sociaux dans le but de mieux comprendre l'importance et la nécessité de pouvoir prévenir et détecter les utilisateurs à risque avant qu'ils ne passent à l'action. C'est exactement ce que permet de faire le système *Protect_UFF* de notre plateforme. Son fonctionnement est expliqué au chapitre 8. Cependant, pour être en mesure de cibler avec précision les utilisateurs à risque tels que les fraudeurs, les psychopathes et les criminels il est essentiel de trouver des caractéristiques communes au niveau de leurs traits de personnalité. D'où l'intérêt du chapitre suivant.

Chapitre 4 : Traits et troubles de personnalité

Ce chapitre est divisé en quatre sections. La première section définit les notions de *personnalité*, *traits de personnalité* et *troubles de personnalité*. La deuxième section rend compte des écrits portant sur la personnalité psychopathique. La troisième section présente les principaux traits de personnalité des fraudeurs tandis que la dernière section évoque certaines caractéristiques des traits de personnalité des criminels en général.

4.1. Définitions

La *personnalité* est une notion étudiée depuis bien longtemps par les chercheurs et possède plusieurs définitions. Elle peut être considérée comme « une structure dynamique, relativement stable et cohérente, qui résulterait de prédispositions génétiques (composantes biologiques et héréditaires manifestées par le tempérament), de la séquence des évènements vécus depuis la petite enfance (composantes environnementales) et de l'interaction de ces deux éléments constitutifs » (Pervin et John 2005). Certains psychologues présentent la personnalité comme étant une organisation psychique dynamique et relativement stable qui influence la manière qu'une personne aura d'interpréter le monde, d'interagir, de réagir et de s'adapter à son milieu ((Allport 1937); (Beck, Freeman et Davis 2003); (Cottraux et Blackburn 1995); (Eysenck 1964); (Kelly 1963); (Organisation Mondiale de la santé 1992)). Cependant, malgré les nombreuses études, il est difficile de déterminer si les émotions, pensées et comportements représentent des éléments constitutifs de la personnalité ou s'ils en découlent (Paquette 2010).

Selon Morizot et Miranda (Morizot et Miranda 2007), les *traits de personnalité* représentent l'unité de mesure de la personnalité la plus fréquemment employée au niveau scientifique. La mise en place de ces traits a facilité le déploiement d'un ensemble de critères permettant de comparer, regrouper et différencier les individus. Les traits de personnalité sont des structures latentes qui manifestent les modes selon lesquels les individus réagissent à leur environnement et s'y adaptent ((Morizot(a) 2003) ; (Rolland 2004)).

Malgré le fait que la personnalité de chaque individu soit unique, il est possible de regrouper les traits de personnalité pour former des catégories et des types ((Davison et Neale 2001); (Morizot(b) et Le Blanc 2005)). Ceci est particulièrement vrai pour l'étude et la compréhension des troubles de la personnalité. Ces derniers constituent des anomalies et des déviances du fonctionnement personnel et social ((Organisation Mondiale de la santé 1992); (Tyrer 1988)). La manifestation de ces anomalies est normalement durable et généralisée à différents contextes et souvent associée à une détresse subjective ((American Psychiatric Association 1994); (Cottraux et Blackburn 1995)).

4.2. La personnalité psychopathique

Le terme *psychopathie* découle du grec ancien. Les racines du mot, *psychê* et *pathos*, signifient respectivement « l'âme, l'esprit, l'état mental ou la psychologie » et « ce qu'on éprouve et ce qui affecte le corps ou l'âme, en bien ou en mal, mais surtout en mal c'est-à-dire, la souffrance » (Paquette 2010). En d'autres termes, le sens premier du mot psychopathie est *anomalie psychologique*. Il existe dans la littérature psychiatrique des études sur la personnalité d'individus considérés aujourd'hui comme psychopathes qui datent des années 1800 ((Hervé 2007); (Côté 2000); (Million 1998); (Werlinder 1978)). Le terme de « psychopathie » a été employé par la suite pour désigner toute maladie mentale aux alentours de la fin du 19e siècle. Il fut ensuite employé en tant que synonyme de *trouble de la personnalité* ((Hervé 2007); (Lalumière et Seto 1998)). L'emploi actuel du terme désigne « un terme diagnostique clinique réservé à un type de personnalité spécifique » (Paquette 2010).

Le psychiatre Hervey Cleckley (Cleckley 1988) décrit un psychopathe comme étant un être qui arrive à se dissimuler derrière un masque artificiel pour faire croire aux autres qu'il ressemble à une personne ordinaire. Cependant, ses comportements dévoilent une attitude antisociale claire très souvent dangereuse pour autrui. Il peut appartenir à n'importe quelle classe sociale.

Selon Yochelson et Samenow (Yochelson et Samenow 1976), un psychopathe est toujours à la recherche du pouvoir, d'admiration et de sensations fortes. Dès qu'il sent qu'il

ne détient plus le pouvoir, un sentiment de colère l'envahit et il cherche très souvent à reprendre ce pouvoir par une agression physique ou verbale. Le narcissisme est un trait marquant dans la personnalité du psychopathe. Selon le DSM-IV (American Psychiatric Association 1994), le narcissisme est caractérisé par une surestimation de soi, par le sentiment d'être unique, par le besoin d'être admiré, par la croyance que tout lui est dû et par le manque d'empathie. D'un autre côté, le besoin de sensations fortes se traduit très souvent par un comportement antisocial, agressif et même criminel.

Certaines études ((Lykken(a) 1957), (Lykken(b) 1995)) stipulent que le tempérament (prédispositions innées) des psychopathes engendre une inaptitude à la socialisation quel que soit leur éducation. Ils arrivent assez souvent à dissimuler leurs comportements antisociaux et à donner une apparence trompeuse d'adaptation sociale. Cette dernière demeure cependant superficielle et utilitaire (pour atteindre des fins égocentriques). Lykken divise les psychopathes en deux catégories : primaires et secondaires. Les psychopathes primaires sont caractérisés par une rareté des sentiments de peur et par une faible anxiété. Ils se retrouvent dans l'ensemble des groupes sociaux (politique, arts, droit, force de l'ordre, etc.). Quant aux psychopathes secondaires, ils sont caractérisés par des prises de risques, par une forte réaction au stress et à l'anxiété et par un repli sur soi.

Hare définit les psychopathes comme étant « des prédateurs sociaux qui charment, manipulent et tracent la voie de leur vie de façon impitoyable en laissant une large traînée de cœurs brisés, de confiance minée et de poches vides » (Hare(a) 1999). Il estime qu'à peu près 1% de la population générale souffre des symptômes de la psychopathie. Hercz croit que 50% des crimes en Amérique du nord sont commis par des personnes psychopathiques (Hercz 2002).

Hare a mis au point en 1980 la première version de la *Psychopathy checklist* (PCL), un instrument diagnostique lui permettant d'identifier les psychopathes. Cette liste a été

révisée en 1991, donnant naissance au *Psychopathy checklist - Revised* (PCL-R) (Hare(b) 1991). Cette dernière est constituée de 20 items[27] :

(1) Loquacité et charme superficiel;

(2) surestimation de soi;

(3) besoin de stimulation et tendance à s'ennuyer;

(4) tendance au mensonge pathologique;

(5) duperie et manipulation;

(6) absence de remords ou de culpabilité;

(7) affects superficiels;

(8) insensibilité et manque d'empathie;

(9) tendance au parasitisme;

(10) faible maitrise de soi;

(11) promiscuité sexuelle et sexualité débridée;

(12) apparition précoce de problèmes de comportement;

(13) incapacité à planifier à long terme de façon réaliste;

(14) impulsivité;

(15) irresponsabilité;

(16) incapacité d'assumer la responsabilité de ses faits et gestes;

(17) instabilité conjugale;

(18) délinquance juvénile;

(19) violation des conditions de remise en liberté conditionnelle et

(20) multiplicité des types de délits.

Ces caractéristiques peuvent être divisées selon deux dimensions[28] ((Hare(c) 1996); (Hare(d), Hart et Harpur 1991); (Harpur, Hakstian et Hare 1988); (Harris, Rice et Quinsey 1994); (Marcus, John et Edens 2004)) :

---

[27] Voir l'annexe 1 pour une description des 20 items de la PCL-R.

[28] Voir l'annexe 1 pour l'identification des items inclus dans chaque dimension de la PCL-R.

- Les relations interpersonnelles et la vie affective : réfère au détachement émotionnel, à l'indifférence affective, au charme superficiel, à l'absence de remords, à la froideur et à l'égocentrisme.

- Le style de vie et les comportements antisociaux : réfère au mode de vie déviant, la chronicité des comportements criminels, l'impulsivité et le faible contrôle de soi.

En conclusion, les individus psychopathiques démontrent des traits de personnalité spécifiques qui se concrétisent notamment à travers leurs relations interpersonnelles, leur vie affective et leurs comportements (Hart, Hare et Forth 1994). Les principaux traits de la psychopathie sont les suivants ((Cooke et Michie 2001); (Harpur, Hart et Hare 2002); (Hervé(b) 2007); (Krueger 2006); (Poythress et Skeem 2006); (Walters 2004)) :

i. *Relations interpersonnelles :*

- Le *narcissisme* : illustré par une surestimation de soi, la recherche de l'admiration et le désir d'être le centre des intérêts.
- Le *charme machiavélique* : illustré par l'usage du mensonge en tant que mode relationnel, la manipulation, l'arrogance, la recherche du pouvoir et du désir de la domination d'autrui, le manque d'authenticité, la malhonnêteté et l'instabilité (soit l'incapacité d'établir des relations stables et profondes).

ii. *Émotions :*

- La *superficialité des affects* ainsi que *la pauvreté des réactions affectives*.
- Le *manque d'empathie* et *l'absence du sentiment de culpabilité*.
- La *faible réactivité aux situations anxiogènes* (sang-froid).

iii. *Comportements :*

- *L'impulsivité, l'irresponsabilité, l'insouciance, la témérité et le désir des sensations fortes*.
- Les *comportements antisociaux* qui vont à l'encontre des normes sociales établies et qui se manifestent par de l'*agressivité*.

iv. *Cognitions :*

- L'*égocentrisme* – les intérêts personnels de l'individu priment sur le reste. Ses comportements et ses pensées sont orientés en fonction de ses propres intérêts.
- L'*obsession du pouvoir.*

4.3. Principaux traits de personnalité des fraudeurs

Cette section introduit le concept de la fraude dans le but de pouvoir déterminer les caractéristiques marquantes de la personnalité des fraudeurs.

4.3.1. La fraude

Dans le but de définir le concept de la fraude, il est important de distinguer entre deux méthodes différentes pour s'emparer des biens d'autrui. La première manière consiste à prendre possession des biens par la force physique, alors que la seconde fait appel à la ruse (Albrecht, Albrecht et Albrecht 2006). La fraude est un crime à caractère non violent dans le sens où il n'y a pas d'agression physique. Le fraudeur agit consciemment et intentionnellement. La réussite de son crime repose en grande partie sur l'innovation et l'intelligence.

Le code criminel canadien définit la fraude comme étant :

« Quiconque, par supercherie, mensonge ou autre moyen dolosif, constituant ou non un faux semblant au sens de la présente loi, frustre le public ou toute personne, déterminée ou non, de quelque bien, service, argent ou valeur est coupable de...» (Article 380.1 du code criminel canadien).

La fraude peut être commise par un particulier ou une compagnie, contre un particulier ou une compagnie, et peut être passible de poursuite criminelle ou civile. Elle est définie généralement en termes juridiques parce qu'elle représente une infraction au code criminel, mais peut aussi être décrite en termes d'escroquerie :

« Contrairement à d'autres délits, l'escroc renverse la situation, il ne s'empare pas d'un bien sous l'effet d'une contrainte ou d'une menace, comme dans le vol, mais il crée les conditions pour [...] parvenir à se faire remettre ce qu'il recherche activement. [...] en obtenant de celui qui le possède la remise volontaire d'un bien (chose, argent), documents de valeur vénale ou juridique par une tromperie caractérisée. La remise du bien par son propriétaire s'avère [...] effectuée sans qu'il ait pu en apprécier les implications péjoratives pour lui » (Puig-Verges et Schweitzer 1996).

Il est possible de décrire l'ampleur de la fraude suivant des indicateurs tels que les résultats des sondages de victimisation auto-révélée et le taux de criminalité. Cependant, il est irréaliste de croire qu'on puisse déterminer l'ampleur réelle de la fraude comme pour la plupart des autres crimes. Ceci remonte en partie au fait qu'un grand nombre de fraudes sont commises sans pour autant qu'elles soient rapportées à la police. Il existerait également une proportion de fraudes commises envers les entreprises qui sont enquêtées par des organismes internes. Ceci diminue le nombre des fraudes rapportées et empêche d'en connaître l'ampleur exacte.

4.3.2. Caractéristiques marquantes

Plusieurs recherches ont porté sur les délinquants dont la criminalité est orientée vers la fraude. Ces recherches ont touché les fraudeurs de toutes les classes socioéconomiques, notamment ceux issus de la classe favorisée très souvent appelés fraudeurs en cols-blancs. Plusieurs aspects de leur personnalité ont été dépeints. Nous nous limitons aux aspects suivants : *les relations interpersonnelles*, *l'impulsivité* et *l'affectivité*.

Les relations interpersonnelles

Les principaux traits de personnalité liés aux interactions interpersonnelles et aptitudes sociales attribués aux fraudeurs sont les suivants (Paquette 2010) :

a) Égocentrisme et narcissisme

L'égocentrisme est la tendance à être absorbé par soi-même et ne voir que ses propres intérêts, besoins et points de vue. Le narcissisme est la tendance à admirer et à glorifier soi-même. Très souvent, un narcissique a le sentiment d'être unique et exceptionnel. Le narcissisme suppose l'égocentrisme, mais l'inverse n'est pas nécessairement vrai. Les narcissiques ont tendance à mentir. En effet, le « besoin de séduire, de paraître, de tromper, la quête d'attention excessive et envahissante sont au cœur de ce trouble de personnalité ».

Plusieurs chercheurs s'entendent pour affirmer que le narcissisme est un trait principal et central de la personnalité des fraudeurs ((Alalehto 2003); (Bromberg 1965); (Collins et Schmidt 1993); (Delord-Raynal 1980); (Duffield et Grabosky 2001); (Jackson 1994); (Maulaz 2001); (Willott, Griffin et Torrance 2001)). Le narcissisme chez les fraudeurs se démarque par un comportement solitaire et une tendance à l'exhibitionnisme. Ils cherchent à susciter une bonne impression aux yeux de tous même au détriment des autres. Ils ne supportent pas les échecs et n'expriment aucune empathie envers autrui (Gauthier 1960).

Les fraudeurs, et plus précisément les fraudeurs à col blanc, sont reconnus pour avoir un état civil traditionnel, qui est caractérisé par le fait d'avoir un conjoint et des enfants, ainsi qu'un réseau social. Cependant leur monde est fragile puisque les fraudeurs n'arrivent pas à établir et maintenir des relations authentiques ((Gauthier 1960); (Lemert 1972); (Levi 1999); (Mergen 1970)). En effet, même lorsque le fraudeur établit une relation, il ne se présentera presque jamais sous sa vraie réalité (Lemert 1972). Par conséquent, les relations qu'il développe demeurent superficielles ainsi qu'utilitaires (Gauthier 1960). Ce comportement amène les chercheurs à qualifier les fraudeurs d'individus solitaires ((Lavoie and Lessard 1987); (Levi 1999); (Mergen 1970)). Cette personnalité solitaire serait à la base de leur attitude égocentrique : n'ayant pas de relations authentiques profondes, ils choisissent de se concentrer sur eux-mêmes.

Le fait qu'ils soient absorbés par eux-mêmes n'implique pas nécessairement une capacité d'introspection. Levi (Levi 1999) dénote chez les fraudeurs égocentriques un

processus « graduel de détachement du vécu intérieur ». Ils ne cherchent plus à savoir qui ils sont.

b) Capacité de communication, de manipulation et de persuasion

Dans le but de parvenir à leurs fins, les fraudeurs usent de la persuasion et de leurs talents de communication qu'ils ont développés au cours des années. Cet atout leur permet de bien paraître et de gagner la confiance de leurs victimes (Jackson 1994). Ceci les rend aptes à l'usage de la manipulation et du mensonge pour servir leurs intérêts ((Lemert 1972); (Maurey 1996)).

Dans son mémoire de maîtrise intitulé « Étude comparative de la capacité interrelationnelle chez deux groupes de fraudeurs », C. Gaudreau-Toutant a affirmé que « le fraudeur, qui a choisi de réaliser son méfait à l'intérieur d'une relation interpersonnelle où il doit appréhender suffisamment l'autre pour connaître ses points faibles et l'amener ainsi à se fourvoyer, donne l'impression d'être plus capable de communication que le voleur qui n'a bien souvent aucun contact avec sa victime au moment du délit » (Gaudreau-Toutant 1969).

Certains chercheurs ((Goulem 1969); (Delord-Raynal 1980); (Maurey 1996)) trouvent que certains fraudeurs éprouvent du plaisir personnel à tromper même en absence de bénéfices immédiats ou lointains. Le plaisir de manipuler et de mentir serait plus intéressant pour eux que les gains obtenus.

c) Extraversion et sociabilité

Très souvent, les fraudeurs sont présentés comme des individus ayant de grandes compétences sociales ((Bromberg 1965); (Gaudreau-Toutant 1969); (Jackson 1994)). Le fait d'avoir des amis et d'être efficace dans les situations sociales peut mener certains fraudeurs à la manipulation (Gagnon 2008).

L'impulsivité

Dans la *Théorie générale du crime* (Hirschi et Gottfredson 1987), l'impulsivité est présentée comme étant la principale cause menant à la délinquance. Le manque de maîtrise

de soi représente le facteur distinctif déterminant entre une personne qui va passer à un acte frauduleux et une autre qui ne va pas le faire.

L'impulsivité peut se manifester par la recherche de sensations fortes et par de l'agressivité dans le but de chasser l'ennui et de ressentir de l'excitation ((Kellens 1977); (Lemert 1972); (Levi 1999); (Piquero et Piquero 2001)).

L'affectivité et les émotions

Certains fraudeurs tendent à montrer une certaine indifférence affective. Ils dévoilent une froideur affective, ainsi qu'un manque d'empathie ((Delord-Raynal 1980); (Kellens 1977)). Le manque d'empathie combiné à l'absence d'émotions authentiques incitent les fraudeurs à passer à l'acte (Puig-Verges et Schweitzer 1996). Cette indifférence affective est directement liée au désir de paraître supérieur aux autres : reconnaître leurs torts serait de reconnaître leur faiblesse (Levi 1999).

4.3.3. En bref

Peu d'études ont évalué expressément la présence de traits psychopathiques chez les fraudeurs. Parmi ces études, on retrouve celle de Richard Blum. Il a analysé la personnalité d'un échantillon de fraudeurs qui ont tous obtenu un score relativement élevé à l'échelle de psychopathie MMPI (Minnesota Multiphasic Personality Inventory) (Blum 1972). De même, les résultats de l'étude menée par Sophie Gagnon ont révélé que les fraudeurs spécialisés de son échantillon avaient une tendance marquée à la déviation psychopathique et au trouble de personnalité antisociale (Gagnon 2008).

Dans le même sens et à l'issu d'une comparaison de la personnalité entre un groupe de délinquants à col blanc et un autre de cols blancs non délinquants, établie par Collins et Schmidt, il en résulte que les délinquants à col blanc présenteraient plus de traits de la personnalité en rapport au manque de fiabilité, à l'irresponsabilité, à l'égocentrisme, à la méfiance ainsi qu'à la prise de risque (Collins et Schmidt 1993). De plus, ces individus seraient contrôlant, sans pour autant pouvoir se contrôler. Ils exerceraient également une

résistance aux normes sociales établies. Ainsi, la personnalité de ces fraudeurs pourrait s'apparenter à celle d'un psychopathe.

4.4. Principaux traits de personnalité des criminels

4.4.1. Le crime

Le terme *crime* dans cette section désigne les « infractions punissables au terme du Code pénal du Québec et causant un dommage évident à autrui » (Cusson 1983). Ainsi, les termes délinquance, délit, infraction et transgression sont considérés comme des synonymes du terme *crime*.

Cette définition tient compte des crimes dirigés contre la personne et contre la propriété. Il peut s'agir par exemple d'un vol (cambriolage), d'un acte de vandalisme, d'une agression avec coups et blessures, d'un viol, d'un enlèvement ou d'un homicide.

Le mot *criminel* est généralement attribué à une personne qui a commis un crime. La responsabilité criminelle d'une personne morale peut découler des actions d'employés occupant des fonctions clés au sein d'une entreprise. Avant mars 2004, la responsabilité criminelle d'une personne morale relevait de la *common law* : « une entreprise ne pouvait être reconnue coupable d'une infraction de *mens rea*[29] que si l'acte répréhensible avait été commis par une personne qui en était l'âme dirigeante » (Leblanc et Renaud 2012). En mars 2004, un changement important a été apporté au Code criminel du Québec : le terme « l'âme dirigeante » a été remplacé par « un cadre supérieur » défini comme un « agent jouant un rôle important dans l'élaboration des orientations de l'organisation visée ou assurant la gestion d'un important domaine d'activités de celle-ci, y compris, dans le cas d'une personne morale, l'administrateur, le premier dirigeant ou le directeur financier » (Article 22.2 du Code criminel du Québec).

[29] Esprit criminel

4.4.2. Caractéristiques marquantes

Les recherches psychologiques sur les conduites criminelles n'ont pas réussi à prouver sans aucun doute l'existence d'une personnalité criminelle spécifique qui expliquerait à elle seule toutes les conduites délinquantes à cause des diverses formes et types de la criminalité (Schuessler et Cressey 1950). Cependant, les recherches empiriques sur les profils psychologiques de certains délinquants chroniques, tels que les criminels sexuels et les tueurs en série, ont permis de mettre en évidence certaines caractéristiques de leurs profils (Luci 1999).

En 1963, les auteurs du *Traité de droit pénal et criminologie* (Pinatel et Bouzat 1963) ont proposé un ensemble de caractéristiques marquantes du profil d'un criminel tels que l'égocentrisme, l'agressivité, le besoin de domination, l'intolérance à la frustration, le manque d'empathie, l'absence d'émotions et la faiblesse au sens moral. Ces caractéristiques sont toujours utilisées actuellement dans les tests psychologiques et les grilles d'évaluation psychiatriques.

4.5. Conclusion

Malgré la complexité et les différentes nuances qui existent au niveau des traits de personnalité des fraudeurs, des psychopathes et des criminels certains points communs persistent tels que le niveau élevé de narcissisme, de manque d'empathie, d'absence d'émotions et d'antisociabilité qui se traduit très souvent par de l'agressivité. C'est en se basant sur ces caractéristiques communes que *Protect_UFF* va chercher à détecter le niveau de risque des « amis » *Facebook* en analysant le contenu des images et des textes de leurs comptes. Les méthodes et les algorithmes utilisés dans l'exploration des textes et des images sont présentés dans le chapitre suivant.

Chapitre 5 : Exploration de textes et d'images

Ce chapitre introduit les méthodes d'exploration de textes et d'images qui ont été appliquées dans notre étude. Il met l'accent sur les étapes de la catégorisation automatique de textes ainsi que sur le rôle des algorithmes d'apprentissage dans l'exploration de données (*Data Mining*). Il présente les algorithmes d'apprentissage utilisés dans notre étude : le C4.5 et les *Machines à Support Vectoriel* (ou *Machines à vecteurs de support*) linéaire de *WEKA* (noté *SMO*). Il expose ensuite deux algorithmes d'exploration d'images : l'algorithme de comparaison d'images *Imgseek* et l'algorithme de détection des formes que nous avons utilisé dans notre étude pour détecter des objets agressifs et antisociaux dans les images.

5.1. L'exploration de textes

Jour après jour, la quantité de textes sur Internet augmente considérablement. Tout ce contenu serait sans intérêt si notre capacité à y accéder et à le traiter n'évolue pas. Pour pouvoir explorer convenablement ces données, nous avons besoin d'outils et de méthodes qui nous permettent de trouver dans un délai raisonnable l'information désirée et d'en extraire des modèles.

L'exploration de textes a pour but de rechercher et d'extraire de l'information utile à partir de gros volumes de textes. Le processus d'exploration comprend normalement plusieurs étapes :

1. Collecte des informations.
2. Organisation et ajustement (nettoyage du bruit) de ces données dans une base de données.
3. Sélection des attributs utiles.
4. Extraction d'information utile de la base donnée.
5. Visualisation des données.
6. Évaluation des résultats de l'extraction de connaissance.

Pour profiter pleinement de l'exploration de textes, la catégorisation automatique se présente très souvent comme une solution utile et efficace.

5.1.1. Catégorisation automatique de textes

L'objectif est de rendre une application informatique capable de déterminer de façon autonome dans quelle catégorie classer du texte à partir de son contenu. La catégorisation du texte repose avant tout sur la signification du texte, ce qui apparente la classification au problème d'extraction de la sémantique du texte. Il s'agit là d'un problème de taille, compte tenu du fait que le traitement de l'aspect sémantique d'un document écrit en langage naturel demeure non résolu.

Il existe deux méthodes distinctes d'aborder le problème de la classification automatique de manière générale. Avant les années 1980, l'approche dominante pour résoudre le problème s'inscrivait dans une optique d'ingénierie des connaissances (« *knowledge engineering* »). On procédait à la construction d'un système dit expert comportant un ensemble de règles définies de façon manuelle par des experts humains. Ce système pouvait alors ensuite procéder à la classification automatique. Les règles évoquées ci-dessus prenaient généralement la forme d'implication logique, où l'antécédent portait sur la présence ou encore l'absence de certains mots, et où le conséquent désignait la catégorie d'appartenance du texte. L'inconvénient de cette approche est que l'édition des règles de décision peut s'avérer très longue. De plus, l'ajout de catégories ou encore l'utilisation du classificateur dans un domaine différent requiert la répétition de l'exercice. Ainsi, l'évolution constante de l'ensemble des règles avec le temps, agit à l'encontre de l'intérêt d'utilisation de cette méthode. Avec le développement de techniques d'apprentissage automatique («*machine learning*»), le problème est vu sous un autre angle. L'introduction en 1961 du classificateur bayésien a laissé sa marque. Il se distingue de l'approche d'édition de règles en se basant plutôt sur un calcul de probabilités. Cette approche n'a connu sa popularité qu'avec le début des années 1990 (Réhel 2005).

L'apprentissage automatique vise à construire des systèmes qui vont apprendre par eux-mêmes à classer les documents. Ainsi, l'accent est mis sur l'automatisation de la

création du classificateur par apprentissage. Le système est dirigé par un ensemble de textes préalablement associés à des catégories. Cette banque de textes libellés doit être préalablement construite par un humain. Ensuite, la machine tente d'apprendre la tâche de classification en observant le travail fait par l'humain. Elle essaie de généraliser les liens entre les textes et les catégories en analysant des exemples. Après la phase d'entraînement, le classificateur peut procéder lui-même au classement de nouveaux textes. Ainsi, l'objectif final est de créer un constructeur automatique de classificateurs. La figure ci-dessous présente le processus général d'entraînement d'un tel système (Réhel 2005) :

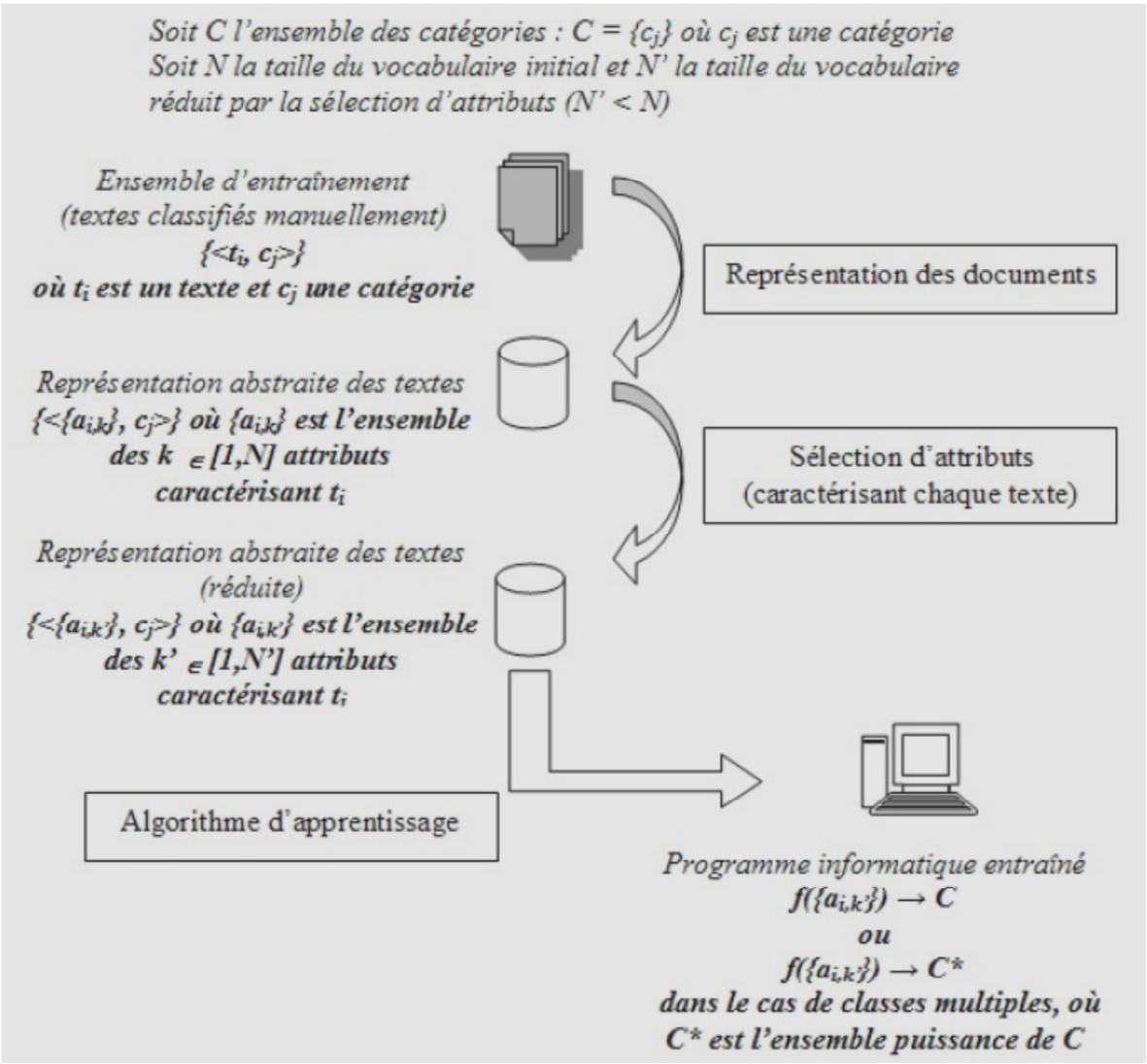


Figure 27. Entraînement d'un système de classification automatique de textes

La figure ci-dessous illustre le processus de classification d'un nouveau document (Réhel 2005) :

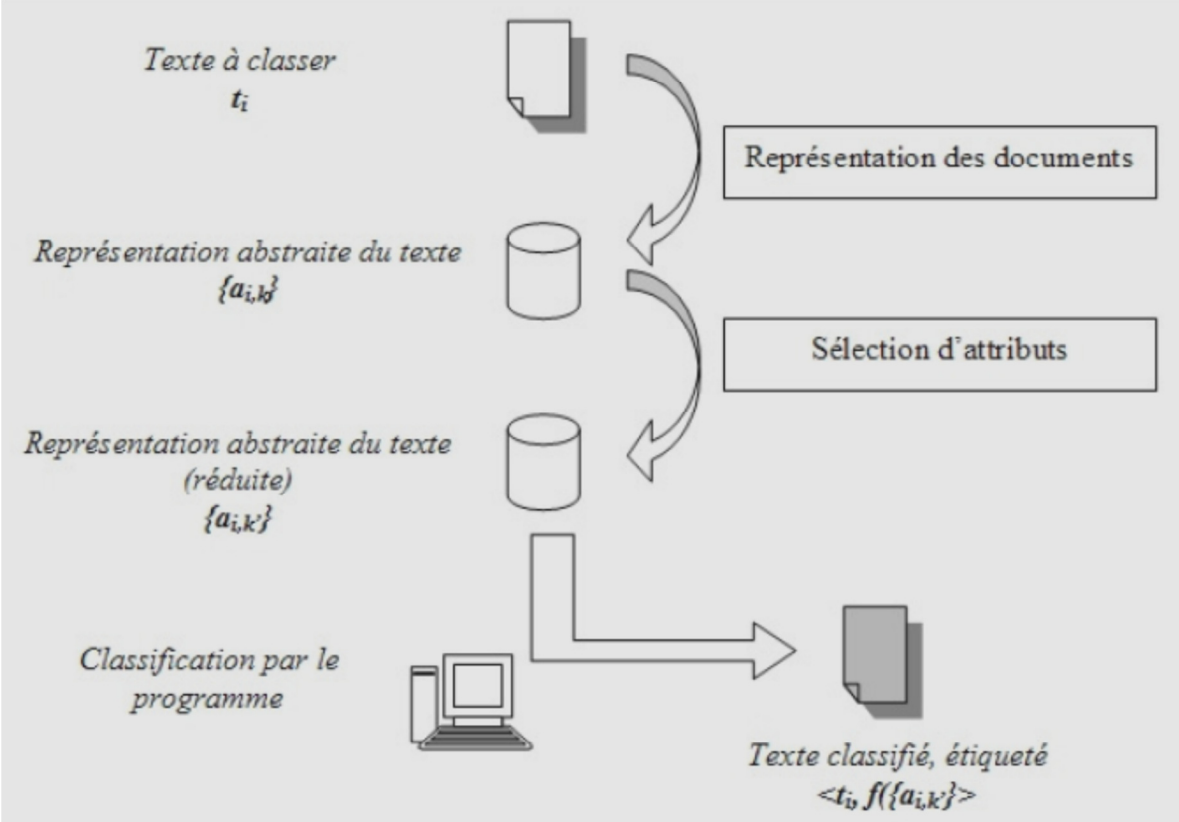


Figure 28. Classification d'un nouveau document

C'est principalement par apprentissage automatique que l'on tente de résoudre le problème de la classification. Dans cette optique, plusieurs algorithmes d'apprentissage ont été mis au point. À titre d'exemple, la section suivante introduit le principe des machines à support vectoriel.

5.1.2. Les machines à support vectoriel

Les machines à support vectoriel ou *SVM* (*Support Vector Machine*) forment une classe d'algorithmes d'apprentissage qui s'appliquent à tout problème qui implique un phénomène *f* qui produit une sortie *y=f(x)* à partir d'un ensemble d'entrées *x*. Le but étant

de trouver *f* à partir de l'observation d'un ensemble de couples entrée/sortie. Dans les cas de classification de textes, les entrées sont des documents et les sorties sont des catégories.

Le problème revient à trouver une frontière de décision, appelée *hyperplan*, qui arrive à séparer correctement les données. Cette frontière doit se placer le plus loin possible de tous les exemples en maximisant la *marge* (distance du point le plus proche de l'*hyperplan*). Ceci est réalisable grâce à des techniques de programmation quadratique.

Figure 29. Maximisation de la marge avec les SVM: (a) Une frontière possible séparant les données

(b) La frontière entraînant une marge maximale

Dans la figure 29, la ligne en trait continu de la partie (b) représente la frontière recherchée entre les données puisque la distance entre cette ligne et les lignes en trait pointillé est maximale. La *marge* de la partie (a) est plus petite. Les points situés sur les lignes en trait pointillé sont appelés les *vecteurs de support*.

L'avantage des *SVM* est qu'ils s'adaptent facilement aux problèmes non linéairement séparables. Avant de procéder à l'apprentissage de la meilleure séparation linéaire, les vecteurs d'entrée sont transformés en vecteurs de caractéristiques de dimension plus élevée. Ainsi, un séparateur linéaire trouvé par un *SVM* dans ce nouvel espace vectoriel devient un séparateur non linéaire dans l'espace original. La figure 30 donne un exemple de séparation non linéaire.

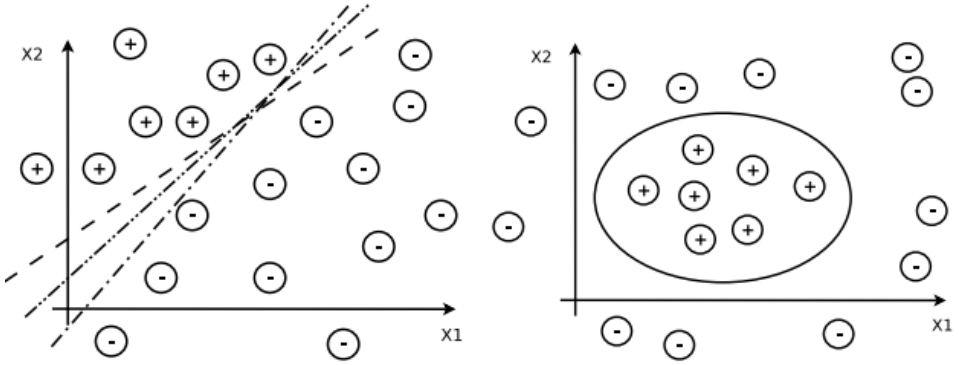


Figure 30. Exemple d'une séparation linéaire (à gauche) et non linéaire (à droite)

5.1.3. Critères d'évaluation des classificateurs

Normalement, il faut diviser la banque de textes déjà classés en deux ensembles : l'ensemble d'entraînement sur lequel le classificateur fait son apprentissage et l'ensemble de test sur lequel on peut évaluer sa performance. Ainsi, l'ensemble de test contient des documents dont on connaît à l'avance les catégories auxquelles ils devraient appartenir. Ceci nous permettra de comparer les décisions prises par le classificateur automatique avec celles des experts et d'obtenir une idée claire sur sa performance. Cette dernière dépend essentiellement des critères suivants :

La précision (*precision*)

C'est le nombre d'assignations correctement attribués à une classe sur le nombre total d'assignations attribués à cette même classe. Le principe est le suivant : quand un utilisateur interroge une base de données, il souhaite que les documents proposées en réponse à son interrogation correspondent à son attente. Tous les documents retournés superflus ou non pertinents constituent du bruit. La précision s'oppose à ce bruit documentaire. Si elle est élevée, cela signifie que peu de documents inutiles sont proposés par le système et que ce dernier peut être considéré comme "précis".

Le rappel (*recall*)

C'est le nombre d'assignations correctement attribués à une classe sur le nombre total d'assignations appartenant à cette classe (nombre d'assignations qui auraient dû être faites). Cela signifie que lorsque l'utilisateur interroge la base il souhaite voir apparaître tous les documents qui pourraient répondre à son besoin d'information. Si le rapport entre le questionnement de l'utilisateur et le nombre de documents présentés est important alors le taux de rappel est élevé. À l'inverse si le système possède de nombreux documents intéressants mais que ceux-ci n'apparaissent pas, on parle de silence. Le silence s'oppose au rappel.

La F-mesure (*F-measure*)

Lors de l'évaluation de la performance d'un classificateur, on ne peut tenir compte de la précision ou du rappel séparément. En effet, on pourrait mettre en place un système qui rejetterait tous les textes : il obtiendrait une précision de 100%, mais un rappel de 0%. À l'inverse, un système qui accepterait tous les textes aurait un rappel de 100%, mais une précision faible. D'où l'utilité de la F-mesure qui permet de combiner la précision et le rappel. Elle est définie ainsi :

$$F = \frac{2 \cdot (\text{précision} \cdot \text{rappel})}{(\text{précision} + \text{rappel})}$$

La F-mesure est optimale lorsque la précision et le rappel sont proches. Notons qu'il est possible de généraliser cette mesure en ajoutant un paramètre qui pondère l'importance relative des deux critères. Il se peut qu'une application particulière nécessite une précision élevée mais puisse se permettre un rappel un peu moins bon et vice versa. Ainsi, on obtient :

$$F_\beta = \frac{(1 + \beta^2) \cdot (\text{précision} \cdot \text{rappel})}{(\beta^2 \cdot \text{précision} + \text{rappel})}$$

Où β est une valeur réelle positive.

5.2. L'exploration d'images

Cette section présente deux algorithmes d'exploration d'images utilisés dans notre recherche.

5.2.1. ImgSeek

ImgSeek est un projet *Open Source* dédié à la comparaison d'images. Notre choix est tombé sur *ImgSeek* parce qu'on cherchait un algorithme qui doit être gratuit et disponible en version client/serveur, qui permet de donner rapidement des résultats pertinents et qui n'occupe pas un grand espace de stockage.

ImgSeek décompose les images en fonctions simples à l'aide des ondelettes de Haar[30] et de l'approche de la pyramide multi-résolution[31] (qui permet de modéliser l'image à différentes résolutions) pour leurs affecter des signatures (des caractéristiques distinguant les images). Il analyse aussi la distribution des couleurs dans les images en se basant sur le système de couleurs XYZ[32]. Ainsi, les images seront représentées par des signatures qui seront stockées dans les bases de données d'images et qui seront utilisées lors du processus de comparaison.

ImgSeek donne de bons résultats lorsqu'on cherche à vérifier si une image figure dans une base de données d'images. Cependant, il est sensible au changement de l'orientation de l'image, à la répartition de ses couleurs ainsi qu'au moindre changement de sa luminosité.

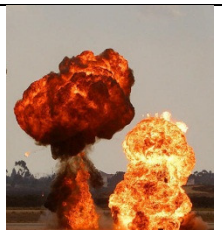
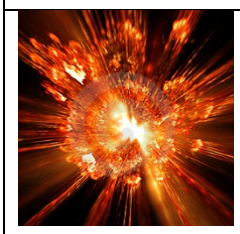
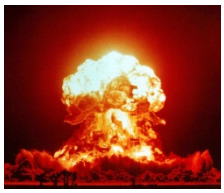
À titre d'exemple, le tableau suivant montre le niveau de ressemblance en pourcentage selon *ImgSeek* entre l'image « Explosion.jpg » et les images se trouvant en-dessous.

---

[30] http://fr.wikipedia.org/wiki/Ondelette_de_Haar
[31] http://fr.wikipedia.org/wiki/Pyramide_%28traitement_d%27image%29
[32] http://fr.wikipedia.org/wiki/CIE_XYZ

Tableau 1. Extrait de quelques résultats générés par *ImgSeek*

Explosion

				
23.9605	27.8064	23.2559	10.5073	18.5494
				
25.4187	8.42575	98.5642	20.1392	18.3514
				
9.50421	16.7224	19.8068	18.8584	29.4355
				
19.8068	14.2589	19.3725	19.5291	5.3096
				
25.8682	22.1803	14.762	9.09022	19.8068

5.2.2. Algorithme de détection de formes (baptisé *GetObject*)

L'algorithme qu'on a utilisé dans la détection des formes a été développé par Monsieur François Destrempes dans le cadre de son doctorat qui était sous la supervision de Monsieur Max Mignotte. MM. Mignotte et Destrempes ont eu la générosité et la gentillesse de nous donner le code pour pouvoir l'exécuter. Comme l'algorithme ne portait pas de nom, on l'a baptisé dans notre thèse sous le nom *GetObject*. Son fonctionnement détaillé est expliqué dans la thèse de Monsieur Destrempes (Destrempes 2006).

En gros, un algorithme d'optimisation stochastique[33] est appliqué pour détecter les contours d'une forme dans des images. Cette forme s'obtient grâce à une phase d'apprentissage durant laquelle l'utilisateur doit délimiter le contour d'un même objet dans une vingtaine d'images en considérant à chaque fois les mêmes positions clés (qui seront marquées par des points-clés). L'image ci-dessous illustre ce fait dans le cas d'une guitare :

Figure 31. Points clés du contour d'une guitare

L'ensemble des tracés va permettre à *GetObject* de générer un modèle standard qui servira de référence lors de la recherche de cet objet dans de nouvelles images.

[33] http://en.wikipedia.org/wiki/Stochastic_optimization

À titre d'exemple, Pour créer le modèle « guitare », il faut calquer le contour d'une vingtaine de guitares qui se trouvent dans des positions différentes et qui ont des orientations variées. L'ensemble des tracés obtenus va permettre à *GetObject* de trouver une forme approximative de la guitare qui peut prendre la forme ci-dessous :

Figure 32. Modèle d'une guitare

Ce modèle servira de base dans la recherche de la guitare dans d'autres images. *GetObject* va prendre en considération dans ses calculs plusieurs paramètres tels que la direction de l'objet, son angle de rotation, sa perspective, sa taille, etc.

Notons que l'efficacité de cet algorithme va grandement dépendre de la forme de l'objet, de sa taille et de sa visibilité (totalement visible ou partiellement visible).

Après avoir introduit dans ce chapitre les techniques et les algorithmes que notre plateforme a utilisés, la première partie du chapitre suivant expose les principaux travaux existants qui sont en lien avec la protection des utilisateurs des réseaux sociaux alors que la deuxième partie se charge de présenter notre façon de protéger ces utilisateurs grâce à nos deux systèmes *Protect_U* et *Protect_UFF*.

Chapitre 6 : Principales études existantes en lien avec notre sujet vs notre approche

Ce chapitre introduit les principaux travaux et systèmes qui ont déjà abordé le sujet de la protection de la vie privée des utilisateurs des réseaux sociaux. La deuxième partie est consacrée à la présentation de notre approche. Elle présente l'architecture globale de notre plateforme et les principaux modules constituants les systèmes *Protect_U* et *Protect_UFF*.

6.1. Principales études existantes

Comme indiqué précédemment, les réseaux sociaux et l'échange de grandes quantités de renseignements personnels et privés sur Internet connaissent une importante expansion. Cette popularité des réseaux sociaux peut représenter une menace importante pour les utilisateurs ((Fogel et Nehmad 1999), (Bonneau et Preibusch 2009)). Elle a conduit à de nombreuses études liées aux différents aspects de la sécurité et de la protection des renseignements personnels et privés des utilisateurs des réseaux sociaux ((Aimeur, Gambs et Ai 2010), (Aimeur, Brassard, et al. 2008), (Yan et Ahmad 2008)). Différents groupes de recherche ont analysé les risques d'atteinte à la vie privée en mettant l'accent sur l'analyse des types de fuites d'informations suite à une attaque ou à une simple divulgation ((Korolova, et al. 2008), (Bin et Jian 2008)).

L'idée de corréler des données provenant de différentes sources pour détecter un profil utilisateur a été étudiée dans différents contextes. Certaines recherches ont montré qu'il est possible de corréler des informations des profils publics pour trouver le nom de jeune fille d'une femme mariée (Griffith et Jakobsson 2005). De même, il a été prouvé que des informations cachées sur le profil d'un utilisateur peuvent également être déduites à l'aide d'informations contextuelles (par exemple, l'appartenance politique d'un utilisateur peut être prédite par l'examen de l'affiliation politique d'amis) (Zheleva et Getoor 2009). D'un autre côté, il est possible de révéler l'identité d'une personne en mettant en corrélation ses profils sur différents réseaux sociaux grâce à son son pseudonyme, son nom réel ou son adresse de courriel ((Irani, et al. 2009), (Balduzzi, et al. 2010)). Les auteurs de l'article *De-*

anonymizing Social Networks (Narayanan et Shmatikov 2009) ont montré que l'analyse des groupes de *Facebook* peut révéler beaucoup d'informations sur les participants.

D'un autre côté, l'approche de *Privacy Feedback and Awarness (PFA)* est appliquée par de nombreuses études afin d'assurer la protection des utilisateurs des réseaux sociaux (Lederer, et al. 2004). Cette approche permet d'attirer l'attention des utilisateurs aux dangers auxquels ils peuvent faire face en leur proposant des recommandations qui les poussent à être plus vigilants. Les utilisateurs divulguent des informations privées sur les réseaux sociaux parce qu'ils ne sont pas conscients des répercussions. La sensibilisation pousse les utilisateurs à prendre des décisions éclairées et les incite à mieux gérer leurs paramètres de confidentialité. Les systèmes de recommandation sont un bon moyen pour sensibiliser les usagers aux dangers imminents.

Dans cette perspective, L. Fang et K. LeFevre présentent l'assistant de confidentialité *Privacy Wizard* (Fang et LeFevre 2010). Il a pour objectif de configurer automatiquement les paramètres de confidentialité de l'utilisateur en exigeant de lui un minimum d'efforts. Pour cela, l'utilisateur commence par configurer manuellement un sous-ensemble simplifié des paramètres de confidentialité avant que *Privacy Wizard* prenne la relève et configure automatiquement le reste des paramètres.

Le système *PViz* a été introduit par A. Mazzia, K. LeFevre et E. Adar. Il permet de représenter graphiquement les groupes d'amis ainsi que les différents niveaux d'interactions entre eux. Ceci aide l'utilisateur à avoir une vue plus globale de son profil et à détecter plus facilement les failles de confidentialité (Mazzia, LeFevre et Adar 2011). La figure 33 donne un aperçu de l'interface de *PViz*. Chaque nœud (ou cercle) représente un sous-groupe d'amis ou tout simplement un ami isolé. Les distances entre ces nœuds reflète le niveau de liens qui les lient.

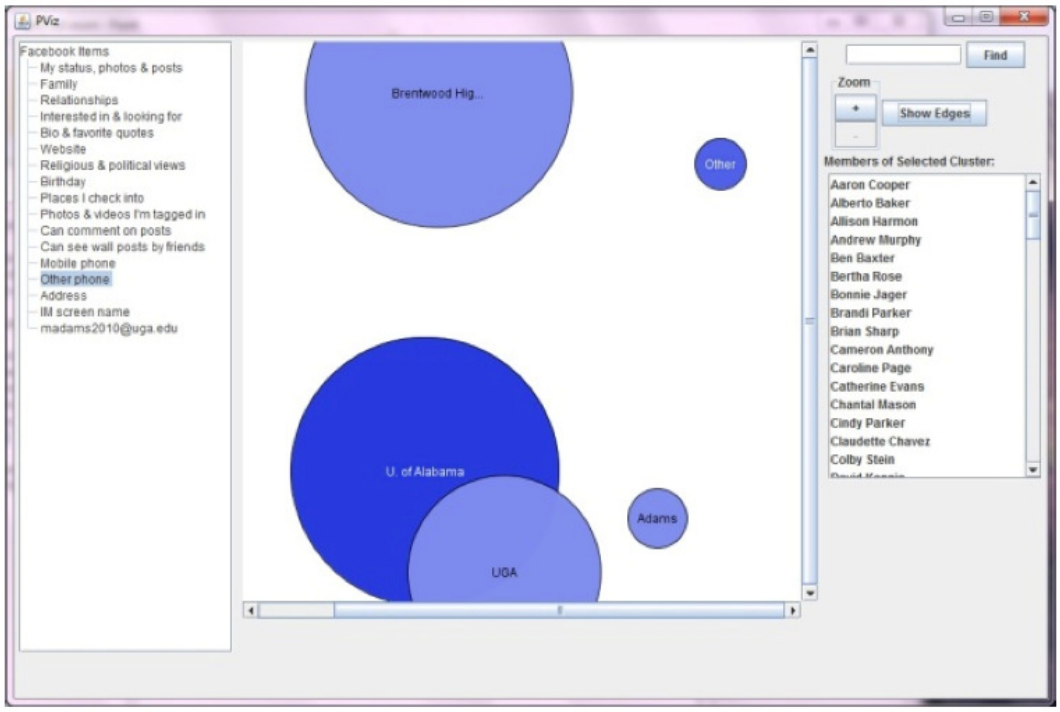


Figure 33. Interface du système *PViz*

S. Patil et A. Kobsa présentent le système *PRISM* (*PRIvacy Sensitive Messaging*). Il fournit aux utilisateurs de messagerie Internet différents niveaux de communication. Il leur donne la possibilité de choisir le niveau de communication qu'ils souhaitent établir avec leurs amis en fonction du lien qui les unit. *PRISM* rend aussi transparent les actions des autres utilisateurs permettant de faciliter les prises de décisions. Il contient également un ensemble de fonctions de gestion des renseignements personnels (Patil et Kobsa 2010).

Le système *Audience View Interface* a été introduit par H. R. Lipford, A. Besmer et J. Watson. Il permet à un utilisateur de visualiser comment son profil apparaît à un ami ou à un groupe d'amis et d'ajuster par conséquent les paramètres de confidentialités de son profil. L'aspect visuel de la présentation des paramètres favorise une meilleure compréhension de l'importance du réglage des paramètres de sécurité (Lipford, Besmer et Watson 2008). La figure 34 donne un aperçu de l'interface *Audience View*.

Figure 34. Interface du système *Audience View*

Facebook applique actuellement cette approche permettant à l'utilisateur du réseau de percevoir la manière dont les amis voient son profil.

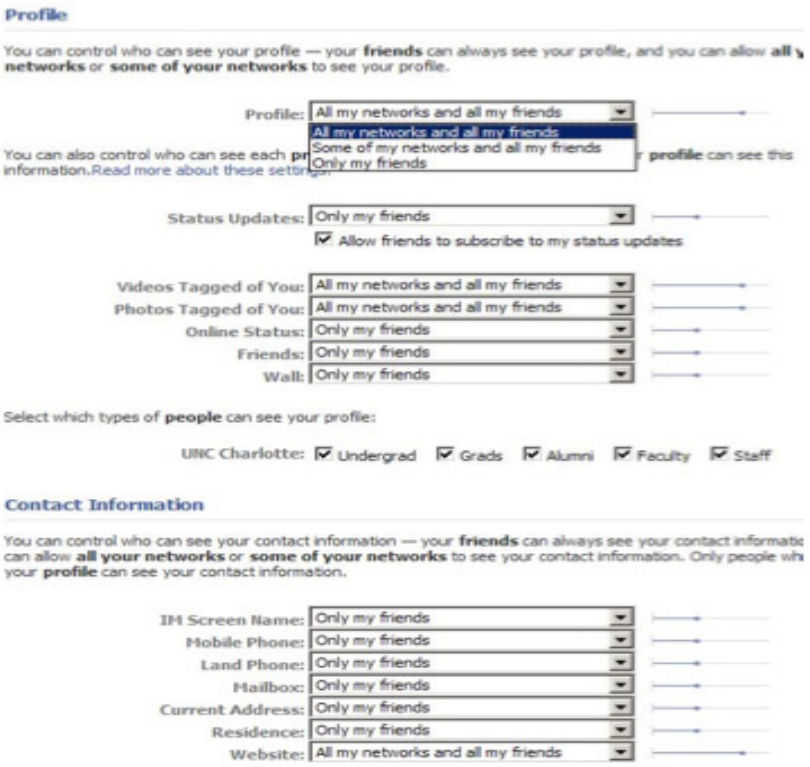


Figure 35. Paramètres de confidentialité d'un profil sur *Facebook*

Les auteurs de l'article *Expandable grids for visualizing and authoring computer security policies* ont proposé l'utilisation d'interfaces extensibles (R. Reeder, et al. 2008).

Leur système permet l'ajustement des paramètres de contrôle d'accès à un profil grâce à une grille extensible riche en couleurs (voir figure 36).

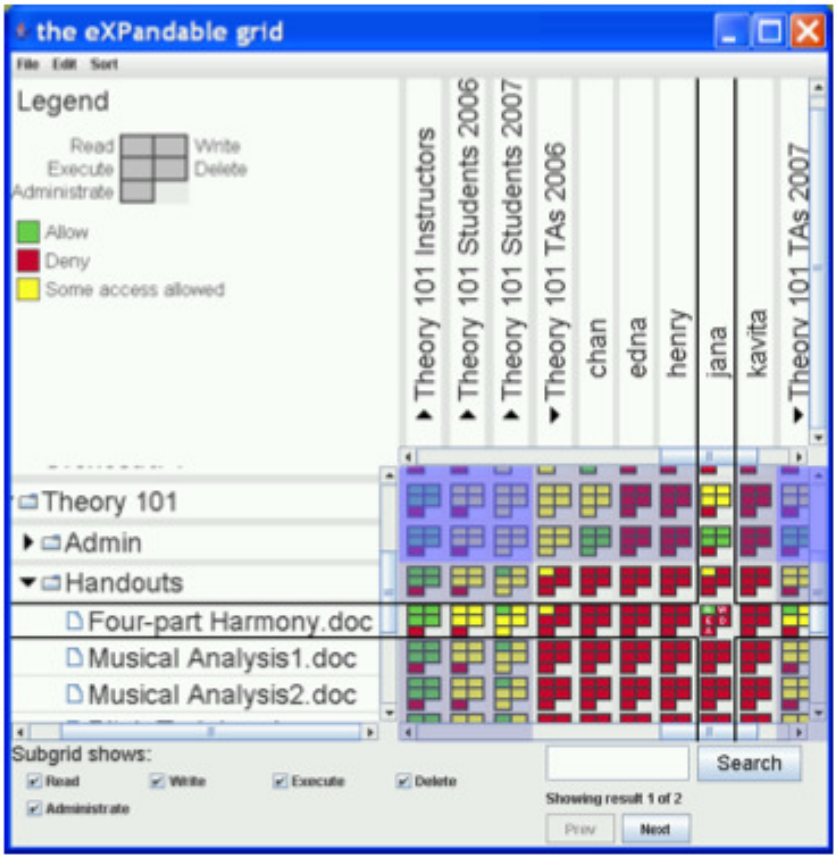


Figure 36. Interface de l'*eXPandable grid*

D'autres chercheurs ont proposé de nouvelles interfaces pour régler les paramètres de confidentialité des profils sur les réseaux sociaux en divisant les amis en « communautés » ((Adu-Oppong, et al. 2008), (Daneziz 2009)).

Suivant la même veine, certaines études ((Liu et Terzi 2010), (Maximilien, et al. 2009)) ont proposé une approche pour quantifier le risque d'un profil en fonction des paramètres de confidentialité. Ainsi, les niveaux de risque des profils pourront être plus facilement comparés. Cependant, cette approche n'aide pas l'utilisateur à consolider le réglage des paramètres de son profil.

D'un autre côté, les auteurs de l'article *A Privacy Preservation Model for Facebook-Syle Social Network Systems* proposent un modèle de protection de la vie privée sur *Facebook* en imposant un mécanisme de contrôle d'accès et de préservation de confidentialité. Ce modèle est implémenté avec le langage Prolog[34] de sorte que le code exécutable du modèle puisse être facilement exécuté sur plusieurs plateformes (Fong, Anwar et Zhao 2010).

Certaines études ont essayé d'exploiter les données qui transitent à travers les réseaux sociaux pour détecter les contenus Web malveillants. L'application *proof-of-concept* (Robertson 2010) s'inscrit dans ce cadre.

La section suivante présente l'approche qu'on propose pour empêcher les utilisateurs de *Facebook* de trop s'exposer aux différents dangers qui les guettent notamment en les invitant à afficher le minimum possible d'informations personnelles sur leurs comptes et en les avisant lorsque le profil de l'un de leurs « amis » présente des symptômes de délinquance.

6.2. Notre approche

Après avoir exposé les principales études qui ont abordées le sujet de la protection des utilisateurs des réseaux sociaux, nous présentons dans cette partie l'approche que nous avons retenue pour protéger les utilisateurs de *Facebook* de deux dangers majeurs : la surexposition des données personnelles et la manifestation de confiance hâtive vis-à-vis des « amis » à risque (psychopathe, fraudeur ou même criminel). Nous avons choisi *Facebook* parce qu'il est actuellement le réseau social le plus populaire comparé à d'autres réseaux sociaux.

---

[34] « Prolog est un langage de programmation utilisé pour explorer les relations entre un ensemble d'objets. Un programme Prolog est généralement exprimé comme un ensemble de relations, et son exécution est simulée en augmentant les requêtes sur les objets et les relations qui sont déduites de ces relations. Compte tenu du fait que le Prolog est à la fois déclaratif et exécutable, il est adapté à la réalisation de l'objectif » (Zhen 2010).

Nous avons créé une plateforme groupant deux systèmes indépendants : *Protect_U* (Hélou, Gandouz et Aïmeur 2012) et *Protect_UFF* (*Protect_U From Friends*). Son architecture générale est représentée par la figure suivante :

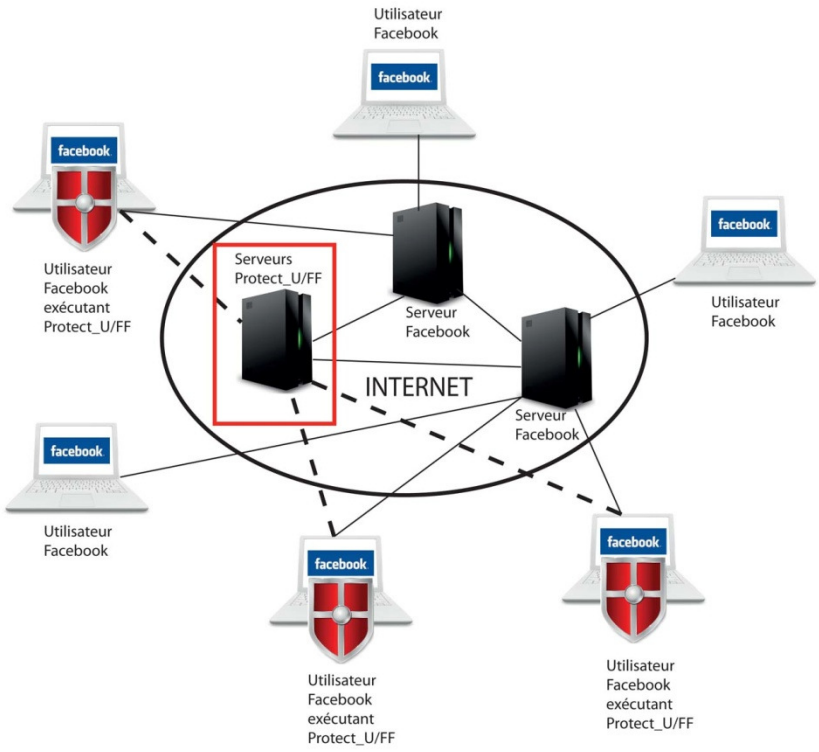


Figure 37. Architecture générale de la plateforme

Ils ont été développés avec les langages *PHP* et *Javascript* en utilisant l'*API* (*Application Programming Interface*) et la plateforme *SDK 3.0* (*Software Development Kit*) que *Facebook* met à la disposition des développeurs. Cette dernière permet l'interaction entre les utilisateurs et le serveur *Facebook*. Nos deux systèmes s'exécutent sur un serveur *PHP* qui est connecté à un serveur « *.NET* ».

Au niveau local, un utilisateur *Facebook* peut exécuter *Protect_U* indépendamment de *Protect_UFF*. Le premier est constitué de deux modules : le *module de classification* et le *module de recommandation*. Le second contient deux modules aussi : le *module d'analyse* et le *module de décision*. Le fonctionnement et l'architecture de ces modules

seront expliqués en détails dans les chapitres suivants. La figure ci-dessous met en valeur ces modules constituants nos deux systèmes :

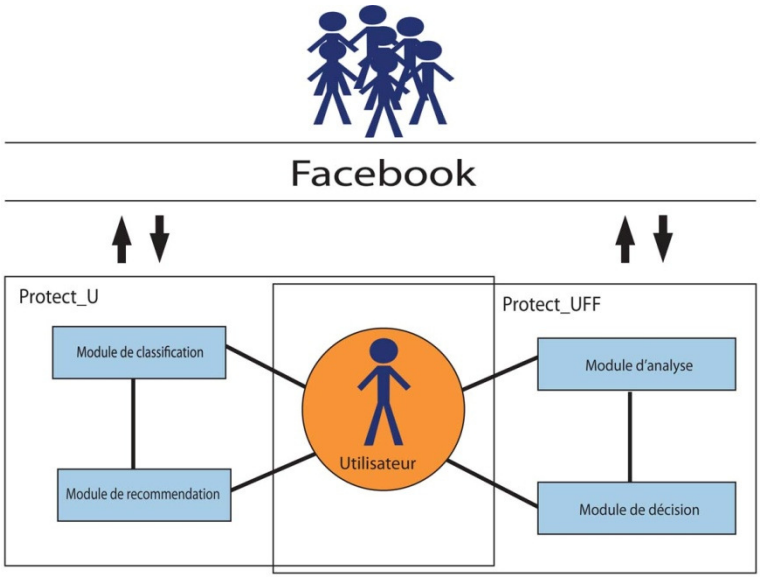


Figure 38. Modules constituant *Protect_U* et *Protect_UFF*

Protect_U a comme mission, dans un premier temps, d'analyser le contenu des comptes des utilisateurs *Facebook* pour déterminer l'étendue de l'exposition de leurs données personnelles. Leurs profils seront classés selon quatre niveaux de risque différents: *peu risqué*, *moyennement risqué*, *risqué* ou *critique*. En se basant sur le niveau de risque des profils, *Protect_U* propose aux différents utilisateurs des recommandations personnalisées pour les inciter à rendre leurs comptes plus sécuritaires. Dans un deuxième temps et dans le but de raffiner les recommandations proposées, *Protect_U* applique un filtre communautaire en faisant appel au cercle d'amis « de confiance » des utilisateurs. Il les invite à évaluer le contenu des comptes de leurs amis en leur demandant de répondre à une liste de questions et en leur donnant la possibilité de proposer leurs propres recommandations. La figure suivante donne un aperçu de l'interface de *Protect_U* :

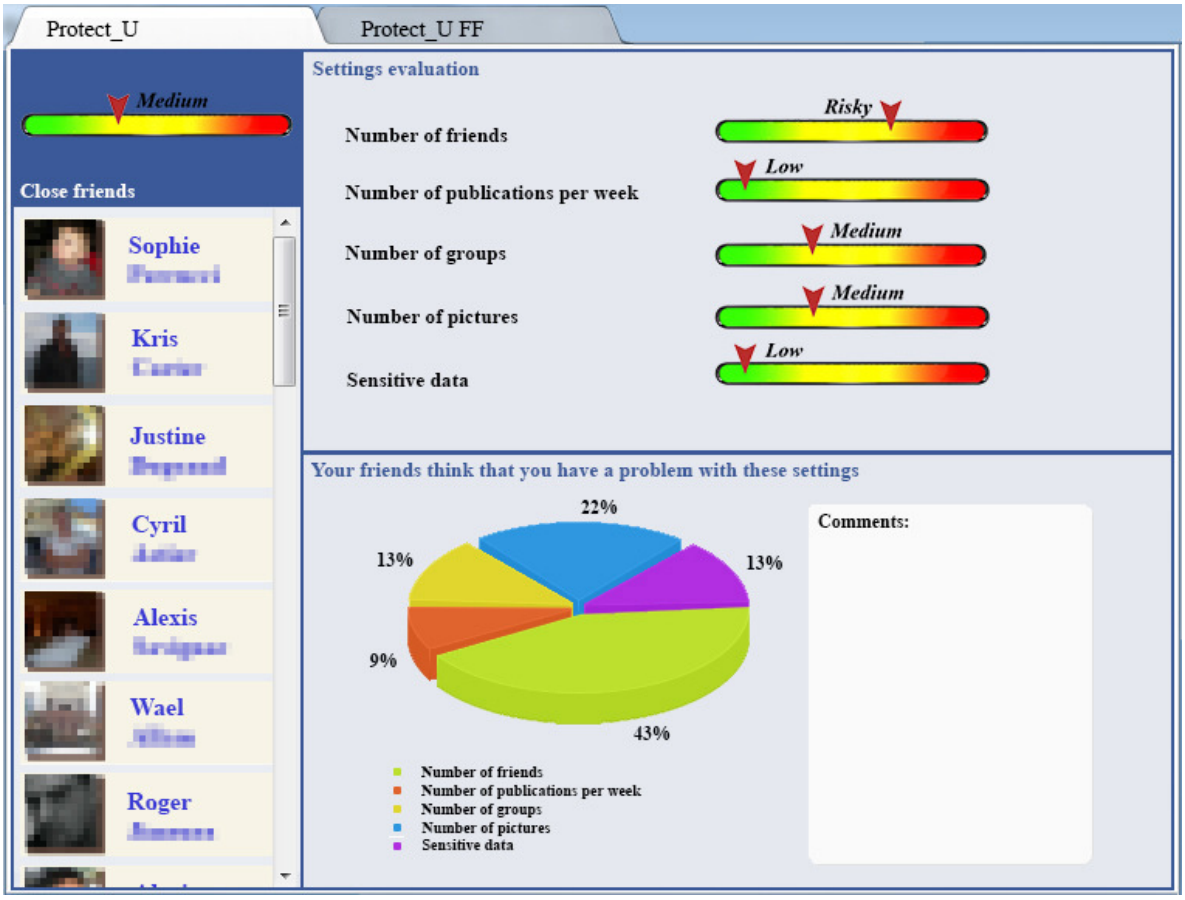


Figure 39. Interface de *Protect_U*

Protect_UFF a comme but de « démasquer » les profils risqués des « amis ». On a vu au chapitre 4 qu'il existe, selon les psychologues, plusieurs points communs entre les traits de personnalité des psychopathes, des fraudeurs et des criminels notamment le côté narcissique, le manque d'empathie et d'émotions, et le comportement antisocial qui se manifeste très souvent par de l'agressivité. C'est la raison pour laquelle nous avons considéré ces trois principaux facteurs comme éléments déterminants dans la détection des profils dangereux. Pour cela, *Protect_UFF* analyse les textes et les commentaires affichés ainsi que les images publiques des « amis » à la recherche de mots, d'expressions et de contenu compromettant afin de mesurer les niveaux de narcissisme, d'émotions et d'agressivité des profils. Dépendamment des résultats obtenus, *Protect_UFF* classe les

profils selon trois niveaux de risque : *faible*, *moyen* ou *élevé*. La figure ci-dessous donne un aperçu de l'interface de *Protect_UFF* :

Figure 40. Interface de *Protect_UFF*

Les chapitres suivants expliquent en détail les architectures, le fonctionnement, la conception et la méthodologie de *Protect_U* et *Protect_UFF*.

Chapitre 7 : Conception et méthodologie du système *Protect_U*

Ce chapitre explique l'approche adoptée avec le système *Protect_U* (Hélou, Gandouz et Aïmeur 2012). Il présente l'architecture générale du système ainsi que le fonctionnement des modules de *classification* et de *recommandation* qui le constitue. Il introduit ensuite les résultats de la validation du système auquel 163 utilisateurs différents de *Facebook* ont participé.

7.1. Choix des paramètres clés

Facebook procure à ses utilisateurs l'opportunité de publier de très grandes quantités d'informations. Le *mur*, par exemple, contient l'ensemble des publications de l'utilisateur ainsi que l'ensemble des activités de ses amis. Ces publications peuvent prendre différentes formes : textes, photos, vidéos, etc. De même, la rubrique *Informations* garde des informations sensibles sur l'utilisateur telles que son sexe, sa date et son lieu de naissance, sa formation, son emploi actuel, ses citations favorites ainsi que ses films, livres et artistes préférés. Il est aussi possible, dans la section *Photos*, d'accéder à toutes les photos et tous les albums que l'utilisateur a publiés, ainsi que toutes les photos dans lesquelles il est « tagué ». Si l'utilisateur ne protège pas convenablement ces informations en ajustant les paramètres de sécurité de son profil, elles pourront être facilement accessibles à tous ses amis, et amis des amis.

Protect_U analyse les informations clés suivantes : *l'âge, le sexe, le nombre d'amis, le nombre de publications par semaine, le nombre de groupes auxquels il est inscrit, le nombre total d'images, le pourcentage d'images privées dans son compte,* de même que certaines données sensibles telles que *la religion et la position politique*. Ces paramètres, groupés ensemble, constituent un bon indice de mesure du niveau d'exposition des profils des utilisateurs (Ninggal et Abawajy 2011).

7.2. Architecture générale

L'architecture générale de *Protect_U* est constituée de deux modules : le *module de classification* et le *module de recommandation*. Elle est représentée par la figure ci-dessous :

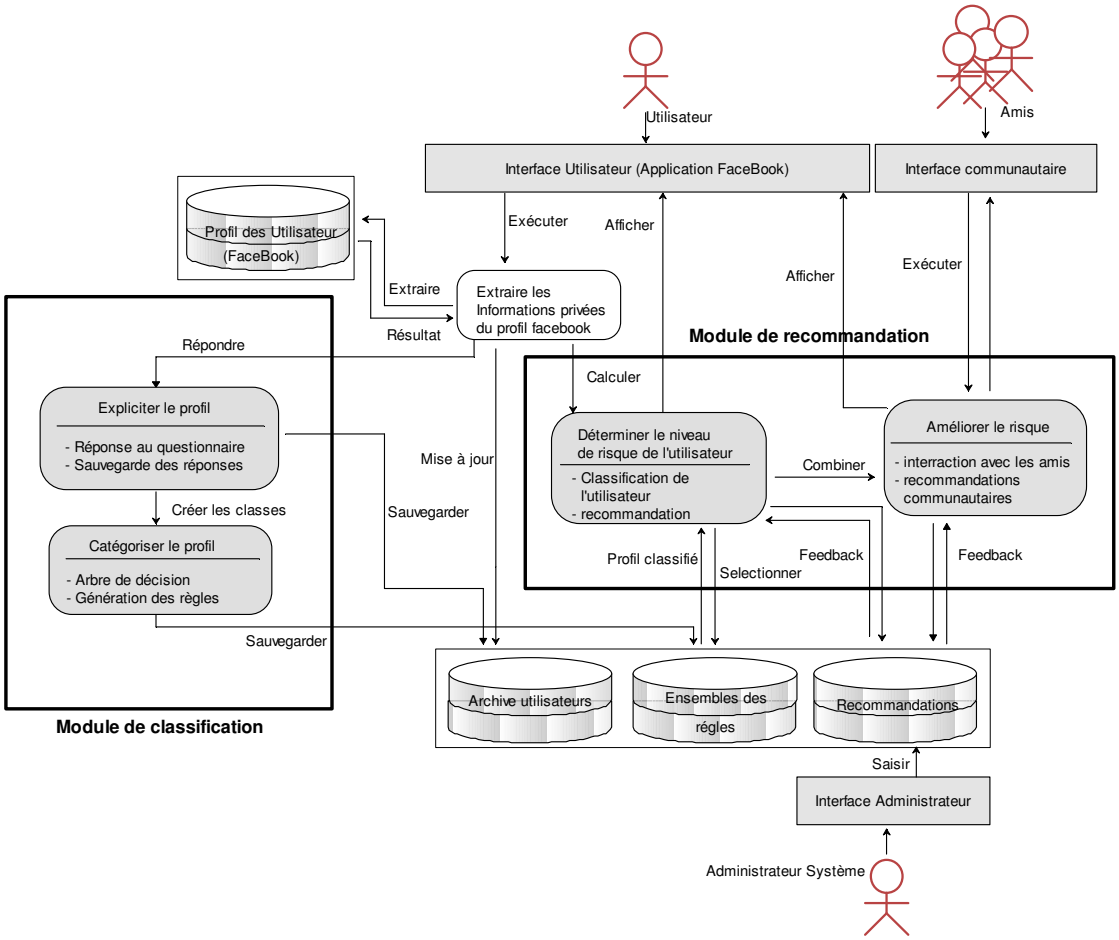


Figure 41. Architecture générale de *Protect_U*

7.2.1. Module de classification

Le *module de classification* détermine les règles de classification à partir desquelles les profils des utilisateurs seront classés selon quatre niveaux : *peu risquée, moyennement risquée, risquée* et *critique*. Ce module comporte essentiellement deux étapes.

L'étape *« Expliciter le profil »* invite les participants à répondre à un questionnaire comportant une série de 18 questions (voir annexe 2) permettant de déterminer le niveau de risque auquel ils ont déjà fait face dans le passé, ainsi que leur façon d'afficher leurs informations privées sur leurs comptes. Ces questions sont divisées en trois groupes :

- Les questions du premier groupe cherchent à déterminer si des informations personnelles délicates sont publiées sur le compte de l'utilisateur. La divulgation de la date de naissance par exemple tombe dans cette catégorie.
- Les questions du deuxième groupe visent à déterminer si l'utilisateur a déjà connu des commentaires agressifs et indécents en réplique à des publications sur son compte.
- Les questions du troisième groupe ciblent à repérer les cas critiques auxquels un utilisateur a dû faire face tels que le vol d'identité, la fraude élaborée ou l'atteinte à sa personne (physique ou morale).

Suite aux réponses recueillies, la classification des profils des participants est effectuée selon le principe suivant :

- Le profil *peu risqué* (*low risk*) est attribué aux participants ayant répondu « non » à l'ensemble des questions.
- Le profil *moyennement risqué* (*medium risk*) est attribué aux participants ayant donné la réponse « oui » à au moins une des questions du premier groupe.
- Le profil *risqué* (*risky*) est attribué aux participants ayant donné la réponse « oui » à au moins une des questions du deuxième groupe.
- Le profil *critique* (*critical*) est attribué aux participants ayant donné la réponse « oui » à au moins une des questions du troisième groupe.

Si un utilisateur répond « oui » à des questions appartenant à plusieurs groupes, on affectera à son profil le plus élevé niveau de risque marqué.

L'étape « *Catégoriser le profil* » fait appel à la technique des arbres de décisions (Kantardzic 2011) pour extraire les règles de classification. Le grand avantage des arbres de décision est la grande lisibilité de leur forme arborescente. Généralement, *ID3* et *C4.5* sont les algorithmes les plus utilisés dans les arbres de décision (Soman et Diwakar 2006). *C4.5* a été retenu vu ses nombreux avantages face à *ID3* notamment lors du traitement simultané des attributs continus et discrets, ainsi que les données avec attributs manquants.

131 utilisateurs de différents âges et emplacements géographiques ont participé à la phase d'entraînement de ce module. Les règles de classification obtenues suite à l'application de l'algorithme *C4.5* sont représentées ci-dessous :

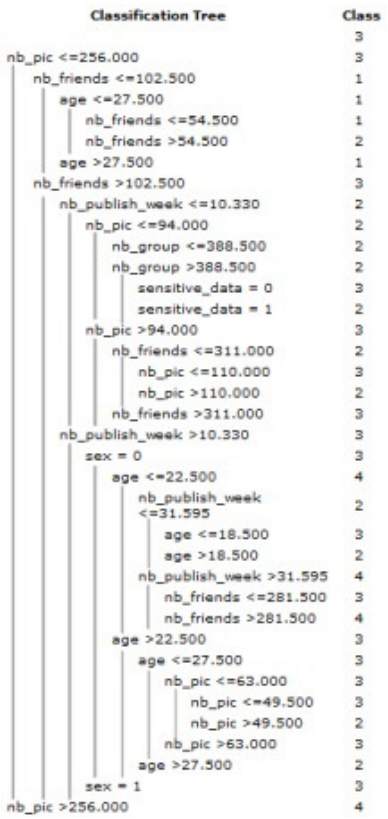
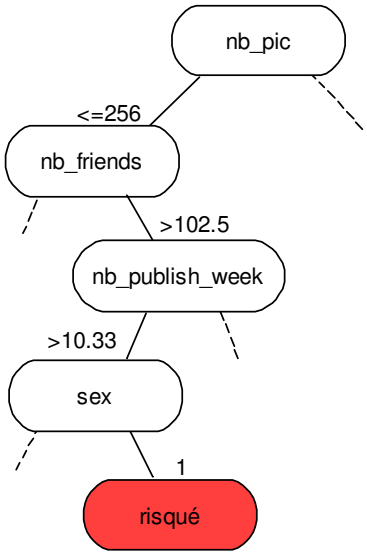


Figure 42. Règles de classification

À titre d'exemple, si pour un profil donné, le nombre des images postées est inférieur ou égal à 256, le nombre d'amis acceptés est supérieur strictement à 102.5, le nombre de publications par semaine est strictement supérieur à 10.33 et le sexe est

masculin alors le profil est risqué. Cette règle est bien représentée dans l'extrait ci-dessous de l'arbre de décision :



SI nb_pic <=256.000 ET nb_friends >102.500 ET
nb_publish_week >10.330 ET sex = 1
ALORS level_risk = 3

Figure 43. Exemple de règle de classification

La méthode « *leave-one-out* » de *validation croisée* (*Cross-validation*) a été appliquée lors de la création du modèle de classification. Cette méthode est la mieux adaptée pour les cas où le nombre des instances dans le fichier d'entraînement est relativement petit (ce qui est notre cas). C'est un cas particulier de la méthode *K-fold cross-validation* qui divise aléatoirement les instances en k sous-ensembles. Elle utilise les k-1 premiers sous-ensembles pour construire le modèle de classification et le valide avec le sous-ensemble restant. La méthode « *leave-one-out* » applique le même principe en prenant k égal au nombre des instances dans l'échantillon (Bramer 2007).

La *matrice de confusion* du tableau 2 donne une idée sur la qualité du modèle de classification obtenu. Elle révèle que 79,3% des profils critiques, 61,7% des profils risqués, 58,5% des profils moyennement risqués et 85,7% des profils peu risqués seraient correctement détectés si le modèle est appliqué à une nouvelle liste d'instances.

Tableau 2. Matrice de confusion

Classes actuelles	Classes prédites: Peu risqué	Moyennement risqué	Risqué	Critique
Peu risqué	85.7%	14.3%	0.0%	0.0%
Moyennement risqué	7.3%	58.5%	31.7%	2.4%
Risqué	0.0%	27.7%	61.7%	10.6%
Critique	0.0%	6.9%	13.8%	79.3%

Le tableau ci-dessous montre que l'estimation moyenne du niveau de certitude de notre modèle de classification est de 67.18%. Il arrive à identifier les résultats positifs à 85,71% des fois, et les résultats négatifs à 97,44% des fois. La probabilité d'obtenir les mêmes résultats dans des conditions similaires est de 80%.

Tableau 3. Évaluation de la précision

	Niveau de certitude	Sensitivité	Spécificité	Précision
C4.5	0.6718	0.8571	0.9744	0.8000

7.2.2. Module de recommandation

Le *module de recommandation* a pour rôle principal de proposer des recommandations adaptées aux différents profils des utilisateurs dans le but de les inciter à mieux protéger leurs comptes. L'annexe 3 contient un exemple de recommandations possibles. La structure de ce module est illustrée par la figure suivante :

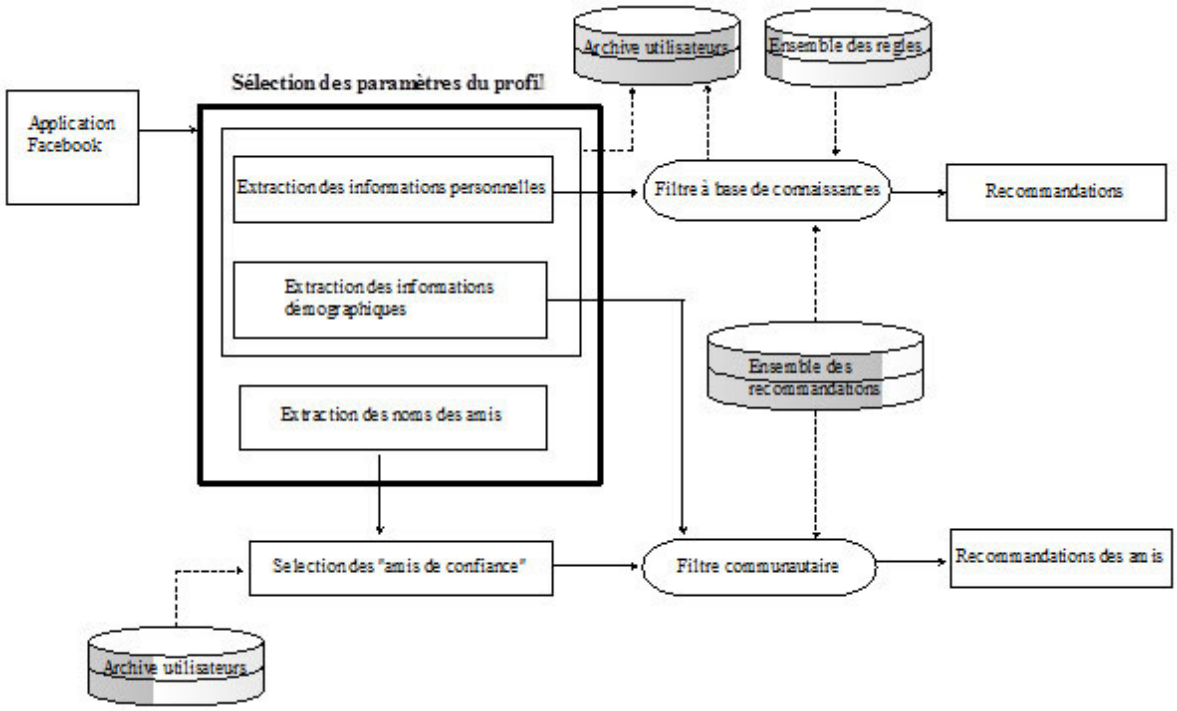


Figure 44. Structure du module de recommandation

Le module de recommandation tire profit des données personnelles récoltées et des règles trouvées au niveau du module de classification pour leur appliquer deux filtres : le *Filtre à base de connaissances* et le *Filtre communautaire*.

Le *Filtre à base de connaissances* a comme mission de trouver la règle qui correspond le plus à un profil donné, afin de trouver les recommandations appropriées. Pour cela, il applique la *fonction de similitude vectorielle*. Pour un profil A et une règle B, la fonction de similitude est donnée par la relation suivante:

$$cosinus\ (A,B) = \sum_{i=1}^{N} \frac{v_{A,i}}{\sqrt{\sum_{i=1}^{N} v_{A,i}^2}} * \frac{v_{B,i}}{\sqrt{\sum_{i=1}^{N} v_{B,i}^2}} \qquad \text{(Équation 1)}$$

N étant le nombre de paramètres considérés, $v_{A,i}$ la valeur du $i^{ème}$ paramètre de l'utilisateur A et $v_{B,i}$ la valeur du $i^{ème}$ paramètre de la règle B. La règle qui donne le plus petit angle du cos (A, B) sera affectée au profil A.

Supposons à titre d'exemple qu'on cherche à trouver la règle qui correspond le plus au profil d'un utilisateur A ayant les paramètres suivants :

$v_{A,1}$ = Âge = 15, $v_{A,2}$ = Nombre d'images publiées = 300, $v_{A,1}$ = Nombre d'amis = 1200.

Supposons aussi qu'on dispose des trois règles suivantes :

Règle B1 : si $v_{B1,1} < 18$, $v_{B1,2} > 100$ et $v_{B1,3} > 50$ alors le profil est *risqué*

Règle B2 : si $v_{B2,1} > 40$, $v_{B2,2} < 20$ et $v_{B2,3} < 10$ alors le profil est *peu risqué*

Règle B3 : si $v_{B3,1} < 25$, $v_{B3,2} > 30$ et $v_{B3,3} > 100$ alors le profil est *risqué*

Le profil de l'utilisateur vérifie les règles B1 et B3. Le cosinus le plus élevé (ou l'angle le plus petit) détermine laquelle de ces deux règles correspond le plus à son profil. Sachant que cos (A, B1) = 0.644 et cos (A, B3) = 0.97 on peut déduire que la règle B3 correspond bien au profil de l'utilisateur A.

À partir de ce moment, le *module de recommandations* va proposer à l'utilisateur les recommandations qui correspondent bien avec la règle B3. Ainsi, il va lui recommander de réduire le nombre d'images publiées et le nombre des amis.

Le *Filtre communautaire* vise à améliorer la qualité des recommandations proposées et ce en faisant appel au réseau d'amis de l'utilisateur. Il favorise par conséquent les *amis de confiance*, soit ceux qui connaissent le mieux l'utilisateur, en leur permettant de participer à l'évaluation de son profil en répondant à un petit questionnaire (voir annexe 4). Par définition, et selon un sondage effectué auprès des 131 participants de la phase d'apprentissage, un *ami de confiance* est un ami qui fait partie d'au moins une des catégories suivantes :

- Un membre de la famille.
- Un ami proche sélectionné par l'utilisateur lui-même.
- Une personne avec qui l'utilisateur communique le plus souvent.

- Une personne « taguée » dans les images affichées dans le profil de l'utilisateur.

Pour être en mesure de sélectionner les *amis de confiance*, on affecte à chaque ami un poids, noté *Poids*, en appliquant la relation suivante :

$$\text{Poids} \quad = C4 \quad \sum_{i=1}^{3} \alpha_i \; P(C_i) \qquad \text{(Équation 2)}$$

Poids dépend des paramètres suivants : *le lien de connaissance avec l'utilisateur* (C_1), *le nombre de messages échangés avec l'utilisateur* (C_2), *le nombre de fois que l'ami est tagué dans le compte de l'utilisateur* (C_3) et *le niveau de risque du profil de l'ami* (C_4).

Si le niveau de risque du profil de l'ami est connu, C_4 prendra la valeur 2 pour un profil *peu risqué*, 1.5 pour un profil *moyennement risqué*, 0.5 pour un profil *risqué* ou 0 pour un profil *critique*. La valeur zéro veut dire que dans le cas d'un profil *critique*, l'ami en question sera écarté de la liste des amis de confiance. Dans le cas où le système n'est pas en mesure de déterminer le niveau de risque du profil, il affectera 1 à C_4 . Ce qui veut dire que ce paramètre ne sera plus pris en considération. Ainsi, C_4 appartient à l'intervalle {0, 0.5, 1, 1.5, 2}.

Afin de donner un plus grand poids aux paramètres les plus importants, nous avons affecté la valeur 3 au *lien de connaissance*, la valeur 2 au *nombre de messages échangés* et 1 *au nombre de fois qu'il est tagué*. Avec ces valeurs, la priorité a été donnée aux usagers qui font partie des membres de la famille ou de la liste des amis proches. D'où les valeurs suivantes:

$$P(C_i) = \begin{cases} 3 \; si \; i = 1 \\ 2 \; si \; i = 2 \\ 1 \; si \; i = 3 \end{cases}$$

Une pondération α_i a été affectée à chaque paramètre C_i (i varie entre 1 et 3) afin d'ajuster la pondération générale. Ainsi, α_1 prend la valeur 2 si l'ami est un membre de la famille, 1 s'il est un ami proche et zéro dans les autres cas. Tandis que α_2 est la partie

entière de $\frac{3*C_2}{MAX}$ (Max étant le nombre maximal de messages échangés par tous les amis) et α_3 est la partie entière de $\frac{3*C_3}{TAG}$ (TAG étant le nombre maximal d'images dans lesquelles l'utilisateur est tagué). Ainsi, $\alpha_1 \in \{0, 1, 2\}$, $\alpha_2 \in$ à $\{0, 1, 2, 3\}$ et $\alpha_3 \in$ à $\{0, 1, 2\}$.

Il est à noter que le *module de recommandations* va retenir seulement les amis qui ont obtenu les dix plus grands poids. Notons aussi que le système donne la possibilité à l'utilisateur de modifier la liste de ses *amis de confiance*. Ceci est important dans la mesure où l'utilisateur préfère parfois ajouter ou retirer des noms de cette liste créée automatiquement.

7.3. Validation du système

Un total de 163 utilisateurs ont participé dans la validation de *Protect_U*. Ces participants sont différents des répondants à qui nous avons fait appel lors de la phase d'entraînement. Après avoir exécuté l'application, ils ont répondu à un petit questionnaire les invitant à commenter la pertinence des recommandations proposées par le système et l'exactitude de leurs *amis de confiance* retenus, et à donner leurs avis sur l'utilité de l'application. La figure 45 présente un aperçu de ce questionnaire.

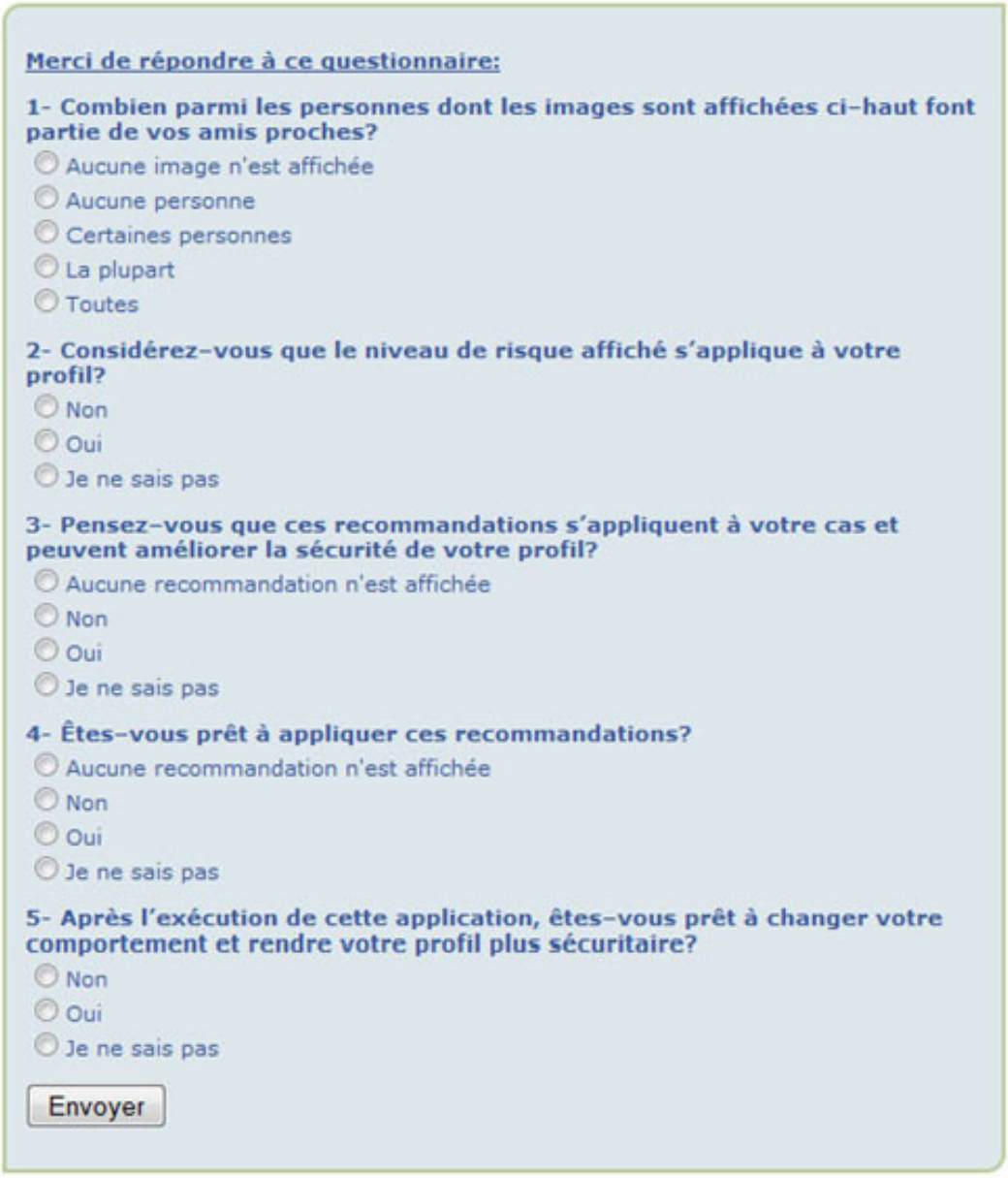


Figure 45. Questionnaire de la validation

Durant la phase de validation, 15,30% des profils des participants ont été classés par *Protect_U* comme *peu risqué*, 40% comme *moyennement risqué*, 25,88% comme *risqué* et 18,82% comme *critique*.

Les réponses à la première question du questionnaire sont représentées à la figure 46.

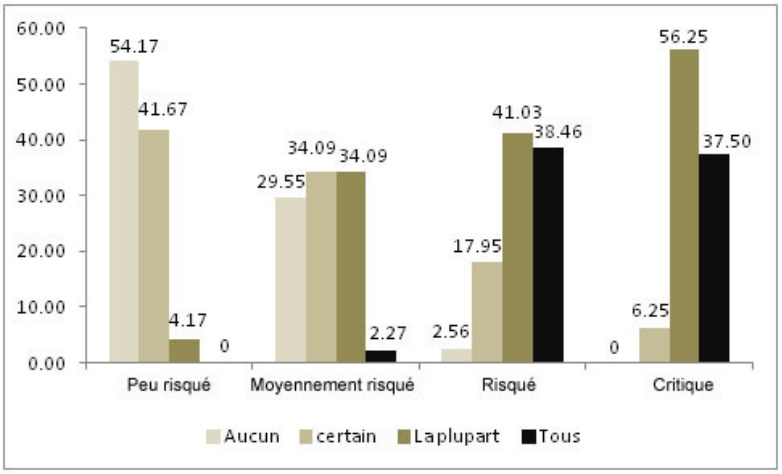


Figure 46. Détection des *amis de confiance* par classe

Ces résultats nous permettent de déduire que le système arrive à repérer la majorité des *amis de confiance* à 79,49% des cas pour les profils *risqué* et à 93,75% des cas pour les profils *critique.* Dans le cas des profils *moyennement risqué* et *peu risqué*, cette majorité tombe à 36,36% et 4,17% des cas respectivement. Ce dernier résultat n'est pas surprenant en soi puisque dans le cas des profils *peu risqué* le nombre d'informations postées et échangées est très limité. De plus, l'avis des *amis de confiance* n'est utile que dans les cas des profils à grands risques. C'est pourquoi, le *module de recommandations* ne doit appliquer le *Filtre communautaire* que dans les cas des profils *risqué* et *critique*.

Les réponses à la deuxième question du questionnaire révèlent cependant que certains répondants ayant obtenu un profil *critique* n'ont pas nécessairement apprécié le résultat obtenu, et tombent dans le déni du résultat. En effet, la figure 47 compare les réponses des participants avec les diagnostics trouvés par *Protect_U*.

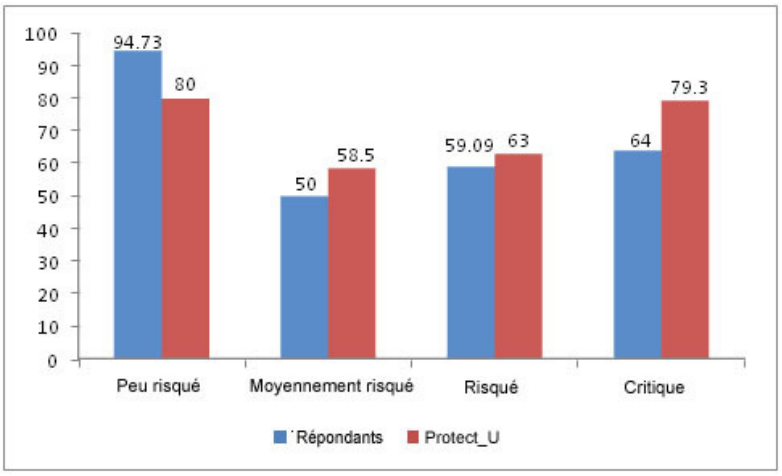


Figure 47. Niveaux de risque des profils selon les répondants vs *Protect_U*

Ces résultats nous laissent croire que beaucoup d'utilisateurs ne sont pas nécessairement conscients du danger auquel ils s'exposent en rendant leurs informations sensibles publiques et refusent d'admettre la réalité.

D'autre part, plus de 70.18% (*en moyenne)* des répondants ont trouvé pertinentes les recommandations proposées. Cependant, seulement 58.38% (*en moyenne)* de ces participants affirment être prêts à les appliquer.

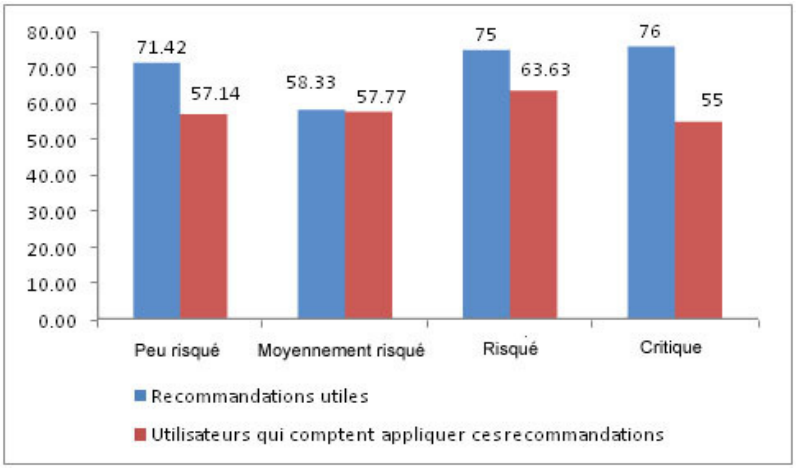


Figure 48. Pourcentages des répondants qui ont trouvé les recommandations utiles et qui sont prêts à appliquer ces recommandations

Il s'avère également que ce sont les utilisateurs ayant obtenu un profil classé *critique* qui démontrent le plus de réticences quant à l'application des recommandations. Cet aspect pourrait être expliqué par le fait que, dans ce cas spécifique, le nombre de recommandations à appliquer était élevé, ce qui a pu décourager certains d'entre eux.

Finalement, 71.93% des répondants (*en moyenne)* avaient affirmés être prêts à modifier leurs habitudes de comportement suite à l'exécution de *Protect_U*.

7.4. Conclusion

Il est évident que la performance de *Protect_U* pourra être améliorée de différentes façons. Nous citons à titre d'exemple :

- Augmenter le nombre de participants au niveau de la phase d'apprentissage pour améliorer la précision du modèle de classification.
- Rendre dynamique la phase d'apprentissage.
- Élargir le cercle des attributs clés considérés.

Le chapitre suivant se veut un complément de *Protect_U*. Il est consacré à introduire le système *Protect_UFF*, une sorte d' « antifraudeurs » qui a comme mission la détection des profils des « amis » *Facebook* qui présentent des symptômes alarmants.

Chapitre 8 : Conception et méthodologie du système *Protect_UFF*

Protect_UFF est une sorte de système d'alarme qui a comme mission de détecter et d'aviser les utilisateurs de *Facebook* de la présence de profils « suspects » parmi leurs « amis ». Un profil est suspect lorsqu'il présente des symptômes qui peuvent être associés à des troubles psychopathiques, à des comportements frauduleux ou à des attitudes criminelles. Il est naïf de croire qu'il existe une seule personnalité problématique responsable à elle seule de toutes les conduites criminelles et délinquantes. Cependant, on a vu au chapitre 4, qu'il existe des caractéristiques communes entre les traits de personnalités des psychopathes, des fraudeurs à cols blanc et les criminels notamment le côté narcissique, le manque d'empathie et d'émotions, et le comportement antisocial qui se manifeste très souvent par de l'agressivité. C'est la raison pour laquelle on a considéré dans notre étude que ces trois attributs constituent de bons indices pour détecter les profils suspects.

Dans cette perspective, *Protect_UFF* analyse le contenu des textes et des images publiques des profils des « amis » *Facebook* d'un utilisateur en tenant compte des trois attributs mentionnés ci-haut. Il attribue trois niveaux possibles pour chaque paramètre : 0,1 ou 2. Il permet ensuite de classer les profils selon trois catégories :

- Classe A pour les profils peu suspects;
- Classe B pour les profils moyennement suspects;
- Classe C pour les profils potentiellement suspects.

Ce chapitre introduit la structure du système *Protect_UFF* qui est constituée essentiellement d'une phase d'apprentissage et de deux modules : le *module d'analyse* et le *module de décision*. Ce chapitre explique aussi le rôle de la phase d'apprentissage qui sert à déterminer les règles de décision qui vont permettre de déterminer le poids des trois paramètres considérés ainsi que le modèle de classification que le système va appliquer pour déterminer à quelle classe appartient chaque profil. Il met en valeur l'importance du *module d'analyse* qui s'occupe de la récolte et de l'analyse du texte des attributs « about_me », « status », « notes » et « comments » des profils des amis ainsi que des

images de l'attribut « photos ». Il introduit ensuite le *module de décision* qui prend en charge l'application du modèle de classification sur les données envoyées par le *module d'analyse* et l'acheminement des résultats vers l'utilisateur. Finalement, il expose les résultats de la validation du système.

8.1. La phase d'apprentissage

L'architecture de la phase d'apprentissage est donnée par la figure ci-dessous :

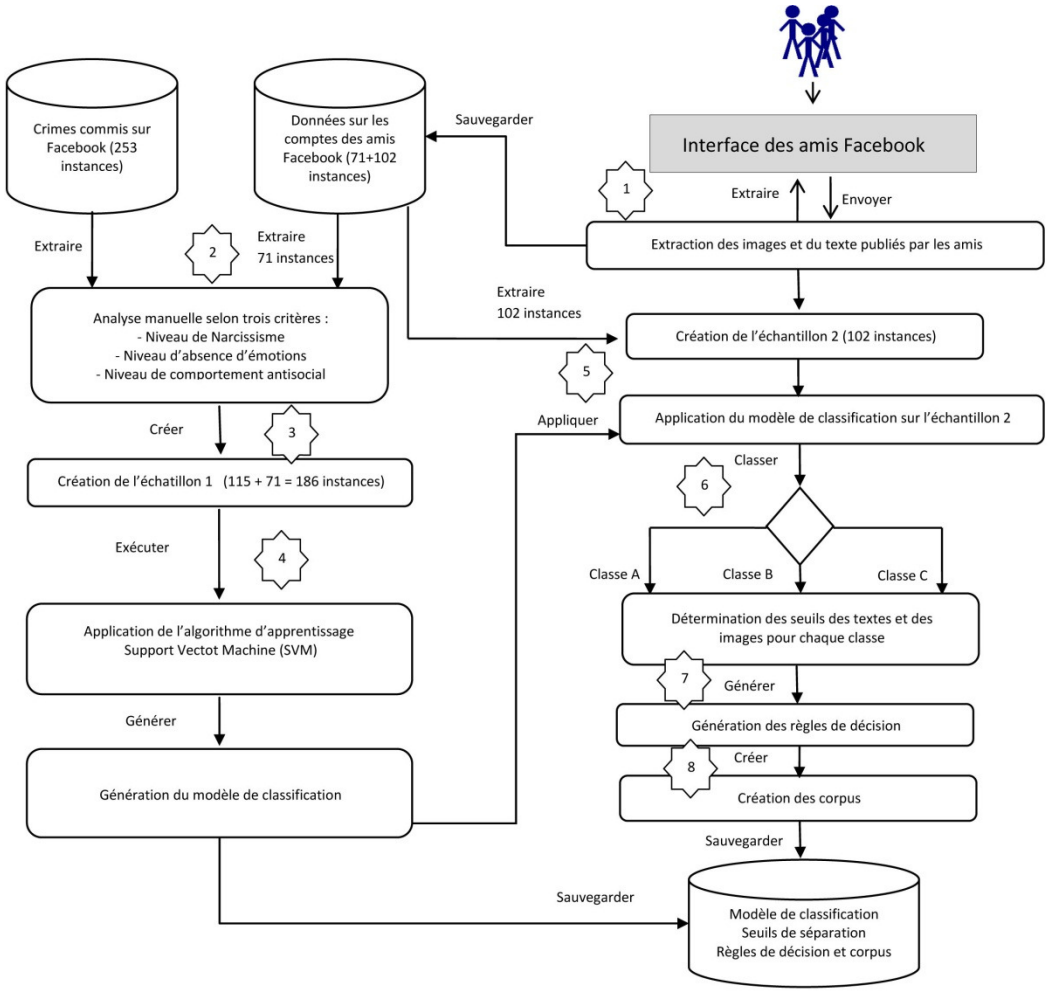


Figure 49. Architecture de la phase d'apprentissage

Toutes les étapes de la phase d'apprentissage doivent être effectuées par un administrateur ou un ensemble de personnes qualifiées pouvant accéder aux serveurs *Protect_U/FF*. Il est souhaitable aussi que ces étapes soient exécutées régulièrement afin d'améliorer constamment la fiabilité du modèle d'apprentissage convoité et de mettre à jour continuellement les seuils de séparation et les corpus de référence du système. Trois paramètres seront étudiés : le *niveau de narcissisme*, le *niveau de manque d'empathie et d'émotions,* et le *niveau de comportement antisocial et agressif.*

Les différentes étapes sont décrites ci-dessous.

Étape 1 : collecte des données

Cette étape consiste à collecter les textes et les images publiés sur les comptes *Facebook* d'un certain nombre d'utilisateurs afin de créer une base de données d'apprentissage. Normalement, pour avoir un modèle de classification précis il faut que le nombre d'instances soit relativement élevé et diversifié pour couvrir tous les cas possibles. Pour notre prototype, on a collecté 173 comptes *Facebook* dont 71 utilisateurs sont potentiellement inoffensifs. De plus, on a retenu 253 comptes-rendus de cas de fraudes et d'actions criminelles commis sur *Facebook* durant une période de vingt-quatre mois (octobre 2008 à septembre 2010)[35], en se basant sur une base de données d'incidents criminels associés à des sites de socialisation en ligne et construite à partir de comptes-rendus médiatiques (puisqu'on n'a pas accès à des comptes *Facebook* de fraudeurs et de criminels). Les actions criminelles incluent, entre autres, des cas de vol d'identités, de vol avec violence, de harcèlement, de menace, de meurtre ou tentative de meurtre, de voie de fait, de suicide, de promotion de la haine, de hameçonnage, d'agressions, etc.

[35] Cette base de données nous a été généreusement donnée par Monsieur Benoît Dupont, professeur à l'école de criminologie de l'Université de Montréal.

Étape 2 : analyse des données

Dans un premier temps, un comité constitué de trois experts dans le domaine ont analysé manuellement le contenu (textes et images) des 71 comptes *Facebook* des utilisateurs potentiellement inoffensifs (puisqu'ils sont connus par au moins l'un des membres du comité et que la probabilité qu'ils aient commis un crime sur Internet est très mince). Ces profils ont été affectés à la classe A.

En se basant sur l'avis des psychologues, décrit au chapitre 4, concernant les symptômes du narcissisme, du manque d'empathie et d'émotions, et du comportement antisocial et agressif, le comité s'est fixé les critères suivants pour l'évaluation des profils :

Tableau 4. Critères de l'évaluation manuelle des profils

Attributs	Critères
Narcissisme:	- Le nombre de fois que l'ami se trouve dans ses photos - La façon avec laquelle il se présente aux autres et les sujets soulevés (combien le moi occupe de place dans ses histoires)
Manque d'émotions:	- L'attitude de l'ami à travers ses photos (présence d'émotions sur le visage ou pas) - Le langage utilisé (émotif ou pas)
Comportement antisocial:	- Nombre d'images vulgaires, agressives et choquantes - Présence de mots offensants dans le texte

Il a attribué ensuite à chaque profil trois notes : une note par paramètre. Les notes varient entre trois valeurs : 0 (faible), 1(moyen) ou 2(élevé). Si on considère par exemple le paramètre *niveau de narcissisme*, le « 0 » veut dire que l'utilisateur présente un signe faible de narcissisme, le « 1 » veut dire qu'il est moyennement narcissique et le « 2 » veut dire qu'il présente des signes élevés de narcissisme.

En second lieu, le comité a analysé le contenu des 253 comptes-rendus. Dans le cas où les comptes-rendus ne permettaient pas de tirer des informations concluantes ou ne faisaient pas l'unanimité parmi les membres du comité, les instances ont été écartées. Dans

d'autres cas, l'interprétation des comptes-rendus était plus évidente. Voici un exemple de compte-rendu qui figure dans notre base de données:

« A popular social networking website helped police catch an arson suspect after the man bragged about the crime online. Police said David Ochoa Delgado, 18, set his neighbor's apartment on fire and then bragged about it on Facebook… »[36]

Ces lignes révèlent clairement que le criminel:

- est très narcissique parce qu'il se vente de son crime devant tous ses amis. (Narcissisme : 2)
- a fait preuve d'un manque flagrant d'empathie vis-à-vis de son voisin. (Manque d'empathie et d'émotions : 2)
- a commis un acte violent et antisocial. (Comportement antisocial : 2)

Finalement, juste 115 comptes-rendus ont été retenus et 138 ont été rejetés. Le comité a divisé ensuite les crimes commis en deux classes : B (grave) et C (très grave). Le « meurtre » par exemple a été considéré comme étant un crime très grave tandis que « une menace » a été considérée comme étant un crime juste grave.

Étape 3 : création de l'échantillon 1

En groupant les 115 instances avec les 71 du départ, on obtient une base de données de 186 instances (échantillon 1) réparties sur trois classes : A, B et C. Le tableau 5 donne un exemple d'instances de l'échantillon 1 :

[36] http://www.kxan.com/dpp/news/man-accused-of-arson-brags-about-it

Tableau 5. Exemples d'instances de l'échantillon 1

Instances	Narcissisme	Manque d'émotions	Comportement antisocial	Classe
1	1	1	1	B
2	2	1	2	C
3	1	2	2	C
4	1	0	1	A

L'échantillon 1 est constitué de 71 instances de classe A, 61 instances de classe B et 54 instances de classe C distribuées comme suit :

Tableau 6. Distribution des instances de la classe A de l'échantillon 1

Niveau	Narcissisme	Manque d'émotions	Comportement antisocial
0	26	33	36
1	35	32	31
2	10	6	4

Tableau 7. Distribution des instances de la classe B de l'échantillon 1

Niveau	Narcissisme	Manque d'émotions	Comportement antisocial
0	52	0	1
1	9	59	59
2	0	2	1

Tableau 8. Distribution des instances de la classe C de l'échantillon 1

Niveau	Narcissisme	Manque d'émotions	Comportement antisocial
0	14	0	1
1	17	10	2
2	23	44	51

Étape 4 : détermination du modèle de classification

On a utilisé l'échantillon 1 comme base d'apprentissage pour trouver le modèle de classification qui nous permettra de classer toute nouvelle instance en fonction des classes A, B ou C. Pour cela, on a utilisé les *Machines à Support Vectoriel* linéaire de *WEKA* (noté *SMO*). Trois classificateurs binaires ont été appliqués pour séparer en premier lieu les classes A et B, puis les classes A et C et ensuite les classes B et C. La matrice de confusion et la précision des résultats sont données ci-dessous :

Tableau 9. Matrice de confusion du modèle de classification de *Protect_UFF*

		Classes Prédites		
		Peu Risqué	**Moyennement risqué**	**Potentiellement Risqué**
Classes actuelles	**Peu risqué**	71,01%	23,19%	5,80%
	Moyennement risqué	16,13%	82,26%	1,61%
	Potentiellement Risqué	5,56%	0,0%	94,44%

Tableau 10. Précision du modèle de classification de *Protect_UFF*

	Précision	**Rappel**	**F-Mesure**
Classe A	0,79	0,71	0,748
Classe B	0,761	0,823	0,791
Classe C	0,911	0,944	0,927
Moyenne	0,816	0,816	0,815

La figure 50 donne un aperçu des fonctions de décisions que le classificateur SMO a généré:

```
=== Classifier model (full training set) ===

SMO

Kernel used:
  Linear Kernel: K(x,y) = <x,y>

Classifier for classes: a, b

BinarySMO

Machine linear: showing attribute weights, not support vectors.

        -2      * (normalized) Narcissique
 +       2      * (normalized) sentiment
 +       2      * (normalized) antisocial
 -       1

Number of kernel evaluations: 581 (46.892% cached)

Classifier for classes: a, c

BinarySMO

Machine linear: showing attribute weights, not support vectors.

         0.0007 * (normalized) Narcissique
 +       0.2497 * (normalized) sentiment
 +       3.7497 * (normalized) antisocial
 -       2.9997

Number of kernel evaluations: 861 (67.522% cached)

Classifier for classes: b, c

BinarySMO

Machine linear: showing attribute weights, not support vectors.

         0.5    * (normalized) Narcissique
 +       0.75   * (normalized) sentiment
 +       3.25   * (normalized) antisocial
 -       3

Number of kernel evaluations: 304 (60.468% cached)
```

Figure 50. Fonctions de décision du classificateur SMO

Étape 5 : création de l'échantillon 2

L'échantillon 1 nous a permis de trouver les fonctions de décisions que le classificateur SMO va appliquer. Cependant, pour pouvoir évaluer automatiquement de nouveaux profils *Facebook*, on a besoin de créer quatre corpus (trois corpus textes et 1 corpus images) pour pouvoir distinguer entre les niveaux 0, 1 et 2 des attributs. De plus, il faut préciser les seuils de séparation à partir desquels le système arrivera à distinguer automatiquement entre les classes. Or, ceci est impossible à réaliser à partir de l'échantillon 1 parce que la majorité des valeurs proviennent d'analyses de contres-rendus et non de comptes *Facebook*.

Pour cette raison, on a fait appel aux 102 instances restant de la première phase, qui proviennent de comptes *Facebook* uniquement, pour créer l'échantillon 2. On a appliqué ensuite à ce dernier le modèle de classification obtenu ultérieurement pour classer les instances. Ainsi, on a obtenu 7 profils *Facebook* de classe C, 15 de classe B et 80 de classe A.

Étape 6 : détermination des seuils de séparation

Pour déterminer les seuils de séparation, on a considéré les profils des classe B et C obtenus au niveau de l'étape 5. On a analysé manuellement les textes écrits par les propriétaires des comptes et leurs images publiques. On a calculé ensuite les pourcentages des mots et des expressions narcissiques, émotionnels, agressifs ou antisociaux rencontrés. On a calculé par la suite pour chaque compte les pourcentages des images dans lesquelles l'utilisateur se trouve, le pourcentage des images dans lesquelles il éprouve de l'émotion et le pourcentage des images agressives et/ou antisociales.

Les tableaux suivants résument les résultats obtenus :

Tableau 11. Moyennes et écarts type des valeurs des classes B et C pour le paramètre Narcissisme

Narcissisme	Classe B		Classe C	
	Moyenne	Écart type	Moyenne	Écart type
Textes	5,303%	1,917%	8,165%	1,392%
Images	44,366%	15,830%	60,085%	16,021%

Tableau 12. Moyennes et écarts type des valeurs des classes B et C pour le paramètre *Émotions*

Émotions	Classe B		Classe C	
	Moyenne	Écart type	Moyenne	Écart type
Textes	7,543%	2,688%	2,797%	1,209%
Images	42,132%	17,808%	65,283%	14,489%

Tableau 13. Moyennes et écarts type des valeurs des classes B et C pour le paramètre *Agressivité*

Agressivité	Classe B		Classe C	
	Moyenne	Écart type	Moyenne	Écart type
Textes	6,186%	1,970%	13,974%	3,221%
Images	33,694%	14,474%	60,638%	8,626%

D’où la répartition suivante :

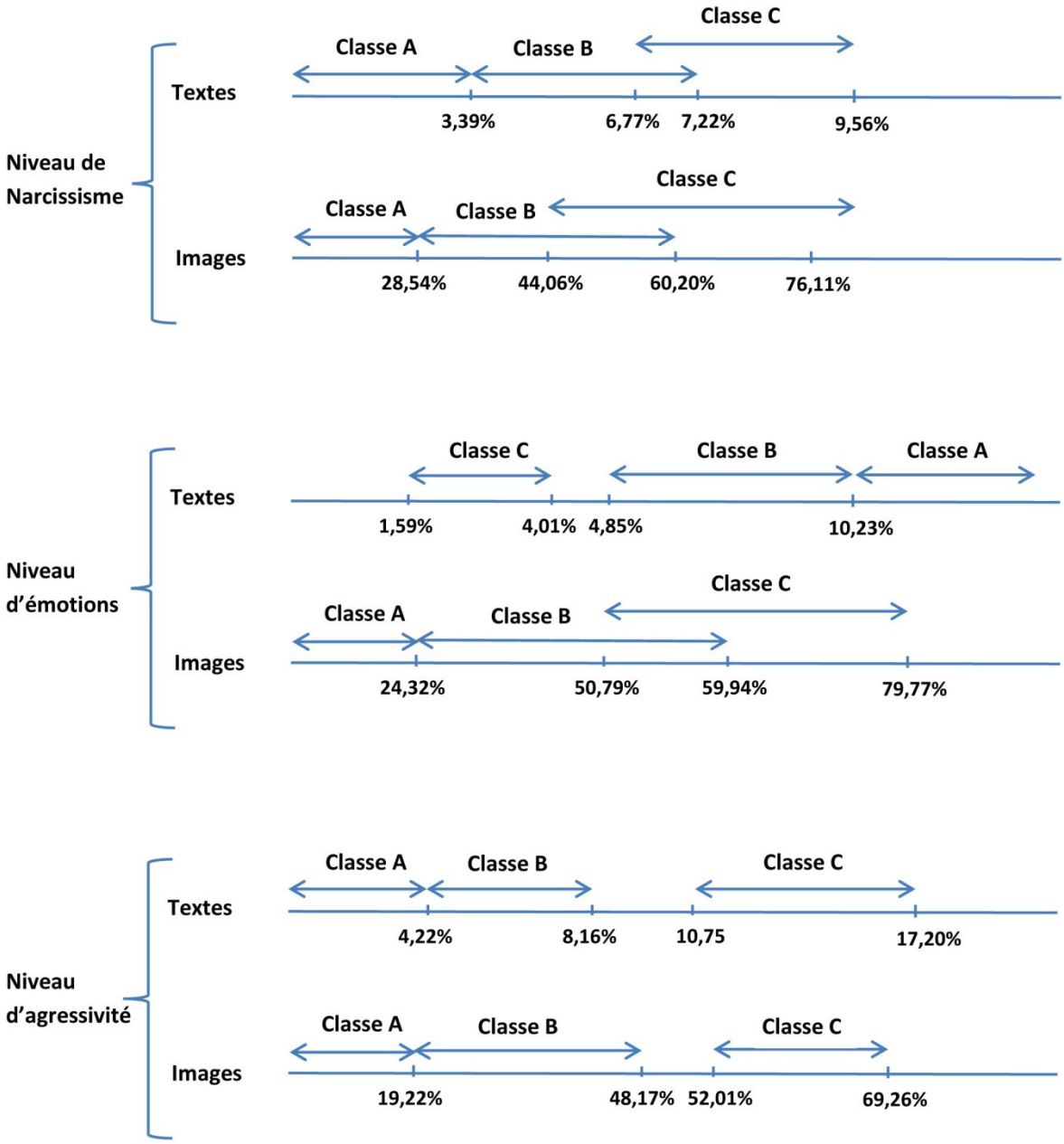


Figure 51. Intervalles obtenus suite à l'analyse des profils des classes B et C

Pour pouvoir déterminer des seuils, on a procédé comme suit :

- Lorsque deux intervalles se chevauchent, on prend le milieu de la partie commune comme seuil pour les deux intervalles. À titre d'exemple, les intervalles [3,39; 7,22[et [6,77; 9,56] auront 7 comme seuil commun (car 6,77+7,22 divisé par 2 donne approximativement 7). Ainsi, on obtient les intervalles [3,39; 7[et [7; 9,56].

- Lorsque deux intervalles ne se chevauchent pas, on prend le milieu de l'espace qui les sépare comme seuil commun. On ajoute alors à la fin du premier intervalle la moitié de la distance qui les sépare et on réduit la borne inférieure du deuxième intervalle de la même valeur. À titre d'exemple, les intervalles [1,59; 4,01[et [4,85; 10,23] ne se coupent pas. La distance qui les sépare est de 0,84 (4,85 moins 4,01). En ajoutant 0,42 à la borne supérieure du premier intervalle et en réduisant de 0,42 la borne inférieure du deuxième intervalle on obtient [1,59; 4,43[et [4,43; 10,23].

Par suite, on obtient la répartition suivante:

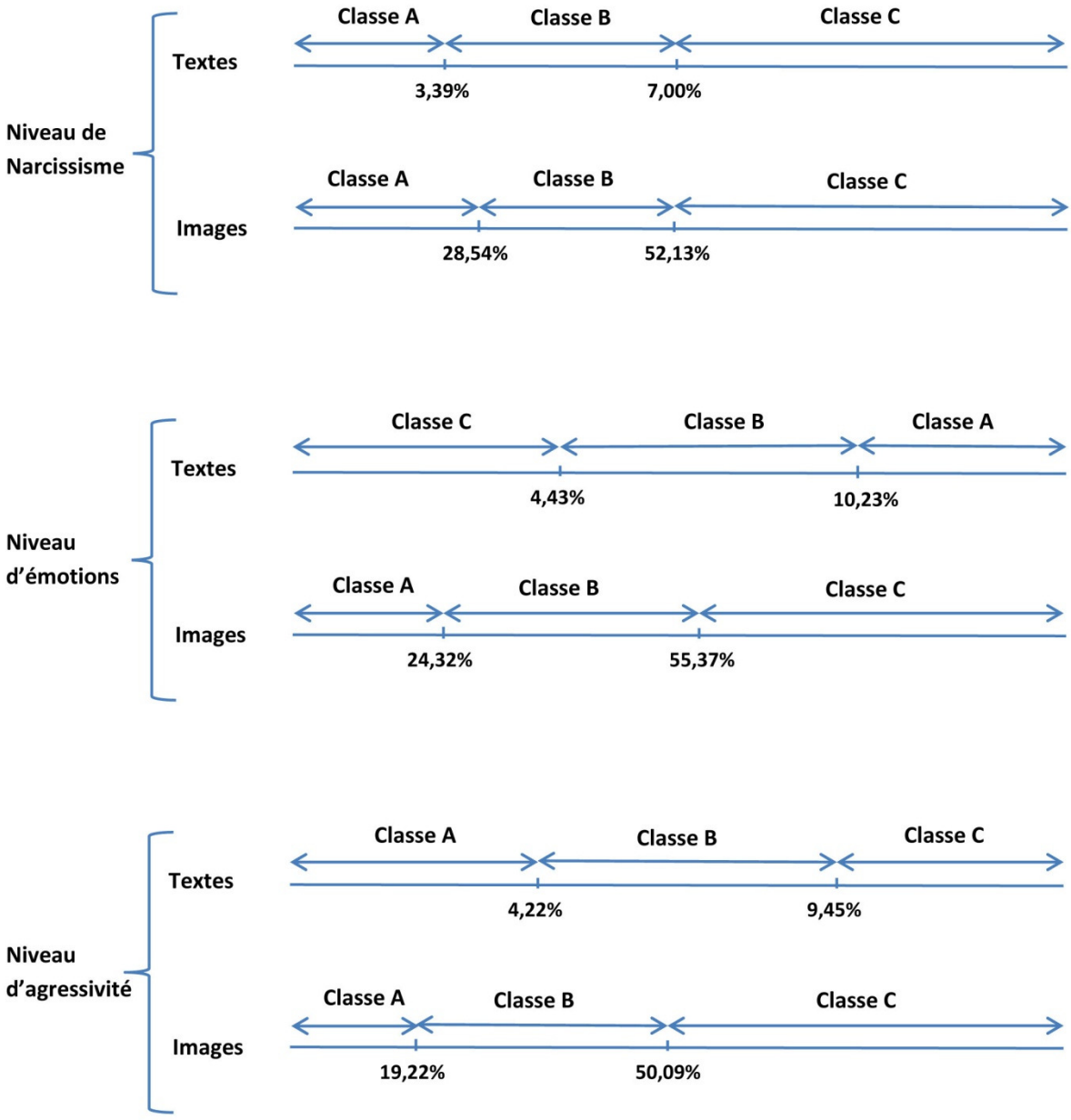


Figure 52. Seuils ajustés séparant les classes A, B et C

Étape 7 : création des règles de décision

Soient les variables suivantes:

PT1 : pourcentage de textes narcissiques
PI1 : pourcentage des images narcissiques

PT2 : pourcentage de textes émotionnels
PI2 : pourcentage des images émotionnelles
PT3 : pourcentage de textes antisociaux et agressifs
PI3 : pourcentage des images antisociales et agressives

NT1 : niveau du texte narcissique
NI1 : niveau des images narcissiques
NT2 : niveau du texte émotionnel
NI2 : niveau des images émotionnelles
NT3 : niveau du texte antisocial et agressif
NI3 : niveau des images antisociales et agressives

Les seuils trouvés au niveau de l'étape précédente donnent naissance aux règles de décision suivantes :

Règles appliquées pour évaluer les textes et les images :

R1 : Si (PT1 $\leq$ 3,39% alors NT1 = 0) sinon
Si (3,39% < PT1 $\leq$ 7,00% alors NT1 = 1) sinon
NT1 = 2

R2 : Si (PI1 $\leq$ 28,54% alors NI1 = 0) sinon
Si (28,54% < PI1 $\leq$ 52,13% alors NI1 = 1) sinon
NI1 = 2

R3 : Si (PT2 $\leq$ 4,43% alors NT2 = 2) sinon
Si (4,43% < PT2 $\leq$ 10,23% alors NT2 = 1) sinon
NT2 = 0

R4 : Si (PI2 $\leq$ 24,32% alors NI2 = 2) sinon
Si (24,32% < PI2 $\leq$ 55,37% alors NI2 = 1) sinon
NI2 = 0

R5 : Si (PT3 $\leq$ 4,22% alors NT3 = 0) sinon
Si (4,22% < PT3 $\leq$ 9,45% alors NT3 = 1) sinon
NT3 = 2

R6 : Si (PI3 $\leq$ 19,22% alors NI3 = 0) sinon
Si (19,22% < PI3 $\leq$ 50,09% alors NI3 = 1) sinon
NI3 = 2

Règles appliquées pour déterminer les trois paramètres en fonction des textes et des images:

R7 : Si (NT1 + NI1 ≤ 1 alors Narcissisme = 0) sinon
Si (NT1 + NI1 ≤ 2 alors Narcissisme = 1) sinon
Narcissisme = 2

R8: Si (NT2 + NI2 ≤ 1 alors Émotions = 2) sinon
Si (NT2 + NI2 ≤ 2 alors Émotions = 1) sinon
Émotions = 0

R9 : Si (NT3 + NI3 ≤ 1 alors Agressivité/Antisociabilité = 2) sinon
Si (NT3 + NI3 ≤ 2 alors Agressivité/Antisociabilité = 1) sinon
Agressivité/Antisociabilité = 0

D'après les trois règles précédentes, on a trois cas possibles pour atteindre chacun des niveaux. La table de vérité ci-dessous énumère ces cas :

Tableau 14. Les différents cas possibles pour atteindre le niveau 0, 1 ou 2

Textes	Images	Niveau
0	0	0
0	1	0
1	0	0
1	1	1
0	2	1
2	0	1
1	2	2
2	1	2
2	2	2

Étape 8 : création des corpus

Durant cette étape, cinq corpus sont créés : trois corpus textes, un corpus images et un corpus de modèles d'objets. Afin de simplifier le traitement, on a uniquement considéré la langue anglaise pour les corpus textes. On a aussi affecté le même poids pour les mots et les expressions rencontrés.

Le premier corpus, noté *textes narcissiques*, groupe un ensemble de termes et d'expressions liés au narcissisme tels que « myself », « I am the one », « I'm the one », « I am the best », « I am the champion », « I am very attractive », « I am the coolest », « I am gorgeous », « I am beautiful », « I am pretty », etc. On a favorisé les expressions au dépend des mots isolés parce qu'il est plus facile de les interpréter : l'existence du mot « best » dans une phrase peut être interprété de différentes façons dépendamment de la sémantique de la phrase, tandis que la présence de l'expression « I am the best » ne laisse pas trop d'ambiguïté quant à son interprétation.

Le deuxième corpus, nommé *textes émotionnels*, contient un ensemble de mots et d'expressions qui permet par leur présence dans un texte de refléter l'état émotionnel de leur auteur. Le but de la création de ce corpus est de détecter les émotions et l'état d'âme des utilisateurs *Facebook* au moment de l'écriture des commentaires (si un texte est pauvre en mots et en expressions émotionnels, ceci montre que l'utilisateur n'exprime pas ou très peu d'émotions qui est très souvent associé au manque d'empathie). L'émotion en question peut être liée à la joie, la tristesse, la peur, la colère, le dégoût, la rage, etc. Là aussi on favorise les expressions telles que « I am afraid », « I hate », « I love », « I am outrageous », « I am amazed », etc. Les icônes émotionnelles sont aussi prises en compte.

Le troisième corpus, intitulé *textes agressifs/antisociaux*, groupe des mots agressifs et offensants (insultes, gros mots, etc.).

Notons que les corpus textes sont basés sur les mots et expressions narcissiques, émotionnels ou agressifs qui ont été collectés dans les textes des 171 comptes *Facebook* de la première étape. Cependant, vu que ce nombre est relativement petit, ils ont été enrichis par des termes et des expressions collectés à plus de 5,000 tweets écrit en anglais. Le premier corpus groupe 267 mots et expressions jugés narcissiques, le deuxième corpus est constitué de 423 mots et expressions émotionnelles tandis que le troisième corpus contient 602 mots et expressions agressifs et/ou offensants.

Le quatrième corpus, noté *images agressives/antisociales*, stocke 438 images agressives et/ou antisociales rencontrées lors de l'analyse des 171 comptes *Facebook*. De

plus, il conserve les modèles de certains objets considérés comme violents ou antisociaux. Ces objets sont extraits à partir des 438 images ci-haut. L'intérêt de créer un tel corpus est de donner au *module d'analyse* la possibilité de détecter automatiquement les images et les objets agressifs et antisociaux lorsqu'ils se faufilent dans les profils des amis. Ceci pourra aider l'utilisateur à mieux les cibler.

La figure ci-dessous présente la distribution des objets agressifs et antisociaux rencontrés :

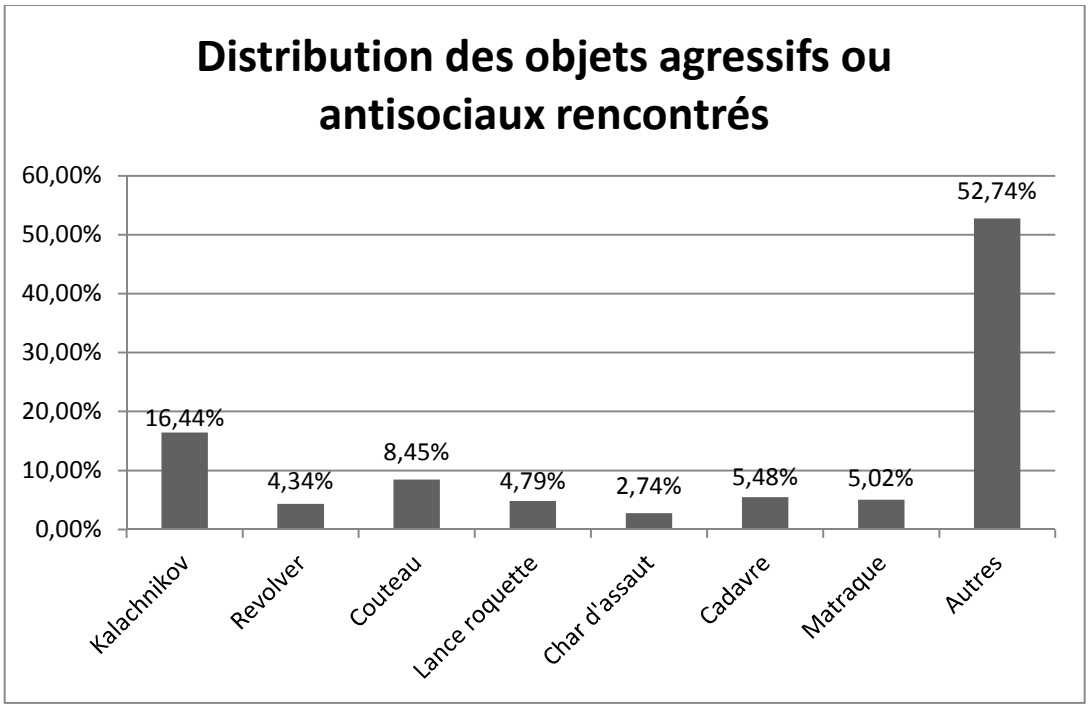


Figure 53. Distribution des objets agressifs ou antisociaux rencontrés

Parmi les images agressives ou antisociales collectées certaines images peuvent être caractérisées par la présence d'un objet agressif ou antisociale tels qu'une arme (kalachnikov, revolver, couteau, lance-roquette ou char d'assaut), un cadavre ou une matraque tandis que d'autres dévoilent plutôt un contexte inamical tels qu'un feu, une explosion, une manifestation, une bagarre, etc. *GetObject* a été utilisé dans la détection des

objets agressifs ou antisociaux tandis qu' *ImgSeek* a été appliqué pour détecter les images à contextes inamicaux.

La figure ci-dessous donne un exemple de modèles d'objets utilisés : kalachnikov, couteau de combat, revolver et char d'assaut.

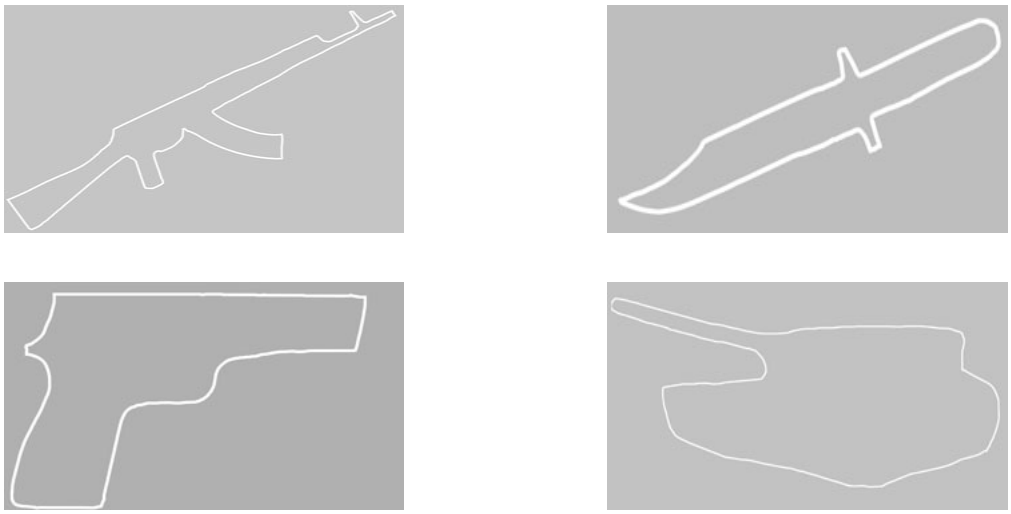


Figure 54. Exemples de modèles d'objets agressifs ou antisociaux

Chaque modèle a été obtenu à partir d'une quinzaine d'images de références dans lesquelles on a déterminé les points de contrôle. La figure 55 indique les points de contrôle qui ont été considérés pour les quatre objets retenus.

Figure 55. Points de contrôle considérés

8.2. Module d'analyse

Le *module d'analyse*, comme son nom l'indique, se charge d'analyser le texte et les images collectés à partir des comptes *Facebook* des « amis ». Il cherche essentiellement les mots, les expressions et les images qui peuvent être comme narcissiques, émotionnels ou agressifs/antisociaux. L'architecture de ce module est donnée par la figure 56.

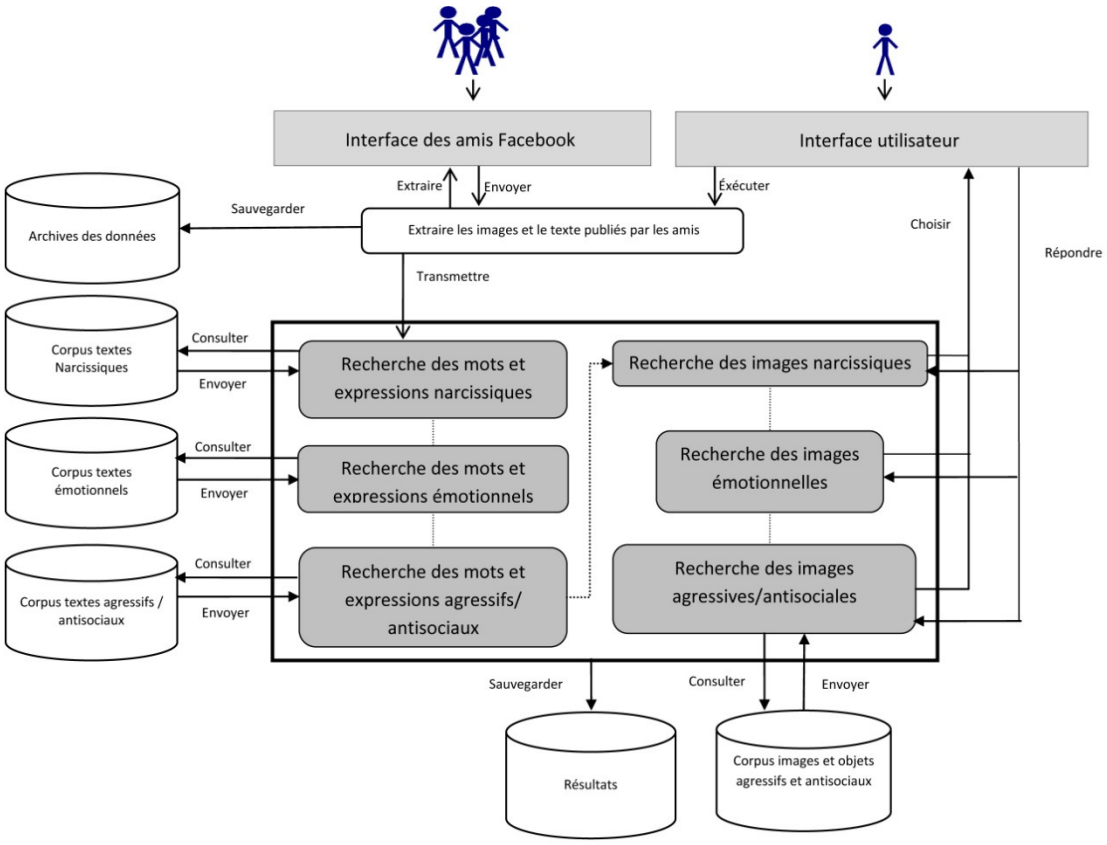


Figure 56. Architecture du *module d'analyse*

En premier lieu, l'utilisateur choisit les « amis » dont il souhaite analyser leurs profils. Lorsqu'il clique sur le bouton « Check » de l'interface de *Protect_UFF*, le module d'analyse va automatiquement extraire le texte des attributs « about_me », « status », « notes » et « comments » (de « status » et « notes ») ainsi que les images de l'attribut « photos » des profils des amis sélectionnés. Il les sauvegarde ensuite dans la base de données « Archives des données ». *Facebook* permet l'accès automatique à uniquement les 25 dernières instances des attributs déjà mentionnés ci-haut.

Analyse du texte

Le *module d'analyse* analyse les textes collectés et leur applique le traitement suivant :

- Il calcule le pourcentage des mots et des expressions narcissiques selon le *corpus textes narcissiques* puis il applique la règle R1.
- Il calcule le pourcentage des mots et des expressions émotionnels selon le *corpus textes émotionnels* puis il applique la règle R3.
- Il calcule le pourcentage des mots et des expressions agressifs/antisociaux selon le *corpus textes agressifs/antisociaux* puis il applique la règle R5.

Analyse des images

Dans cette partie, le *module d'analyse* implique l'utilisateur directement :

- Pour calculer le pourcentage des images narcissiques de chaque ami, le *module d'analyse* exécute la bibliothèque graphique libre *OpenCV* [37] pour détecter les images contenant un ou plusieurs visages. La figure 57 ci-dessous donne un exemple de détection de visage. Il affiche ensuite à l'utilisateur l'ensemble des images sélectionnées pour lui demander d'indiquer manuellement dans quelles images son ami se trouve. La règle R2 sera ensuite appliquée.

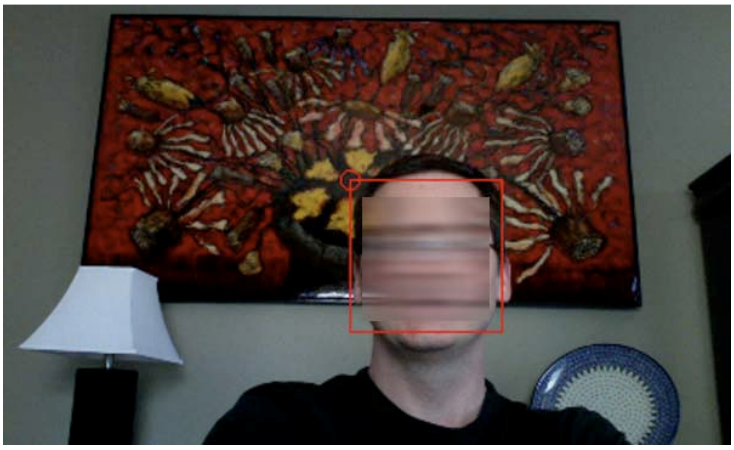


Figure 57. Exemple de détection de visage avec *OpenCV*

[37] http://docs.opencv.org/modules/core/doc/intro.html

- Pour déterminer le pourcentage des images dans lesquelles l'ami n'éprouve pas d'émotions, le *module d'analyse* affiche à l'utilisateur les images qu'il avait choisi à l'étape précédente et lui demande de préciser les images dans lesquelles le visage de l'ami n'inspire aucun sentiment ou aucune émotion. La règle R4 sera ensuite appliquée.
- Pour calculer le pourcentage des images agressives/antisociales, le *module d'analyse* va procéder comme suit :
    - Comparer l'ensemble des images d'un ami avec la liste des images du *corpus images agressives/antisociales* en utilisant l'algorithme *Imgseek*[38] (voir chapitre 6). Le seuil de décision a été réglé à 90%. Si le pourcentage de similitude retourné par *Imgseek* est supérieur ou égal à ce seuil, l'image correspondante sera considérée comme agressive ou antisociale.
    - Fouiller le contenu des images à la recherche d'objets agressifs ou antisociaux en se basant sur la liste des modèles des objets trouvés ultérieurement au niveau de la phase d'entraînement. Cette recherche s'effectue à l'aide de l'algorithme *GetObject* (voir chapitre 6). L'efficacité de cette application dépend de la nature, de la taille et de la forme de l'objet recherché. La figure 58 donne un exemple de réussite de l'algorithme pour l'objet « kalachnikov ».

[38] http://www.imgseek.net/home

Figure 58. Résultats réussis de l'exécution de *GetObject* pour l'objet « kalachnikov »

- Afficher les images avec les résultats des deux étapes précédentes à l'utilisateur pour qu'il confirme lui-même quelles sont les images qu'il considère réellement agressives ou antisociales. Le but des deux étapes précédentes est d'aider l'utilisateur dans son choix. Elles permettent d'attirer son attention sur des images qui ont été déjà jugées agressives ou antisociales par autrui.
- Appliquer la règle R6.

Finalement tous les résultats obtenus dans ce module seront sauvegardés dans la base de données *Résultats*.

8.3. Module de décision

Le *module de décision* va se baser sur les résultats obtenus au niveau du *module d'analyse* pour appliquer les règles 7, 8 et 9 afin de déterminer le niveau de narcissisme,

d'émotions et d'agressivité/antisociabilité. Il applique par la suite le modèle de classification trouvé au niveau de la phase d'entraînement pour déterminer le niveau de risque de chaque « ami ». Finalement, il sauvegarde les résultats dans la base de données « Résultats » et les transmet à l'utilisateur. L'architecture du *module de décision* est donnée par la figure ci-dessous :

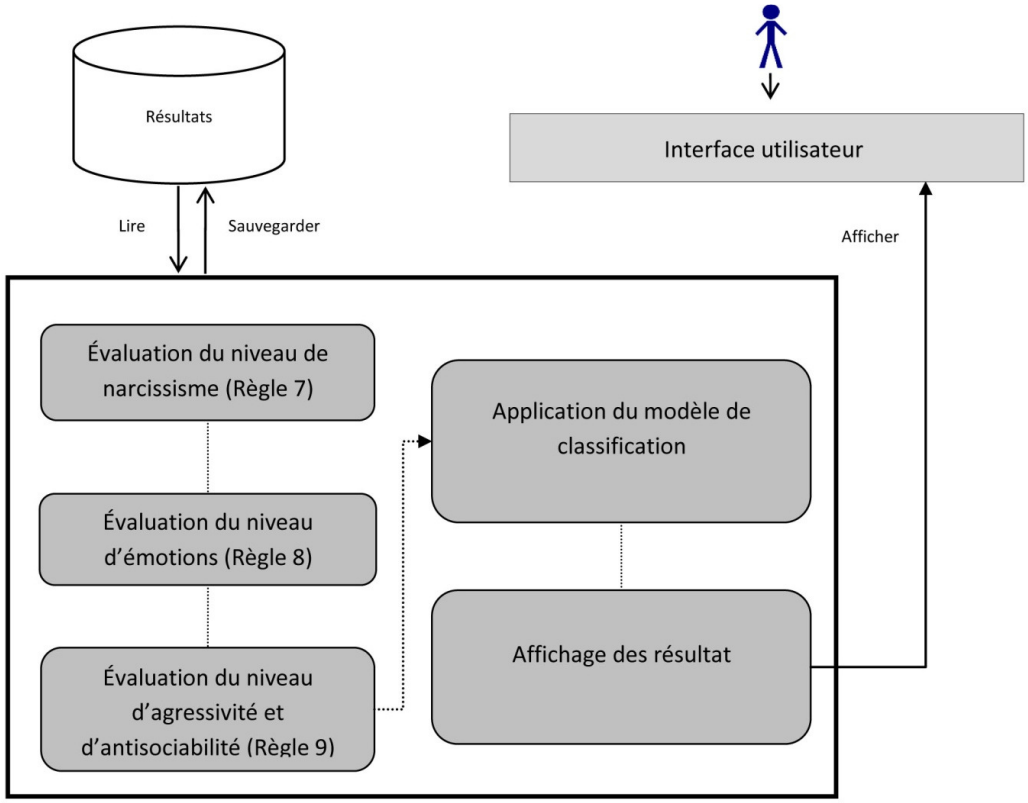


Figure 59. Architecture du module de décision

8.4. Test du bon fonctionnement du système

Afin de tester le bon fonctionnement du système on a créé 24 comptes *Facebook* dont 4 appartiennent à la classe C, 13 à la classe B et 7 à la classe A. La répartition des niveaux des paramètres pour chaque classe est donnée par le tableau suivant :

Tableau 15. Répartition des classes sur les 24 comptes *Facebook*

Narcissisme			Émotions			Agressivité/ Antisociabilité			Classe
Textes	Images	Niveau	Textes	Images	Niveau	Textes	Images	Niveau	
2	2	2	2	2	2	2	2	2	C
1	1	1	2	1	2	1	2	2	C
2	1	2	2	0	1	1	2	2	C
1	2	2	1	2	2	2	0	1	C
1	2	2	1	1	1	1	1	1	B
0	2	1	1	2	2	0	2	1	B
2	0	1	0	2	1	2	2	2	B
0	1	0	2	1	2	2	1	2	B
2	1	2	0	0	0	2	1	2	B
2	2	2	2	2	2	0	0	0	B
1	1	1	2	0	1	0	2	1	B
1	2	2	1	1	1	0	0	0	B
2	2	2	0	1	0	1	1	1	B
2	0	1	2	2	2	1	0	0	B
0	2	1	1	0	0	1	2	2	B
1	0	0	0	2	1	2	1	2	B
0	0	0	1	2	2	2	0	1	B
0	1	0	0	0	0	0	1	0	A
1	0	0	0	1	0	1	1	1	A
0	0	0	1	1	1	1	0	0	A
1	1	1	1	0	0	0	1	0	A
0	0	0	2	0	1	0	2	1	A
2	0	1	0	0	0	2	0	1	A
0	2	1	0	2	1	0	0	0	A

Les 24 comptes *Facebook* ont été enrichis par du texte et des images de façon à respecter la répartition des niveaux indiqués dans le tableau ci-dessus. Les mots-clés narcissiques, émotionnels et agressifs/antisociaux du texte ajouté ont été choisis aléatoirement à partir des 3 corpus textes qu'on avait déjà créés au niveau de la phase d'apprentissage. La moitié des images agressives/antisociales ont été choisies aléatoirement à partir d'Internet et le reste a été choisi à partir du corpus images qui a été déjà créé avec les 3 corpus textes. On s'est assuré aussi d'inclure des images contenant certains des objets agressifs/antisociaux dont on a créé déjà les modèles.

Les 24 comptes ont été ajoutés comme « ami » à l'utilisateur Michel Tremb (nom fictif qu'on avait créé nous-mêmes). Ce dernier a exécuté *Protect_UFF* et a obtenu les résultats suivants :

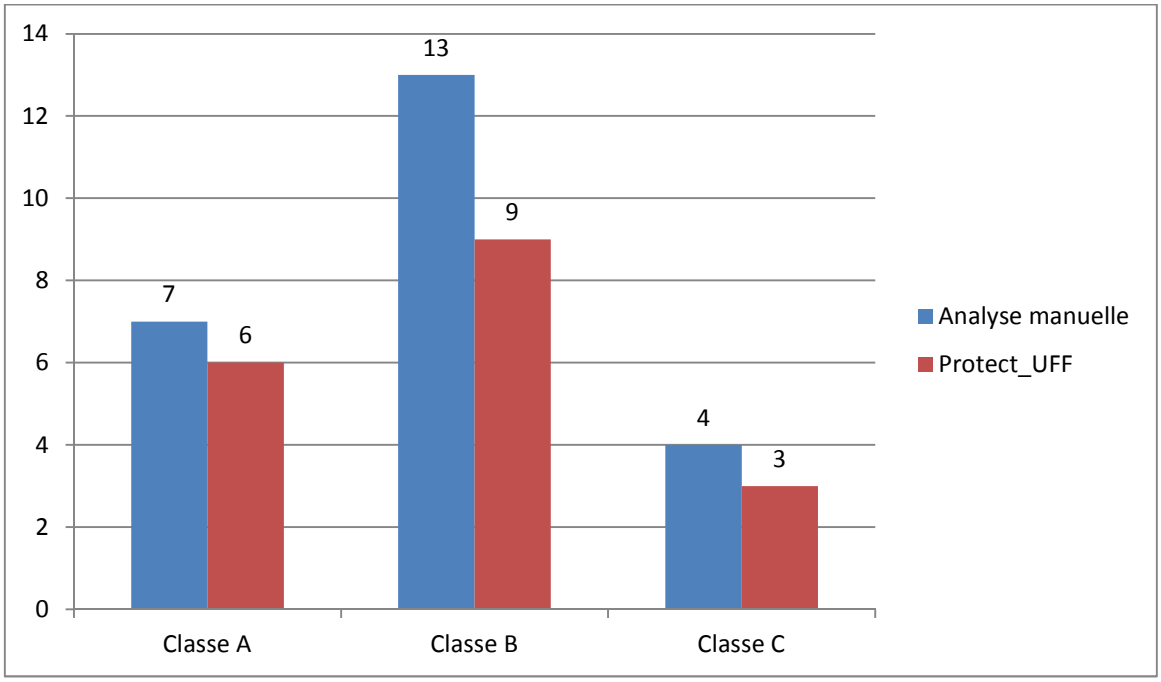


Figure 60. Taux de réussite de *Protect_UFF*

Les résultats montrent que *Protect_UFF* a résussi à classer correctement 75% des profils de classe C, 69,23% des profils de classe B et 87,5% des profils de classe A. L'ambiguïté se trouve surtout au niveau de la classe B. Afin de mieux comprendre les résultats, on les a comparés avec ceux de la validation manuelle du tableau précédent :

Tableau 16. Résultats détaillés de l'exécution de *Protect_UFF*

Narcissisme			Émotions			Agressivité/ Antisociabilité			Validation manuelle	Protect_UFF
Textes	Images	Niveau	Textes	Images	Niveau	Textes	Images	Niveau		
2	2	2	2	2	2	2	2	2	C	C
1	1	1	2	1	2	1	2	2	C	C
2	1	2	2	0	1	1	2	2	C	C
1	2	2	1	2	2	2	0	1	C	B
1	2	2	1	1	1	1	1	1	B	B
0	2	1	1	2	2	0	2	1	B	B
2	0	1	0	2	1	2	2	2	B	B
0	1	0	2	1	2	2	1	2	B	B
2	1	2	0	0	0	2	1	2	B	B
2	2	2	2	2	2	0	0	0	B	A
1	1	1	2	0	1	0	2	1	B	B
1	2	2	1	1	1	0	0	0	B	A
2	2	2	0	1	0	1	1	1	B	B
2	0	1	2	2	2	1	0	0	B	A
0	2	1	1	0	0	1	2	2	B	B
1	0	0	0	2	1	2	1	2	B	B
0	0	0	1	2	2	2	0	1	B	A
0	1	0	0	0	0	0	1	0	A	A
1	0	0	0	1	0	1	1	1	A	A
0	0	0	1	1	1	1	0	0	A	A
1	1	1	1	0	0	0	1	0	A	A
0	0	0	2	0	1	0	2	1	A	A
2	0	1	0	0	0	2	0	1	A	B
0	2	1	0	2	1	0	0	0	A	A

Cette comparaison nous laisse croire que le modèle de classification donne plus de poids au paramètre « Agressivité/Antisociabilité » que « Narcissisme » et « Émotions » lorsqu'il doit choisir entre la classe A ou B. Ce même paramètre a aussi causé la dévaluation du profil de la quatrième ligne de la classe C à la classe B.

Notons cependant que les résultats risquent d'être moins reluisants lorsque *Protect_UFF* sera appelé à évaluer de vrais comptes *Facebook* dans lesquels se trouvent des images, des mots et des expressions qui n'ont pas encore été ajoutés au quatre corpus. En fait, plus il y a d'utilisateurs qui vont l'utiliser plus le résultat de la classification sera précis parce que ça va contribuer à enrichir davantage les corpus. L'aspect dynamique de la phase d'entraînement est aussi crucial pour obtenir des résultats pertinents.

D'un autre côté, on a remarqué en exécutant *Protect_UFF*, que l'algorithme *Imgseek* donne de bons résultats lorsqu'il s'agit de comparer des images identiques. Son taux de réussite avoisine les 92% (en prenant un pourcentage de similitude supérieur à 80%). Il est donc surtout utile à détecter les images qui ont été déjà classées dans le corpus des images. Cependant, l'algorithme *GetObject* arrive à repérer les objets agressifs/antisociaux à presque 52% des fois. Sa réussite va dépendre de la taille de l'objet, sa direction et sa visibilité (s'il est totalement ou partiellement visible). Quant à *OpenCV*, il a réussi à détecter 63% des visages dans les images. Là aussi, la clarté et la visibilité de tout le visage est nécessaire pour avoir un bon résultat.

8.5. Conclusion

L'intérêt et la précision de *Protect_UFF* va dépendre de plusieurs facteurs. Nous distinguons :

- La mise à jour continue de l'échantillon 1 utilisé dans la création du modèle de classification;
- Le raffinement régulier des seuils de séparation de l'étape 6 de la phase d'entraînement;
- L'enrichissement fréquent des corpus textes et images;
- L'augmentation continue de la liste des modèles des objets agressifs/antisociaux.

En absence de ces modifications, *Protect_UFF* perdra rapidement son efficacité et son potentiel et par suite sa raison d'être.

Malgré le fait que *Protect_UFF* et *Protect_U* sont conçus pour faire un travail complémentaire, ils divergent sur plusieurs aspects :

- *Protect_U* s'exécute plus rapidement que *Protect_UFF* vu que ce dernier doit exécuter plusieurs algorithmes et doit donner la main à l'utilisateur pour la sélection des images. L'utilisateur doit consacrer beaucoup de son temps

s'il décide d'évaluer les profils de tous ses amis en même temps. Par contre, l'exécution de *Protect_U* est quasi instantanée. Les résultats et les recommandations sont affichés en quelques secondes.

- *Protect_U* est destiné à être exécuter au moins une fois par mois par un utilisateur dépendamment de sa fréquence d'utilisation de *Facebook*, tandis que *Protect_UFF* est appelé à être exécuté plus régulièrement (idéalement à chaque fois qu'un utilisateur ajoute un nouveau ami à sa liste).
- Il est beaucoup plus simple de maintenir et de mettre à jour *Protect_U* que *Protect_UFF*. La phase d'entrainement de ce dernier est beaucoup plus complexe et de loin plus exigeante.

Chapitre 9: Conclusion

Nous avons présenté dans cette thèse une plateforme de protection pour les utilisateurs de *Facebook*. Cette plateforme est basée sur deux systèmes : *Protect_U* et *Protect_UFF*. Le premier permet essentiellement de protéger la vie privée et les informations personnelles des utilisateurs novices. Il permet de classer leurs profils selon quatre niveaux de risque : « peu risqué », « moyennement risqué », « risqué » et « crucial ». Il leur propose ensuite des recommandations pour leur permettre de rendre leurs profils plus sécuritaires. Il fait aussi appel à l'aide de leurs « amis de confiance » pour qu'ils contribuent eux-aussi à leur protection. Le deuxième système cherche à détecter les profils suspects des « amis » des utilisateurs *Facebook*. Un profil est suspect si son utilisateur présente des symptômes élevés de narcissisme, de manque d'émotion et d'agressivité ou d'antisociabilité. Selon les psychologues, ces attributs sont des points marquants des traits de personnalité des psychopathes, des fraudeurs à cols blancs et des criminels. *Protect_UFF* se charge alors d'analyser le texte et les images publiques affichés sur les comptes des « amis » pour mesurer le niveau de ces trois attributs. Il applique ensuite un modèle de classification pour classer les profils selon trois niveaux : peu risqué (classe A), moyennement risqué (classe B) et potentiellement risqué (classe C). Ainsi, les utilisateurs seront plus avisés et pourront prendre leurs précautions vis-à-vis des « amis ». Ils pourront choisir, d'une façon éclairée, avec qui ils aimeraient communiquer et avec qui ils préfèreraient garder une certaine distance.

Cette étude couvre un aspect important dans la lutte contre les menaces qui guettent les utilisateurs des réseaux sociaux. Notre plateforme pourra contribuer davantage à défendre ces utilisateurs en l'enrichissant de différentes façons :

- Il sera intéressant d'étudier l'impact de l'ajout de nouveaux attributs à ceux qu'on a déjà considérés dans notre recherche.
- Il est primordial d'avoir des corpus textes constitués de plusieurs langues afin que des utilisateurs de différents pays puissent exécuter notre plateforme.

- Il sera nécessaire dans le futur d'allouer des poids différents pour les mots et les expressions des corpus textes selon leur niveau d'impact. Ce qui pourra mener à de meilleurs résultats.
- Il sera souhaitable d'analyser et de mesurer automatiquement le niveau d'agressivité des images sans avoir recours à chaque fois à l'utilisateur. Pour cela, il faut travailler à améliorer les algorithmes de comparaison d'images et de fouille d'objets. Il y a beaucoup de travail à faire à ce niveau-là.
- Il sera indispensable, pour avoir de bons résultats, d'inclure aussi dans notre plateforme l'analyse des vidéos qui sont devenus un élément incontournable dans les réseaux sociaux.
- Il sera aussi important d'enrichir notre plateforme pour tenir compte des traits de personnalité d'autres prédateurs tels que les agresseurs sexuels.
- La plateforme peut être enrichie en créant un nouveau système qui sera dédié à protéger les utilisateurs de commerces électroniques: compagnies, acheteurs, clients potentiels, etc. Dans ce contexte, l'analyse des traits de personnalité des fraudeurs économiques sera sûrement un plus.
- Il sera important de rendre notre plateforme polyvalente de façon à ce que plusieurs réseaux sociaux puissent l'exécuter et ne pas se limiter à seulement *Facebook*.

Pour terminer, cette étude a établi une nouvelle base pour protéger les utilisateurs de *Facebook*. Il est sûrement prétentieux de dire qu'elle va repérer les profils de tous les fraudeurs, cependant elle peut constituer un obstacle de taille contre ces utilisateurs virtuels qui sont constamment à la recherche de victimes innocentes et de proies faciles sur Internet. Très souvent, les guerres justes se gagnent en termes de centimètres et non en termes de kilomètres…

Bibliographie

Adu-Oppong, Gardiner, Kapadia, and Tsang. "Social Circles: Tackling Privacy in Social Networks ." *The 4th Symposium on Usable Privacy and Security (SOUPS)*, 2008.

Aimeur, E., G. Brassard, J. M. Fernandez, F.S. Mani Onana, et Z. Rakowski. «Experimental Demonstration of a Hybrid Privacy-Preserving Recommender System.» ARES, 2008. p. 161-170.

Aimeur, E., S. Gambs, et H. Ai. «Towards a Privacy-Enhanced Social Networking Site.» *Availability, Reliability and Security ARES'10 International Conference*, 2010: p. 172-179.

Alalehto, T. "Economic Crime: Does Personality Matter?" *International Journal Of Offender Therapy And Comparative Criminology*, 2003: 335-355.

Albrecht, S. W., C. C. Albrecht, et C. O. Albrecht. *Fraud Examination.* Mason, OH: Thomson South-Western, 2006.

Allport, G.W. "Personality: a psychological interpretation." New York: Holt, 1937.

American Psychiatric Association. "Diagnostic and statistical manual of mental disorders (4 ed)." *American Psychiatric Association*, 1994.

Anti-Phishing Working Group. «Phishing Activity Trends Report.» 2011.

Balduzzi, Platzer, Holz, Kirda, Balzarotti, and Kruegel. "Abusing social networks for automated user profiling ." *RAID'2010, 13th International Symposium on Recent Advances in Intrusion Detection*, 2010.

Beck, A. T., A. Freeman, et D. D. Davis. «Cognitive therapy of personnality disorders (2 ed.).» *Guilford Press*, 2003.

Bin, and Jian. "Preserving Privacy in Social Networks Against Neighborhood Attacks." *Data Engineering IEEE 24th International Conference.* ICDE, 2008. p. 506-515.

Blum, R. H. *Deceivers and deceived; observations on confidence men and.* Springfield: Thomas, 1972.

Bonneau, J., et S. Preibusch. *The Privacy Jungle: On the Market for Data Protection in Social Networks.* WEIS '09, 2009.

Bramer, M.A. *Principles of Data Mining.* London, UK: Springer, 2007.

Bromberg, W. *Crime and the mind; a psychiatric analysis of crime and punishment.* New York: Macmillan, 1965.

CBC news. *Justice for Nadia.* 01 14, 2011. http://www.cbc.ca/fifth/2010-2011/justicefornadia/.

Centre antifraude du Canada. "Activités de fraude par marketing de masse et de vol d'identité." Rapport statistique annuel, 2011.

Cleckley, H. "The mask of insanity." 1988, 6 ed.

Collée, Laurent. *Sécurité et vie privée sur les réseaux sociaux.* Mémoire, Université du Luxembourg, 2009.

Collins, J.M., et F.L. Schmidt. «Personality, integrity and white collar crime: A construct validity study.» *Personal Psychology*, 1993: 295-311.

Cooke, D.J., et C. Michie. «Refining the construct of psychopathy: Towards a hierarchical model.» *Psychological Assessment*, 2001: 171-188.

Coronges, K., R. Dodge, et et al. «The Influences of Social Networks on Phishing Vulnerability.» *45th Hawaii International Conference on System Sciences.* 2012.

Côté, G. "Vers une définition de la psychopathie." Edited by Presses universitaires du Septentrion. *Psychopathie: Théorie et Recherche*, 2000: 21-46.

Cottraux, J., et I.M. Blackburn. «Thérapies cognitives des troubles de la personnalité.» 1995.

Cusson, Maurice. *Le contrôle social du crime.* Paris: Les presses universitaires de France, 1983.

Daneziz. "Inferring privacy policies for social networking services." *Proceedings of the 2nd ACM workshop on Security and artificial intelligence*, 2009: p. 5-10.

Davison, G. C., et J. M. Neale. «Abnormal Psychology.» New York, 2001.

Delord-Raynal, Y. "Le délinquant d'affaires : son profil psychologique à partir de l'observation d'audiences de jugement." *Revue internationale de criminologie et de police technique*, 1980: 271-288.

Destrempes, F. *Estimation de paramètres de champs markoviens cachés avec applications à la segmentation d'images et la localisation de formes.* Thèse, Montréal: Université de Montréal, 2006.

Duffield, G., et P. Grabosky. «The Psychology of Fraud.» *Trends & Issues in crime and criminal justice*, 2001.

Eysenck, H. J. "Crime and personality." New York: Houghton Mifflin, 1964.

Fang, L., et K. LeFevre. «Privacy Wizards for social networking sites.» *Proceeding of the 19th international conference on World wide web*, 2010: p. 351-360.

Fogel, and Nehmad. "Internet social network communities: Risk taking, trust, and privacy concerns." *Computers in Human Behavior 25*, 2009: p. 153-160.

Fogel, J., et E. Nehmad. «Internet social network communities : Risk, trust, and privacy concerns.» *Computers in Human Behavior* 25, n° 1 (1999): 153-160.

Fong, P.W.L., M. Anwar, et Z. Zhao. «A privacy preservation model for facebook-style social network systems.» 2010.

Gagnon, S. *L'évaluation de la structure de personnalité d'un échantillon de fraudeurs québécois judiciarisés.* Mémoire de Maîtrise, École de criminologie, Université de Montréal, 2008.

Gaudreau-Toutant, C. *Étude comparative de la capacité interrelationnelle chez deux groupes de fraudeurs.* Mémoire de Maîtrise, École de criminologie, Université de Montréal, 1969.

Gauthier, M. "La psychologie du faussaire récidiviste." *Thèse de Doctorat.* Université de Montréal, 1960.

Goulem, P. *Étude sur la manipulation du fraudeur incarcéré.* Mémoire de Maîtrise, École de criminologie, Université de Montréal, 1969.

Griffith, and Jakobsson. "Messin' with Texas Deriving Mother's Maiden Names Using Public Records." *ACNS*, 2005: p. 91-103.

Hare(a), R.D. "Without conscience: the disturbing world of the psychopaths among us." (The Gilford Press) 1999.

Hare(b), R.D. "The Hare Psychopathy Checklist-Revised." Edited by Multi-Health Systems. 1991.

Hare(c), R.D. "Psychopathy: a clinical construct whose time has come." *Criminal Justice and Behavior* 23, no. 1 (1996): 25-54.

Hare(d), R.D., S.D. Hart, et T.J. Harpur. «Psychopathy and the DSM-IV. Criteria for Antisocial Personality Disorder.» *Journal of Abnormal Psychology* 100, n° 3 (1991): 391-398.

Harpur, T. J., S. D. Hart, et R. D. Hare. «Personality of the psychopath.» *Personality Disorders and the Five-Factor Model of Personality*, 2002: 149-173.

Harpur, T.J., A.R. Hakstian, et R.D. Hare. «Factor structure of the Psychopathy checklist.» *Journal of Consulting and clinical Psychology* 56, n° 5 (1988): 741-747.

Harris, G.T., M.E. Rice, et V.L. Quinsey. «Psychopathy as a taxon: Evidence that psychopaths are discrete class.» *Journal of consulting and clinical psychology* 62, n° 2 (1994): 387-397.

Hart, S.D., R.D. Hare, et A.E. Forth. «Psychopathy as a risk marker for violence: Development and validation of a screening version of the revised psychopathy checklist.» *Violence and mental disorder: developments in risk assessment*, 1994: 81-98.

Heller, S. *Comment protéger les jeunes internautes ?* Université de Lausanne, 2000.

Hélou, C., A. E. Gandouz, et E. Aïmeur. «A Privacy Awareness System for Facebook Users.» *Journal of Information Security Research*, 2012: 15-29.

Hercz, R. "Psychopathes parmi nous? Entrevue avec Robert Hare." *Revue de l'Association Canadienne des policiers "Express"*, 2002: 55-56, 4-7, 14-16.

Hervé(b), H. "Psychopathic Subtypes: Historical and Contemporary Perspectives." *The psychopath : theory, research, and practice*, 2007: 431-460.

Hervé, H. "Psychopathy Across the Ages: A History of the Hare Psychopath." Edited by Lawrence Erlbaum Associates. *The psychopath: theory, research and practice*, 2007: 31-55.

Hirschi, T., et M. Gottfredson. «Causes of White-Collar Crime.» *Criminology* 25, n° 4 (1987): 949-974.

Industrie Canada. "Freinons le pourriel." 2005.

Irani, Webb, Li, and Pu. "Large Online Social Footprints-An Emerging Threat." *International Conference on Computational Science and Engineering.* Vancouver, BC, Canada: IEEE, 2009. p. 271-276.

Jackson, J. E. "Fraud Masters: Professional Credit Card Offenders and Crime." *Criminal Justice Review*, 1994: 24-55.

Kantardzic, M. *Data Mining: Concepts, Models, Methods and Algorithms.* New York, NY, USA: John Wiley & Sons, Inc., 2011.

Kaspersky Lab. *Spam report: April 2012.* 17 05 2012. http://www.securelist.com/en/analysis/204792230/Spam_Report_April_2012#9.

Kellens, G. "La criminalité des affaires. Aspects sociologiques et psychologiques." *Aspects criminologiques de la délinquance des affaires, Études relatives à la recherche criminologique* (Conseil de l'Europe) 4 (1977): 73-118.

Kelly, G. "A theory of personality: the psychology of personal constructs." New York: Norton, 1963.

Korolova, A., R. Motwani, S.U. Nabar, et Y Xu. «Link privacy in social networks.» *17th conference on Information and knowledge management.* Napa Valley, California USA: ACM, 2008. p. 289-298.

Krueger, R. F. "Perspectives on the conceptualization of psychopathy: Toward an integration." *Handbook of psychopathy*, 2006: 193-202.

Lalumière, M., et M. Seto. «Qui a-t-il d'anormal chez les psychopathes? Définition des causes et des effets de la psychopathie.» *Psychiatrie: Conférences scientifiques*, 1998.

Lavoie, M., and S. Lessard. *La fraude: approche clinique et socio-criminologique.* Montréal: Ministère de la sécurité publique, 1987.

Leblanc, Marilyn, et Madeleine Renaud. *McCarthy Tetrault.* 12 juin 2012. news.mccarthy.ca (accès le juillet 16, 2012).

Lederer, Hong, Dey, and Landay. "Personal privacy through understanding and action: five pitfalls for designers." *Personal Ubiquitous Computer*, 2004: p. 440-454.

Lemert, E. M. "Human deviance, social problems, and social control." *Englewood Cliffs*, 1972.

Levi, M. "Motivations and Criminal Careers of Long-Firm Fraudsters." *Fraud: Organization, Motivation and Control*, 1999.

Lipford, Besmer, et Watson. «Understanding privacy settings in facebook with an audience view.» *Proceedings of the 1st Conference on Usability, Psychology, and Security*, 2008: p. 1-8.

Lipford, H.R., A. Besmer, et J. Watson. «Understanding privacy settings in facebook with an audience view.» *Proceedings of the 1st Conference on Usability, Psychology, and Security*, 2008: p. 1-8.

Liu, et Terzi. «A Framework for Computing the Privacy Scores of Users in Online Social Networks.» *ACM trans. Knowl. Discov. Data 5*, 2010: p. 1-30.

Liu, K., et E. Terzi. «A Framework for Computing the Privacy Scores of Users in Online Social Networks.» *ACM trans. Knowl. Discov. Data 5*, 2010: p. 1-30.

Luci, M. "La psychopathologie des conduites criminelles." *Sciences Humaines*, 1999.

Lykken(a), D.T. «A study of anxiety in the sociopathic personality.» *Journal of Abnormal and Social Psychology*, 1957, éd. 55: 6-10.

Lykken(b), D.T. "The Antisocial Personalities." Edited by Lawrence Erlbaum Associates. N.J.: Hillsdale, 1995.

Marcus, D.K., S.L. John, et J.F. Edens. «A Taxometric Analysis of Psychopathic Personality.» *Journal of Abnormal Psychology* 113 (2004): 626-635.

Maulaz, E. "Approche psychopathologique de l'escroc : Étude menée au moyen du Rorschach et du TAT." *Bulletin de Psychologie*, 2001: 535-542.

Maurey, G. *Mentir : bienfaits et méfaits.* Bruxelles: De Boeck Université, 1996.

Maximilien, Grandison, Sun, Richardson, Guo, et Liu. «Privacy-as-a-Service: Models, Algorithms, and Results on the Facebook Platform.» *Proceedings of W2SP 2009: Web 2.0 Security and Privacy*, 2009.

Mazzia, A., K. LeFevre, et E. Adar. «The pviz compregension tool for social network privacy settings.» *UMTech Report*, 2011.

Mergen, A. "La personnalité du "criminel à col blanc." *Revue internationale de criminologie et de police technique*, 1970.

Milligan, Shelly. *Le harcèlement criminel au Canada.* Statistique Canada, 2009.

Million, T. «Historical conceptions of psychopathy in the United States and Europe.» Édité par Guilford. *Psychopathy: Antisocial, criminal, and violent behavior*, 1998: 50-68.

Morizot(a), J. «Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie: Approches centrées sur les variables et sur les personnes.» Montréal, 2003.

Morizot(b), J., et M. Le Blanc. «Searching for a developmental typology of personality and its relations to antisocial behavior.» *Journal of Personality* 73 (2005): 139-182.

Morizot, J., et D. Miranda. «Développement des traits de personnalité au cours de la vie : continuité ou changement?» *Canadian psychology* 48, n° 3 (2007): 156-173.

Namestnikov, Yuri. *Viruslist.* 05 30, 2012. file:///C:/Users/Ghost/Desktop/THESE/statistiques_codes_malveillants.htm (accessed 08 22, 2012).

Nansel, T., M. Overpeck, R. Pilla, W. Ruan, B. Simons-Morton, and P. Scheidt. "Bullying Behaviors among US Youth: Prevalence and Association with Psychosocial Adjustment." *Journal of the American Medical Association* 285 (2001): 2094-2100.

Narayanan, A., et V, Shmatikov. «De-anonymizing social networking sites.» *Proceedings of the 19th international conference on World wide web.* Raleigh, North Carolina, USA: ACM, 2009. p. 173-187.

Ninggal, M., et J. Abawajy. «Privacy Threat Analysis of Social Network Data.» *Algorithms and Architectures for Parallel Processing* (springer) 7017 (2011): 165-174.

Office québécois de la langue française. *Le grand dictionnaire terminologique.* 2012. http://www.gdt.oqlf.gouv.qc.ca/ (accès le 08 02, 2012).

Olweus, D. "School-yard bullying-grounds for." *School Safety* 6 (1987): 4-11.

Organisation Mondiale de la santé. "Classification statistique Internationale des maladies et des problèmes de santé connexes." Genève, 1992.

Paquette, Ève. *Des pensées criminelles et des traits de personnalité de fraudeurs incarcérés, sous l'angle de la psychopathie.* Montréal, 2010.

Patil, S., et A. Kobsa. «Enhancing privacy management support in instant messaging.» *Interact Computer*, 2010: p. 206-217.

Pervin, L.A., et O.P. John. *Personnalité Théorie et recherche.* Bruxelles: De Boeck, 2005.

Pinatel, J., et P. Bouzat. *Traité de droit pénal et de criminologie.* Paris: Dalloz, 1963.

Piquero, N. L., et A. Piquero. «Characteristics and Sources of White Collar Crime.» *Crimes of Privilege* (Oxford University Press), 2001: 329-341.

Poythress, N. G., et J. L. Skeem. «Disaggregating psychopathy: Where and how to look for subtype.» *Handbook of Psychopathy*, 2006: 172-192.

Puig-Verges, N., et M.-G. Schweitzer. «Escrocs, escroqueries et psychopathologie de l'expression.» *Annales médico-psychologiques*, 1996: 132-136.

Reeder, et al. "Expendable grids for visualizing and authoring computer security policies." *Proceedings of the twnty-sixth annual SIGCHI conference on Human factors in computing systems*, 2008: p. 1473-1482.

Reeder, R.W., et al. «Expendable grids for visualizing and authoring computer security policies.» *Proceedings of the twnty-sixth annual SIGCHI conference on Human factors in computing systems*, 2008: p. 1473-1482.

Réhel, S. *Catégorisation automatique de textes et cooccurrence de mots provenant de documents non étiquetés.* Maîtrise en informatique, Québec: Université de Laval, 2005.

Rhône, c., et Peter A.B. Widener. «1 R.C.S.» 1993.

Rigby, K. "Countering bullying in schools." *CAFHS Forum* 1, no. 2 (1993).

Robertson, Michael J. " A social Approach to Security using social networks to help detect malicious web content." 2010.

Rolland, J.- P. *L'évaluation de la personnalité : le modèle en cinq facteurs.* Sprimont Mardaga, 2004.

Ryan, N., P. Lavoie, B. Dupont, et Fortin F. «Fraude via médias sociaux.» Note de recherche no. 13, 2011.

Schuessler, K.F., et D.R. Cressey. «Personality characteristics of criminals.» *American journal of Sociology* (University of Chicago Press) 55, n° 5 (1950).

Sécurité publique Canada. *Stratégie de cybersécurité du Canada.* 3 10 2011. http://www.securitepublique.gc.ca/prg/ns/cbr/ccss-scc-fra.aspx.

Sevgi, K. "Vol, fraude et autres infractions semblables et Internet." *Lex Electronica*, 2007.

Skinner, Robin, et Stephen McFaull. «Suicide among children and adolescents in Canada: trends and sex differences, 1980–2008.» *Canadian Medical Association* 184 (2012): 1015-1016.

Smith, P.K., J. Mahdavi, M. Carvalho, S. Fisher, and N. Tippett. "Cyberbullying: its nature and impact in secondary school pupils." *J Child Psychol Psychiatry* 49, no. 4 (2008): 376-385.

Soman, K.P., et S. Diwakar. *Insight into Data Mining: Theory and Practice.* India: Prentice-Hall of India Pvt.Ltd., 2006.

Statistique Canada. *Les crimes haineux déclarés par la police, 2010.* 12 04 2012. http://www.statcan.gc.ca/daily-quotidien/120412/dq120412b-fra.htm (accès le 08 22, 2012).

Stone-Gross, B., et T. Holz. «The underground economy of spam: a botmaster's perspective of coordinating large-scale spam campaigns.» *Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats.* Boston, MA: USENIX, 2011.

theguardian. *Facebook sues 'clickjacking' firm.* 01 27, 2012. http://www.guardian.co.uk/technology/2012/jan/27/facebook-sues-clickjacking-firm-adscend (accessed 08 2012).

Timm, C., et R. Perez. *Seven Deadliest Social Network Attacks.* Burlington, USA: Elsevier, 2010.

Tyrer, P. J. "Personality disorders : diagnosis, management, and course." *London Wright*, 1988.

Up2social. *Le social et l'avenir de la communications.* 2011. http://up2social.com/livre-blanc-social-avenir-communication/ (accès le 2012).

Vandebosch, H., and K. van Cleemput. "Defining cyberbullying: A qualitative research into the perceptions of youngsters." *CyberPsychology & Behavior* 11 (2008): 499–503.

Walters, G. D. "The Trouble with Psychopathy as a General Theory of Crime." *International Journal of Offender Therapy and Comparative Criminology*, 2004: 133-148.

Werlinder, H. "Psychopathy: A History of the Concepts. Analysis of the origin and development of a family of concepts in psychopathology." Edited by Almquist and Wiksell. Stockholm, Sweden, 1978.

Willott, S., C. Griffin, et M. Torrance. «Snakes and Ladders: Upper-Middle Class Male Offenders Talk About Economic Crime.» *Criminology*, 2001: 441-466.

Yan, and Ahmad. "A low-cost attack on a Microsoft captcha." *15th ACM conference on Computer and communications security.* Virginia, USA: ACM, 2008. p. 543-554.

Yochelson, S., et S.E. Samenow. «The criminal personality.» 1976.

Zheleva, and Getoor. "To join or not to join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles." *18th International World Wide Web conference (WWW)*, 2009.

Zhen, Zhao. "A privacy preservation model for facebook-style social network systems." 2010.

Annexe 1 – Description de l'échelle de psychopathie révisée (PCL-R) de Robert Hare

	Critères	Description
1.	Loquacité - charme superficiel	Parle avec aisance. Il a la répartie facile et se présente sous un jour favorable. Il tente d'impressionner avec des mots recherchés.
2.	Surestimation de soi	Écart entre l'image qu'il a de lui et ses réalisations passées.
3.	Besoin de stimulation - tendance à s'ennuyer	Drogue, loisir extrême. Les activités routinières sont insupportables.
4.	Tendance au mensonge pathologique	Le mensonge est une habitude. Il est convaincant, il sait jouer avec les sentiments.
5.	Duperie - Manipulation	Il manipule dans un but précis.
6.	Absence de remords ou de culpabilité	Il nie les conséquences de ses actes. Il peut prétendre avoir des remords. (Situation toujours spécifique à un délit particulier).
7.	Affects superficiels	Émotions exprimées de façon théâtrale, voire exagérée. Sauf pour la colère et la haine qui sont authentiques.
8.	Insensibilité – manque d'empathie	Incapable de comprendre les besoins et les souffrances d'autrui. Concerne toutes les sphères de la vie.
9.	Tendance au parasitisme	Il a les aptitudes pour travailler, mais il ne le fait pas. Il vit au dépend de quelqu'un d'autre, soit des fruits de sa criminalité.
10.	Faible maîtrise de soi	Se met en colère facilement (explosion). Ceci inclut l'impulsivité.
11.	Promiscuité sexuelle - Sexualité débridée	Sexualité sans implication émotive. Pas nécessairement de comportements hors normes. Multiplicité des partenaires.
12.	Apparition précoce de problème de comportement	Avant 12 ans. Absentéisme élevé à l'école, suspension scolaire, renvois de l'école et même placement dans un centre de rééducation.
13.	Incapacité à planifier à long terme de façon réaliste	N'a pas de projet à long terme et vit bien avec cette situation. Change fréquemment de projets de vie et ceux-ci ne sont pas réalisables.
14.	Impulsivité	Il prend des décisions sur l'impulsion du moment. Ce concept est inclus dans la faible maîtrise de soi.
15.	Irresponsabilité	Aspect comportemental. Par exemple, ne paie pas ses dettes. Également, il peut émettre des comportements qui peuvent mettre la vie des autres en danger.
16.	Incapacité d'assumer la responsabilité de ses faits et gestes	Aspect cognitif. Ce n'est pas de sa faute s'il commet des crimes ou s'il ne peut prendre soin de sa famille. Lieu de contrôle externe.
17.	Instabilité conjugale	Nombreuses cohabitations de courte durée.
18.	Délinquance juvénile	Entre 13 et 17 ans.
19.	Violation des conditions de remise en liberté conditionnelle.	
20.	Multiplicité des types de délits	

Description de l'échelle de psychopathie révisée (PCL-R) (Suite)

- Cotation
  0 : ne caractérise pas le sujet.
  1 : le définit bien à certains égards mais sous réserves ou doute.
  2 : caractérise dans l'ensemble assez bien.

  Score total possible entre 0 et 40.

- Point de coupure pour la majorité des études (Pham 2000)
  Diagnostic de psychopathie: entre 29 et 40.
  Psychopathie mixte : entre 20 et 28.
  Absence de psychopathie : moins de 20.

- Les deux principaux facteurs
  Facteur 1 - Traits de personnalité: items 1, 2, 4, 5, 6, 7, 8, 16.
  Facteur 2 - Comportements antisociaux: items 3, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20.

Annexe 2 – Questions de la première étape du module de classification de *Protect_U*

Groupe 1

1- Affichez-vous dans votre profil votre date de naissance ?
2- Affichez-vous dans votre profil l'adresse de votre résidence ou celle de votre lieu de travail ?
3- Rendez-vous publique votre adresse électronique ?
4- Affichez-vous vos opinions politiques ?
5- Affichez-vous votre orientation sexuelle ?
6- Avez-vous précisé votre religion ?

Groupe 2

7- Avez-vous reçu des commentaires agressifs ?
8- Avez-vous reçu des commentaires discriminatoires?
9- Avez-vous reçu des commentaires indécents?
10- Avez-vous reçu des invitations suspectes (suicide, magie noire, …) ?
11- Avez-vous senti que votre vie privée était en danger ?
12- Avez-vous reçu des commentaires haineux ?

Groupe 3

13- Avez-vous été victime de vol d'identité ?
14- Avez-vous été victime de fraude ?
15- Avez-vous été attaqué physiquement à cause de votre compte?
16- Avez-vous été victime de harcèlement ?
17- Avez-vous subi des dommages moraux à cause de votre compte ?
18- Avez-vous reçu des menaces ?

Annexe 3 – Exemples de recommandations que peut proposer le module de recommandations de *Protect_U*

1) **Vous avez trop de contacts dans votre liste d'amis.** Supprimez ceux avec qui vous ne communiquez pas, ou bien créez des listes d'amis avec des droits d'accès restreins. Plus le nombre d'amis est élevé plus il y a un risque d'atteinte à votre vie privée.

2) **Trop de publications sont affichées sur votre mur.** Un grand nombre d'informations publiées sur votre mur augmente le risque de récolte de renseignements personnels de votre profil. Faites attention au contenu de ce que vous partagez avec vos amis.

3) **Vous êtes abonné à un grand nombre de groupes.** Récolter le nom de ces groupes permet d'extraire beaucoup d'informations sur vos habitudes, vos préférences et votre vie privée. Essayez de vous désabonner du maximum de ces sites, surtout si vous n'êtes plus intéressé par leurs contenus ou s'il n'y a plus d'activités sur ces sites.

4) **Vos images sont accessibles à beaucoup de monde.** Il est impératif de les protéger surtout s'il s'agit de photos personnelles. Limitez l'accès aux photos potentiellement compromettantes pour qu'elles ne soient visibles que par les personnes concernées.

5) **Certaines données saisies dans votre compte peuvent porter atteinte à votre vie privée.** Ne remplissez pas tous les champs disponibles sur *Facebook* et surtout les informations liées à votre vie privée comme votre opinion politique, votre religion, votre adresse ou votre numéro de téléphone.

6) **Votre compte contient beaucoup d'images.** Mettre beaucoup d'images de voyages que vous avez faits, de soirées entre amis ou d'évènements familiaux peut vous nuire. Réduisez le nombre de vos images privées et supprimez toutes les images qui donnent des détails sur votre vie personnelle. Dans le cas où vous cherchez quand même à partager certaines images personnelles, limitez l'accès à la personne concernée par ces images.

7) Félicitations! Vos données personnelles sont bien protégées. Aucune action n'est exigée de votre part.

Annexe 4 – Questions auxquelles un *ami de confiance* doit répondre

Trouvez-vous que votre ami:

- possède beaucoup d'amis?
- a un nombre élevé de publications?
- adhère à trop de groupes?
- partage beaucoup d'images?
- partage des données sensibles?

Commentaires :