Rapport de synthèse sur le projet :

Protection de la vie privée sur les réseaux sociaux

Juin 2018

1 Introduction

Afin de bénéficier du pouvoir social des réseaux sociaux, les utilisateurs ont tendance à être plus actifs et à partager plus de contenus dans une quête de renommée, de richesse, d'emploi ou simplement d'interactions sociales. Cependant, ils sont incapables d'évaluer les risques que des informations sensibles soient déduites sur eux-mêmes. Même les utilisateurs avertis qui se soucient de leur vie privée peuvent être exposés au risque de divulgation des informations personnelles (telles que leur opinion politique et leur orientation sexuelle) en se basant sur des informations inoffensives et anodines mais corrélées comme, par exemple, les couleurs, les musiques et les auteurs préférés.

Dans ce rapport de synthèse, nous allons présenter les techniques développées au cours de notre projet afin d'aider les utilisateurs des réseaux sociaux à évaluer le risque qu'un tiers découvre leur réseau d'amis et déduise des valeurs de leurs attributs sensibles. La sensibilité d'un attribut est une notion subjective qui peut différer d'un utilisateur à l'autre. Certains utilisateurs peuvent considérer les opinions politiques ou l'origine ethnique comme sensibles alors que d'autres les considèrent comme inoffensives. Nous avons identifié à travers une enquête les attributs les plus sensibles pour un échantillon situé en France.

Les utilisateurs doivent manipuler avec soin leurs publications concernant les attributs corrélés à l'attribut sensible afin de protéger leur vie privée. Par exemple, si la musique est très corrélée à la politique dans le réseau social, les utilisateurs doivent prêter attention aux musiques qu'ils publient afin de préserver le secret de leurs opinions politiques. La corrélation peut être complexe à comprendre et/ou inattendue pour les utilisateurs standards. C'est pourquoi nous proposons un outil pour les aider à contrôler leurs publications : une fois qu'un utilisateur a défini l'attribut sensible qu'il souhaite cacher autant que possible, l'outil vérifie si d'autres attributs publiés donnent des indications sur cet attribut sensible et quels attributs sont les plus révélateurs. Si tel est le cas, et comme recommandations de protection, l'utilisateur peut supprimer ses préférences concernant cet attribut afin de diminuer ou annuler la corrélation. Et par conséquent, ces actions de protection permettent de réduire ou éliminer le risque sur la vie privée. L'outil est conçu pour fonctionner raisonnablement avec les ressources limitées d'un ordinateur personnel, en collectant et traitant une partie relativement petite des données sociales.

Dans la section 2, nous proposons une définition des sujets sensibles. Cette définition est basée sur le comportement des utilisateurs des réseaux sociaux qui ont participé à notre enquête par questionnaire en 2015. Dans la Section 3, nous détaillons l'architecture et les fonctionnalités de notre système d'audit de la vie privée sur les réseaux sociaux : SONSAI. SONSAI évalue la vulnérabilité des utilisateurs de Facebook face aux attaques de prédiction des liens d'amitié et aux attaques d'inférence de valeur d'attributs sensibles. Il est constitué de deux outils : (i) un collecteur qui explore le réseau social et échantillonne des données pour les extraire ; (ii) un analyseur qui examine les données collectées et affiche les résultats des attaques tout en indiquant les informations qui ont joué un rôle important dans la réalisation de ces attaques et ce pour aider l'utilisateur à se protéger contre ces attaques. Cette section présente également les expériences menées sur des profils Facebook pour tester les algorithmes de SONSAI. Nous concluons ce rapport en rappelant nos contributions et en présentant ensuite quelques travaux futurs qui devraient permettre l'amélioration des résultats

obtenus durant ce projet.

2 Définition des sujets sensibles

Afin de lutter contre les fuites des informations sensibles, il est important de définir quelles informations personnelles sont sensibles. Certains chercheurs considèrent que toutes les informations non publiées par un utilisateur donné sont sensibles pour lui [Ryu et al., 2013, Vidyalakshmi et al., 2016]. Alors que d'autres choisissent quelques informations et les considèrent comme sensibles, comme l'affiliation politique [Heatherly et al., 2013, Conover et al., 2011], l'âge [Perozzi and Skiena, 2015] et l'orientation sexuelle [Heatherly et al., 2013]. La CNIL (la loi relative à l'informatique, aux fichiers et aux libertés) donne aussi une définition sur l'information sensible. Cependant, les réseaux sociaux évoluent plus vite que la loi. Par exemple, les données de santé n'ont pas été considérées sensibles par la loi française du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (version 1978). Elles ont été jugées sensibles beaucoup plus tard.

Il est également possible de s'appuyer sur une définition de sujet sensible donnée par les médias sociaux eux-mêmes. Par exemple, selon Google, les données sensibles sont ≪*relatives à des faits médicaux confidentiels, à des origines raciales ou ethniques, à des croyances politiques ou religieuses ou à la sexualité*≫ [Google, 2018]. Mais comment pouvons-nous faire confiance aux réseaux sociaux dans la définition de ce qui est sensible ou non, sachant qu'ils tirent le meilleur parti de leurs profits en utilisant des informations personnelles pour une publicité ciblée ?

Nous avons mené une enquête par questionnaire pour définir des sujets sensibles en fonction du comportement des internautes français. Cette méthode a l'avantage d'être rapide, précise et peut facilement être mise à jour. Les sujets sensibles sont définis par les utilisateurs eux-mêmes au lieu d'être imposés par les réseaux sociaux ou les lois. Il est possible de prendre en compte les résultats statistiques de nouvelles enquêtes plus récentes afin de mettre à jour la définition sans répéter l'ensemble du processus. En outre, l'enquête nous a permis d'évaluer la vulnérabilité des internautes à certaines attaques de confidentialité.

Notre échantillon compte 232 utilisateurs de médias sociaux qui ont fourni des réponses valides et cohérentes. Ces utilisateurs sont situés dans 21 régions françaises et ont suivi plus de 18 disciplines différentes d'étude.

Nous avons classé les sujets discutés sur les médias sociaux selon quatre critères : le taux de discussion sur les réseaux sociaux, le taux de discussion sur les forums et les sites web, le taux de publication anonyme et les sujets évités. Sur la base de ces critères, nous avons proposé une définition des sujets sensibles. Ensuite, nous avons calculé le coefficient de sensibilité des sujets étudiés dans notre enquête. Parmi les 25 sujets analysés dans l'enquête, nous avons défini 6 sujets sensibles comme représenté par le Tableau 1. Les sujets délicats sont évités **ou** [1] dont le taux de publication anonyme sur les forums et sites web est supérieur à la moyenne de toutes les publications. Les sujets épineux sont délicats **et** dont le taux de discussion sur les forums et sites web **ou** les réseaux sociaux sont en dessous du seuil de la moyenne de toutes les discussions. Les

[1] Ou indique une disjonction inclusive.

Sujets	délicat	épineux	controversé	sensible
Argent	×	×	×	×
Religion, Libre-pensée	×	×	×	×
Achats	×	×	×	×
Rencontre	×	×		×
Santé	×		×	×
Politique	×		×	×
Famille	×			
Actualité	×			
Travail	×			
Voyages	×			
Sport	×			
Art	×			
Jeux	×			
Cuisine	×			
Mode				
Émission de télévision				
Études				
Technologie				
Musique	×	**Information insuffisante**		
Film	×			
Humour	×			
Livre	×			
Maison	×			
Astuce	×			
Sexe	×			

Table 1: Sujets analysés par notre enquête.

sujets controversés sont les sujets évités **et** dont le taux de publication anonyme sur les forums et sites web est supérieur à la moyenne. Les sujets sensibles sont épineux **ou** controversés.

Notre méthode basée sur la sensibilité peut être enrichie par d'autres études statistiques pour analyser la sensibilité de plus de sujets (par exemple, la sexualité). Le coefficient de sensibilité varie de 0 à 4. Plus le coefficient de sensibilité est élevé, plus le sujet correspondant est sensible. Plus le taux de discussion d'un sujet donné est élevé, plus son coefficient de sensibilité est faible. Cependant, plus le taux de publication anonyme d'un sujet donné est élevé et plus il est évité, plus son coefficient de sensibilité est élevé. Le Tableau 2 trie les sujets sensibles par ordre décroissant du plus sensible au moins sensible sur les réseaux sociaux.

Enfin, nous avons analysé le comportement des internautes, actifs dans des médias sociaux, afin d'identifier certaines vulnérabilités sur la vie privée. Environ 76% des parents confirment qu'ils ne contrôlent pas ce que leurs enfants peuvent découvrir sur les réseaux sociaux. De plus, environ 77,63% des internautes sont vulnérables aux attaques de croisement de profils entre différents médias sociaux car ils utilisent des e-mails ou des pseudos similaires. Par ailleurs, environ 65,25% des internautes sont exposés au risque de fuite d'informations sensibles sur le même réseau social. Enfin, notre étude montre que plus de 70% des utilisateurs de médias sociaux sont exposés au risque de fuite d'informations

Sujet x	Coefficient de sensibilité $C(x)$
Religion	2.25
Argent	2.18
Politique	2.08
Rencontre	2.00
Achats	1.85
Santé	1.63

Table 2: Ordre décroissant des sujets sensibles

sensibles, principalement dû à une utilisation maladroite des médias sociaux et à une méconnaissance des problèmes liés à la vie privée.

3 SONSAI : Outil de sensibilisation et d'aide à la protection de la vie privée

Dans ce travail, nous visons à fournir aux utilisateurs des réseaux sociaux un outil pour protéger leurs données personnelles. À cette fin, nous étudions les attaques potentielles sur la vie privée. Nous analysons la faisabilité ainsi que l'impact de chaque attaque. Cette approche nous permet de mettre la main sur l'origine des menaces. Concrètement, nous concevons des attaques en ligne sur le plus grand réseau social du monde, Facebook. Plusieurs expériences ont été réalisées pour tester les attaques en ligne sur plusieurs profils issus de Facebook.

Afin de lutter efficacement contre les fuites de l'information sensible, il est très important de prendre en compte la combinaison d'attaques (prédiction de lien et prédiction d'attribut). En fait, ces attaques sont étroitement liées et lorsqu'elles sont combinées, elles présentent des menaces plus importantes pour la vie privée. Par exemple, un adversaire peut effectuer des attaques de prédiction de lien afin de dévoiler le réseau local de sa cible (les amis et groupes de la cible). Ensuite, il peut effectuer une attaque de prédiction d'attribut basée sur le réseau local découvert.

3.1 Attaque de prédiction de liens en ligne

Un réseau social peut être défini comme un site web qui permet aux utilisateurs de créer des pages personnelles afin de partager des informations avec leurs amis et connaissances. Ces pages sont généralement appelées profils et contiennent des informations personnelles. Les profils sont connectés les uns aux autres par le biais de liens d'amitié qui peuvent être symétriques ou asymétriques, selon la politique du réseau. Pour imiter les interactions sociétales réelles (c'est-à-dire non cybernétiques), certains réseaux sociaux tels que Facebook, Linkedin et Viadeo permettent la création de groupes en plus de la création de profils. Les profils sont connectés à des groupes via des liens d'adhésion.

Une attaque de prédiction de liens en ligne permet d'établir avec certitude/incertitude une relation entre deux utilisateurs ou une adhésion d'un utilisateur à un groupe de discussion. Afin d'effectuer cette attaque, il est important de prendre en compte sa faisabilité. Par exemple, pour trouver des liens sur

Facebook, un adversaire peut être tenté de vérifier les listes d'amis publiques des utilisateurs de Facebook dans l'espoir de trouver la cible dans ces listes. Cependant, le moteur de recherche de Facebook fournit des résultats incomplets et aléatoires. Facebook compte environ 2 milliards d'utilisateurs actifs par mois et une recherche aléatoire peut durer des années. De plus, Facebook est très dynamique. Par exemple, chaque seconde, 5 nouveaux profils sont créés et 8 500 commentaires sont affichés [Noyes, 2018]. Ainsi, les attaques basées sur la reconnaissance de motif de réseau sont assez difficiles à réaliser. Le but de ces attaques est d'identifier une partie du réseau où la structure des connexions entre les utilisateurs est connue, par exemple une structure de connexions similaire au réseau de connaissance dans la vraie vie.

Nous avons conçu une stratégie d'attaque de divulgation de liens en ligne (avec certitude). La stratégie proposée est passive : l'adversaire n'a pas besoin d'interagir avec sa cible pour éviter d'attirer son attention. Notre attaque est réalisée sur un réseau social supportant plusieurs niveaux de visibilité (secret, amis seulement, amis et leurs amis, tout le monde) et a été testée en ligne sur des profils Facebook contrairement à l'inférence de liens hors ligne (avec incertitude) proposée dans [Wang et al., 2015, Gao et al., 2015] et l'attaque active par divulgation de liens dans [Jin et al., 2013] effectuée sur un réseau social supportant deux niveaux de visibilité (secret et amis directs). L'attaque décrite dans [Jin et al., 2013] divulgue l'amitié par le biais d'une requête d'ami commun mais elle n'a pas été testée en ligne. En outre, en explorant efficacement le réseau social, notre stratégie est capable d'effectuer des attaques de révélation de groupes, d'amitié et d'amis communs en minimisant le nombre de requêtes. Seules les requêtes légitimes sont utilisées pour effectuer des attaques (c'est-à-dire des requêtes et des outils fournis par le réseau social ciblé). Notre étude exploite la relation intrinsèque entre les communautés (habituellement représentées en tant que groupes) et les amitiés entre individus. Pour développer une attaque efficace, nous avons analysé les distributions de groupes, les densités et les paramètres de visibilité d'un échantillon d'utilisateurs Facebook.

Le résultat de nos tests effectués sur 14 517 profils Facebook montre que la probabilité pour un utilisateur de rejoindre au moins un groupe réunissant moins de 50 membres et de publier son adhésion à celui-ci est de 0,49. Ainsi, environ la moitié des profils analysés sont exposés au risque de divulgation de liens d'amitié par des groupes auxquels ils adhèrent et qui regroupent moins de 50 membres. Le nombre espéré de liens d'amitié publiés entre un membre donné et tous les autres membres du même groupe g est taille(g)$\times$ densité(g) (où densité(g) est le rapport entre le nombre de liens existants et le nombre total de liens). L'analyse de 1 100 groupes Facebook, dont la taille est comprise entre 2 et 80, montre que le nombre espéré de liens divulgués entre les membres cibles et les groupes augmente de 2 lorsque la taille des groupes augmente de 10. Les groupes et les membres peuvent choisir de publier ou cacher la relation d'adhésion. Nous avons conçu une attaque pour dévoiler des groupes autour de plusieurs cibles données. Le réseau de groupe autour de chaque cible est ensuite utilisé pour divulguer les liens d'amitié et d'appartenance à des groupes qu'elle cache. Les résultats des attaques effectuées sur les profils Facebook actifs montrent que 5 liens d'amitié différents sont divulgués en moyenne pour chaque requête.

3.2 Attaque de prédiction d'attributs

L'attaque de prédiction d'attribut permet de révéler avec une certaine probabilité la valeur d'un attribut même si celle-ci a été cachée. Cette attaque comprend deux étapes: (i) la collecte et (ii) l'analyse des données. La collecte des données doit être rapide, sélective, passive et indétectable. En effet, les réseaux sociaux sont très dynamiques et contiennent de gros volumes de données.

Pour l'analyse de données, nous avons conçu un algorithme rapide et précis qui peut examiner des informations incomplètes à cause des contraintes de la collecte. En effet, le collecteur échantillonne les données à collecter pour limiter le temps de collecte et diminuer le nombre de requête de collecte. Un grand trafic de collecte peut facilement être signalé par le réseau comme une attaque. L'ensemble du processus d'analyse ne dépasse pas quelques minutes afin d'identifier rapidement le périmètre des menaces et concevoir des contremesures dans un travail future. Le processus d'analyse comprend deux étapes: (i) quantifier l'importance de chaque attribut collecté et (ii) les utiliser pour déduire les valeurs secrètes de l'attribut sensible de la cible.

Pour quantifier l'importance des attributs collectés, nous avons conçu un algorithme pour évaluer et trier les attributs. Cet algorithme peut rapidement détecter et quantifier la corrélation entre les attributs. Par exemple, il peut détecter la corrélation entre les préférences politique et musicale. De plus, l'amitié est considérée comme un attribut parmi d'autres. Par conséquent, l'influence de l'amitié est également prise en compte lors de l'évaluation de l'importance des attributs. D'autre part, nous avons conçu un algorithme pour regrouper des valeurs similaires d'attributs afin d'accélérer le processus d'inférence et traiter des informations incomplètes. Nous distinguons deux types d'attributs. Le premier type comprend des attributs pouvant avoir plusieurs valeurs arbitraires. Les profils peuvent être connectés à plusieurs valeurs du même attribut telles que des pages sur les livres. Le deuxième type concerne des attributs qui ont des valeurs prédéfinies : les profils peuvent être connectés à au plus une valeur du même attribut, comme le genre. Par conséquent, nous avons défini deux méthodes pour analyser les données selon le type de l'attribut sensible. Lorsque l'attribut sensible est un attribut qui admet des valeurs arbitraires, l'analyseur quantifie la corrélation entre les attributs en fonction de la similarité des valeurs préférées par les utilisateurs. Par exemple, si les utilisateurs qui aiment le même genre de musique aiment le même groupe de politiciens, alors la corrélation entre politicien et musique est élevée. Cependant, lorsque l'attribut sensible est un attribut qui possède un petit ensemble de valeurs prédéfinies, l'analyseur quantifie la corrélation entre les attributs en fonction du pouvoir de discrimination des valeurs préférées des utilisateurs. Par exemple, si la plupart des utilisateurs qui aiment les pages de football sont des hommes, alors la discrimination entre le genre et le football est élevée.

Les attributs les plus corrélés à l'attribut sensible sont ensuite utilisés pour inférer les préférences de la cible concernant l'attribut sensible. L'analyse consiste à calculer les probabilités de préférence de la cible concernant les valeurs de l'attribut sensible en comparant ses autres préférences aux préférences des autres utilisateurs qui ont publié leurs valeurs sensibles.

Pour tester nos algorithmes, nous avons mené plusieurs expériences sur de grands ensembles de données collectées à partir des profils Facebook. Pour chaque profil collecté, nous avons extrait la liste des pages qu'il aime, la liste

de ses amis, son genre et son état civil. Pour générer le premier ensemble de données D1, nous avons exploré le réseau d'amitié de 100 profils Facebook des utilisateurs du Nord-Est de la France. Les données sont collectées en 2016. D1 contient 1 926 types de pages différents, 1 022 847 pages différentes et 15 012 profils Facebook différents. La Table 3 détaille l'ensemble de données D1.

♯ Profils collectés	15 012	♯ Pages	1 022 847
♯ état civil	11	♯ Types de pages	1 926
♯Pages de politiciens	4 589	♯Profil qui publient leurs politiciens préférés	2 554
♯Profil qui publient leur genre	11 141	♯Profil qui publient leur état civil	2 395

Table 3: Détails sur l'ensemble de données D1.

Pour générer le deuxième ensemble de données D2, nous avons exploré le réseau d'amitié de 17 profils Facebook d'utilisateurs résidant en Île-de-France. Les données sont collectées en 2017. D2 contient 1 296 types de pages différents, 298 604 pages différentes et 6 550 profils Facebook différents. La Table 4 détaille l'ensemble de données D1.

♯ Profils collectés	6 550	♯ Pages	298 604
♯ état civil	11	♯ Types de pages	1 296
♯Profil qui publient leur genre	4 597	♯Profil qui publient leur état civil	991

Table 4: Détails sur l'ensemble de donnée D2.

La Table 5 détaille les 23 attributs les plus corrélés à l'attribut sensible "pages des politiciens". L'expérience a été menée sur D1 qui contient 1 929 attributs. Facebook définit 11 état civils différents. Pour simplifier la présentation, nous définissons deux classes d'état civils comme suit :

$E1$= {Celibataire, Divorcé, Sépare, Veuf, Compliqué}
$E2$= {Partenariat domestique, Marié, Engage, Relation, Union civile, Relation ouverte}

La Table 6 donne des détails sur les 20 attributs les plus corrélés à l'attribut sensible "état civil". L'expérience a été menée sur D2 qui contient 1 299 attributs. Le score de corrélation prend en compte le pourcentage de statut de classe d'état civil des utilisateurs ainsi que le taux d'utilisateurs qui publient à la fois leur état civil et leurs préférences. Nous remarquons que la plupart des attributs corrélés à la classe $E1$ sont axés sur les formations et les loisirs. D'autre part, la plupart des attributs corrélés à la classe $E2$ sont axés sur les entreprises. La Table 7 donne des détails sur les 20 attributs les plus corrélés à l'attribut sensible " genre". L'expérience a été menée sur D1. Le score de corrélation prend en compte le pourcentage de genre des utilisateurs ainsi que

le taux d'utilisateurs qui publient à la fois leur genre et leurs préférences. Nous constatons que la plupart des attributs corrélés aux hommes sont axés sur les sports, les jeux et les logiciels. De plus, la plupart des attributs corrélés aux femmes sont axés sur la santé, la maison et le luxe.

Attributs	♯ Profiles qui publient leurs valeurs	♯ Valeurs d'attribut
Utilisateurs	13 155	15 012
Communautés	8 118	137 338
Musiciens/Bande de musiciens	7 141	84 762
Figures publiques	6 455	28 289
Associations à but non lucratif	6 180	25 847
Artistes	5 970	31 681
Entreprises	5 939	20 750
Sites Internet	5 829	17 931
Émissions de télévision	5 778	11 876
Sites Web de divertissement	5 669	8 319
Médias/Nouvelles	5 871	14 042
Produits/Services	5 496	15 986
Sites Web d'actualités/médias	5 550	9 247
Organisations	5 328	14 738
Films	5 171	16 282
Entreprises locales	5 111	17 321
Vêtements	4 729	16 090
Gastronomies	4 763	8 422
Acteurs/Réalisateurs	4 785	10 425
Magazines	4 733	9 955
Athlètes	4 583	14 123
Pages d'application	4 396	4 244
Équipes sportives	4 309	10 433

Table 5: Les 23 attributs les plus corrélés à l'attribut sensible "pages des politiciens" dans l'ensemble de données D1.

Nous avons réalisé plusieurs expériences sur les deux ensembles de données D1 et D2. Pour chaque expérience, nous générons un nouvel ensemble de données auxiliaires à partir de l'ensemble de données original en sélectionnant tous les profils utilisateurs (cibles) qui publient leurs préférences concernant l'attribut sensible et au moins un autre attribut (les amis sont également considérés comme un attribut). Ensuite, nous supprimons toutes les préférences concernant l'attribut sensible de 10 % des profils sélectionnés (cibles). Les expériences ont consisté ensuite à inférer les préférences supprimées en analysant l'ensemble de données auxiliaires. L'analyseur utilise des techniques d'intelligence artificielle pour trier les valeurs de l'attribut sensible en fonction de leur probabilité d'être les vraies valeurs de la cible. Les valeurs suggérées de l'attribut sensible généré par l'analyseur sont ensuite comparées aux vraies valeurs de la cible pour calculer la précision de l'inférence.

Politiciens. Nous avons réalisé une expérience sur l'ensemble de données D1. L'analyseur a sélectionné les 23 attributs les plus corrélés aux attributs "pages des politiciens", comme détaillé dans la Table 5. Seules les préférences concernant les attributs sélectionnés sont ensuite analysées pour déduire les préférences

Attributs	Corrélations	Discrimination
Éducation	2.75	88.41 % E1
Collège communautaire	2.74	90.02 % E1
Agence de consultation	2.71	90.70 % E2
Site web de loisirs & sports	2.56	91.18 % E2
Site web de Maison & jardin	2.49	91.89 % E2
Automobile, Avion & Bateau	2.48	92.86 % E2
Localité	2.47	92.59 % E2
Siège social	2.46	91.18 % E2
Sites Web d'actualités/médias	2.42	90.32 % E2
Service financier	2.41	90.00 % E2
Société industrielle	2.40	89.29 % E2
Conseillère pédagogique	2.02	75.00 % E1
Cour de récréation	1.80	66.67 % E1
Téléphone/Tablette	1.70	63.64 % E1
Chirurgien plastique	1.60	60.00 % E1
Consulat & Ambassade	1.60	60.00 % E2
Équipe sportive scolaire	1.53	52.00 % E1
Bar de plongée	1.45	54.55 % E1
Vidéo	1.44	51.00 % E1
Playlist (musique)	1.41	53.04 % E1

Table 6: Les 20 attributs les plus corrélés à l'attribut sensible "État civil" dans l'ensemble de données D2

Attributs	Corrélations	Discrimination
Ligue sportive	4.22	75.97 % Mâle
Site de loisirs & sports	3.80	77.09 % Mâle
Jeux vidéo	3.66	80.16 % Mâle
Voitures	3.25	73.15 % Mâle
Équipes de sport amateur	3.03	72.86 % Mâle
Sport	2.80	73.07 % Mâle
Bijoux & Montres	2.72	56.26 % Femelle
Électronique	2.68	73.19 % Mâle
Logiciels	2.52	77.23 % Mâle
Produits de plein air et de sport	2.35	77.19 % Mâle
Magasin de vêtements féminins	2.35	77.28 % Femelle
Décoration de maison	2.29	54.60 % Femelle
Stade & Arena	2.28	74.45 % Mâle
Articles pour bébés / articles pour enfants	2.14	66.61 % Femelle
Cuisine	2.08	55.93 % Femelle
Sacs/Bagages	2.04	59.16 % Femelle
Beauté, Cosmétique et Soins Personnels	2.03	60.59 % Femelle
Magasin de cosmétiques	1.98	66.25 % Femelle
Salon de coiffure	1.92	61.44 % Femelle
Site web de Maison & jardin	1.72	55.18 % Femelle

Table 7: Les 20 attributs les plus corrélés à l'attribut sensible "genre" dans l'ensemble de données D2.

des cibles concernant l'attribut sensible "pages de politiciens". La précision de l'inférence est égale à 79 %. En d'autres termes, en moyenne, l'ensemble inféré de pages de politiciens par l'analyseur est 79% semblable à celui réellement aimé

par la cible. Cependant, la précision d'inférence lorsque les 23 attributs corrélés sont sélectionnés de façon aléatoire est seulement de 41%. Nous avons effectué plus de tests en sélectionnant manuellement 3 attributs sémantiquement proches de la politique: organisations politiques, partis politiques et idéologies politiques. Bien que les attributs sélectionnés semblent prometteurs, la précision de l'inférence n'est que de 46%. Cela peut s'expliquer par le fait que de nombreux utilisateurs sont vigilants et ne publient pas leurs préférences concernant ces attributs. Par conséquent, l'algorithme d'apprentissage ne peut pas les exploiter correctement car il ne dispose pas d'informations suffisantes sur les préférences.

Comme la musique et la politique sont empiriquement connues pour être corrélées [Street, 2012], nous vérifions la capacité de nos algorithmes à déduire les politiciens préférés des utilisateurs de Facebook en se basant uniquement sur les attributs relatifs à la musique. Nous avons sélectionné seulement deux attributs : genres musicaux et musiciens/bandes de musiciens. La précision de l'inférence dans cette expérience est égale à 62 %, qui est donc significative. Nous notons que l'attribut musiciens/bandes de musiciens a été automatiquement sélectionné par l'analyseur. La Table 8 résume les résultats des expériences conduites.

Sélection d'attributs corrélés	Précisions en %	♯ Cibles	♯ Préférences supprimées
Sélection automatique par algorithme (23 attributs)	79%	252	409
Sélection des attributs de musiques (2 attributs)	62%	233	379
Sélection des attributs politiques (3 attributs)	46%	123	297
Sélection aléatoire (23 attributs)	41%	204 (moyenne)	351 (moyenne)

Table 8: Résultats expérimentaux d'inférence des pages des politiciens.

État civil. Nous menons cette expérience sur les deux ensembles de données D1 et D2. Le Tableau 9 donne plus de détails sur les utilisateurs qui publient leur état civil dans les deux ensembles de données.

Classe d'état civil		E1	E2
♯ profils qui publient	D1 (15 012 profils)	1 114	1 281
	D2 (6 550 profils)	208	783

Table 9: État civil des utilisateurs dans D1 et D2.

La précision de l'inférence de l'état civil est supérieure à 70% dans les deux ensembles de données D1 et D2 dès que la cible publie ses préférences concernant au moins 4 attributs parmi les 20 attributs les plus corrélés à l'état civil.

Genres. Nous avons mené cette expérience sur les deux ensembles de données D1 et D2. La Table 10 donne plus de détails sur les utilisateurs qui publient leur genre dans les deux ensembles de données.

Genres		Femelle	Mâle
♯ profils qui publient	D1 (15 012 profils)	4 491	6 650
	D2 (6 550 profils)	1 606	2 991

Table 10: Genres des utilisateurs dans D1 et D2.

La précision de l'inférence du genre est supérieure à 83% dans D1 et supérieure à 67% dans D2 dès que la cible publie ses préférences concernant au moins deux attributs parmi les 20 principaux attributs les plus corrélés au genre.

Le temps de traitement La Table 11 affiche les temps de traitement. La vitesse d'horloge du processeur est de 2,3 GHz. La machine ne dispose que de 8 Go de mémoire RAM. Grâce aux algorithmes de détection de corrélation, seules les préférences concernant les attributs importants sont analysées. Par conséquent, les tâches d'inférence sont accélérées et la précision est améliorée en écartant les informations non pertinentes.

	Attributs sensibles	Données	Corrélation	Analyse
Temps	État civil Genre	D1	7m3s	1m24s
		D2	4m30s	55s
	Politiciens	D1	13m2s	24m7s
		D2	9m34s	19m44s

Table 11: Le temps de traitement.

3.3 Mode d'emploi

Dans le système développé, SONSAI, que nous détaillons dans cette partie, un attribut sensible est initialement spécifié par l'utilisateur du système et choisi dans la liste des attributs découverts par le collecteur dans le réseau social de l'utilisateur. L'attribut sensible peut concerner des pages de politiciens ou tout autre attribut jugé sensible par l'utilisateur. SONSAI aide les utilisateurs à simuler une attaque consistant à déduire des informations sensibles les concernant afin d'évaluer leur niveau de protection. De plus, cela aide les utilisateurs à comprendre d'où vient la menace en générant une liste triée en fonction de l'importance des attributs corrélés. SONSAI est composé de deux outils principaux : un collecteur et un analyseur des données.

Collecteur. L'interface graphique de l'outil Collecteur est représentée dans la Figure1. Pour se connecter au réseau Facebook et collecter des données, l'utilisateur doit fournir son identifiant et son mot de passe. Le collecteur explorera ensuite le réseau Facebook via le compte d'utilisateur. L'utilisateur doit ensuite définir la durée de la collection sur une valeur non nulle. Il peut continuer la collection qu'il a précédemment commencée en cliquant sur le bouton "Collecter". L'utilisateur peut également mettre à jour les données déjà collectées en cliquant sur le bouton "Mettre à jour". Les données les plus anciennes sont mises à jour en premier. Enfin, il peut arrêter la collection en cliquant sur le bouton "Arrêter".

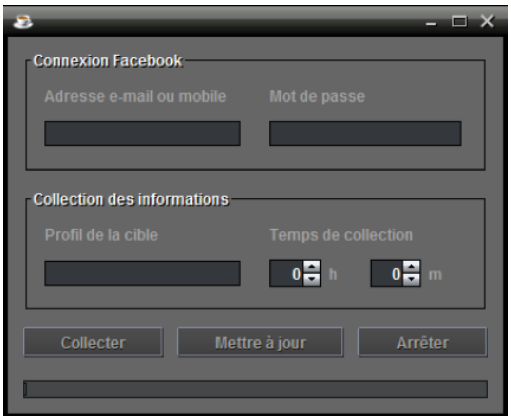

Figure 1: Interface du collecteur.

Analyseur. L'interface graphique de configuration de l'outil Analyseur est représentée dans la Figure 2. Afin de prendre en compte les données collectées récemment par le collecteur, l'utilisateur doit cliquer sur le bouton "Mettre à jour le dataset". Il peut ensuite sélectionner l'attribut sensible dans la liste de tous les attributs découverts autour de son profil. Il peut choisir la précision de l'analyse. La précision est en fait donnée comme le pourcentage d'attributs corrélés, à considérer pour l'inférence de l'attribut sensible, à partir de la collection et l'analyse de tous les attributs disponibles dans le réseau de l'utilisateur ; ces attributs sélectionnés sont ceux utilisés pour déduire les valeurs les plus proches de l'attribut sensible pour l'utilisateur. Si le nombre de valeurs concernant l'attribut sensible est énorme, l'utilisateur peut construire des groupes de valeurs, en cochant sur "Grouper les valeurs de l'attribut sensible", afin d'accélérer l'étape de l'inférence. Lorsque l'utilisateur clique sur le bouton "Analyser", la page de résultats s'affiche. Les résultats de l'analyseur sont représentés dans les Figures 3 et 4.

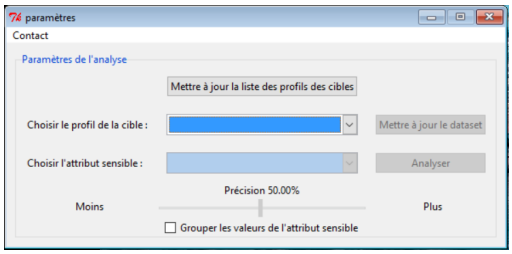

Figure 2: Configuration de l'analyseur.

Le tableau en haut à gauche de l'écran (Figure 3) résume la liste des attributs sélectionnés durant l'analyse. Chaque attribut est muni de son taux d'importance pour contribuer à la révélation des valeurs de l'attribut sensible ainsi que le nombre de valeurs publiées. Les lignes correspondant aux attributs dont l'utilisateur cible a publié certaines valeurs sont de couleur orange.

Le deuxième tableau affiché dans la partie supérieure à droite de l'écran (Figure 4) trie les valeurs de l'attribut sensible en fonction de leur proximité

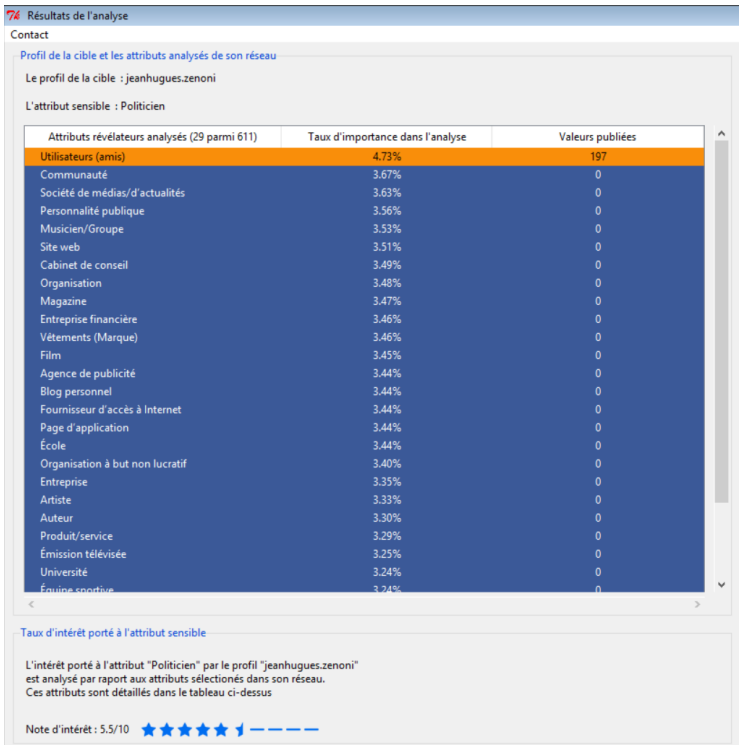

Figure 3: Résultats de l'analyseur (tableau à gauche).

avec l'utilisateur. Plus la proximité est élevée, plus la probabilité que la valeur soit la valeur réelle de l'utilisateur est élevée. Les valeurs sont regroupées dans des classes de tailles similaires d'une manière qui maximise la similitude entre les valeurs à l'intérieur d'une classe et la minimise entre les valeurs de différentes classes. La similitude est mesurée en pourcentage. Si un utilisateur aime une valeur dans une classe «c», il a tendance à aimer "x %" des valeurs de la même classe «c». L'utilisateur peut ouvrir les classes pour afficher les valeurs qu'elles contiennent. Il peut double-cliquer sur la valeur pour ouvrir sa page Facebook et modifier ses paramètre de confidentialité. Lorsque l'utilisateur publie ces valeurs, les cases correspondantes dans la colonne "Valeurs publiées" sont cochées. Dans la partie inférieure droite de l'écran (Figure 4), l'utilisateur peut évaluer le risque que ses valeurs sensibles soient déduites (par un tiers). Il doit d'abord spécifier le nombre de ses vraies valeurs dans chaque classe même s'il ne les a pas publiées sur Facebook. Puis il clique sur le bouton "Évaluer le risque". L'algorithme mesure la précision du classement. Nous définissons trois niveaux de risque d'inférence comme suit :

Si la précision est supérieure à 0,65, le risque d'inférence est élevé. En revanche, si elle est inférieure à 0,5, le risque d'inférence est faible. Si elle est comprise entre 0,5 et 0,65, le risque d'inférence est considéré comme modéré.

Par ailleurs, la partie inférieure à gauche de l'écran (Figure 3) contient l'évaluation de l'intérêt de l'utilisateur pour l'attribut sensible. Le score affiché est le produit de la moyenne des proximités de l'attribut sensible (Figure 4) par dix.

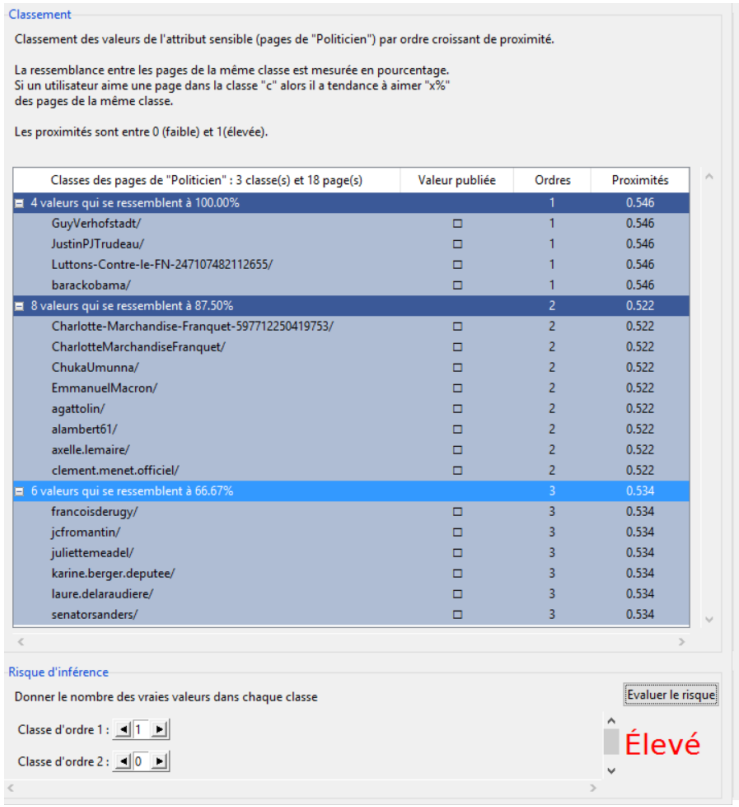

Figure 4: Résultats de l'analyseur (tableau à droite).

Recommandations de protection. Les recommandations de protection seront déduites à partir des résultats de l'analyseur comme décrits précédemment par les tableaux des Figures 3 et 4. Pour minimiser le risque sur la vie privée et rendre lointaines les valeurs inférées dans le tableau de la Figure 4, les recommandations que nous proposons à l'utilisateur est de supprimer manuellement de son profil quelques informations (par exemple, les pages les plus pertinentes qui sont en tête de liste) indiquées dans le tableau de la Figure 3.

4 Conclusion

Contributions. Dans ce projet, nous avons analysé les risques de fuite des informations sensibles sur les réseaux sociaux. Premièrement, nous avons introduit une mesure de la sensibilité des sujets discutés sur les réseaux sociaux. Les sujets les plus sensibles selon les comportements des participants français à notre étude sont *Religion, Argent, Politique, Rencontres, Achats et Santé*. Afin de déduire des informations sensibles sur une cible donnée, nous dévoilons d'abord son réseau local (à 1 saut de la cible). À cette fin, nous avons conçu et testé des attaques de divulgation de liens en ligne avec certitude. Nous avons réalisé plusieurs attaques sur de vrais profils Facebook. Nous concluons que l'adversaire peut facilement et rapidement divulguer des liens cachés (amitié

et appartenance à un groupe) avec certitude en utilisant des API des réseaux sociaux. Les données collectées autour de la cible sont ensuite traitées pour déduire les valeurs des attributs sensibles. Nos algorithmes montrent qu'il est possible de détecter et quantifier la corrélation entre les attributs. Les algorithmes que nous proposons sont intégrés dans un système appelé SONSAI qui peut être installé sur n'importe quel ordinateur commercial équipé de Windows. Il contient très peu de paramètres à définir et est conçu pour être utilisé avec des connaissances informatiques de base. Il permet aux utilisateurs de lancer un audit sur leurs réseaux locaux et détecter rapidement les fuites potentielles d'information sensible avec une bonne précision.

Par ailleurs, la CNIL a récemment approuvé et autorisé nos activités de recherche sur la protection de l'information personnelle sur les réseaux sociaux. En effet, notre projet a réussi à sensibiliser la CNIL, après un long parcours, sur l'apport de la recherche au développement d'outils permettant d'assister les internautes dans la protection de leur vie privée sur les réseaux sociaux. Pour la communication scientifique des résultats du projet, nous avons publié des articles dans des conférences nationales et internationales, à savoir :

1. Y. Abid, A. Imine, A. Di Napoli, C. Raissi and M. Rusinowitch. "Two-Phase Preference Disclosure in Attributed Social Networks". *The 28th International Conference of Database and Expert Systems Applications (DEXA)*, (LNCS 10438), Lyon, France, August, 2017.

2. Y. Abid, A. Imine, A. Di Napoli, C. Raissi and M. Rusinowitch. "Online link disclosure strategies for social networks". *The 11th International Conference on Risks and Security of Internet and Systems (CRiSIS)*, (LNCS 10158), Roscoff, France, September, 2016.

3. Y. Abid, A. Imine, A. Di Napoli, C. Raissi, M. Rigolot and M. Rusinowitch. "Analyse d'activité et exposition de la vie privée sur les médias sociaux". *16èmes journées Francophones Extraction et Gestion des Connaissances (EGC)*, Reims, France, Janvier 2016.

4. Y. Abid, A. Imine, A. Di Napoli, C. Raissi and M. Rusinowitch. "Stratégies de divulgation de lien en ligne pour les réseaux sociaux". *32ème Conférence sur la Gestion de Données (BDA)*, 15-18 Novembre, Poitiers, France, 2016.

Quant à la vulgarisation de nos travaux de recherche, nous avons participé à plusieurs activités : rédaction d'articles (Revue Préventique, News de la Fondation Maif), participation à des reportages animés par la Fondation Maif, et animation de conférences pour l'Université Populaire et Participative de Vandœuvre (UP2V).

Travaux futurs. Plusieurs pistes intéressantes méritent d'être explorées. Il serait judicieux de mener une étude d'usage de notre application SONSAI pour identifier les difficultés de son utilisation et collecter les besoins des utilisateurs pour l'améliorer. Il faudra également tester les techniques de prédiction de liens et d'attributs sur d'autres échantillons de données pour mieux les affiner. Nous avons remarqué que certains types d'attribut sont très proches et peuvent être regroupés. Par exemple, une classe de santé peut inclure des magasins

d'équipement médical, des acupuncteurs et des services médicaux. Aussi, il serait utile d'introduire des techniques de traitement du langage naturel pour aider à la classification des attributs. Nos résultats peuvent être exploités pour concevoir des contre-mesures efficaces et minimales afin de lutter contre les fuites d'informations sensibles sur les réseaux sociaux. Deux techniques principales peuvent être étudiées. La première technique consiste à supprimer des informations et des liens afin d'éviter les inférences dues au manque de données. La deuxième technique consiste à ajouter de l'information et des liens afin de modifier la précision de l'inférence en raison du désaccord sur les données. Les principaux défis dans les deux techniques sont d'équilibrer l'utilité des réseaux sociaux, la vie privée de l'utilisateur ainsi que celle de ses voisins.

References

[Conover et al., 2011] Conover, M., Gonçalves, B., Ratkiewicz, J., Flammini, A., and Menczer, F. (2011). Predicting the political alignment of twitter users. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 192–199.

[Gao et al., 2015] Gao, F., Musial, K., Cooper, C., and Tsoka, S. (2015). Link prediction methods and their accuracy for different social networks and network metrics. *Scientific Programming*, 2015:172879:1–172879:13.

[Google, 2018] Google (2018). Google privacy and terms.

[Heatherly et al., 2013] Heatherly, R., Kantarcioglu, M., and Thuraisingham, B. M. (2013). Preventing private information inference attacks on social networks. *IEEE Trans. Knowl. Data Eng.*, 25(8):1849–1862.

[Jin et al., 2013] Jin, L., Joshi, J. B. D., and Anwar, M. (2013). Mutual-friend based attacks in social network systems. *Computers & Security*, 37:15–30.

[Noyes, 2018] Noyes, D. (2018). The top 20 valuable facebook statistics.

[Perozzi and Skiena, 2015] Perozzi, B. and Skiena, S. (2015). Exact age prediction in social networks. In *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*, pages 91–92.

[Ryu et al., 2013] Ryu, E., Rong, Y., Li, J., and Machanavajjhala, A. (2013). curso: protect yourself from curse of attribute inference: a social network privacy-analyzer. In *Proceedings of the 3rd ACM SIGMOD Workshop on Databases and Social Networks, DBSocial 2013, New York, NY, USA, June, 23, 2013*, pages 13–18.

[Street, 2012] Street, J. (2012). *Music & Politics*. Polity Press.

[Vidyalakshmi et al., 2016] Vidyalakshmi, B. S., Wong, R. K., and Chi, C. (2016). User attribute inference in directed social networks as a service. In *IEEE International Conference on Services Computing, SCC 2016, San Francisco, CA, USA, June 27 - July 2, 2016*, pages 9–16.

[Wang et al., 2015] Wang, P., Xu, B., Wu, Y., and Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *SCIENCE CHINA Information Sciences*, 58(1):1–38.