UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Quels antécédents peuvent influencer les inquiétudes des utilisateurs de Facebook quant à la protection de leur vie privée et l'accès qu'ils donnent à leurs données personnelles ?

Focus sur le Big Data, Cambridge Analytica et le RGPD

Auteur : **Thibaud Aerts**
Promoteur : **Isabelle Schuiling**
Année académique 2018-2019

Je tiens tout d’abord à remercier mon promoteur, Madame Schuiling, d’avoir accepté de m’encadrer dans la réalisation de ce mémoire.

Ma reconnaissance s’adresse également à toutes les personnes ayant contribué à mon enquête en ligne.

Enfin, je remercie ma famille et mes amis pour leur soutien tout au long de ce mémoire.

II.

Table des matières

INTRODUCTION

Selon un rapport publié en 2018 par Hootsuite et WeAreSocial, le réseau social Facebook comptait à lui seul au 1er janvier 2018 plus de 2 milliards d'utilisateurs réguliers dans le monde. En Belgique ils étaient 7,5 millions, soit 65% de la population belge, ce qui représente un taux de pénétration impressionnant.

L'essentiel des revenus de Facebook proviennent de la vente d'espaces publicitaires sur son site. Pour attirer les annonceurs, Facebook peut mettre en avant l'immense quantité de données personnelles qu'il récolte en permanence sur ses utilisateurs, en fonction de leurs activités sur Internet (géolocalisation, pages visitées, achats réalisés, temps de connexion, likes, commentaires, tags, …) ou de l'information qu'ils ajoutent sur leur profil (photos, vidéos, listes d'amis, informations de statut, …). L'analyse sans cesse perfectionnée de cette masse de données permet aux annonceurs de mieux cibler les publicités qu'ils adressent de la manière la plus individualisée possible à chaque utilisateur de Facebook, pour un maximum d'impact.

Ces pratiques marketing sont de plus en plus répandues : une étude récente de Statbel (2018) indique qu'en 2018, 55% des grandes entreprises belges ont analysé du Big Data, notamment en provenance de réseaux sociaux.

Avant le scandale Cambridge Analytica, la plupart des utilisateurs de Facebook étaient plus ou moins conscients que les données personnelles qu'ils posaient sur le réseau social révélaient un peu d'eux-mêmes et qu'elles pouvaient être utilisées à certaines fins commerciales. Pour protéger leur vie privée, ils considéraient qu'ils pouvaient maîtriser la liste d'amis qui avait accès aux informations les plus sensibles ainsi que quelques autres paramètres de confidentialité. Les avantages perçus de l'utilisation d'un réseau aussi vaste et aussi convivial étaient bien plus grands que les éventuels petits inconvénients liés à des publicités parfois invasives. Ils ne pensaient pas être exposés à des faits aussi graves que le vol de numéros de cartes de crédit ou l'usurpation d'identité numérique, constatés auparavant sur Internet.

En 2018, l'affaire Cambridge Analytica a mis en évidence que des données personnelles parfois considérées comme anodines disent tellement de choses sur nous qu'elles peuvent être utilement exploitées pour construire des stratégies de manipulation électorale. Mais surtout l'affaire Cambridge Analytica force à constater que Facebook a géré de manière a

minima laxiste les données personnelles qui lui étaient confiées et qu'à notre insu ces données ont pu être utilisées par des tiers et détournées à des fins potentiellement inappropriées. Et les choses n'en sont pas restées là puisque Facebook fait désormais face à d'autres accusations d'usage abusif des données personnelles qui lui ont été confiées. Ces atteintes à la confiance qu'ont les utilisateurs de Facebook dans le réseau social pourraient laisser des traces.

2018 est également l'année où le Règlement Général sur le Protection des Données est entré en vigueur. Tous les résidents européens ont été largement informés de la panoplie d'outils légaux dont ils disposent désormais pour conserver la maîtrise de leurs données personnelles. Mais dans le même temps, ils ont été envahis de demandes en provenance de tous les sites Web qu'ils consultaient pour autoriser ceux-ci à conserver les accès à leurs données personnelles dont ils disposaient précédemment, sans avoir d'autre choix que de l'accepter pour continuer à visiter le site Web en question...

Cette conjonction d'événements est à l'origine de la réflexion poursuivie dans ce mémoire. La problématique de notre mémoire est ainsi la suivante :

Dans quelle mesure les pratiques du Big Data, le scandale Cambridge Analytica et l'entrée en vigueur du RGPD ont-ils une influence sur les inquiétudes des utilisateurs de Facebook quant à la protection de leur vie privée et l'accès qu'ils donnent à leurs données personnelles ?

La méthodologie suivie au cours de ce mémoire repose, à la fois, sur une analyse théorique réalisée au travers de diverses ressources bibliographiques, mais aussi sur une étude quantitative accomplie au moyen de la diffusion d'un questionnaire en ligne.

La structure de ce mémoire comporte ainsi trois parties.

Dans la première partie, nous réaliserons une revue de la littérature afin de présenter la mise en contexte de l'étude. Nous examinerons dans un premier temps les pratiques du Big Data et plus particulièrement celles du marketing prédictif. Dans un deuxième temps, nous nous intéresserons aux réseaux sociaux et à Facebook en particulier. Cela nous amènera à aborder les récents scandales autour de la vie privée, tels que celui de Cambridge Analytica. Enfin, au sein du dernier chapitre de cette partie, nous développerons en détail le concept de vie privée et nous nous intéresserons au nouveau Règlement Général sur la Protection des Données (RGPD). Tout ceci, afin de poser les

termes de notre problématique et ainsi pouvoir choisir le cadre théorique de référence le plus approprié.

Dans la deuxième partie, nous nous inspirerons du modèle APCO de Smith, Dinev et Xu (2011) afin de déterminer les relations entre les antécédents potentiels et les inquiétudes en matière de vie privée des utilisateurs de Facebook. Ce cadre de référence nous permettra également de déterminer les relations entre ces inquiétudes en matière de vie privée et les comportements des utilisateurs quant au partage et à la protection de leurs données personnelles. Nous disposerons alors d'une base solide nous permettant non seulement d'appuyer notre argumentation mais aussi de développer le modèle conceptuel ainsi que la série d'hypothèses à tester.

Dans la troisième partie, il sera question de notre étude quantitative. Tout d'abord, la méthodologie de la recherche sera explicitée. Nous poursuivrons ensuite avec l'analyse des résultats ainsi que l'interprétation de ceux-ci. Enfin, nous conclurons avec les recommandations, les limites de notre étude et les pistes pour de futures recherches.

4.

PARTIE 1 : REVUE DE LA LITTÉRATURE

Chapitre 1 : Big Data et marketing

1. Évolution du marketing

La pratique du marketing est ancienne car ses racines remontent à la première révolution industrielle, avec l'ambition croissante des industriels d'étendre les débouchés pour leur production et de stimuler la consommation (Volle, 2011). Le début du 20ème siècle marque l'avènement des techniques de marketing modernes. Au travers d'importants efforts publicitaires et la réalisation d'études de marché, les stratégies marketing des entreprises reposent sur l'acquisition de consommateurs cibles appartenant à un segment de marché bien défini (Dussart & Nantel, 2007).

Dans les années 60, le marketing traditionnel fait place à un marketing plus relationnel et individuel. En effet, l'augmentation des capacités de traitement et de stockage des ordinateurs permet le passage d'un marketing passif vers un marketing actif. Il ne s'agit plus seulement d'acquérir de nouveaux clients mais bien de les conserver et de les satisfaire (Dussart & Nantel, 2007).

À partir de la fin du 20ème siècle, l'explosion d'Internet et du numérique bouleversent les habitudes des consommateurs. Ceux-ci, ayant accès à de plus en plus d'informations, sont plus attentifs et critiques dans leurs comportements d'achat. Les objectifs du marketing évoluent afin de se concentrer exclusivement sur le client. En permettant de relier l'identité physique et digitale des clients, les données permettent la reconnaissance de ceux-ci à chaque point de contact ainsi que la personnalisation de l'interaction avec ceux-ci (Hirth, 2017).

Le marketing de masse tend à disparaître avec tout ce flot de données accessibles sur le client. Malgré cette explosion des données, leur exploitation n'en reste pas moins laborieuse, le volume des données récoltées étant trop grand et les techniques d'analyse n'existant pas encore. À partir de 2005, une nouvelle génération d'algorithmes fait son apparition et permet le traitement de gros volumes de données à des vitesses mille fois supérieures qu'auparavant (Babinet, 2015).

La segmentation, le ciblage et le profilage des consommateurs deviennent ainsi de plus en plus précis. Il est alors question de révolution *Big Data* avec émergence d'un marketing

contextualisé et hyper-personnalisé pouvant aboutir à la prédiction des envies et des besoins des consommateurs (Crié & Willart, 2016).

2. Définition du Big Data

Medioni et Benmoyal-Bouzaglo (2018, pp.21-22) définissent le terme *Big Data* de la manière suivante : *« de façon globale, on considère le Big Data (ou mégadonnées) comme toutes les données que l'on peut collecter quelle que soit leur provenance : vidéos publiées, informations médicales, signaux GPS, données de climat, achats effectués et montants dépensés dans un magasin ou site en ligne… ».*

Cette définition s'articulant surtout sur l'aspect volumineux des données disponibles, nous mentionnons la définition plus complète du *Big Data* proposée par Le Nagard et Giannelloni (2016, p.3) : *« plus qu'à de gros volumes de données, l'expression 'big data' (BD) se réfère avant tout à des technologies logicielles qui permettent de traiter en temps réel des données issues de sources hétérogènes et complexes, selon des protocoles qualifiés de 'massivement parallèles' ».*

Laney (2001) suggère que les trois dimensions du *Big Data* sont le volume, la variété et la vélocité. En outre, deux autres dimensions sont souvent mentionnées dans la littérature. Il s'agit de la véracité et de la valeur. Toutes ces dimensions permettent de mieux cerner le concept du *Big Data*.

a. Volume

Cette dimension correspond à l'ampleur des données. Elle est la traduction du '*big*'. La taille minimale d'une base de données pour être considérée comme une architecture *Big Data* est actuellement de 1 téra-octet (Lee, 2017). Un téra-octet de données peut, par exemple, stocker près de 16 millions de photos Facebook (Gandomi & Haider, 2015). Cette notion évolue avec le temps et les développements technologiques. Par exemple, le volume des données n'a cessé d'augmenter ces dernières années avec la multiplication des objets connectés (Lee, 2017).

b. Variété

Cette dimension fait référence aux différents types de données possibles. Les avancées technologiques actuelles permettent en effet aux entreprises d'utiliser à la fois des données structurées et des données non-structurées. Les données structurées font référence aux données trouvées dans la majorité des bases de données traditionnelles. Par

exemple, un ticket de caisse est une donnée structurée. Selon Cukier (2010), elles ne représentent que 5% de toutes les données disponibles. Les données non-structurées ne sont quant à elles pas toujours pourvues d'une structure leur permettant d'être manipulées correctement (Gandomi & Haider, 2015). Les images, vidéos et textes sont des exemples de données non-structurées. Elles sont souvent récoltées au travers des réseaux sociaux.

Au fil du temps, les techniques d'analyse s'étant améliorées, le fait qu'une donnée soit de type non-structuré est de moins en moins devenu un obstacle à son analyse (Lee, 2017). Selon Crié et Willart (2016), le courant *Big Data* traite surtout des données non-structurées et attitudinales en comparaison avec l'approche 'données traditionnelles' où les données utilisées étaient majoritairement de nature structurée et transactionnelle.

c. Vélocité

Cette dimension correspond à la vitesse à laquelle les données sont générées et analysées. Le nombre grandissant de données générées par les smartphones et les applications mobiles accroît la nécessité de pouvoir les analyser en temps réel (Gandomi & Haider, 2015). Ces analyses en temps réel peuvent en effet permettre aux entreprises d'adresser des offres plus personnalisées à leurs consommateurs. Cela est rendu possible par l'amélioration croissante de la vitesse de traitement des données. Alors que le décodage du génome humain, qui consiste à analyser plus de 3 milliards de paires de nucléotides, a pris près de dix ans la première fois qu'il a été réalisé en 2003, il peut maintenant s'opérer en une semaine seulement (Cukier, 2010).

d. Véracité

Cette dimension souligne qu'il est important que les données traitées soient de qualité. Cette problématique a été introduite par IBM (2018), qui a démontré qu'un dirigeant d'entreprise sur trois n'avait pas confiance dans les données qu'il utilisait pour prendre ses décisions. Les sentiments qui sont exprimés par les consommateurs sur les réseaux sociaux sont, par exemple, peu fiables car ils impliquent un jugement humain (Gandomi & Haider, 2015). Des outils technologiques ont ainsi été mis au point afin de pouvoir manipuler ces données jugées incertaines (Lee, 2017) car celles-ci ont beaucoup de valeur aux yeux des responsables du marketing.

e. Valeur

Cette dimension fait référence au fait qu'il est important que les données traitées soient utiles. Selon une enquête de Pure Storage (2015) auprès de 308 entreprises européennes, 72% des entreprises collectent des données qui ne sont ensuite pas utilisées. Pour Lee (2017), cette dimension traduit donc qu'il est nécessaire, pour les entreprises, de choisir des sources de données à valeur ajoutée, d'évaluer les bénéfices et les coûts de la collecte et du traitement des données et de construire des techniques d'analyse permettant de rendre ces données utiles.

3. Pratiques du Big Data

Selon Le Nagard et Giannelloni (2016, p.4), *« le Big Data est désormais partout, tout le temps »*. Les domaines d'application des techniques du *Big Data* sont en effet très nombreux. À titre d'exemples, nous citerons les domaines de la santé, de la sécurité, de l'entreprise en général et du marketing en particulier.

Tout d'abord, dans le domaine de la santé, les données sont souvent de types hétérogènes, ce qui complique leur manipulation et leur traitement. Il s'agit en effet de pouvoir croiser des données telles que l'âge, le sexe, les compte rendus d'intervention, les ordonnances, les imageries ou encore les médicaments prescrits. Aujourd'hui, avec l'émergence du *Big Data* et de ses techniques, il est, par exemple, possible de suivre des patients à distance, ce qui relevait de l'impossible par le passé (Vouters & Lavielle, 2018). Un autre exemple est celui de la création d'un outil par des scientifiques de l'Université de Stanford permettant de détecter les premiers signes d'un cancer de la peau avec une précision équivalente à celle d'un dermatologue (Cynober, 2018).

Ensuite, dans le domaine de la sécurité, le *Big Data* peut également s'avérer être un outil précieux dans la prévention des crimes. Le croisement de multiples données telles que l'historique des infractions, les images des caméras de vidéosurveillance ou encore la structure démographique d'un quartier peuvent en effet permettre aux forces de l'ordre de déterminer où de prochains crimes ont le plus de chances de se produire. C'est le cas, par exemple, à Santa Cruz, aux États-Unis, où un logiciel en test utilisant un algorithme capable de prédire quand et où un crime pouvait se réaliser, a permis de réduire de 21% les crimes commis dans la ville (Le Monde, 2013).

De plus, dans le monde des entreprises, le *Big Data* offre également d'énormes potentiels comme, par exemple, dans la réduction des coûts (Lee, 2017). C'est ainsi que Tesco, le leader des supermarchés britanniques, a réussi à réduire drastiquement ses coûts énergétiques en utilisant les techniques du *Big Data* pour contrôler toutes les trois secondes la température de ses frigos, ce qui leur a permis d'économiser près de 25 millions de dollars chaque année (Van Rijmenam, 2014). Selon une étude réalisée par les analystes de Dresney Advisory (2017), 53% des entreprises interrogées déclarent faire usage des technologies du *Big Data*. Elles n'étaient que 17% en 2015 (Columbus, 2017).

Dans le domaine du marketing, l'émergence du *Big Data* a révolutionné le marketing prédictif. Auparavant essentiellement utilisées dans le monde financier et les banques, les analyses prédictives font désormais partie intégrante des décisions marketing prises au sein des entreprises (Hirth, 2017). C'est sur cette notion que nous allons nous concentrer dans la section suivante.

4. Marketing prédictif

Selon Hirth (2017, p.271), le marketing prédictif est désigné comme *« la discipline visant à anticiper le comportement d'un consommateur, grâce à la modélisation mathématique de celui-ci »*. Le but recherché est d'offrir l'offre la plus personnalisée possible à la bonne personne, au bon moment et au bon endroit.

Afin de mieux comprendre ce que représente le marketing prédictif, nous mentionnons une définition plus complète de celui-ci établie par Bathelot (2017, para.1) : *« le marketing prédictif regroupe les techniques de traitement et de modélisation des comportements clients et prospects qui permettent d'anticiper leurs actions futures à partir du comportement présent et passé »*.

Les algorithmes prédictifs existent depuis toujours. Néanmoins, l'accès de plus en plus facile aux données provenant de bases de données publiques ainsi que celles provenant de la géolocalisation des individus a permis à ces pratiques anciennes de s'offrir une seconde jeunesse (Le Nagard & Giannelloni, 2016). En outre, aujourd'hui, avec des logiciels toujours plus conviviaux et interactifs qui arrivent sur le marché, ce domaine n'est plus seulement réservé aux statisticiens et mathématiciens mais s'est considérablement ouvert aux analystes du milieu des affaires en général (SAS, 2018).

Quant aux acteurs majeurs dans le domaine du marketing prédictif, nous distinguons, d'un côté, les principaux fournisseurs de solutions en matière d'analyse prédictive et de l'autre, leurs principaux utilisateurs.

a. Fournisseurs

Selon le rapport de Forrester Wave (2018), les leaders mondiaux en matière de solutions d'analyse prédictive sont actuellement SAS et IBM. Parmi les treize fournisseurs analysés dans cette étude, ce sont en effet ces deux entreprises qui auraient, à la fois, les meilleures offres, les stratégies les plus performantes et la présence la plus étendue sur le marché.

SAS est reconnu comme le leader de l'analytique. Cette entreprise commercialise de nombreux outils classiques d'analyse prédictive permettant notamment de réduire des risques, d'optimiser des campagnes marketing, d'améliorer des processus mais aussi de détecter des fraudes (SAS, 2018).

IBM, de son côté, est reconnu pour son intelligence artificielle nommée Watson qui a les capacités de comprendre tous les types de données, d'apprendre, de raisonner mais aussi d'interagir de façon naturelle avec les êtres humains (IBM, 2018). C'est ainsi qu'IBM s'est, par exemple, appuyé sur Watson pour créer les premières publicités cognitives, c'est-à-dire des publicités moins intrusives, plus intelligentes et mieux ciblées (Jaimes, 2016).

En plus de ces leaders en la matière, le secteur comporte aussi une multitude de start-ups présentes sur le marché. Nous avons identifié quelques exemples intéressants. Ainsi, certaines start-ups sont spécialisées dans la détection des fraudes. C'est par exemple le cas de la start-up française, Datategy, qui facilite la détection de la fraude dans les transports en commun via son outil de prédiction OctoCity. En analysant de nombreuses données telles que la structure du réseau, les historiques de fraude ou encore la localisation des arrêts, cet outil est capable d'optimiser l'itinéraire de tournée des contrôleurs (De Carme, 2018). D'autres start-ups sont spécialisées dans les données de géolocalisation, comme la start-up française, Roofstreet, qui se targue d'être capable de prédire les déplacements d'une personne jusqu'à douze heures à l'avance (French Web, 2018). Enfin, certaines ont comme spécialité l'optimisation des campagnes marketing, comme la start-up Tinyclues qui cherche, via ses algorithmes d'intelligence artificielle, à optimiser l'envoi des bonnes campagnes marketing aux bonnes personnes (Raynal, 2018).

b. Utilisateurs

De plus en plus d'entreprises font usage des techniques du marketing prédictif. Les plus importantes sont les GAFA, c'est-à-dire Google, Apple, Facebook et Amazon. Celles-ci utilisent en effet le marketing prédictif depuis de nombreuses années (Launay, 2017).

Ainsi, Google a élaboré un logiciel intitulé '*the predictive marketing engine'*. Celui-ci lui permet d'attribuer un score à chaque internaute en fonction de ses comportements multi-canaux, c'est-à-dire à partir de ses visites en magasin ou en ligne, de ses e-mails ou encore de ses canaux de vente utilisés (Mauchamp, 2015). Tous ces scores récoltés permettent ensuite à Google de prédire les actions futures de ses internautes en ligne. Tout ceci lui permet ensuite de faire usage, de manière plus efficace et rentable, de la publicité programmatique, qui consiste principalement à la vente et à l'achat d'espaces publicitaires au travers d'un système informatique automatisé (Blavier, 2016).

Apple, via son assistant personnel Siri, est capable de reconnaître les commandes prononcées par quelqu'un en comparant celles-ci avec plusieurs millions d'autres commandes effectuées par d'autres utilisateurs, qui sont stockées au sein de ses différentes plateformes d'analyse prédictive (Marr, 2015). De cette manière, Siri peut alors effectuer les commandes désirées par le consommateur.

Quant à Facebook, le réseau social ayant le plus d'utilisateurs au monde, il est, par exemple, capable de prédire quelles publications sont les plus susceptibles d'intéresser un individu sur son fil d'actualité (Siegel, 2016). Facebook met ainsi certains contenus en avant afin d'inciter ses utilisateurs à prolonger leurs activités sur son site. Tout ceci dans le but d'engranger un maximum de publicités vues par ces mêmes utilisateurs.

Facebook utilise également les techniques du marketing prédictif dans les suggestions d'amis. Un brevet a même été déposé par Facebook en 2014 où il est question d'une technique permettant de déterminer si deux personnes se font face ou si celles-ci marchent côte-à-côte lorsqu'elles sont très proches l'une de l'autre. De cette manière, après une visite dans un bar, Facebook serait capable d'uniquement vous suggérer comme ami les personnes avec qui vous auriez parlé en face-à-face et non pas toutes les autres que vous auriez simplement croisées et avec qui vous n'auriez pas échangé de conversation. Ce serait en effet possible grâce à l'analyse des relevés du gyroscope et de l'accéléromètre de chacun des smartphones appartenant aux deux personnes (Hill & Mattu, 2018).

Enfin, Amazon, le leader mondial du commerce en ligne, utilise notamment les techniques d'analyse prédictive afin de recommander les meilleurs produits à ses clients et ce, de manière toujours plus précise (Fredenucci, 2017). Visant toujours plus loin, Amazon a même déposé un brevet en 2014 qui lui permettrait de livrer des colis avant même que ceux-ci ne soient commandés par leurs consommateurs (Bensinger, 2014).

En outre, de nombreuses autres entreprises font également usage du marketing prédictif. On cite souvent l'exemple de Target, cette entreprise américaine de grande distribution, qui a réussi à concevoir en 2012 un algorithme permettant de prédire la grossesse de ses clientes dans le but de pouvoir leur proposer les produits les plus pertinents (Siegel, 2016). Cela a été révélé par le New York Times qui a expliqué comment un père de famille américain mécontent de voir que sa fille recevait de la part de Target des bons de réduction pour des produits que les jeunes mamans avaient tendance à acheter, a finalement appris quelques jours plus tard que sa fille était enceinte (Hirth, 2017).

Plus récemment, en France, on a beaucoup parlé de SFR qui a développé un outil de marketing prédictif lui permettant de repérer parmi ses clients ceux qui seraient sur le point de changer d'opérateur téléphonique et d'entamer à leur égard des stratégies de fidélisation. SFR utiliserait à cette fin des informations relatives au nombre de pages vues par ces clients, la durée de leurs visites sur le site et les mots-clés utilisés dans leurs recherches sur Internet (Brossas, 2018).

Chapitre 2 : Réseaux sociaux

1. Définition des réseaux sociaux

Selon Kaplan et Haenlein (2010), un réseau social en ligne est défini comme *« un groupe d'applications Internet qui se fondent sur l'idéologie et la technologie du Web 2.0 et qui permettent la création et l'échange du contenu généré par les utilisateurs »* (cité dans Stenger & Coutant, 2013, p.109). Cette définition fait ressortir la notion de Web 2.0 dont la paternité est attribuée à Tim O'Reilly (2005), qui n'en a pas livré une définition exacte mais plutôt de nombreux principes fondateurs comme la participation au lieu de la simple publication, les utilisateurs en tant que contributeurs, le Web en tant que plateforme ou encore l'intelligence collective (cité dans Fuchs, 2017). Selon Scibetta, Kempf et Alves (2012, p.152), le concept du Web 2.0 sert donc à *« désigner les nouvelles applications rendues possibles par les évolutions technologiques, le Web devenant plus simple et plus interactif »*.

Une définition plus précise des réseaux sociaux est donné par Mercklé (2016, p.81) qui déclare que ceux-ci ont *« pour caractéristiques communes d'offrir à leurs utilisateurs :*

(1) la possibilité de créer un espace personnel de présentation de soi, où ils peuvent mettre à disposition de tous les informations et les images qu'ils souhaitent, leur permettant de dire ce qu'ils sont et ce qu'ils font, ce qui définit leur 'profil' ;
(2) la possibilité d'accéder, selon des modalités et à des degrés variés, aux 'profils' mis en ligne selon le même principe par les autres membres du réseau ;
(3) la possibilité de nouer des relations avec des membres du réseau, relations qui font ensuite partie de leurs 'profils' respectifs, et contribuent donc ainsi à les définir, puisque la liste des 'amis' peut être affichée dans le profil au même titre que les goûts musicaux ou littéraires ».

2. Histoire des réseaux sociaux

Le premier site web reconnu en tant que réseau social apparait en 1997 sous le nom de SixDegrees. Bien qu'il attire plusieurs millions d'utilisateurs, il ne réussit pas à perdurer et doit fermer en 2000. Selon son fondateur, Andrew Weinreich, SixDegrees était simplement trop en avance sur son temps (Boyd & Ellison, 2007). Dans les années suivantes, d'autres réseaux sociaux font leur apparition comme Friendster en 2002 ou MySpace en 2003. Ceux-ci connaissent un véritable succès et les utilisateurs sont

nombreux à les rejoindre. Toutefois, de multiples problématiques liées à la sécurité et à la confiance finissent par détourner définitivement les utilisateurs de ces deux plateformes.

Au fil du temps et malgré les difficultés à continuer à exister dans un monde digital en perpétuel renouvellement, certains réseaux sociaux ont poursuivi leur expansion et sont devenus des acteurs majeurs d'Internet. Ceux qui ont réussi à se faire une place dans la cour des grands sont essentiellement Facebook, Youtube, WhatsApp, Messenger, Wechat ou encore Instagram (Statista, 2018a). Notons qu'à l'exception de Wechat, qui appartient à un groupe chinois, et Youtube, qui a été racheté par Google en 2006, tous ces réseaux sociaux font désormais partie de la galaxie Facebook.

3. Facebook

a. Origines

À l'origine, Facebook est créé en 2004 par Mark Zuckerberg et d'autres étudiants d'Harvard. Sa croissance contrôlée est ce qui le différencie en grande partie des autres réseaux sociaux créés durant la même période. Celui-ci est en effet d'abord réservé aux étudiants du campus d'Harvard. Ensuite, Facebook s'étend à d'autres universités américaines. Pour s'inscrire, un utilisateur doit posséder une adresse électronique universitaire associée à ces établissements. Cette exigence permet de garder le site relativement fermé et contribue à développer la perception d'appartenir à une communauté intime et privée (Boyd & Ellison, 2008). Son design épuré et non-personnalisable lui permet également de s'offrir la réputation d'un espace sécurisé. Ce n'est qu'à partir de 2006 que Facebook devient accessible à toute personne de plus de 13 ans possédant une adresse e-mail (Press, 2018).

b. Caractéristiques

Facebook est aujourd'hui le réseau social qui compte le plus d'utilisateurs avec en moyenne 1,49 milliard de personnes actives chaque jour (Facebook, 2018). Au fil des ans, Facebook s'est en effet propagé dans le monde entier et selon Cosenza (2018), c'est le premier réseau social dans 152 des 167 pays étudiés. En Belgique, selon une étude menée par Hootsuite et WeAreSocial (2018), il y a près de 7,5 millions de Belges actifs sur Facebook.

Afin de rester au sommet, Facebook utilise plusieurs stratégies d'acquisition. La première consiste à acheter d'autres entreprises numériques qui sont en concurrence directe avec la

plateforme (Fuchs, 2017). Instagram et WhatsApp ont ainsi été respectivement acquis en 2012 et en 2014. La deuxième stratégie d'acquisition de Facebook consiste à acquérir des entreprises dont les technologies pourraient permettre d'améliorer la plateforme et les services offerts par Facebook (Fuchs, 2017). C'est ainsi que la société de réalité virtuelle, Oculus, a été acquise en 2014 par Facebook.

Avec près de 500 milliard de dollars de capitalisation boursière, Facebook est la cinquième entreprise la plus valorisée en bourse après Apple, Amazon, Google et Microsoft (Aflalo, 2018).

c. Fonctionnement

À l'origine, Facebook était une plateforme biface avec d'un côté des personnes envoyant des messages et de l'autre des personnes les recevant. Très rapidement après son lancement, Facebook a entamé la vente d'espaces publicitaires et s'est rapidement transformé en une plateforme triface (Evans & Schmalensee, 2017). Au fil des ans, le nombre d'utilisateurs grandissant en masse, de plus en plus d'annonceurs sont arrivés sur le site. Facebook a dès lors entrepris le développement de nombreux outils spécifiquement élaborés pour ces annonceurs. À chaque fois qu'ils se connectent sur Facebook, les utilisateurs produisent de multiples données liées à leur profil et à leurs comportements en ligne. Ces données sont vendues aux annonceurs. Ceux-ci les utilisent alors afin d'améliorer le ciblage des publicités à afficher en fonction du profil de l'utilisateur (Fuchs, 2017). C'est ainsi que les recettes publicitaires de Facebook ont grimpé jusqu'à atteindre près de 34 milliards de dollars en 2018 (Statista, 2018b).

Dans le courant de l'année 2007, Facebook a ajouté une quatrième face à son réseau en lançant sa plateforme de développement d'applications spécialement conçue pour les développeurs afin que ceux-ci puissent produire leurs propres applications. Des jeux tels que Farmville, développé par Zynga, ont ainsi connu un grand succès et ont attiré beaucoup d'utilisateurs (Evans & Schmalensee, 2017). Plus tard en 2007, Facebook a activé une autre fonction permettant à toute organisation de créer sa propre page. Les organisations ont donc non seulement reçu la possibilité d'envoyer des messages aux utilisateurs de Facebook mais aussi d'en recevoir de la part de ceux-ci. Tout ceci a donc ajouté deux faces à l'architecture de Facebook et c'est ainsi que ce réseau social est considéré comme une plateforme à six faces par Evans et Schmalensee (2017).

4. Scandale Cambridge Analytica

Facebook connaît depuis quelques mois un scandale de très grande ampleur touchant au vol de données personnelles de plus de 50 millions de ses utilisateurs. Tout débute le 17 mars 2018 lorsque le Guardian, le New York Times et le London's Observer publient conjointement une enquête sur Cambridge Analytica dans laquelle Christophe Wylie, un ex-employé, explique comment cette société a obtenu de manière illégale de nombreuses données privées.

Cambridge Analytica, entreprise de communication stratégique, spécialiste dans le profilage d'électeurs a été fondée en 2013 à Londres par Steve Bannon et Robert Mercer. Elle est principalement connue pour avoir offert ses services à Ted Cruz pendant la course à l'investiture en 2015 et à Donald Trump lors de la campagne présidentielle en 2016 (Audureau, 2018).

Entre 2014 et 2015, Aleksandr Kogan, jeune chercheur de l'université de Cambridge, a conçu sur Facebook une application de quiz nommée *'thisisyourdigitallife'* qui permettait à ses utilisateurs d'être rémunérés quatre dollars en répondant à un questionnaire servant soi-disant à des recherches universitaires en psychologie. Plus de 270 000 personnes ont téléchargé cette application et ont ainsi offert à Kogan de précieuses données telles que leur profession, âge et pages qu'ils ont aimées. Toutefois, sans le savoir, ces personnes ont aussi donné accès aux données de leurs 'amis' sur Facebook, ce qui porte à près de 87 millions le nombre d'utilisateurs concernés (Drouglazet, 2018). En outre, sans en avertir Facebook, Kogan a transmis toutes les données qu'il a récoltées à Cambridge Analytica qui les a ensuite utilisées afin d'identifier la personnalité des électeurs et ainsi tenter d'influencer leurs votes par le biais de publicités ciblées.

Après ces révélations, Cambridge Analytica s'est déclarée en faillite et a cessé ses activités en mai 2018. De son côté, Facebook a été condamné en octobre 2018 par le Royaume-Uni à l'amende maximale, c'est-à-dire près de 565 000 euros, pour avoir laissé des sociétés tierces recueillir les données de ses utilisateurs sans qu'ils en soient préalablement avertis (Marchand, 2018). Le Parlement européen a également voté, en octobre 2018, une résolution non contraignante enjoignant Facebook de permettre à l'Union européenne de réaliser un audit complet pour évaluer les risques de sécurité du réseau social pour les données à caractère personnel des utilisateurs et prévenir les manipulations électorales (Braun, 2018). Facebook s'est engagé envers le Parlement

européen à être plus proactif dans le futur pour la protection de la sécurité de ces données. Aux États-Unis, le Washington Post a annoncé que l'Attorney General du district de Columbia a entamé des poursuites contre Facebook le 19 décembre 2018 pour avoir permis à la société de consultance politique Cambridge Analytica d'avoir accès aux données personnelles de millions d'américains sans leur permission. La Federal Trade Commission (FTC) a également entamé des investigations sur le même sujet (Romm et al., 2018).

Il ne s'agit pas du premier scandale auquel est mêlé Facebook en matière de violation de la vie privée. En effet, le réseau social a déjà fait face à un premier incident majeur en 2007 dans le cadre du lancement de son système publicitaire intitulé *Beacon.* Celui-ci lui permettait de publier automatiquement sur le profil des utilisateurs les achats que ceux-ci avaient réalisés sur un site partenaire (Le Monde, 2012). Un grand nombre de personnes s'est alors indigné et une plainte en action collective a été déposée. Facebook a finalement décidé de supprimer totalement ce service en 2009. En 2013, c'est l'affaire Snowden qui a impliqué Facebook dans un autre scandale. Les révélations d'Edward Snowden ont en effet démontré que Facebook était au cœur du vaste réseau d'espionnage mis en place par la NSA (Smyrnaios, 2018).

Très récemment, deux nouvelles affaires ont éclaté. Tout d'abord, on a appris à la fin du mois de septembre 2018, que plus de 50 millions de comptes Facebook venaient d'être piratés, les pirates ayant exploité une faille du réseau social afin de se connecter aux comptes des utilisateurs à leur place et d'ainsi récupérer leurs données personnelles (Ronfaut, 2018). Ensuite, le 18 décembre 2018, le New York Times a révélé que, depuis 2010, Facebook autorise plusieurs de ses partenaires majeurs à avoir un accès privilégié sur les données censées être confidentielles de ses utilisateurs. Les partenaires cités sont entre autres Microsoft, Netflix, Amazon ou encore Spotify (Dance et al., 2018).

Chapitre 3 : Vie privée

1. Définition du concept

Le concept de vie privée a été défini de nombreuses manières différentes. Ce que ce concept représente exactement est en effet toujours débattu entre représentants du gouvernement, décideurs politiques, entreprises privées et consommateurs (Lainier et Saini, 2008). Selon Warren et Brandeis (1890, p.195), le droit à la vie privée est défini comme *« le droit d'être laissé seul »*. Leur article est apparu en 1890 dans une revue juridique et est considéré comme un des tout premiers articles diffusés sur le thème de la vie privée. Selon Sevignani (2012), c'est l'intrusion de journalistes lors du mariage de la fille de Warren qui aurait poussé les auteurs à écrire cet article. À cette époque, la frontière entre la sphère privée et publique était beaucoup plus forte qu'aujourd'hui. Une personne, même une famille, pouvait se retirer de la société et un déplacement d'une centaine de kilomètres pouvait être suffisant pour effacer son passé (Stewart, 2017).

Les avancées technologiques que nous connaissons en matière de transports, et de télécommunications notamment, le développement d'Internet et des pratiques du *Big Data*, mentionnées dans le chapitre 1, ont rendu quasi impossible cet isolement caractéristique de la vie privée. Pour protéger sa vie privée, il faut désormais poser des actes positifs spécifiques, alors même que les demandes de partage d'informations sont toujours plus nombreuses et conditionnent dans certains cas l'appartenance à la société en général (Stewart, 2017). Afin d'intégrer cette évolution, une autre définition du concept de vie privée a alors été utilisée dans de nombreux travaux dans le domaine. Il s'agit de celle décrite par Westin (1967, p.7) qui définit la protection de la vie privée *« comme la revendication d'individus, groupes ou institutions de pouvoir déterminer eux-mêmes quand, comment et jusqu'où leurs informations personnelles sont communiquées à d'autres ».* Cette définition se focalise donc sur la notion de contrôle des individus quant à l'accès et à la propagation de leurs informations personnelles.

2. Approches théoriques

Selon Miltgen (2011), il existe deux approches théoriques au sein de la recherche dans le domaine de la vie privée. La première s'intéresse aux inquiétudes en matière de vie privée alors que la seconde se concentre sur le concept de *'privacy calculus'*.

a. Première approche théorique : Inquiétudes en matière de vie privée

La première approche se concentre sur les relations entre les inquiétudes en matière de vie privée des individus et leurs comportements pour y répondre. Ainsi, Smith, Milberg et Burke (1996) identifient les quatre dimensions du concept d'inquiétudes en matière de vie privée.

Ils prennent d'abord en considération les préoccupations des individus face à la collection de leurs données, c'est-à-dire leurs inquiétudes par rapport à la quantité considérable de données d'identification personnelle qui sont collectées et stockées dans des bases de données.

Ensuite, ce sont les préoccupations face aux erreurs possibles commises par les entreprises collectant les données qui forment la deuxième dimension. Ces préoccupations représentent le manque de confiance des individus dans les mesures de protection proposées par ces entreprises face aux erreurs dans la gestion de leurs données personnelles, qu'elles soient délibérées ou accidentelles.

Les inquiétudes face à l'usage secondaire de leurs données privées constituent la troisième dimension. Cela traduit les craintes liées au fait que, sans autorisation préalable, des informations soient collectées dans un but précis mais finalement utilisées pour un tout autre dessein.

Enfin, la dernière dimension définie par Smith, Milberg et Burke (1996) est celle des préoccupations des individus liées à l'accès potentiel de leurs données personnelles par certaines personnes qui ne seraient pas préalablement autorisées à voir ou travailler avec ces données.

Ce modèle à quatre dimensions a, par la suite, été vérifié empiriquement par Stewart et Segars (2002) et divers auteurs s'en sont inspirés dans leurs travaux (Angst & Agarwal, 2009 ; Lowry et al., 2011).

Toujours au sein de cette première approche, Son et Kim (2008) expliquent que les utilisateurs très inquiets quant au respect de leur vie privée ont souvent tendance à penser que les entreprises se montrent opportunistes avec leurs informations personnelles, ce qui signifie que ces individus vont rarement répondre aux sollicitations de celles-ci.

De leur côté, Hong et Thong (2013) font la distinction entre les inquiétudes et les attentes des consommateurs quant au respect de leur vie privée. Ils définissent en effet les inquiétudes en matière de vie privée comme le résultat des perceptions qu'ont les individus sur la façon dont leurs informations personnelles sont traitées par certaines entités, ce qui, selon eux, est tout à fait différent des attentes que ceux-ci ont envers les pratiques de ces entités. En effet, le fait qu'un individu s'attende à ce qu'une entreprise protège ses informations personnelles de façon adéquate ne veut pas dire que celui-ci est d'office préoccupé au sujet de sa vie privée.

b. Seconde approche théorique : Le 'privacy calculus'

La seconde approche considère que les individus agissent de manière rationnelle lorsqu'ils prennent la décision de fournir des données personnelles, en évaluant la valeur perçue de l'échange, c'est-à-dire la différence entre ce qu'ils perçoivent et ce qu'ils offrent (Miltgen, 2011). Les conséquences futures de leurs actions sont comparées aux coûts actuels engendrés, et si celles-ci sont plus avantageuses, alors ils seront plutôt enclins à divulguer des données personnelles.

C'est la notion de *'privacy calculus'* qui est utilisée pour identifier ce modèle coût-bénéfice auquel auraient recours les individus (Culnan & Armstrong, 1999 ; Dinev & Hart, 2006). Cette notion trouve ses origines dans les travaux de Laufer et Wolfe (1977) qui l'avaient préalablement nommé *'calculus of behavior'* en faisant plutôt référence aux comportements dans le monde réel et non pas sur internet (cité dans Trepte et al., 2017).

Selon Gomez-Barroso, Feijoo et Martinez (2018), dans tout *'privacy calculus'*, les coûts sont évidents et sont représentés par les risques associés au fait que les données personnelles récoltées puissent engendrer des inconvénients à l'avenir. Les bénéfices peuvent quant à eux prendre différentes formes. Selon Hui, Tan et Goh (2006), ceux-ci peuvent par exemple inclure des réductions de prix, des gains de temps, une amélioration de l'estime de soi, plus de plaisir ou encore la découverte de nouveautés.

Selon Miltgen (2011), c'est la confrontation entre ces deux approches qui va faire naître le concept de 'paradoxe de la vie privée' que nous allons aborder dans la section suivante.

3. Paradoxe de la vie privée

Au sein de la littérature sur la vie privée, de nombreux auteurs se sont intéressés au concept de ‘paradoxe de la vie privée’ et à ses conséquences potentielles (Barnes, 2006 ; Dienlin & Trepte, 2015 ; Norberg et al., 2007 ; Taddicken, 2014).

Celui-ci se produit lorsque des individus, qui expriment pourtant des inquiétudes en matière de vie privée, agissent de manière contraire à celles-ci en acceptant de dévoiler des informations personnelles sans réel contrôle, en échange de certains avantages (Dinev, 2014). C’est la divergence entre les inquiétudes affichées et les comportements effectifs des individus qui caractérise le ‘paradoxe de la vie privée’ (Estienne, 2011).

L’apparition et le succès des réseaux sociaux au cours des dernières années, comme nous l’avons évoqué dans le chapitre précédent, ont rendu ce paradoxe de plus en plus apparent (Boyd & Ellison, 2007). En effet, le mode de fonctionnement des réseaux sociaux fait en sorte que leurs utilisateurs se sentent à l’aise de fournir des données personnelles car ils peuvent directement en ressentir les avantages, tels qu’une plus grande visibilité en ligne par exemple (Estienne, 2011).

Aujourd’hui, ce paradoxe n’est pas encore totalement expliqué par les chercheurs. Il semblerait néanmoins que le manque de sensibilisation des individus aux risques ainsi que le manque de connaissance de ceux-ci par rapport aux possibilités qu’ils leur sont offertes pour protéger leur vie privée pourraient jouer un rôle important, ainsi qu’une tendance générale à sous-estimer les risques liés à la divulgation d’informations privées (Taddicken, 2014). Dans ce contexte, il reste pertinent de s’interroger sur la perception qu’ont les utilisateurs de réseaux sociaux de l’étendue de la collecte et de l’utilisation qui peut être faite de leurs données personnelles pour mieux comprendre par la suite ce qui peut les freiner à divulguer des données personnelles sur un réseau social (Miltgen, 2011).

4. Règlement Général sur la Protection des Données

Le Règlement Général sur la Protection des Données (RGPD) est entré en vigueur le 25 mai 2018 après avoir été définitivement adopté deux ans plus tôt, le 14 avril 2016, par le Parlement européen. Il remplace ainsi la directive européenne du 24 octobre 1995 sur la protection des données personnelles. Celle-ci était en effet dépassée par l’évolution des normes technologiques. Les directives européennes sont des actes législatifs fixant des objectifs à tous les pays membres de l’Union tout en leur laissant la liberté de mettre en

place leurs propres mesures pour les atteindre. Au contraire, les règlements européens, comme le RGPD, sont des actes législatifs contraignants directement applicables à tous les pays membres de l'Union Européenne (Europa, 2018). Il n'y a donc pas de transposition nécessaire dans les droits nationaux comme c'était le cas auparavant. Toutefois, les pays membres bénéficient d'une marge de manœuvre leur permettant de préciser certaines règles ou conditions prévues par le RGPD.

C'est ainsi qu'en Belgique, une loi-cadre exécutant le nouveau règlement européen a été instaurée le 5 septembre 2018 (Salovic et al., 2018). Celle-ci remplace la loi belge du 8 décembre 1992. Une nouvelle autorité de contrôle a également été créée. Il s'agit de l'Autorité belge de Protection des Données (APD) qui, depuis le 25 mai 2018, aurait déjà reçu le signalement de plus de 317 incidents relatifs à des données personnelles en Belgique, avec comme secteurs les plus touchés, ceux des soins de santé et des assurances (APD, 2018).

De par son cadre homogène et cohérent à tous les pays européens, le RGPD cherche avant tout à renforcer la confiance des individus envers les entreprises ou administrations qui traitent leurs données personnelles. L'objectif est aussi de simplifier la circulation des données entre les pays membres.

Selon Calder (2016), ce règlement apporte des changements majeurs au cadre législatif existant précédemment afin d'atteindre ces objectifs.

a. Champ d'application

Les premiers changements se situent au niveau du champ d'application du règlement. Tout d'abord, l'harmonisation des règles entre tous les pays membres de l'Union Européenne est très avantageuse pour les entreprises opérant dans plusieurs pays en Europe. En effet, le règlement permet dorénavant à ces organisations de n'avoir à traiter qu'avec une seule autorité de contrôle. Ensuite, le RGPD oblige toutes les entreprises ou administrations collectant, stockant ou traitent des données personnelles sur des résidents européens à se conformer aux nouvelles règles. Cette portée territoriale élargie a pour conséquence que des entreprises basées en dehors de l'Europe peuvent être affectées par le RGPD dès lors qu'elles manipulent des données personnelles sur des résidents européens. C'est pour cette raison que des entreprises comme les GAFA, Airbnb ou encore Uber y sont également soumises.

b. Droits des personnes concernées

Les deuxièmes changements touchent aux droits des personnes concernées. Tout d'abord, deux nouveaux droits ont été introduits au sein de ce règlement. Le premier est le droit à l'oubli. En effet, les consommateurs ont maintenant le droit de demander à toute entreprise ou administration d'effacer toutes les données personnelles qu'ils détiennent sur eux. Le deuxième est le droit à la portabilité des données. Ainsi, toute personne peut à présent demander une copie de ses données personnelles afin de les transférer vers une autre entité.

Ensuite, le règlement renforce les règles à respecter en matière de profilage, c'est-à-dire la collecte d'informations sur un individu afin de les analyser pour en tirer des conclusions intéressantes. En effet, il est dorénavant obligatoire d'avertir les individus, au moyen de cookies par exemple, lorsqu'une activité de profilage est déclenchée par toute entreprise ou administration.

Enfin, les règles en matière de consentement sont renforcées. C'est ainsi que le consentement n'est plus valable en cas de silence, de cases pré-cochées ou d'inactivité de la part des individus comme c'était le cas auparavant. Toute entreprise ou administration doit non seulement demander de manière explicite le consentement des individus avant de pouvoir traiter leurs données personnelles mais doit aussi tenir un registre démontrant le consentement de ces mêmes individus.

c. Protection des données

Les troisièmes changements se situent au niveau de la protection des données personnelles. Le concept de *'privacy by design'* est ainsi introduit au sein du règlement. Il est en effet obligatoire de se mettre en conformité avec les règles liées à la protection des données dès la conception initiale de tout système de traitement de données personnelles. Toute entreprise ou administration doit ainsi prendre l'habitude de s'assurer du respect continu des exigences du règlement en la matière.

En outre, lorsque le traitement de données est réalisé à grande échelle ou de manière automatique, un outil d'évaluation d'impact du traitement des données, nommé DPIA pour *'Data Protection Impact Assessment'*, est requis par le règlement.

Enfin, le règlement oblige toute entreprise ou administration à tenir un registre d'activités de traitement de données. Celui-ci doit être directement mis à disposition lorsque

l'autorité de contrôle le demande. Dans ce registre, les entreprises doivent à la fois montrer quelles sont les données traitées, mais aussi, où, comment et pourquoi celles ont été traitées.

d. Responsabilité

Les quatrièmes changements touchent à la responsabilité des organisations. Il est dorénavant obligatoire de nommer un délégué à la protection des données dans les entreprises où de grandes quantités de données sont traitées. Celui-ci doit s'assurer que tout est en conformité avec le règlement. Grâce à ses connaissances en la matière, le délégué à la protection des données peut également servir de conseiller au responsable du traitement des données.

En outre, le règlement oblige les entreprises ou administrations à signaler toute violation à la protection des données personnelles à l'autorité compétente endéans les 72 heures après en avoir pris connaissance. Lorsque la violation représente un risque élevé pour les droits et les libertés des individus, ceux-ci doivent être contactés dans les meilleurs délais par l'organisation.

Par ailleurs, tout traitement de données à caractère privé en infraction avec le nouveau règlement peut amener les entreprises concernées à être lourdement sanctionnées. En effet, le montant des amendes peut aller jusqu'à 20 millions d'euros ou 4% du chiffre d'affaires annuel de l'exercice précédent (Perreau, 2018).

PARTIE 2 : CADRE DE RÉFÉRENCE – MODÈLE APCO

Pour rappel, la problématique choisie dans ce mémoire est la suivante :

Dans quelle mesure les pratiques du Big Data, le scandale Cambridge Analytica et l'entrée en vigueur du RGPD ont-ils une influence sur les inquiétudes des utilisateurs de Facebook quant à la protection de leur vie privée et l'accès qu'ils donnent à leurs données personnelles ?

Dans la première partie de ce mémoire, nous avons défini les concepts clés de notre problématique ainsi que les différentes controverses qui les entourent. À présent, nous allons mobiliser un modèle théorique de référence afin d'élaborer notre cadre conceptuel et les hypothèses que nous testerons de manière statistique dans la troisième partie de ce mémoire.

Chapitre 1 : Présentation du modèle APCO

Comme nous l'avons vu dans la partie précédente, la littérature autour de la protection des données personnelles est abondante. La majorité des travaux ont été publiés à partir de la deuxième moitié des années 90, période correspondant à la démocratisation d'Internet et à l'apparition des tout premiers navigateurs Web.

Depuis lors, certains auteurs ont essayé de concevoir un cadre général permettant de théoriser ce domaine de recherche. C'est ainsi que durant l'année 2011, trois articles de synthèse ont été rédigés dans ce sens (Bélanger & Crossler, 2011 ; Li, 2011 ; Smith et al., 2011). Bien qu'ils n'aboutissent pas totalement aux mêmes conclusions, ces articles se ressemblent fortement.

Dans le cadre de ce mémoire, nous avons décidé de nous concentrer sur l'article de synthèse rédigé par Smith, Dinev et Xu (2011), considérant celui-ci comme le plus pertinent pour notre recherche étant donné qu'il est généralisable à plusieurs domaines.

Ces trois auteurs identifient et analysent de multiples publications consacrées à la question de la protection de la vie privée parues entre 1960 et 2011. Ces publications relèvent de disciplines aussi diverses que les sciences sociales, le marketing, l'économie, les sciences politiques, la philosophie, l'économie, le droit, la psychologie ou encore les systèmes d'information. Ils s'appuient ainsi sur un total de 320 articles et 128 livres et

sections de livres. La conclusion de leur travail est que presque toutes les recherches empiriques relatives à la protection de la vie privée peuvent être considérées dans un cadre, qu'ils nomment 'modèle APCO' et synthétisent comme suit :

Antecedents (A) → Privacy Concerns (PC) → Outcomes (O)

Figure 1 : Modèle APCO (Smith et al., 2011)

Dans leur modèle APCO, Smith, Dinev et Xu (2011) considèrent que ce sont des antécédents (A) qui amènent les individus à former des inquiétudes en matière de vie privée (PC) qui, à leur tour, provoquent des conséquences (O).

Les auteurs font des inquiétudes en matière de vie privée (PC) le pivot de leur modèle. En ce qui concerne les antécédents (A), ils identifient les expériences en matière de non-respect de la vie privée, la sensibilisation à la protection de la vie privée, les différences de personnalité, les différences démographiques ou encore les différences culturelles et climatiques. Quant aux conséquences potentielles de ces inquiétudes en matière de vie privée, Smith, Dinev et Xu (2011) identifient certaines réactions comportementales telles que la moindre divulgation de données personnelles, les réticences à contracter et le jeu du *'privacy calculus'*. Ils envisagent également l'éventuelle nécessité de l'intervention d'un régulateur.

Dans leur article de synthèse, Smith, Dinev et Xu (2011) soulignent que beaucoup d'études se sont intéressées aux intentions des individus d'adopter des comportements donnés plutôt qu'à leurs comportements réels. En outre, ils notent que la majorité de la recherche existante se concentre surtout sur la dernière portion du modèle, c'est-à-dire sur l'examen des relations entre le degré des inquiétudes en matière de vie privée et les différents comportements associés à ces inquiétudes. Ils constatent par contre que la première portion du modèle, qui met l'accent sur les antécédents des inquiétudes en matière de vie privée, a reçu moins d'attention de la part des chercheurs.

C'est dans ce cadre que notre mémoire intervient puisque nous cherchons à mesurer les impacts respectifs de la familiarité du *Big Data*, de l'exposition aux médias sur la question de la vie privée et de la connaissance du RGPD sur les comportements réels des utilisateurs de Facebook quant à la divulgation de leurs données personnelles et à l'adoption de mesures de protection de leur vie privée.

De manière à répliquer parfaitement le modèle APCO, nous insérons les inquiétudes en matière de vie privée en tant qu'intermédiaire entre les antécédents et les comportements.

La problématique étudiée par ce mémoire se divisera ainsi en deux questions de recherche distinctes :

1. *Quels antécédents potentiels peuvent influencer les inquiétudes des utilisateurs de Facebook quant au respect de leur vie privée ?*

2. *Comment les inquiétudes en matière de vie privée peuvent-elles influencer les comportements des utilisateurs de Facebook quant au partage de leurs données personnelles et quant à la protection de leur vie privée ?*

Considérant que beaucoup d'auteurs ont utilisé et testé avec succès le modèle APCO dans leurs recherches (Alashoor et al., 2017 ; Benamati et al., 2017 ; Heravi et al., 2018 ; Kuo & Talley, 2014 ; Ozdemir et al., 2017 ; Zhang et al., 2013), nous le considérons comme suffisamment robuste pour guider notre étude sur les comportements des individus en matière de vie privée dans le contexte des réseaux sociaux, et plus particulièrement de Facebook.

Chapitre 2 : Antécédents

Nous abordons ici notre première question de recherche :

Quels antécédents potentiels peuvent influencer les inquiétudes des utilisateurs de Facebook quant au respect de leur vie privée ?

Nous avons retenus les antécédents suivants : la familiarité avec le Big Data, l'exposition aux médias sur la question de la vie privée, la connaissance des nouvelles règles mises en place par le RGPD et les variables sociodémographiques que sont l'âge, le genre et le niveau d'éducation.

1. Familiarité avec le Big Data

Comme indiqué précédemment, l'émergence des pratiques du *Big Data*, ainsi que celles du marketing prédictif, a d'énormes implications sur la vie privée des consommateurs.

De nombreux auteurs se sont penchés sur la question. Ainsi, Watson (2014) montre que plus les individus sont au courant des pratiques du *Big Data*, plus ils sont enclins à manifester des inquiétudes concernant l'utilisation de leurs données privées. Mais beaucoup d'entre eux ne sont pas du tout conscients du nombre incalculable d'agents et d'algorithmes qui sont actuellement en train de collectionner et stocker leurs données personnelles pour une utilisation ultérieure (Boyd & Anderson, 2012). Selon Clemons, Wilson et Jin (2014), les consommateurs n'ont au départ que très peu de connaissance de ces pratiques.

Alashoor, Han et Joseph (2017) divisent la familiarité avec le *Big Data* en deux catégories s'opposant ; la familiarité avec le terme *Big Data* et ses avantages et la familiarité avec les implications du *Big Data*. D'un côté, les individus qui sont au courant des avantages résultant des pratiques du *Big Data*, comme la personnalisation des services, sont moins susceptibles d'être préoccupés par le respect de la confidentialité de leurs données parce qu'ils perçoivent tout à fait que pour pouvoir profiter des avantages des technologies actuelles, il faut être prêt à sacrifier une partie de sa vie privée (Alashoor et al., 2017). D'un autre côté, le fait d'être au courant des utilisations omniprésentes de leurs données personnelles peut amener les consommateurs à être beaucoup plus susceptibles quant au respect de leur vie privée (Alashoor et al., 2017).

Dans ce contexte, nous poserons les hypothèses suivantes :

H1a : La familiarité avec le terme *Big Data* et ses bénéfices tend à réduire les inquiétudes en matière de vie privée.

H1b : La familiarité avec les implications du *Big Data* tend à augmenter les inquiétudes en matière de vie privée.

2. Exposition aux médias sur la question de la vie privée

Dans les médias, presque tous les faits relatifs à la vie privée se focalisent sur la violation de celle-ci. Le scandale Cambridge Analytica, évoqué dans la première partie de ce mémoire, en est un exemple récent. Il concerne l'accès non-autorisé aux données personnelles de près de 87 millions d'utilisateurs de Facebook et de l'utilisation qui a pu être faite de ces données pour influer sur le résultat de l'élection présidentielle américaine. Il en a résulté une couverture médiatique de très grande ampleur à l'échelle mondiale. Depuis mars 2018, ces révélations ont ramené le sujet de la vie privée sur Internet au premier plan dans les yeux des individus.

Westin (1990) a été le premier auteur à remarquer que l'exposition médiatique à des faits relatifs aux atteintes de la vie privée pouvait augmenter le degré de préoccupation en matière de vie privée des individus (cité dans Benamati et al., 2017). C'est ensuite Smith, Milberg et Burke (1996) qui ont confirmé cette relation de manière empirique dans leurs premiers travaux concernant les préoccupations liées à la vie privée. Plus tard, Malhotra, Kim et Agarwal (2004) n'ont plus considéré l'exposition médiatique comme un antécédent mais bien comme une de leurs sept covariables pouvant influencer les réactions des consommateurs face aux problèmes de confidentialité des données. De manière similaire, Li, Sarathy et Xu (2010) utilisent l'exposition médiatique comme une de leurs six covariables mais aucune d'entre elles ne s'avèrent finalement significative.

Dans leur modèle APCO, Smith, Dinev et Xu (2011) considèrent l'exposition médiatique comme un des facteurs permettant de déterminer la connaissance qu'ont les individus des risques relatifs à la confidentialité de leurs données. Benamati, Ozdemir et Smith (2017) vont plus loin en posant que ce sont les expositions médiatiques ainsi que les expériences passées d'atteinte à la vie privée qui forment le degré de connaissance des individus par rapport à la problématique du respect de la vie privée.

L'exposition aux médias sur la question de la vie privée et en particulier sur les menaces liées à celle-ci et qui est donc de nature négative, peut pousser les individus à être plus inquiets par rapport à la confidentialité de leurs données personnelles.

L'hypothèse suivante sera alors posée :

H2 : L'exposition aux médias sur la question de la vie privée tend à augmenter les inquiétudes en matière de vie privée.

3. Connaissance du RGPD

Comme mentionné dans la revue de la littérature, le nouveau Règlement Général sur la Protection des Données (RGPD) est depuis le 25 mai 2018 applicable à toutes les entités collectant et traitant des données sur des résidents européens. Les consommateurs ont été mis au courant de ces nouvelles règles par le biais de nombreux mails, articles ou encore cookies à l'entrée des sites Web. Leurs connaissances en la matière ont donc potentiellement augmenté.

Dommeyer et Gross (2003) font partie des premiers auteurs à s'être aperçus que les connaissances sur les réglementations touchant à la protection des données et sur les options de protection disponibles peuvent amplifier le sentiment de contrôle des individus et ainsi diminuer leurs inquiétudes liées à la vie privée. De leur côté, Lwin, Wirtz et Williams (2007) mentionnent que l'efficacité perçue des réglementations en vigueur ainsi que l'efficacité perçue de leur exécution réduisent les préoccupations des consommateurs vis-à-vis de leur vie privée. En effet, selon eux, les utilisateurs d'Internet se reposent encore beaucoup sur les garanties institutionnelles et les lois protégeant leurs droits civils car ils n'ont souvent que des connaissances et des moyens limités pour évaluer la sécurité réelle de leurs données personnelles.

Néanmoins, dans leur article introduisant le modèle APCO, Smith, Dinev et Xu (2011) notent que peu d'études se sont intéressées aux connaissances des individus en matière de réglementation sur la protection des données et sur la manière dont ces connaissances peuvent influencer les inquiétudes et les comportements de ces individus. Depuis lors, Miltgen et Smith (2015) ont démontré que le niveau de connaissance en matière de réglementation relative à la protection de la vie privée a une influence positive sur la perception qu'ont les individus d'être protégés par ces règles.

Cette impression de protection est ensuite considérée par ces auteurs comme un facteur déterminant de la confiance des individus envers les entités impliquées dans le traitement de leurs données privées. Dans ce travail, nous considérons la connaissance de la règlementation actuelle relative à la protection des données personnelles comme un antécédent réduisant les inquiétudes en matière de vie privée.

Nous poserons ainsi l'hypothèse suivante :

H3 : La connaissance des nouvelles règles mises en place par le RGPD tend à réduire les inquiétudes en matière de vie privée.

4. Données démographiques

L'article de Smith, Dinev et Xu (2011) considère les différences démographiques comme un des antécédents permettant d'expliquer les inquiétudes des individus vis-à-vis du respect de leur vie privée. Peu de recherches ont été effectuées afin de comprendre les facteurs précis influençant ces différences. Dans le cadre de ce mémoire, nous nous intéresserons aux influences respectives de l'âge, du genre et du niveau d'éducation sur les inquiétudes en matière de respect de la vie privée.

a. Age

Selon Morris, Venkatesh et Ackerman (2005), l'âge a une influence certaine sur la capacité des individus à traiter l'information mais aussi sur leurs interactions avec la technologie, ce qui rend cette variable démographique particulièrement intéressante (cité dans Miltgen et al., 2014). Miltgen et Peyrat-Guillard (2014) prouvent que les préoccupations en matière de vie privée diffèrent entre personnes de groupes d'âge différents. En effet, selon eux, même si les jeunes expriment actuellement plus d'aspirations à une meilleure protection de leur vie privée, il n'en reste pas moins que leurs inquiétudes sont moins fortes que celles des personnes plus âgées.

Nous postulerons donc l'hypothèse suivante :

H4 : Les jeunes sont moins inquiets quant au respect de leur vie privée que les personnes plus âgées.

b. Genre

Des différences de genre en matière de préoccupations relatives à la vie privée ont été identifiées à de multiples reprises (Lin et al., 2013 ; Rowan & Behlinger, 2014 ; Sheehan,

1999). Benamati, Ozdemir et Smith (2017) postulent ainsi que les femmes sont plus préoccupées par la protection de leur vie privée que les hommes. En effet, les hommes et les femmes partagent des types d'information différents en ligne, les femmes ayant tendance à échanger des données qu'elles perçoivent comme étant plus sensibles.

L'hypothèse suivante sera ainsi utilisée :

H5 : Les femmes sont plus inquiètes quant au respect de leur vie privée que les hommes.

c. Éducation

Zhang, Chen et Lee (2013) montrent que le niveau d'éducation est positivement corrélé avec le niveau des préoccupations en matière de vie privée, ce qui veut dire que lorsque l'un des deux augmente, l'autre diminue. En effet, selon eux, une personne ayant bénéficié d'un niveau d'instruction élevé, a plus de chance d'être mieux informée à propos des problèmes de confidentialité et de sécurité des données personnelles, ce qui va lui permettre de mieux en appréhender les circonstances et ainsi réduire ses craintes.

L'hypothèse suivante sera donc :

H6 : Les individus ayant un niveau d'éducation plus élevé sont moins inquiets quant au respect de leur vie privée.

Chapitre 3 : Conséquences

Nous arrivons ici à notre deuxième question de recherche :

Comment les inquiétudes en matière de vie privée peuvent-elles influencer les comportements des utilisateurs de Facebook quant au partage de leurs données personnelles et quant à la protection de leur vie privée ?

Nous avons retenu les conséquences suivantes : la divulgation de données personnelles et l'adoption de mesures de protection de la vie privée.

1. Divulgation de données personnelles

Le dévoilement de soi est une notion définie par McCroskey et Richmond (1997). Il s'agit du fait de communiquer toute information à propos de soi-même de manière intentionnelle ou involontaire à une autre personne. Dans le contexte des réseaux sociaux, Heravi, Mani, Choo et Mubarak (2017) définissent la divulgation de données personnelles comme étant le fait de fournir toute information personnelle sur son profil, que ce soit son nom, sa photo, ses coordonnées, sa situation amoureuse, ses affiliations politiques ou religieuses, son emploi ou encore son niveau d'éducation mais aussi dans ses communications avec les autres, comme lors de mises à jour de son statut ou de publications de commentaires ou de *'likes'*.

Metzger (2004) est l'un des premiers auteurs à avoir démontré que les inquiétudes à propos de la protection des données personnelles amènent de nombreux consommateurs à être plus réticents sur le dévoilement de soi en ligne. De manière similaire, Milne et Culnan (2004) déclarent que les préoccupations liées à la vie privée réduisent le dévoilement de soi et contraignent les utilisateurs à adopter des comportements leur permettant de faire face aux risques de violation de leurs données personnelles (cité dans Chen, 2018). Selon Dienlin et Trepte (2015), les utilisateurs de Facebook qui sont inquiets quant au respect de leur vie privée ont plutôt tendance à divulguer moins d'informations personnelles, que ce soit sur leur profil ou dans leurs publications. De son côté, Taddicken (2014) détermine que la relation directe entre les inquiétudes liées à la vie privée et le dévoilement de soi n'est pas significative. Il explique en effet que la divulgation de données personnelles dépend plutôt du caractère sensible de l'information à dévoiler.

Malgré des résultats contradictoires, comme ceux de Taddicken (2014), la plupart des études empiriques soutiennent que les inquiétudes en matière de vie privée sont négativement associées au dévoilement de soi en ligne (Krasnova et al., 2009 ; Xu, Michaels & Chen, 2013).

Nous poserons ainsi l'hypothèse suivante :

H7 : Les inquiétudes en matière de vie privée ont un impact négatif sur la divulgation de données personnelles sur Facebook.

2. Mesures de protection de la vie privée

Depuis le scandale Cambridge Analytica, Facebook a décidé de rassembler l'ensemble de ses paramètres de confidentialité en un seul et même endroit afin de faciliter le contrôle de ses utilisateurs par rapport à leurs données personnelles. Dans la littérature, les mesures de protection de la vie privée prises par les utilisateurs de Facebook sont généralement évaluées sur base de l'utilisation et de la familiarité avec les paramètres de confidentialité mais aussi en fonction des comportements relatifs à l'ajout d'amis et à l'adhésion à de nouveaux groupes (Heravi et al., 2017).

Pour ce qui est tout d'abord des mesures de protection en lien avec les paramètres de confidentialité, Kokolakis (2017) explique, par exemple, que les utilisateurs ayant vécu personnellement une violation de leur vie privée sont plus susceptibles de modifier leurs paramètres de confidentialité. De son côté, Chen (2018) déclare que les consommateurs inquiets du respect de leur vie privée et qui ont la conviction d'être capable de faire attention au partage de leurs données personnelles sur Facebook, sont souvent les plus prompts à modifier leurs paramètres de confidentialité afin de limiter la visibilité de leur profil, par exemple.

En ce qui concerne ensuite le fait de limiter le nombre d'amis ou l'adhésion à des groupes, Young et Quan-Haase (2013) démontrent que les individus qui n'acceptent que de façon limitée l'ajout d'étrangers dans leur groupe d'amis sur Facebook ont plutôt tendance à être plus inquiets en ce qui concerne leurs données personnelles. Stutzman et Kramer-Duffield (2010) trouvent également qu'un niveau élevé d'inquiétudes par rapport au respect de la vie privée est positivement associé au fait d'avoir un profil restreint aux amis uniquement (cité dans Heravi et al., 2017).

L’hypothèse suivante sera ainsi posée :

H8 : Les inquiétudes en matière de vie privée ont un impact positif sur l’utilisation de mesures de protection de la vie privée sur Facebook.

Chapitre 4 : Modèle à tester

Compte tenu de tout ce qui précède, nous avons construit notre propre cadre conceptuel. Celui-ci s'inspire du modèle APCO de Smith, Dinev et Xu (2011). Comme antécédents aux inquiétudes en matière de vie privée, les impacts respectifs de la familiarité avec le *Big Data*, de l'exposition aux médias sur la question de la vie privée, de la connaissance des nouvelles règles du RGPD, de l'âge, du genre et de l'éducation seront testés. Les conséquences des inquiétudes en matière de vie privée seront quant à elles évaluées en examinant la relation entre, d'une part, les comportements relatifs au dévoilement de soi et, d'autre part, les comportements relatifs à l'utilisation de mesures de protection de la vie privée.

Notre cadre conceptuel comprend donc sept antécédents et deux conséquences potentielles aux inquiétudes en matière de vie privée et est ainsi représenté par la figure suivante :

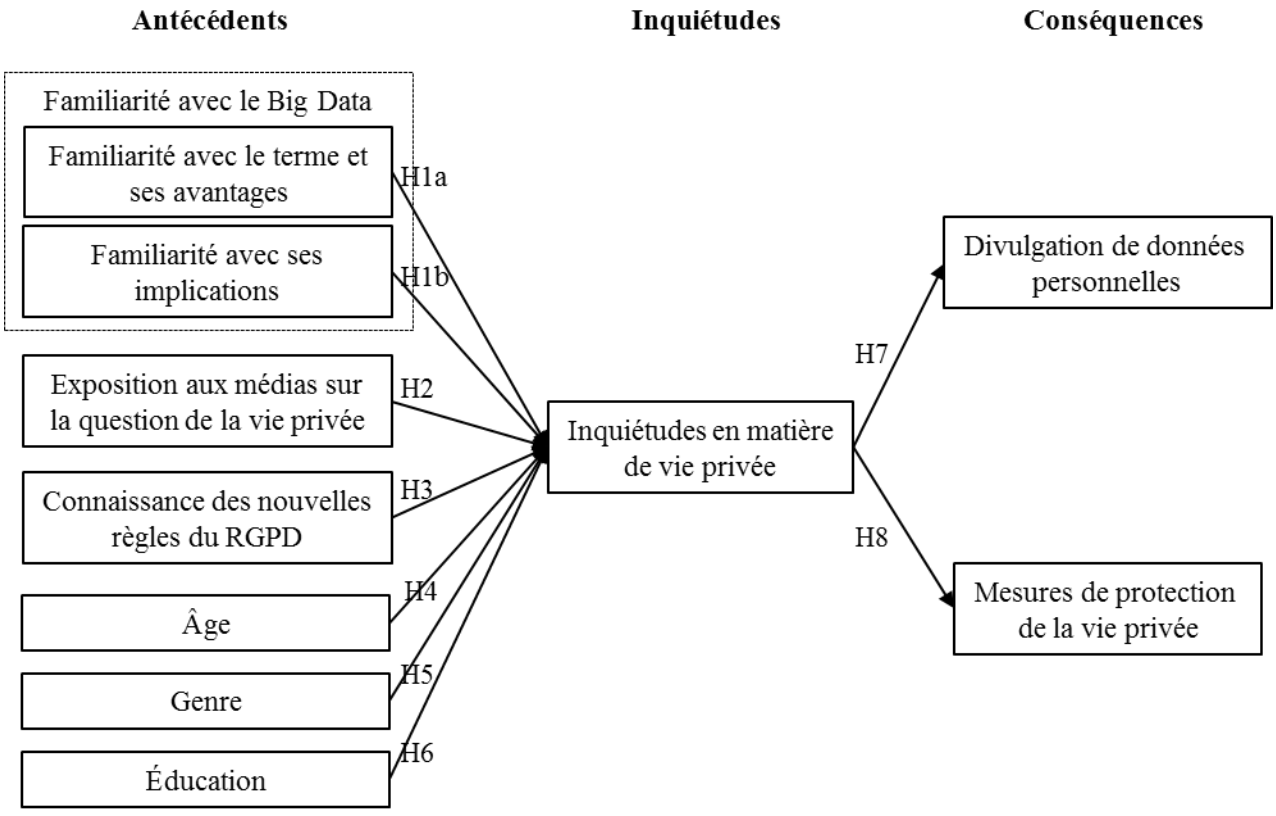


Figure 2 : Modèle conceptuel

1. Liste des hypothèses

Voici la liste des hypothèses que nous avons posées afin de répondre à notre première question de recherche qui, pour rappel, est la suivante :

Quels antécédents potentiels peuvent influencer les inquiétudes des utilisateurs de Facebook quant au respect de leur vie privée ?

H1a : La familiarité avec le concept *Big Data* et ses avantages tend à réduire les inquiétudes en matière de vie privée.

H1b : La familiarité avec les implications du *Big Data* tend à augmenter les inquiétudes en matière de vie privée.

H2 : L'exposition aux médias sur la question de la vie privée tend à augmenter les inquiétudes en matière de vie privée.

H3 : La connaissance des nouvelles règles mises en place par le RGPD tend à réduire sur les inquiétudes en matière de vie privée.

H4 : Les jeunes sont moins inquiets quant au respect de leur vie privée que les personnes plus âgées.

H5 : Les femmes sont plus inquiètes quant au respect de leur vie privée que les hommes.

H6 : Les individus ayant un niveau d'éducation plus élevé sont moins inquiets quant au respect de leur vie privée.

Ensuite, voici les hypothèses posées nous aidant à répondre à notre deuxième question, qui est pour rappel :

Comment les inquiétudes en matière de vie privée peuvent-elles influencer les comportements des utilisateurs de Facebook quant au partage de leurs données personnelles et quant à la protection de leur vie privée ?

H7 : Les inquiétudes en matière de vie privée ont un impact négatif sur la divulgation de données personnelles sur Facebook.

H8 : Les inquiétudes en matière de vie privée ont un impact positif sur l'utilisation de mesures de protection de la vie privée sur Facebook.

38.

PARTIE 3 : ÉTUDE QUANTITATIVE

Dans la troisième partie de ce mémoire, nous présentons l'étude quantitative qui a été réalisée afin de tester les hypothèses associées à notre modèle de recherche, inspiré des travaux de Smith, Dinev et Xu (2011). Nous allons ainsi développer notre contribution au domaine de recherche relatif à la protection de la vie privée.

L'approche quantitative a été privilégiée car celle-ci permet non seulement d'investiguer des comportements, des opinions, des faits ou encore des critères sociodémographiques mais aussi de tirer des conclusions généralisables à la population (Pleyers, 2017). Comme indiqué dans les chapitres précédents, le sujet de la vie privée a fait l'objet de nombreux travaux. Nous jugeons par conséquent que la base théorique disponible pour la construction d'une étude quantitative est suffisante et qu'il n'est pas nécessaire de réaliser une étude qualitative.

L'objectif de cette étude quantitative est de mesurer les relations existant entre les différentes variables que nous avons retenues. À cette fin, nous avons procédé à l'envoi d'un questionnaire en ligne. Une fois les données récoltées, nous avons effectué une série de manipulations au travers du logiciel SPSS.

Nous concluons cette partie de notre mémoire par une discussion générale sur les résultats de notre étude quantitative.

Chapitre 1 : Méthodologie de la recherche

Afin de récolter les données primaires nécessaires à la vérification de nos hypothèses, nous avons conçu un questionnaire en ligne à l'aide du logiciel Google Forms. Nous avons diffusé celui-ci par e-mail et via les réseaux sociaux. La diffusion en ligne a été privilégiée par rapport à d'autres méthodes de diffusion parce que celle-ci permet d'atteindre beaucoup de personnes en peu de temps et à faible coût. Cette méthode de récolte offre une forme d'anonymat aux répondants qui les pousse à être plus honnêtes dans leurs réponses. Elle élimine les biais liés à la présence d'un enquêteur tout en fournissant certaines possibilités de contrôle comme la mise en place de questions filtres ou encore l'obligation de respecter l'ordre des questions (Pleyers, 2017). Grâce à cette technique, les réponses sont encodées de manière systématique dans une base de données et les erreurs de saisie sont réduites (Lambin & de Moerloose, 2012).

Néanmoins, cette méthode comporte certaines limites dont il faut être conscient lors de l'analyse des résultats. La représentativité de l'échantillon est plus faible que celle résultant d'autres méthodes, surtout lorsque c'est le réseau de contacts de l'enquêteur qui est mobilisé pour la diffusion du questionnaire (Lambin & de Moerloose, 2012). De plus, les taux de réponse sont assez bas, le risque d'abandon en cours de route élevé et il est impossible de contrôler le taux d'implication réel des répondants (Pleyers, 2017).

Dans ce chapitre, nous allons détailler la structure de notre questionnaire, le pré-test effectué, la matrice hypothèses-questions et le choix de l'échantillon.

1. Conception du questionnaire

Le questionnaire débute par un petit mot d'introduction motivant les répondants à participer à l'enquête et les remerciant d'avance pour leur collaboration.

Afin d'éviter d'avoir des valeurs manquantes au sein de notre base de données, nous avons rendu toutes les questions obligatoires. Nous avons fait en sorte que le questionnaire soit le plus concis possible afin d'éviter que certains répondants ne prennent la décision d'arrêter de répondre avant la fin. De plus, nous avons fait attention à ce que nos questions ne soient pas orientées et que les termes utilisés soient les plus simples possibles à comprendre pour les répondants (Pleyers, 2017).

Pour ce qui est de l'ordre des questions, nous avons appliqué le principe de l'entonnoir qui consiste à poser des questions plus générales au début du questionnaire et des questions plus précises à la fin (Lambin et de Moerloose, 2012). C'est ainsi que les questions relatives aux variables sociodémographiques se trouvent dans la dernière section de notre questionnaire. En posant des questions trop personnelles dès le début, certains répondants pourraient en effet renoncer à répondre à la suite du questionnaire. Nous avons également fait attention à ce que les questions soient posées dans un ordre logique (Malhotra, 2014). Enfin, nous avons choisi de poser certaines questions après d'autres de manière à ce que celles-ci ne puissent pas influencer les réponses des répondants. C'est ainsi que les questions relatives à l'exposition aux médias, au sein desquelles il est fait mention du récent scandale Cambridge Analytica, ont été placées dans l'avant-dernière section du questionnaire. Nous avons estimé que le rappel explicite de cette affaire aurait pu affecter les réponses aux questions relatives à la connaissance du RGPD et à la familiarité avec le Big Data.

Nous avons décidé de scinder le questionnaire en 7 sections afin de le rendre le plus fluide possible aux yeux des répondants.

La première section du questionnaire comporte une question filtre permettant d'exclure tous les répondants n'acceptant pas que leurs données personnelles soient enregistrées et analysées de façon anonyme dans un cadre purement scientifique. Cette question répond à l'exigence du Règlement Général sur la Protection des Données (RGPD) selon laquelle il est obligatoire de recueillir le consentement univoque de tout répondant avant de pouvoir traiter ses données personnelles. Une seule personne a refusé de donner son accord.

La deuxième section du questionnaire inclut une autre question filtre permettant d'exclure tous les répondants ne possédant pas de compte Facebook. Cette question est nécessaire étant donné que nous nous intéressons aux comportements spécifiques des utilisateurs de ce réseau social dans la section suivante. Cinq personnes ont mentionné ne pas avoir de compte Facebook.

La troisième section du questionnaire se concentre sur les comportements des répondants sur Facebook. Une question sur la fréquence d'utilisation de ce réseau social est d'abord posée. Celle-ci servira à analyser l'échantillon de manière plus précise dans le prochain chapitre. Ensuite, ce sont les conséquences étant étudiées dans le cadre de notre modèle conceptuel qui sont mesurées avec, dans un premier temps, les questions relatives à la divulgation de données personnelles sur Facebook et, dans un second temps, celles relatives à l'utilisation de mesures de protection de la vie privée sur Facebook.

La quatrième section du questionnaire comporte les questions relatives aux inquiétudes en matière de respect de la vie privée.

La cinquième section du questionnaire s'intéresse au premier des antécédents aux inquiétudes en matière de vie privée étudiés dans notre modèle conceptuel. Il s'agit de celui relatif à la familiarité avec le *Big Data*. Deux séries de questions sont posées. Les premières sont relatives à la familiarité avec le concept et ses avantages et les secondes à la familiarité avec les implications du *Big Data*.

La sixième section du questionnaire se concentre sur deux autres des antécédents aux inquiétudes en matière de vie privée étudiés dans notre modèle conceptuel. Les questions relatives à la connaissance des nouvelles règles mises en place par le RGPD sont d'abord

posées. Ensuite viennent les questions relatives à l'exposition aux médias sur la question de la vie privée.

Enfin, la dernière section du questionnaire identifie les caractéristiques sociodémographiques des répondants à savoir leur genre, âge, niveau d'éducation et statut professionnel.

Le questionnaire se termine par un petit mot de remerciement à l'adresse des répondants.

Pour chacune des variables étudiées, nous avons veillé à mobiliser à chaque fois au moins quatre items de mesure. Nous nous sommes inspirés de la littérature existante pour le choix de presque toutes les échelles de mesure utilisées. Celles-ci ont été traduites en français et légèrement modifiées afin d'assurer une certaine cohérence entre les questions. Pour mesurer la connaissance des répondants aux nouvelles règles mises en place par le RGPD, nous avons décidé de nous baser sur certains articles du nouveau règlement touchant aux droits des personnes concernées afin de constituer nos différents items de mesure.

La liste complète des items de mesure ainsi que leurs sources respectives sont présentées dans la figure 3.

Nous avons pris la décision d'utiliser les échelles d'évaluation de type Likert en 7 points pour mesurer tous les items de mesure de nos variables. À chaque question, les répondants ont le choix entre plusieurs étiquettes allant de 1 = 'pas du tout d'accord' à 7 = 'tout à fait d'accord' pour indiquer leur accord ou désaccord avec les affirmations qui leur sont proposées. Le nombre impair de propositions offre la possibilité aux répondants d'exprimer un avis neutre, ce qui permet de ne pas forcer leur réponse. Cet avis neutre est représenté par le point 4 = 'ni d'accord, ni pas d'accord' présent dans la liste. Nous avons pris soin d'insérer une description complète de toutes les catégories de l'échelle de Likert au tout début du questionnaire afin de s'assurer de la bonne compréhension de celles-ci par les répondants.

Ces échelles de Likert sont très répandues car elles possèdent beaucoup d'avantages, comme le fait d'être simples à concevoir et à manipuler pour l'enquêteur ou encore le fait de pouvoir être rapidement compréhensibles pour les répondants (Malhotra, 2014). Le fait de n'avoir employé que ce type d'échelle dans le corps du questionnaire permet aux répondants de s'habituer plus facilement à la manière d'y répondre et augmente ainsi la

fluidité du questionnaire. Néanmoins, il faut tenir compte du fait que cela peut aussi engendrer des effets de halo poussant certains répondants à sélectionner de manière systématique les mêmes réponses.

Le tableau suivant présente les différents items de mesure sélectionnés pour chaque variable ainsi que les sources dont ils proviennent :

Nom de la variable	**Liste des items de mesure**	**Sources**
Familiarité avec le concept Big Data et ses avantages	(1) J'ai déjà entendu parler du Big Data (2) Je suis conscient des avantages du Big Data pour les individus, notamment en ce qui concerne la personnalisation des services offerts (3) Je suis conscient de la valeur du Big Data pour les entreprises, notamment en ce qui concerne l'exploitation plus rapide des données collectées (4) Je connais la signification du terme Big Data	Alashoor et al. (2017)
Familiarité avec les implications du Big Data	(1) Je suis conscient que mes données personnelles sur les réseaux sociaux puissent être collectées (2) Je suis conscient que mes données personnelles sur les réseaux sociaux puissent être stockées pendant de nombreuses années (3) Je suis conscient que mes données personnelles sur les réseaux sociaux puissent être analysées de différentes façons (4) Je suis conscient que mes données personnelles sur les réseaux sociaux puissent être utilisées pour cibler les services ou les informations les plus susceptibles de m'intéresser (5) Je suis conscient que mes données personnelles sur les réseaux sociaux puissent être partagées avec différentes tierces parties	Krasnova et al. (2009) Alashoor et al. (2017)
Exposition aux médias sur la question de la vie privée	(1) Presque tous les jours, j'entends ou je lis des nouvelles sur la violation de la confidentialité des données personnelles (2) Beaucoup d'informations sont publiées sur la façon dont nos données personnelles peuvent être utilisées ou détournées par les entreprises (3) Le sujet du respect de la vie privée fait beaucoup parler de lui dans l'actualité (4) J'ai entendu ou lu beaucoup d'informations sur l'affaire Cambridge Analytica au cours de ces derniers mois	Smith et al. (1996) Benamati et al. (2017)
Connaissance des nouvelles règles mises en place par le RGPD	(1) J'ai le droit de demander à toute entreprise ou administration d'avoir accès aux données personnelles que celle-ci possède sur moi (2) J'ai le droit de demander à toute entreprise ou administration la rectification de mes données personnelles si celles-ci sont inexactes (3) J'ai le droit de demander à toute entreprise ou administration que mes données personnelles soient effacées (4) J'ai le droit de demander à toute entreprise ou administration une copie de mes données personnelles afin de transférer celles-ci vers toute autre entreprise ou administration (5) J'ai le droit de demander à toute entreprise ou administration de ne pas traiter mes données personnelles	Basé sur: Art. 15 Art. 16 Art. 17 Art. 20 Art. 21 du RGPD (2016)

Inquiétudes en matière de vie privée	(1) Je crains que les réseaux sociaux collectent trop d'informations personnelles sur moi (2) Je crains que des personnes non autorisées puissent accéder à mes informations personnelles sur les réseaux sociaux (3) Je crains que les informations que je partage sur les réseaux sociaux puissent être utilisées à mauvais escient (4) Je crains que les réseaux sociaux utilisent mes informations personnelles à des fins qui ne me sont pas communiquées	Dinev & Hart (2004) Zlatolas et al. (2015)
Divulgation de données personnelles	(1) Mon profil Facebook en dit long sur moi (2) Je tiens mes amis au courant de ce qui se passe dans ma vie à travers Facebook (3) Je révèle beaucoup d'informations sur moi sur Facebook, notamment par mes publications (4) Cela ne me dérange pas de mettre des informations personnelles sur Facebook	Krasnova et al. (2009) Zlatolas et al. (2015) Chen (2018)
Mesures de protection de la vie privée	(1) Je fais attention à ce que je publie sur les réseaux sociaux (2) Je limite (au-delà des paramètres par défaut) l'accès à mes informations personnelles, à mes publications ou à mes pages likées sur Facebook (3) J'évite d'utiliser des applications tierces, des jeux ou des tests sur Facebook (4) Je fais attention avec qui je suis ami sur Facebook (5) J'empêche Facebook d'avoir accès à ma localisation	Benamati et al. (2017)

Figure 3 : Liste des items de mesure

2. Pré-test

Avant d'utiliser notre questionnaire sur le terrain, nous avons réalisé un pré-test. C'est une étape cruciale dans l'élaboration d'un bon questionnaire. L'objectif est de détecter d'éventuels problèmes présents dans la conception du questionnaire tout en s'assurant du bon fonctionnement de celui-ci (Malhotra, 2014). Nous avons administré ce pré-test à cinq personnes, présentes dans notre entourage et cercle d'amis proches.

Suite à ce pré-test, certaines difficultés ont été identifiées par les répondants. Cela nous a permis d'améliorer le questionnaire. Tout d'abord, certaines questions ont été modifiées afin de les rendre plus compréhensibles pour les répondants ou d'éviter d'orienter les réponses dans un sens précis. Certains termes traduits de l'anglais vers le français portaient également à confusion. Il a donc été décidé d'utiliser des synonymes afin d'assurer plus de clarté. De plus, pour certains items relatifs aux comportements de divulgation de soi et de protection de la vie privée sur Facebook, les répondants ont manifesté quelques hésitations. Ceux-ci ne savaient pas toujours s'ils devaient répondre par rapport à leur comportement spécifique sur Facebook ou sur les réseaux sociaux en

général. Nous avons clarifié ces items en faisant attention à inclure à chaque fois le terme Facebook dedans.

La version finale du questionnaire après pré-test se trouve à l'annexe 1.

3. Matrice hypothèses-questions

Le tableau suivant présente les différentes questions du questionnaire et indique à chaque fois le lien avec les différentes hypothèses auxquelles elles tentent de répondre. Il permet ainsi de vérifier que toutes les hypothèses sont au moins couvertes par une question.

Toutefois, certaines remarques s'imposent. Tout d'abord, comme les deux premières questions de notre questionnaire sont des questions filtres, elles ne correspondent à aucune des hypothèses et ne sont pas reprises dans ce tableau. Ensuite, étant donné que la troisième et la dernière question de notre questionnaire ne servent qu'à analyser l'échantillon de manière plus précise, elles ne sont pas non plus intégrées à ce tableau. Il s'agit de la question sur la fréquence d'utilisation de Facebook et de la question sur le statut professionnel des répondants.

	Q4	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
H1a			X	X						
H1b			X		X					
H2			X				X			
H3			X			X				
H4			X						X	
H5			X					X		
H6			X							X
H7	X		X							
H8		X	X							

Figure 4 : Matrice hypothèses-questions

4. Choix de l'échantillon

Pour ce qui est de la population cible de notre questionnaire, nous avons choisi de couvrir tous les résidents belges présents sur Facebook. Nous considérons en effet que les problèmes touchant à la protection des données personnelles peuvent potentiellement toucher tous les utilisateurs de ce réseau social en Belgique.

Ce mémoire étant soumis à des contraintes de temps et de budget, notre questionnaire en ligne a été partagé par e-mail et via les deux réseaux sociaux, Facebook et LinkedIn. Nous avons fait attention à atteindre les profils les plus variés possibles en termes de genre, âge et éducation. La sélection des participants s'est principalement reposée sur des choix personnels et non sur le hasard.

Nous avons fait usage d'une technique d'échantillonnage non probabiliste, l'échantillonnage de convenance. Il s'agit de la technique d'échantillonnage la plus rapide, la moins chère et la plus pratique qui existe (Malhotra, 2014). Néanmoins, l'interprétation des résultats doit être prise avec précaution, les échantillons non-probabilistes ne permettant pas l'évaluation objective de la précision des résultats (Malhotra, 2014).

Notre questionnaire en ligne a été diffusé durant une période de 10 jours dans le courant du mois de décembre 2018. Nous avons arbitrairement fixé la taille minimale de l'échantillon de participants à obtenir à 150 personnes afin de disposer d'une puissance de test suffisante lors de nos futures analyses. À la fin de la période de diffusion de notre questionnaire, notre objectif était atteint puisque la barre des 191 participants a été franchie.

Chapitre 2 : Analyse des résultats

Après la collecte des données, nous entamons le traitement de celles-ci et la vérification des hypothèses à tester. Nous utiliserons un intervalle de confiance à 95% pour l'ensemble de nos analyses, ce qui veut dire que le risque d'erreur admis sera de 5%.

Dans ce chapitre, nous décrirons d'abord toutes les opérations préliminaires effectuées afin de préparer nos données à la suite des analyses. Ensuite, une analyse descriptive de notre échantillon sera effectuée. Troisièmement, nous vérifierons la représentativité de notre échantillon en fonction du genre et de l'âge. Quatrièmement, des analyses de validité et de fiabilité seront exécutées pour chacune de nos variables. Enfin, nous pourrons passer à la vérification de notre modèle conceptuel et des hypothèses associées au travers de multiples tests statistiques.

1. Opérations préliminaires

Nos données ont été exportées sur Microsoft Excel. Ce logiciel est souvent recommandé pour le nettoyage et la préparation des données avant importation dans SPSS (Weiss & Townsend, 2005). Notre base de données initiale comprend 191 questionnaires complétés. Néanmoins, nous observons qu'un répondant n'a pas accepté que ses données personnelles soient traitées dans le cadre de ce mémoire. La ligne de la base de données lui correspondant a donc été supprimée. De plus, nous remarquons que 5 répondants ont notifié ne pas posséder de compte Facebook. Ceux-ci sont aussi éliminés de notre base de données. Au final, cela nous fait 185 réponses exploitables pour la suite de nos analyses.

En outre, nous décidons de regrouper de manière logique certaines catégories de nos variables. Il s'agit de toutes celles comportant un nombre d'effectifs inférieur à 5, car celui-ci est insuffisant pour la réalisation de certains tests statistiques. C'est ainsi que pour les catégories relatives au niveau d'éducation des répondants, les catégories 'Enseignement primaire' et 'Enseignement secondaire' sont rassemblées. Ensuite, en ce qui concerne les catégories relatives au statut professionnel des répondants, nous regroupons 'Étudiant(e)' et 'Stagiaire' ensemble ainsi que 'Retraité(e)' et 'Sans emploi'.

Nous devons encore transformer certaines de nos données en variables numériques afin de les rendre exploitables pour la suite de nos analyses. Une fois toutes ces opérations réalisées, notre fichier Excel peut être exporté sur le logiciel d'analyse statistique SPSS. Le plan de codage utilisé dans SPSS est disponible à l'annexe 2.

2. Analyse de l'échantillon

Toutes les sorties analysées dans cette section sont reprises dans l'annexe 3.

Nous débutons nos analyses sur SPSS par la description de notre échantillon. Celui-ci est composé de 50,3% d'hommes et 49,7% de femmes. Toutes les catégories d'âge sont représentées. Près de la moitié de l'échantillon appartient à la catégorie '18-24 ans' (48,6%). Les catégories '45-54 ans' et '25-34 ans' suivent avec, respectivement, 17,8% et 13%. Plus de 80% de l'échantillon possède un diplôme d'enseignement supérieur (84,4%). En ce qui concerne le statut professionnel, l'échantillon est assez varié. Les répondants sont en majorité étudiants (31,9%) et employés (23,2%). En outre, notre échantillon est en grande partie composé d'utilisateurs journaliers de Facebook (85,5%). Seuls 14,5% affirment ne pas l'utiliser quotidiennement. Près de la moitié de ceux-ci affirment même passer plus de 30 minutes par jour sur Facebook (44,3%) et 28,6% entre 10 et 30 minutes par jour.

Nous poursuivons nos analyses en évaluant la représentativité de notre échantillon par rapport à la population cible, c'est-à-dire celle des utilisateurs belges de Facebook. À cette fin, les données recueillies pour la Belgique dans l'étude de Hootsuite et WeAreSocial (2018) ont été utilisées. Cette étude comporte des données sur le nombre d'utilisateurs belges de Facebook par catégorie d'âge et par genre.

Celles-ci sont présentées à l'annexe 4.

Afin de tester si la proportion d'observations en fonction du genre correspond à la proportion donnée par les statistiques, nous réalisons un test d'ajustement Chi-carré. L'hypothèse nulle de ce test stipule que les effectifs observés correspondent aux effectifs théoriques.

Genre

	Effectif observé	N théorique	Résidus
Homme	93	92,5	,5
Femme	92	92,5	-,5
Total	185		

Figure 5 : Fréquences observées et théoriques en fonction du genre

Étant donné qu'aucun effectif théorique n'est inférieur à 5, les conditions d'application au test sont respectées.

Tests statistiques

	Genre
Khi-carré	,005
ddl	1
Sig. asymptotique	,941

Figure 6 : Test d'ajustement Chi-carré pour le genre

Nous observons une p-valeur supérieure à 0,05 (0,941), ce qui veut dire que l'hypothèse nulle n'est pas rejetée et que notre échantillon est représentatif de la population cible sur base du genre.

Nous avons réalisé un autre test d'ajustement Chi-carré afin de tester si la proportion d'observations en fonction des catégories d'âge correspond à la proportion donnée par les statistiques. L'hypothèse nulle de ce test stipule que les effectifs observés correspondent aux effectifs théoriques.

Âge

	Effectif observé	N théorique	Résidus
-18 ans	5	11,0	-6,0
18-24 ans	90	33,0	57,0
25-34 ans	24	43,0	-19,0
35-44 ans	14	36,0	-22,0
45-54 ans	33	29,0	4,0
55-64 ans	13	20,0	-7,0
+65 ans	6	13,0	-7,0
Total	185		

Figure 7 : Fréquences observées et théoriques en fonction des catégories d'âge

Étant donné qu'aucun effectif théorique n'est inférieur à 5, les conditions d'application au test sont respectées.

Tests statistiques

	Âge
Khi-carré	130,338
ddl	6
Sig. asymptotique	,000

Figure 8 : Test d'ajustement Chi-carré pour les catégories d'âge

Nous observons une p-valeur inférieure à 0,05 (0,0001), ce qui veut dire que l'hypothèse nulle est rejetée et que notre échantillon n'est pas représentatif de la population cible sur base des catégories d'âge. En effet, la catégorie '18-24 ans' est surreprésentée dans notre échantillon au détriment des catégories '25-34 ans' et '35-44 ans'. Ceci s'explique par le

fait que c'est le réseau de contacts de l'enquêteur qui a été le plus largement mobilisé pour l'enquête.

En conclusion, nous devrons rester prudent dans la généralisation de nos résultats à l'ensemble de la population cible étant donné que notre échantillon n'est pas représentatif sur base de tous les critères.

3. Analyse de la validité et de la fiabilité des échelles de mesure

Avant de procéder à la validation de nos hypothèses, nous devons encore vérifier la validité et la fiabilité de nos échelles de mesure. En réalisant des analyses de validité, nous vérifions que nos items de mesure mesurent bien les phénomènes qu'ils sont censés appréhender (Gobert et al., 2010) tandis qu'en effectuant des analyses de fiabilité, nous nous assurons que nos items de mesure sont cohérents et reflètent bien le même concept sous-jacent (Swaen, 2017).

Au total, notre base de données comporte 31 items qui mesurent les variables suivantes : la divulgation de données personnelles sur Facebook ('DIV'), l'utilisation de mesures de protection de la vie privée ('PRO'), les inquiétudes en matière de vie privée ('INQ'), la familiarité avec le concept du Big Data et ses avantages ('AVA'), la familiarité avec les implications du Big Data ('IMP'), la connaissance des nouvelles règles mises en place par le RGPD ('REG') et l'exposition aux médias sur la question de la vie privée ('EXP').

a. Vérification de la validité des échelles de mesure

La première étape des vérifications consiste à réaliser des analyses factorielles en composantes principales pour nous assurer que tous les items choisis pour une variable sont suffisamment corrélés entre eux et qu'ils ne mesurent qu'une seule dimension lorsqu'ils sont rassemblés. Selon Malhotra (2004), il est nécessaire que la taille de l'échantillon soit au minimum 4 à 5 fois plus grande que le nombre d'items pour qu'une analyse factorielle soit réalisable. Avec 31 items, cela veut dire qu'il nous faut au minimum 155 répondants, ce qui est respecté puisque notre échantillon est composé de 185 individus au total.

Nous lançons la procédure pour l'analyse factorielle de tous nos items sur SPSS et nous examinons les sorties obtenues. Tout d'abord, nous souhaitons vérifier l'existence de corrélations suffisantes entre les items pour pouvoir réaliser notre analyse factorielle. À cette fin, nous observons d'abord le test de sphéricité de Bartlett. L'hypothèse nulle de ce

test stipule qu'il n'y a pas de corrélations entre les items. Ensuite, nous examinons l'indice de Kaiser-Meyer-Olkin (KMO). Au plus cet indice est proche de 1, au plus l'analyse factorielle est pertinente. Il est d'usage de considérer qu'un indice KMO supérieur à 0,5 est nécessaire (Malhotra, 2004).

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,776
Test de sphéricité de Bartlett	Khi-carré approx.	3186,972
	ddl	465
	Signification	,000

Figure 9 : Test de sphéricité de Bartlett et indice KMO pour le modèle global

Nous remarquons une p-valeur inférieure à 0,05 (0,0001) pour le test de sphéricité de Bartlett, ce qui veut dire que l'hypothèse nulle est rejetée et que les corrélations entre les items sont suffisantes. De plus, l'indice KMO est supérieur à 0,5 (0,776), ce qui veut dire que l'analyse factorielle est pertinente.

Une fois les conditions d'application respectées, nous nous intéressons au nombre de composantes à retenir dans l'analyse factorielle. SPSS est paramétré par défaut pour suivre le critère de Kaiser lorsqu'il s'agit de déterminer le nombre de facteurs. Ce critère indique que tous les facteurs avec des valeurs propres supérieures à 1 sont retenus. Cependant, SPSS offre aussi la possibilité de déterminer à l'avance le nombre de facteurs à inclure dans l'analyse, ce que nous faisons puisque notre modèle conceptuel comporte 7 variables au total. Nous utiliserons la règle du coude sur le tracé d'effondrement afin de confirmer notre choix. Celle-ci indique qu'il ne faut tenir compte que des facteurs présents à gauche du point d'inflexion du tracé (Field, 2013).

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,761	15,360	15,360	4,761	15,360	15,360	4,217	13,602	13,602
2	4,445	14,338	29,698	4,445	14,338	29,698	3,632	11,716	25,318
3	3,523	11,365	41,063	3,523	11,365	41,063	3,208	10,347	35,665
4	2,587	8,344	49,407	2,587	8,344	49,407	2,882	9,295	44,960
5	2,210	7,130	56,538	2,210	7,130	56,538	2,352	7,586	52,546
6	1,795	5,790	62,328	1,795	5,790	62,328	2,282	7,363	59,909
7	1,489	4,802	67,130	1,489	4,802	67,130	2,238	7,220	67,130
8	1,027	3,313	70,443						

Méthode d'extraction : Analyse en composantes principales.

Figure 10 : Variance totale expliquée pour le modèle global

Dans le tableau ci-dessus (figure 10), nous observons qu'en réalité 8 composantes ont une valeur propre supérieure à 1. Ce qui veut dire que selon le critère de Kaiser, nous devrions ajouter une variable à notre modèle global. Néanmoins, comme la valeur propre est très proche de 1 (1,027) et que le pourcentage de la variance totale expliquée en plus par l'ajout de cette composante est relativement faible (3,313%), nous décidons de ne pas en tenir compte. Le tracé d'effondrement pour le modèle global, présenté à l'annexe 5, confirme notre décision puisque nous observons que le point d'inflexion se situe au 8ème facteur. Cela veut dire que seuls les 7 premiers facteurs à la gauche de celui-ci doivent être considérés.

À présent, nous nous intéressons à la matrice des qualités de représentation du modèle global. Celle-ci se trouve également à l'annexe 5. Elle comporte une colonne 'extraction' qui nous renseigne sur la valeur de la variance qu'un item partage avec tous les autres items appartenant à la même dimension (Swaen, 2017). Il est courant de vérifier que celles-ci soient supérieures à 0,5. Nous faisons de même et nous remarquons que même si la grande majorité de nos 31 items ont des valeurs d'extraction au-dessus du seuil, les valeurs de 5 d'entre eux sont tout de même légèrement inférieures à 0,5 ('DIV4', 'PRO1', 'PRO2', 'PRO3', 'PRO4').

Finalement, nous observons la matrice des composantes après rotation Varimax (figure 11). Celle-ci nous indique les corrélations entre les items et les composantes, ce qui nous permet de vérifier que tous les items d'une variable mesurent bien le même concept. Selon Swaen (2017), ces valeurs doivent idéalement être supérieures à 0,5.

Dans le tableau présenté ci-dessous, nous remarquons que l'item 'DIV4' ('Cela ne me dérange pas de mettre des informations personnelles sur Facebook') possède une corrélation inférieure à 0,5. Comme celui-ci fait également partie des 5 items avec une valeur d'extraction inférieur à 0,5, nous décidons de le supprimer.

Les résultats de l'analyse factorielle pour le modèle global sans l'item 'DIV4' se trouve à l'annexe 5. Certains de nos items ont toujours des valeurs d'extraction en-dessous de 0,5 (« PRO1 », « PRO2 », « PRO3 », « PRO4 »). Il s'agit de la plupart des items liés à la variable relative à l'adoption de mesures de protection de la vie privée. Nous décidons tout de même de les conserver car ceux-ci sont très proches du seuil et sont importants pour la suite de nos analyses. Par contre, toutes les corrélations présentes dans la nouvelle matrice des composantes sont à présent bien supérieures à 0,5.

	Composante						
	1	2	3	4	5	6	7
IMP3	,944						
IMP2	,925						
IMP4	,889						
IMP1	,877						
IMP5	,877						
AVA1		,915					
AVA4		,909					
AVA3		,907					
AVA2		,837					
REG1			,811				
REG3			,791				
REG4			,779				
REG2			,778				
REG5			,713				
INQ2				,835			
INQ3				,802			
INQ1				,776			
INQ4				,769			
PRO5					,704		
PRO4					,637		
PRO2					,620		
PRO1					,594		
PRO3					,571		
DIV3						,870	
DIV1						,766	
DIV2						,765	
DIV4					-,375	,481	
EXP2							,783
EXP3							,774
EXP1				,307			,715
EXP4		,440					,506

Figure 11 : Matrice des composantes après rotation Varimax du modèle global

Après avoir effectué l'analyse factorielle de notre modèle global, nous réalisons la même opération sur SPSS pour chacune de nos variables afin de nous assurer que celles-ci sont bien valides. Les sorties SPSS de ces différentes analyses factorielles en composantes principales sont disponibles à l'annexe 5. Pour faciliter la lecture, nous avons ensuite regroupé les principaux résultats obtenus au sein de la figure 8.

Nous observons que tous les indices KMO sont supérieurs à 0,5 et que le pourcentage de la variance expliquée par chaque facteur est supérieur à 50% sauf pour la variable 'Protection'.

Variable	Nom	Nb. items	Indice KMO	% de variance expliquée	Alpha de Cronbach
Divulgation de données personnelles sur Facebook	Divulgation	3	0,639	67,326%	0,741
Utilisation de mesures de protection de la vie privée sur Facebook	Protection	5	0,755	42,696%	0,634
Inquiétudes en matière de vie privée	Inquiétudes	4	0,753	68,703%	0,847
Familiarité avec le concept du Big Data et ses avantages	Avantages_BD	4	0,830	82,976%	0,931
Familiarité avec les implications du Big Data	Implications_BD	5	0,885	82,149%	0,944
Connaissance des règles mises en place par le RGPD	Règles_RGPD	5	0,795	61,773%	0,839
Exposition aux médias sur la question de la vie privée	Exposition_Médias	4	0,737	55,403%	0,710

Figure 12 : Tableau récapitulatif des résultats des analyses de validité et de fiabilité pour chaque variable

b. Vérification de la fiabilité des échelles de mesure

La deuxième étape consiste à mesurer la fiabilité de nos échelles de mesure. C'est l'alpha de Cronbach qui va nous permettre de déterminer si chacune de nos variables sont fiables, c'est-à-dire si elles sont composées d'items mesurant bien le même concept. Il est courant de considérer qu'un alpha de Cronbach supérieur à 0,7 est nécessaire pour affirmer qu'un facteur est fiable (Swaen, 2017). Les résultats des analyses de fiabilité pour chaque variable se trouvent à l'annexe 5. Par souci de simplicité, nous avons repris les valeurs de l'alpha de Cronbach pour chaque variable dans la figure 8 ci-dessus.

Nous pouvons ainsi constater que toutes nos variables ont un alpha de Cronbach au-dessus de 0,7 sauf la variable relative à l'adoption de mesures de protection de la vie privée sur Facebook. Nous nous sommes alors demandé s'il n'était pas judicieux de supprimer un des items mesurant celle-ci. Néanmoins, après avoir testé plusieurs possibilités, nous sommes arrivés à la conclusion que cela n'améliorait pas les résultats. Comme l'alpha de Cronbach de cette variable est proche du seuil (0,634) et que celle-ci est importante pour notre analyse, nous avons donc décidé de conserver la variable en question pour la suite de nos analyses.

Maintenant que nous avons vérifié la validité et la fiabilité de toutes nos échelles de mesure, nous pouvons créer de nouvelles variables en calculant la moyenne arithmétique des différents items retenus pour chaque facteur. Ce sont ces nouvelles variables que nous utiliserons dans la partie suivante de notre analyse.

4. Vérification des hypothèses

Dans cette partie, nous allons procéder à la vérification de nos hypothèses. Nous effectuerons à la fois des régressions linaires simples et multiples.

Toutes les sorties analysées dans cette section sont reprises dans l'annexe 6.

a. Antécédents → Inquiétudes

Pour commencer, nous réalisons une régression linéaire multiple afin de tester nos hypothèses **H1a**, **H1b**, **H2**, **H3**, **H4**, **H5** et **H6**. En faisant cela, nous testons la première partie de notre modèle conceptuel, c'est-à-dire la relation entre les antécédents et les inquiétudes.

Pour cette régression, la variable dépendante sera les inquiétudes en matière de vie privée ('Inquiétudes') alors que les variables indépendantes seront la familiarité avec le terme *Big Data* et ses avantages ('Avantages_BD'), la familiarité avec les implications du *Big Data* ('Implications_BD'), l'exposition aux médias sur la question de la vie privée ('Exposition_Médias'), la connaissance des nouvelles règles mises en place par le RGPD ('Règles_RGPD') ainsi que l'âge, le genre et le niveau d'éducation.

Il est important de noter que ces trois dernières variables sont catégorielles. L'âge et le niveau d'éducation sont des variables mesurées sur des échelles ordinales. Néanmoins, pour pouvoir les intégrer à notre modèle de régression multiple, nous les traiterons comme des variables quantitatives d'intervalle. Ensuite, comme le genre est une variable mesurée sur une échelle nominale, nous devons la transformer en une variable *'dummy'* (0 ou 1) afin de pouvoir l'intégrer à notre régression multiple. Cette opération est réalisée au travers du logiciel SPSS.

Avant de se lancer dans l'analyse des résultats de cette régression multiple, nous devons d'abord nous assurer que les conditions d'application sont vérifiées. Pour vérifier l'indépendance des résidus, nous regardons si la valeur du test de Durbin-Watson est proche de 2 (Field, 2013), ce qui est bien le cas puisqu'elle vaut 2,053. Ensuite, pour

vérifier la normalité des résidus, nous examinons à la fois si l'histogramme des résidus standardisés suit la courbe de Gauss et si toutes les données se rapprochent de la droite de référence dans le tracé P-P plot. Pour vérifier l'homoscédasticité des résidus, nous observons si les résidus forment un nuage de points, sans forme apparente, autour de 0. Enfin, pour vérifier l'absence de multicolinéarité entre les variables explicatives, nous regardons si tous les indicateurs VIF sont inférieurs à 3 (Swaen, 2017), ce qui est bien le cas. Toutes les conditions sous-jacentes étant respectées (annexe 6), nous passons à l'analyse des résultats de la régression.

Dans le tableau récapitulatif des modèles (annexe 6), nous observons que le R^2 ajusté du modèle est de 0,121, ce qui signifie que 12,1% de la variation totale d''Inquiétudes' est expliquée par la variation de toutes les variables explicatives prises en compte. La p-valeur dans le tableau ANOVA étant inférieure à 0,05 (0,0001), nous pouvons rejeter l'hypothèse nulle ($R^2 = 0$) et affirmer l'existence d'un modèle de régression linéaire entre les variables.

Dans le tableau des coefficients (annexe 6), nous remarquons que seules les variables 'Exposition_Médias', 'Règles_RGPD' et 'Genre' ont des p-valeurs inférieures à 0,05 (0,013, 0,010 et 0,012). Nous observons néanmoins que la p-valeur de la variable 'Avantages_BD' est proche du seuil (0,083). L'influence des autres variables n'étant pas significative, nous réalisons à nouveau la régression en ne conservant que les variables mentionnées ci-dessus.

Celle-ci nous donne les résultats suivants : le R^2 ajusté prend une valeur de 0,127, la p-valeur de la variable 'Avantages_BD' se rapproche plus du seuil (0,067), ce qui nous incite à la considérer comme significative. Les coefficients des autres variables retenues sont toujours quant à eux significatifs. Nous remarquons également que l'influence de la variable 'Avantages_BD' est négative puisque le β vaut -0,087 alors que les influences des variables 'Exposition_médias', 'Règles_RGPD' et 'Genre' ont une influence positive puisque leurs β valent respectivement 0,209, 0,182 et 0,484. L'équation de régression linéaire peut ainsi s'écrire :

$$\text{Inquiétudes} = 3{,}762 - 0{,}087\ (\text{Avantages\_BD}) + 0{,}209\ (\text{Exposition\_Médias}) + 0{,}182\ (\text{Règles\_RGPD}) + 0{,}484\ (\text{Genre}) + \varepsilon$$

Nous pouvons maintenant parcourir les résultats de la régression multiple hypothèse par hypothèse.

H1a : La familiarité avec le terme *Big Data* et ses avantages tend à réduire les inquiétudes en matière de vie privée.

Considérant que le coefficient d''Avantages_BD' vaut -0,087 et que celui-ci est considéré comme étant significatif dans le modèle de régression final, nous pouvons conclure que plus un utilisateur de Facebook sera familiarisé avec le terme *Big Data* et de avantages, moins il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H1a est donc validée.

H1b : La familiarité avec les implications du *Big Data* tend à augmenter les inquiétudes en matière de vie privée.

Considérant que le coefficient d'« Implications_BD » n'est pas significatif dans le modèle de régression initial, nous ne pouvons pas conclure que plus un utilisateur de Facebook sera familiarisé avec les implications du *Big Data*, plus il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H1b n'est donc pas validée.

H2 : L'exposition aux médias sur la question de la vie privée tend à augmenter les inquiétudes en matière de vie privée.

Considérant que le coefficient d''Exposition_Médias' vaut 0,209 et que celui-ci est significatif dans le modèle de régression final, nous pouvons conclure que plus un utilisateur de Facebook sera exposé aux médias sur la question de la vie privée, plus il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H2 est donc validée.

H3 : La connaissance des nouvelles règles mises en place par le RGPD tend à réduire sur les inquiétudes en matière de vie privée.

Considérant que le coefficient de 'Règles_RGPD' vaut 0,182 et que celui-ci est significatif dans le modèle de régression final, nous pouvons conclure que plus un

utilisateur de Facebook sera au courant des nouvelles règles mises en place par le RGPD, plus il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H3 n'est donc pas validée. En effet, la connaissance des nouvelles règles mises en place par le RGPD ne tend pas à réduire les inquiétudes en matière de vie privée. Au contraire, celle-ci tend à les augmenter.

H4 : Les jeunes sont moins inquiets quant au respect de leur vie privée que les personnes plus âgées.

Considérant que le coefficient d''Âge' n'est pas significatif dans le modèle de régression initial, nous ne pouvons pas conclure que moins un utilisateur de Facebook sera âgé, moins il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H4 n'est donc pas validée.

H5 : Les femmes sont plus inquiètes quant au respect de leur vie privée que les hommes.

Considérant que la variable 'Genre' a été transformée en une variable *'dummy'* (0 ou 1), le coefficient obtenu dans la régression multiple nous montre l'écart entre la catégorie codée 1 ('Femme') et la catégorie de référence codée 0 ('Homme'). Nous pouvons donc conclure que les inquiétudes moyennes des femmes en matière de vie privée sont plus élevées de 0,484 par rapport à celles des hommes.

L'hypothèse H5 est donc validée.

H6 : Les individus ayant un niveau d'éducation plus élevé sont moins inquiets quant au respect de leur vie privée.

Considérant que le coefficient de 'Niveau d'éducation' n'est pas significatif dans le modèle de régression initial, nous ne pouvons pas conclure que plus un utilisateur de Facebook aura un niveau d'éducation élevé, moins il sera préoccupé par de possibles atteintes à sa vie privée.

L'hypothèse H6 n'est donc pas validée.

b. Inquiétudes → Conséquences

Pour la vérification de nos deux dernières hypothèses (**H7** et **H8**), nous allons effectuer deux régressions linéaires simples. En faisant cela, nous testons la seconde partie de notre modèle conceptuel, c'est-à-dire la relation entre les inquiétudes en matière de vie privée et les conséquences sur les comportements des utilisateurs de Facebook.

H7 : Les inquiétudes en matière de vie privée ont un impact négatif sur la divulgation de données personnelles sur Facebook.

Pour la vérification de cette hypothèse, nous effectuons une régression linéaire simple avec les inquiétudes en matière de vie privée ('Inquiétudes') comme variable indépendante et la divulgation de données personnelles sur Facebook ('Divulgation') comme variable dépendante.

Avant de nous lancer dans l'analyse des résultats de cette régression simple, nous devons d'abord nous assurer que les conditions d'application sont respectées. Pour vérifier l'indépendance des résidus, nous observons si la valeur du test de Durbin-Watson est proche de 2 (Field, 2013), ce qui est bien le cas puisqu'elle vaut 1,958. Ensuite, pour vérifier la normalité des résidus, nous examinons à la fois si l'histogramme des résidus standardisés suit la courbe de Gauss et si toutes les données se rapprochent de la droite de référence dans le tracé P-P plot. Enfin, pour vérifier l'homoscédasticité des résidus, nous observons si les résidus forment un nuage de points, sans forme apparente, autour de 0. Toutes les conditions sous-jacentes étant respectées (annexe 6), nous passons à l'analyse des résultats de la régression.

Dans le tableau récapitulatif des modèles (annexe 6), nous observons que le R^2 du modèle est de 0,001, ce qui veut dire que 0,1% de la variation totale de 'Protection' est expliquée par la variation d''Inquiétudes'. La p-valeur dans le tableau ANOVA étant supérieure à 0,05 (0,619), nous ne pouvons pas rejeter l'hypothèse nulle ($R^2 = 0$), ni affirmer l'existence d'un modèle de régression linéaire entre les deux variables. L'analyse de régression s'arrête donc là pour cette hypothèse.

Le fait d'être inquiet quant à la protection de ses données personnelles n'influence pas, négativement ou positivement, la divulgation de soi sur Facebook.

L'hypothèse H7 n'est donc pas validée.

H8 : Les inquiétudes en matière de vie privée ont un impact positif sur l'adoption de mesures de protection de la vie privée sur Facebook.

Pour la vérification de cette hypothèse, nous effectuons une régression linéaire simple avec les inquiétudes en matière de vie privée ('Inquiétudes') comme variable indépendante et l'utilisation de mesures de protection de la vie privée sur Facebook ('Protection') comme variable dépendante.

Avant de nous lancer dans l'analyse des résultats de cette régression simple, nous devons d'abord nous assurer que les conditions d'application sont respectées. Pour vérifier l'indépendance des résidus, nous observons si la valeur du test de Durbin-Watson est proche de 2 (Field, 2013), ce qui est bien le cas puisqu'elle vaut 2,105. Ensuite, pour vérifier la normalité des résidus, nous examinons à la fois si l'histogramme des résidus standardisés suit la courbe de Gauss et si toutes les données se rapprochent de la droite de référence dans le tracé P-P plot. Enfin, pour vérifier l'homoscédasticité des résidus, nous observons si les résidus forment un nuage de points, sans forme apparente, autour de 0. Toutes les conditions sous-jacentes étant respectées (annexe 6), nous passons à l'analyse des résultats de la régression.

Dans le tableau récapitulatif des modèles (annexe 6), nous observons que le R^2 du modèle est de 0,114, ce qui veut dire que 11,4% de la variation totale de 'Protection' est expliquée par la variation d''Inquiétudes'. La p-valeur dans le tableau ANOVA étant inférieure à 0,05 (0,0001), nous pouvons rejeter l'hypothèse nulle ($R^2 = 0$) et affirmer l'existence d'un modèle de régression linéaire entre les deux variables.

Dans le tableau des coefficients (annexe 6), nous remarquons que la variable 'Protection' est significativement influencée par la variable 'Inquiétudes' car la p-valeur est inférieure à 0,05 (0,0001). Cette influence est positive puisque le β vaut 0,275. L'équation de régression linéaire peut ainsi s'écrire :

$$\text{Protection} = 4{,}089 + 0{,}275\ (\text{Inquiétudes}) + \varepsilon$$

Nous pouvons conclure que plus un utilisateur de Facebook sera préoccupé par de possibles atteintes à sa vie privée, plus il prendra des mesures de protection de la vie privée.

L'hypothèse H8 est donc validée.

En résumé, la vérification des hypothèses de notre modèle conceptuel aboutit aux résultats suivants :

Hypothèses		Conclusion
H1a	La familiarité avec le concept Big Data et ses avantages tend à **réduire** les inquiétudes en matière de vie privée	✓
H1b	La familiarité avec les implications du Big Data tend à **augmenter** les inquiétudes en matière de vie privée	✗
H2	L'exposition médiatique aux médias sur la question de la vie privée tend à **augmenter** les inquiétudes en matière de vie privée	✓
H3	La connaissance des nouvelles règles mises en place par le RGPD tend à **réduire** sur les inquiétudes en matière de vie privée *Relation inverse* : La connaissance des nouvelles règles mises en place par le RGPD tend à **augmenter** sur les inquiétudes en matière de vie privée	✗ ✓
H4	Les jeunes sont **moins** inquiets quant au respect de leur vie privée que les personnes plus âgées	✗
H5	Les femmes sont **plus** inquiètes quant au respect de leur vie privée que les hommes	✓
H6	Les individus ayant un niveau d'éducation plus élevé sont **moins** inquiets quant au respect de leur vie privée	✗
H7	Les inquiétudes en matière de vie privée ont un impact **négatif** sur la divulgation de données personnelles sur Facebook	✗
H8	Les inquiétudes en matière de vie privée ont un impact **positif** sur l'utilisation de mesures de protection de la vie privée sur Facebook	✓

Figure 13 : Tableau récapitulatif de la vérification des hypothèses

Chapitre 3 : Discussion

Dans le chapitre précédent, nous avons réalisé la vérification de nos hypothèses et certaines analyses supplémentaires à l'aide de différents tests statistiques. À présent, nous allons reprendre les principaux résultats obtenus et tenter de les comprendre. Par la suite, les recommandations établies sur base de nos résultats seront présentées. Nous exposerons enfin les limites de notre étude ainsi que les différentes pistes proposées pour de futures recherches.

1. Interprétation des résultats

Comme plusieurs de nos hypothèses n'ont pas été validées, nous allons essayer de comprendre pourquoi.

En ce qui concerne la première partie de notre modèle conceptuel, c'est-à-dire l'influence des antécédents sur les inquiétudes en matière de vie privée, rappelons que nous avions choisi, sur base de notre revue de littérature, d'intégrer les antécédents suivants au sein de notre modèle conceptuel : la familiarité avec le Big Data, l'exposition aux médias sur la question de la vie privée, la connaissance des nouvelles règles mises en place par le RGPD et des variables sociodémographiques telles que l'âge, le genre et l'éducation.

Pour le premier de nos antécédents, c'est-à-dire la familiarité avec le Big Data, nous pouvons constater que les résultats de notre étude quantitative ne concordent pas totalement avec la littérature sur le sujet (Alashoor et al., 2017). En effet, si la familiarité avec le terme Big Data et ses avantages s'est avérée pertinente pour réduire les inquiétudes en matière de vie privée, par contre, aucune conclusion n'a pu être tirée de la familiarité avec les implications du Big Data. La différence se situe peut-être dans le fait qu'Alashoor, Han et Joseph (2017) se sont appuyés sur un échantillon composé uniquement d'étudiants afin de réaliser leur étude alors que nous avons décidé de faire parvenir notre questionnaire à des profils beaucoup plus variés. Ils ont motivé leur décision en expliquant qu'ils voulaient s'assurer de s'adresser à des personnes ayant certaines notions sur le sujet, en l'occurrence ici des étudiants. De plus, nous pensons à posteriori que nos questions posées pour mesurer cette variable étaient peut-être trop générales.

Pour le deuxième de nos antécédents, c'est-à-dire l'exposition aux médias sur la question de la vie privée, nous pouvons constater que les résultats de notre étude quantitative

concordent avec la littérature sur le sujet. En effet, plus un individu sera exposé dans la presse, les journaux ou encore sur Internet à des faits d'atteinte à la confidentialité des données personnelles, plus ses inquiétudes quant à la protection de sa vie privée seront élevées. Cambridge Analytica est l'un des scandales majeurs à ce sujet et nous avons pu voir qu'un peu plus de la moitié des répondants affirmaient être au courant de cette affaire.

Pour le troisième de nos antécédents, c'est-à-dire la connaissance des nouvelles règles mises en place par le RGPD, nous pouvons constater que c'est la relation inverse qui a été vérifiée au travers de notre étude quantitative. En effet, le fait que les individus soient au courant des droits qui leur sont offerts par les nouvelles règles du RGPD ne tend pas à diminuer leurs inquiétudes en matière de vie privée, mais au contraire, tend à les augmenter. Ceci est donc en contradiction avec la littérature à ce sujet (Lwin et al., 2007). La différence se situe peut-être dans le fait que dans l'étude de Lwin, Wirtz et Williams (2007), leur mesure de cette variable était réalisée de manière générale sur les lois existant en la matière, alors que nous avons choisi de traiter spécifiquement de certains droits résultant du nouveau règlement européen (RGPD) dans notre questionnaire. De plus, nous estimons que la plupart des individus considèrent encore actuellement les droits garantis par le RGPD comme théoriques et compliqués à défendre. Ils ont également constaté que la plupart des entreprises s'y sont prises au dernier moment pour le mettre en œuvre, alors que le texte existait déjà depuis deux ans. Enfin, depuis la mise en application du RGPD, ces individus n'ont cessé de recevoir des demandes d'autorisation d'accès à leurs données personnelles souvent bien plus larges que par le passé. Tout cela nous permet d'identifier certaines explications possibles à cette relation inverse (Lemaire, 2018).

Ensuite, nous nous concentrons sur la deuxième partie de notre modèle conceptuel, c'est-à-dire sur l'influence des inquiétudes sur les conséquences. Sur base de notre revue de la littérature, nous avons choisi d'intégrer les conséquences suivantes au sein de notre modèle conceptuel : la divulgation de données personnelles et la prise de mesures de protection de la vie privée.

Pour la première de nos conséquences, c'est-à-dire la divulgation d'informations personnelles sur Facebook, nous pouvons constater que les résultats de notre analyse quantitative ne concordent pas avec la littérature sur le sujet. Nous n'observons en effet

pas de relation significative entre les deux variables. Cela contredit la plupart des résultats trouvés dans la littérature. Néanmoins, nous avions également fait état des analyses de Taddicken (2014), qui arrivait aussi à la conclusion que cette relation n'était pas significative. Il expliquait cette situation par le 'paradoxe de la vie privée', décrit précédemment dans ce mémoire et qui joue peut-être un rôle plus important qu'envisagé au départ. Le souci de la protection de la vie privée reste un sentiment diffus, instinctif, impacté par de nombreux biais cognitifs et difficile à adresser de manière logique, alors que les avantages retirés des services fournis à titre gratuit par Facebook sont tangibles et visibles tous les jours (Glance, 2018).

À l'opposé, pour la seconde de nos conséquences, c'est-à-dire l'adoption de mesures de protection de la vie privée sur Facebook, nous pouvons constater que les résultats de notre analyse concordent avec la littérature sur le sujet. En effet, les individus les plus inquiets quant à la protection de leur vie privée auront quand même plus tendance à protéger l'accès qu'ils donnent à leurs données personnelles.

2. Recommandations

Avec toutes les connaissances accumulées au travers de la littérature analysée et les résultats de notre étude quantitative, nous allons procéder à la formulation de quelques recommandations s'adressant à la fois aux entreprises et aux utilisateurs du réseau social.

a. Recommandations à l'attention des entreprises

L'accès aux données personnelles de leurs clients est vital pour les entreprises du 21ème siècle et l'avenir des services et produits qu'elles ont à leur vendre, d'autant plus que l'accès à ces informations est largement gratuit, ce qui rend leur modèle économique particulièrement attractif. Mais ce mémoire a révélé que les inquiétudes en matière de vie privée sont relativement élevées, quels que soient l'âge, le genre ou le niveau d'éducation.

Nous recommandons par conséquent à toute entreprise recueillant des données personnelles sur ses clients de mettre en œuvre d'initiative des procédés permettant de mieux protéger leur vie privée et de communiquer proactivement à ce sujet pour réduire leurs inquiétudes. Cela pourrait notamment impliquer la mise en œuvre effective des bénéfices garantis par le RGPD, plutôt que le contournement auquel nous assistons aujourd'hui, avec des demandes d'accès obscures mais toujours plus larges, sans réelle possibilité de refus. Elles pourraient aussi offrir plus d'options à leurs clients quant à la

gestion de leurs paramètres de confidentialité et encourager ceux-ci à les utiliser, au lieu de simplement les mettre à disposition d'une manière souvent peu conviviale.

Par ailleurs les entreprises devraient investir lourdement dans des systèmes de sécurité robustes. Cela présenterait le double avantage de convaincre leurs clients de l'importance qu'elles accordent à leurs précieuses données personnelles, mais aussi d'éviter d'être victimes de piratage de données, largement commentés dans la presse et négatifs pour leur réputation. Elles pourraient enfin tenter de limiter les données qu'elles recueillent en masse auprès de leurs clients et qui n'ont pas toujours énormément de valeur et se montrer plus sélectives dans leur choix. Les entreprises pourraient ainsi s'engager sur la voie d'une meilleure personnalisation de l'expérience du client, en expliquant de manière transparente à celui-ci quelles données ont de la valeur pour elles, la personnalisation du service rémunérant en quelque sorte le client pour les données qu'il a fournies (Glance, 2018).

Avec des inquiétudes plus faibles, les individus auront tendance à utiliser des mesures de protection de la vie privée avec plus de discernement et ainsi maintenir un accès large à leurs données personnelles.

b. Recommandations à l'attention des utilisateurs des réseaux sociaux

Il faut que les utilisateurs des réseaux sociaux perçoivent la valeur des données personnelles qu'ils abandonnent sur les réseaux sociaux et tendent d'adopter un comportement raisonné à cet égard : ne confier des données personnelles que dans la mesure du nécessaire, que ce soit pour conserver de bonnes relations avec amis et connaissances ou obtenir de meilleurs services de la part d'opérateurs économiques.

Ensuite, il faut que les utilisateurs de réseaux sociaux se maintiennent informés sur les dernières évolutions et adaptent en permanence leurs comportements face à une technique qui évolue très vite. Notre enquête a démontré que les jeunes ne sont pas suffisamment au courant du scandale Cambridge Analytica, ni des nouvelles règles du RGPD. L'enquête a également démontré qu'ils mettent en œuvre moins de mesures de protection de leur vie privée, alors qu'ils sont des utilisateurs les plus actifs de Facebook.

3. Limites de l'étude et pistes pour de futures recherches

Notre mémoire contient certaines limites qu'il est important de prendre en compte dans la discussion de nos résultats. Au vu de chaque limite qui sera présentée, des voies

d'amélioration et pistes pour de futures recherches dans le domaine seront également proposées.

Tout d'abord, la taille de notre échantillon n'est pas optimale. Nous aurions dû utiliser la formule théorique suivante pour calculer la taille idéale de celui-ci :

$$n = \left(\frac{Z_\alpha \ * \ s}{E} \right)^2$$

Où n est la taille de l'échantillon, α est le niveau de signification voulu, Z est une constante extraite des tables de la loi normale, s est l'écart-type des observations et E est le niveau de précision recherché.

Dans le cadre de ce mémoire, nous souhaitions avoir un niveau de précision E de 10% et un niveau de signification α de 5%, ce qui veut dire que la Z_α valait 1,96. Pour calculer l'écart-type des observations s, dans la mesure où nous utilisons des échelles de mesure de type Likert en 7 points, il pouvait être estimé au moyen de la formule suivante : s = (valeur maximum – valeur minimum) / (valeur maximum – 1). Nous obtenions donc que s vaut 1.

Ainsi, nous pouvons calculer que la taille d'échantillon la plus prudente serait n = 385. Afin d'atteindre ce nombre théorique idéal, il nous aurait encore fallu obtenir la réponse de 200 individus supplémentaires.

Ensuite, en ce qui concerne la représentativité de notre échantillon, nous avons établi que celui-ci était représentatif du point de vue du genre mais pas du point de vue des catégories d'âge. La majorité de nos répondants se situe en effet dans la même tranche d'âge (18-24 ans). Cela peut s'expliquer par le fait que nous avons utilisé une méthode d'échantillonnage non-probabiliste. En effet, notre échantillon n'a pas été constitué aléatoirement. Celui-ci a été partagé au sein de notre réseau de connaissances via les réseaux sociaux. Dans le cadre de futures recherches, notre étude pourrait être reproduite avec cette fois-ci un échantillon plus large et plus diversifié afin que les résultats puissent être généralisables à la population cible.

Enfin, un autre limite à ce mémoire est de n'avoir retenu que certains des antécédents et des conséquences présents dans le modèle APCO (Smith et al., 2011). Il se pourrait donc que certaines variables non reprises dans le modèle influencent les résultats. De futures

recherches pourraient ainsi inclure d'autres antécédents ou conséquences dans le modèle à tester afin d'obtenir des résultats plus complets et plus exhaustifs. D'autres futures recherches pourraient également se servir de la méthode des équations structurelles pour la validation du modèle global et des hypothèses associées. Elles pourraient ainsi tenir compte de la possible présence de variables non-observées et les résultats seraient certainement plus riches qu'avec l'approche traditionnelle que nous avons utilisée.

68.

CONCLUSION

Nous sommes partis initialement d'une réflexion sur le développement exponentiel que connaissent actuellement les pratiques du marketing prédictif et la nécessaire récolte massive de données personnelles en provenance des réseaux sociaux pour alimenter les analyses sous-jacentes à celui-ci.

L'actualité récente, liée à la multiplication d'épisodes de piratage de données personnelles, du scandale Cambridge Analytica et de l'entrée en vigueur très médiatisée du RGPD nous a incité à nous interroger sur l'impact potentiel des inquiétudes grandissantes des utilisateurs des réseaux sociaux en matière de protection de leur vie privée sur le développement des pratiques de marketing prédictif.

En effet, les utilisateurs de Facebook sont les indispensables pourvoyeurs des données personnelles qui alimentent le marketing prédictif et ces inquiétudes pourraient potentiellement avoir des conséquences importantes : moins de données personnelles, des données de moindre valeur, des données dont l'accès serait restreint, voire même plus de données du tout si certains utilisateurs décidaient de sortir définitivement de Facebook.

Nous avons par conséquent concentré notre réflexion sur ce qui pouvait motiver ces inquiétudes, les développer ou les réduire et examiné quels impacts concrets elles ont déjà aujourd'hui sur l'accès des réseaux sociaux aux données personnelles de leurs utilisateurs. A cette fin, nous avons utilisé le modèle APCO développé par Smith, Dinev et Xu (2011) pour construire notre propre modèle conceptuel.

Notre recherche a été ardue car même s'il existe beaucoup de littérature, scientifique ou non, sur le sujet, les concepts restent mal définis, en particulier celui de la vie privée. En effet, il s'agit d'une notion qui reste assez abstraite et qui évolue fortement dans le cadre de l'omniprésence des réseaux sociaux. Par conséquent, la sensibilité à la protection de la vie privée s'avère plus instinctive que logique et il semble difficile d'en déduire des comportements raisonnés.

Nous avons néanmoins réussi à démontrer que la connaissance du terme *Big Data* et de ses avantages tend à diminuer les inquiétudes des utilisateurs de Facebook quant à la protection de leurs données personnelles, tandis que l'exposition aux médias sur la question de la vie privée et la connaissance des règles du RGPD récemment entré en

vigueur tendent à augmenter ces mêmes inquiétudes. Nous avons également pu démontrer que les inquiétudes en matière de vie privée avaient un impact positif sur l'adoption de mesures de protection par les utilisateurs de Facebook. Ainsi, par exemple, ils limitent l'accès à leurs informations personnelles et publications au-delà des paramètres par défaut proposés par Facebook, ils font attention avec qui ils sont amis sur le réseau, ils évitent d'utiliser des applications tierces et, mais dans une moindre mesure, ils empêchent Facebook d'avoir accès à leur localisation.

Il est exact que certaines hypothèses, et non des moindres, n'ont pas pu être validées. Dans certains cas, cela s'explique par la taille et le type d'échantillon que nous avons utilisé pour nos tests statistiques. Nous avons sans doute également sous-estimé ce que les auteurs appellent le paradoxe de la vie privée, qui n'incite pas les utilisateurs de Facebook à modifier leurs comportements de divulgation de soi sur le réseau social malgré les inquiétudes qu'ils affichent pour la protection de leurs données personnelles.

Nous considérons cependant que notre travail apporte une contribution intéressante à des réflexions bien actuelles et aux impacts potentiellement importants pour le développement du modèle économique propre aux réseaux sociaux.

BIBLIOGRAPHIE

Aflalo, A. (2018). *Apple, Amazon, Alphabet... les 10 entreprises les plus valorisées en bourse*. Retrieved from http://www.leparisien.fr/economie/business/apple-amazon-facebook-les-10-entreprises-les-mieux-capitalisees-en-bourse-03-08-2018-7842417.php.

Alashoor, T., Han, S. & Joseph, R. (2017). Familiarity with Big Data, privacy concerns, and self-disclosure accuracy in social networking websites: An APCO model. *Communications of the Association for Information Systems*, 41, 62-96.

Angst, C. & Agarwal, R. (2009). Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly*, 33(2), 339-370.

APD (2018). *Le RGPD après six mois : bilan.* Retrieved from https://www.autoriteprotectiondonnees.be/news/le-rgpd-apres-six-mois-bilan.

Audureau, W. (2018). *Ce qu'il faut savoir sur Cambridge Analytica, la société au cœur du scandale Facebook*. Retrieved from https://www.lemonde.fr/pixels/article/2018/03/22/ce-qu-il-faut-savoir-sur-cambridge-analytica-la-societe-au-c-ur-du-scandale-facebook_5274804_4408996.html.

Babinet, G. (2015). *Big Data, penser l'homme et le monde autrement*. Paris: Le Passeur.

Barnes, S. (2006). A privacy paradox: Social networking in the United States. *First Monday*, 11(9).

Bathelot, B. (2017). *Définition : marketing prédictif*. Retrieved from https://www.definitions-marketing.com/definition/marketing-predictif/.

Bélanger, F. & Crossler, R. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4), 1017-1041.

Benamati, J., Ozdemir, Z. & Smith, H. (2017). An empirical test of an Antecedents – Privacy Concerns – Outcomes model. *Journal of Information Science*, 43(5), 583-600.

Bensinger, G. (2014). *Amazon wants to ship your package before you buy it*. Retrieved from https://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/.

Blavier, A. (2016). *Le marketing digital toujours plus mobile.* Retrieved from http://www.hecexecutiveschool.be/le-marketing-digital-toujours-plus-mobile/.

Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679.

Boyd, D. & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

Braun, E. (2018). Cambridge Analytica: le Parlement européen réclame un audit complet de Facebook. Retrieved from http://www.lefigaro.fr/secteur/high-tech/2018/10/25/32001-20181025ARTFIG00243-cambridge-analytica-le-parlement-europeen-reclame-un-audit-complet-de-facebook.php

Brossas, V. (2018). *Marketing prédictif : 3 cas concrets d'utilisation en e-commerce !* Retrieved from https://www.leptidigital.fr/big-data/marketing-predictif-e-commerce-13055/

Calder, A. (2016). *Règlement général européen sur la protection des données – Guide de conformité*. Ireland: IT Governance.

Chausson, C. (2013). *Analyse prédictive : SAP rachète Kxen*. Retrieved from https://www.lemagit.fr/actualites/2240205521/Analyse-predictive-SAP-rachete-Kxen.

Chen, H. (2018). Revisiting the privacy paradox on social media with an extended privacy calculus model: The effect of privacy concerns, privacy self-efficacy, and social capital on privacy management. *American Behavioral Scientist*, 62(10), 1392-1412.

Clemons, E., Wilson, J. & Jin, F. (2014). Investigations into consumers preferences concerning privacy: An initial step towards the development of modern and consistent privacy protections around the globe. *Proceedings of the 47th Hawaii International Conference on System Sciences*, 4083-4092.

Columbus, L. (2017). *53% of companies are adopting Big Data analytics*. Retrieved from https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/.

Cosenza, V. (2018). *World map of social networks*. Retrieved from http://vincos.it/world-map-of-social-networks/.

Crié, D. & Willart, S. (2016). Le marketing et les mégadonnées. *Statistique et société*, 4(3), 13-18.

Cukier K. (2010). *Data, data everywhere*. Retrieved from https://www.economist.com/special-report/2010/02/25/data-data-everywhere.

Culnan, M. & Armstrong, P. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1), 104-115.

Cynober, T. (2018). Enjeux des Big Data en santé. *Actualités Pharmaceutiques*, 57(578), 25-29.

Dance, G., LaForgia, M. & Confessore, N. (2018). *As Facebook raised a privacy wall, it carved an opening for tech giants.* Retrieved from https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html.

De Carme, J. (2018). *Un Français crée un algorithme pour en finir avec la fraude dans les transports.* Retrieved from https://detours.canal.fr/francais-cree-algorithme-eliminer-fraude-transports/.

Dienlin, T. & Trepte, S. (2015). Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. *European Journal of Social Psychology*, 45(3), 285-297.

Dinev, T. & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61-80.

Dinev, T. (2014). Why would we care about privacy? *European Journal of Information Systems*, 23(2), 97-102.

Dinev, T., McConnell, A. & Smith, H. (2015). Informing privacy research through information systems, psychology, and behavioral economics: Thinking outside the APCO box. *Information Systems Research*, 26(4), 639-655.

Dommeyer, C. & Gross, B. (2003). What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing*, 17(2), 34-51.

Dresner Advisory Services (2017). *2017 Big Data analytics market study*. New Hampshire: Dresner Advisory Services.

Drouglazet, K. (2018). *Ce qu'il faut savoir sur le scandale Cambridge Analytica qui fait vaciller Facebook*. Retrieved from https://www.usine-digitale.fr/article/ce-qu-il-faut-savoir-sur-le-scandale-cambridge-analytica-qui-fait-vaciller-facebook.N669244.

Dussart, C. & Nantel, J. (2007). L'évolution du marketing : retour vers le futur. *Gestion*, 32(3), 66-74.

Estienne, Y. (2011). Un monde de verre : Facebook ou les paradoxes de la vie privée (sur)exposée. *Terminal*, 108, 65-84.

Europa (2018). *Règlements, directives et autres actes législatifs*. Retrieved from https://europa.eu/european-union/eu-law/legal-acts_fr.

Evans, D. & Schmalensee, R. (2017). *De précieux intermédiaires: Comment Blablacar, Facebook et Uber créent de la valeur.* Paris: Odile Jacob.

Facebook (2018). *Company info*. Retrieved from https://newsroom.fb.com/company-info/.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics (4th Edition)*. London: SAGE Publication.

Forrester Research (2017). *The Forrester Wave: Predictive analytics and machine learning solutions, Q1 2017*. Cambridge: Forrester Research.

Fredenucci, C. (2017). *Les 4 étapes qu'a suivies Amazon pour dominer l'e-commerce.* Retrieved from https://alphalyr.fr/blog/4-etapes-qua-suivies-amazon-dominer-commerce/.

French Web (2018). *Roofstreet, la solution de géomarketing prédictif.* Retrieved from https://www.frenchweb.fr/fw-radar-roofstreet-la-solution-de-geomarketing-predictif/322247.

Fuchs, C. (2017). *Social media: A critical introduction*. London: SAGE Publications.

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.

Glance, D. (2018). *Facebook et le « paradoxe de la vie privée »*. Retrieved from https://theconversation.com/facebook-et-le-paradoxe-de-la-vie-privee-94684

Giannelloni, J. & Le Nagard, E. (2016). Big Data et marketing prédictif. Vers un consommateur nu ? *Décisions Marketing*, 82, 5-15.

Gobert, M., Piron C., Filion N., Bulteel L., Daem M., Vanderwee K., Caillet O., Lardennois M., Folens B., & Defloor T. (2010). BeST – Belgian screening tools : mise à disposition d'outils d'aide à la décision en soins infirmiers. *Pratiques et Organisation des Soins*, 41(2), 167-179.

Gomez-Barroso, J., Feijoo, C. & Martinez, I. (2018). Privacy calculus: Factors that influence the perception of benefit. *El Profesional de la Informacion*, 27(2), 341-348.

Heravi, A., Mani, D., Choo, K. & Mubarak, S. (2017). Making decisions about self-disclosure in online social networks. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 1922-1931.

Heravi, A., Mubarak, S. & Choo, K. (2018). Information privacy in online social networks: Uses and gratification perspective. *Computers in Human Behavior*, 84, 441-459.

Hill, K. & Mattu, S. (2018). *Facebook knows how to track you using the dust on your camera lens.* Retrieved from https://gizmodo.com/facebook-knows-how-to-track-you-using-the-dust-on-your-1821030620.

Hirth, J. (2017). *Le data marketing*. Paris: Eyrolles.

Hong, W. & Thong, J. (2013). Internet privacy concerns: An integrated conceptualization and four empirical studies. *MIS Quarterly*, 37(1), 275-298.

Hootsuite & WeAreSocial (2018). *Digital in 2018 in Western Europe.* Retrieved from https://amadeusconcept.be/wp-content/uploads/2018/03/digitalin2018_belgium.pdf.

Hui, K., Tan, B.& Goh, C. (2006). Online information disclosure: Motivators and measurements. *ACM Transactions on Internet Technology*, 6(4), 415-441.

IBM (2016). *The four V's of Big Data*. Retrieved from https://www.ibmbigdatahub.com/infographic/four-vs-big-data.

IBM (2018). *Watson*. Retrieved from https://www.ibm.com/watson/ca-fr/.

Jaimes, N. (2016). *IBM Watson met de l'intelligence artificielle dans ses pubs*. Retrieved from https://www.journaldunet.com/ebusiness/publicite/1187637-ibm-watson-met-de-l-intelligence-artificielle-dans-ses-pubs/.

JDN (2018). *À quoi sert le clustering des données ?* Retrieved from https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203345-clustering-definition/.

Kaplan, A. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59-68.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122-134.

Krasnova, H. Kolesnikova, E., & Gunther, O. (2009). It won't happen to me!: Self-disclosure in online social networks. *Proceeding of the 15th Americas Conference on Information Systems*, 343, 1-10.

Kuo, K. & Talley, P. (2014). An empirical investigation of the privacy concerns of social network site users in Taiwan. *International Journal of Scientific Knowledge*, 5(2), 1-19.

Lambin, J., & de Moerloose, C. (2012). *Marketing stratégique et opérationnel: Du marketing à l'orientation-marché (8e éd.)*. Paris: Dunod.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. Connecticut: META Group.

Lanier, C. & Saini, A. (2008). Understanding consumer privacy: A review and future directions. *Academy of Marketing Science Review*, 12(2), 1-45.

Laufer, R. & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, 33(3), 22-42.

Launay, E. (2017). *L'utilité du marketing prédictif personnalisé pour les PME.* Retrieved from https://solutions.lesechos.fr/com-marketing/c/lutilite-marketing-predictif-personnalise-pme-8766/.

Le Monde (2012). *Facebook, de nombreuses polémiques sur la vie privée.* Retrieved from https://www.lemonde.fr/technologies/article/2012/09/25/facebook-de-nombreuses-polemiques-sur-la-vie-privee_1765127_651865.html.

Le Monde (2013). *Le logiciel qui prédit les délits*. Retrieved from https://www.lemonde.fr/ameriques/article/2013/01/04/le-logiciel-qui-predit-les-delits_1812195_3222.html.

Lee, I. (2017). Big Data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303.

Lemaire, B. (2018). *Inquiétude des consommateurs : seulement 1% des sites web sont conformes au RGPD*. Retrieved from https://www.cio-online.com/actualites/lire-inquietude-des-consommateurs-seulement-1-des-sites-web-sont-conformes-au-rgpd-10853.html

Li, H., Sarathy, R. & Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems*, 51(1), 62-71.

Li, Y. (2011). Empirical studies on online information privacy concerns: Literature review and an integrative framework. *Communications of the Association for Information Systems*, 28, 453-496.

Lin, X., Li, Y., Califf, C. & Featherman, M. (2013). Can social role theory explain gender differences in Facebook usage? *Proceedings of the 46th Hawaii International Conference on System Sciences*, 690-699.

Lowry, P., Cao, J. & Everard, A. (2011). Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures. *Journal of Management Information Systems*, 27(4), 163-200.

Lwin, M., Wirtz, J. & Williams, J. (2007). Consumer online privacy concerns and responses: A power responsibility equilibrium perspective. *Journal of the Academy of Marketing Science*, 35(4), 572-585.

Malhotra, N. (2014). *Études marketing (6ème édition)*. Paris: Pearson France.

Malhotra, N., Kim, S. & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4), 336-355.

Marchand, L. (2018). *Cambridge Analytica : Facebook écope d'une amende symbolique*. Retrieved from https://www.lesechos.fr/tech-medias/hightech/0600035549808-cambridge-analytica-facebook-ecope-dune-amende-symbolique-2216689.php.

Marr, B. (2015). *How Apple uses Big Data to drive success.* Retrieved from https://www.datasciencecentral.com/profiles/blogs/how-apple-uses-big-data-to-drive-success.

Mauchamp, P. (2015). *Marketing prédictif : une révolution portée par le Big et le Smart Data*. Retrieved from http://archives.lesechos.fr/archives/cercle/2015/02/24/cercle_124613.htm#.

McCroskey, J. & Richmond, V. (1977). Communication apprehension as a predictor of self-disclosure. *Communication Quarterly*, 25(4), 40-43.

Medioni, S. & Benmoyal-Bouzaglo, S. (2018). *Marketing digital*. Paris: Dunod.

Mercklé, P. (2016). *Sociologie des réseaux sociaux*. Paris: La Découverte.

Metzger, M. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication*, 9(4).

Milne, G. & Culnan, M. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15-29.

Miltgen, C. (2011). Vie privée et marketing: Étude de la décision de fournir des données personnelles dans un cadre commercial. *Réseaux*, 167(3), 131-166.

Miltgen, C. & Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: A qualitative study in seven European countries. *European Journal of Information Systems*, 23(2), 103-125.

Miltgen, C. & Smith, H. (2015). Exploring information privacy regulation, risks, trust, and behavior. *Information and Management*, 52(6), 741-759.

Morris, M., Venkatesh, V. & Ackerman, P. (2005). Gender and age differences in employee decisions about new technology: An extension to the theory of planned behavior. *IEEE Transactions on Engineering Management*, 52(1), 69-84.

Norberg, P., Horne, D. & Horne, D. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1), 100-126.

O'Reilly, T. (2005). *What is Web 2.0*. Retrieved from https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

Ozdemir, Z., Smith, H. & Benamati, J. (2017). Antecedents and outcomes of information privacy concerns in a peer context: An exploratory study. *European Journal of Information Systems*, 26(6), 642-660.

Pavlou, P. (2011). State of the information privacy literature: Where we are now and where should we go? *MIS Quarterly*, 35(4), 977-988.

Perreau, C. (2018). *RGPD : texte, signification, rôle de la Cnil...* Retrieved from https://www.journaldunet.com/economie/services/1208625-rgpd-texte-signification-role-de-la-cnil/.

Pleyers, G. (2017). *LSMS2000 : Études et modèles de marché*. Document non publié, Louvain School of Management, Louvain-la-Neuve.

Pras, B. (2012). Entreprise et vie privée : le privacy paradox et comment le dépasser ? *Revue française de gestion*, 5(224), 87-94.

Press, G. (2018). *Why Facebook triumphed over all other social networks*. Retrieved from https://www.forbes.com/sites/gilpress/2018/04/08/why-facebook-triumphed-over-all-other-social-networks/#795bde8b6e91.

Pure Storage (2015). *Le big échec du Big Data : comment les entreprises luttent pour avoir accès aux informations dont elles ont besoin*. Paris: Pure Storage.

Rachels, J. (1975). Why privacy is important. *Philosophy and Public Affairs*, 4(4), 323-333.

Raynal, J. (2018). *Tinyclues injecte 14,6 millions d'euros dans sa solution de marketing prédictif*. Retrieved from https://www.usine-digitale.fr/article/digital-retail-tinyclues-injecte-14-6-millions-d-euros-dans-sa-solution-de-marketing-predictif.N655374.

Romm, T., Fung, B., Davies, A., & Timberg, C. (2018). *'It's about time': Facebook faces first lawsuit from U.S. regulators after Cambridge Analytica scandal*. Retrieved from https://www.washingtonpost.com/technology/2018/12/19/dc-attorney-general-sues-facebook-over-alleged-privacy-violations-cambridge-analytica-scandal/?noredirect=on&utm_term=.53354c5dc139

Ronfaut, L. (2018). *Piratage de Facebook: ce qu'il faut savoir*. Retrieved from http://www.lefigaro.fr/secteur/high-tech/2018/09/28/32001-20180928ARTFIG00351-facebook-pirate-50-millions-de-comptes-concernes.php.

Rowan, M. & Dehlinger, J. (2014). Observed gender differences in privacy concerns and behaviors of mobile device end users. *Procedia Computer Science*, 37, 340-347.

Salovic, B., Guerguinov, O. & Léonard, T. (2018). *La Belgique transpose (enfin) le RGPD (GDPR).* Retrieved from https://www.droit-technologie.org/actualites/belgique-transpose-enfin-rgpd-gdpr/.

SAP (2018). *Analyse prédictive*. Retrieved from https://www.sap.com/france/products/analytics/predictive-analytics.html.

SAS (2018). *Analyse prédictive : rôle et atouts*. Retrieved from https://www.sas.com/fr_fr/insights/analytics/predictive-analytics.html.

Schermer, B. & Wagemans, T. (2010). Freedom in the days of the Internet: Regulating, legislating and liberating the Internet while protecting rights and unlocking potentials. *European View*, 9(2), 287-293.

Sevignani, S. (2012). Privacy on social networking sites within a culture of exchange. In *Media, knowledge and education: Cultures and ethics of sharing* (pp. 89-104). Autriche: Innsbruck University Press.

Sheehan, K. (1999). An investigation of gender differences in online privacy concerns and resultant behaviors. *Journal of Interactive Marketing*, 13(4), 24-38.

Scibetta, C., Kempf, A. & Alves, C. (2012). *Projets de communication : conduite et vente*. Paris: Armand Colin.

Siegel, E. (2016). *Predictive analytics: The power to predict who will click, buy, lie or die*. New Jersey: John Wiley & Sons.

Smith, H., Dinev, T. & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, 35(4), 989-1015.

Smith, H., Milberg, S. & Burke, S. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly*, 20(2), 167-196.

Smyrnaios, N. (2018). *Du réseau de l'élite aux scandales en série : brève histoire de Facebook*. Retrieved from https://www.inaglobal.fr/numerique/article/du-reseau-de-l-elite-aux-scandales-en-serie-breve-histoire-de-facebook-10275.

Son, J. & Kim, S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly*, 32(3), 503-529.

Statbel (2018). *Une entreprise belge sur cinq analyse des big data*. Retrieved from https://statbel.fgov.be/fr/themes/entreprises/ict-dans-les-entreprises

Statista (2018a). *Most famous social network sites worldwide as of October 2018, ranked by number of active users (in millions)*. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Statista (2018b). *Facebook's U.S. and non-U.S. advertising revenue from 2014 to 2018*. Retrieved from https://www.statista.com/statistics/544001/facebooks-advertising-revenue-worldwide-usa/.

Stenger, T. & Coutant, A. (2013). Médias sociaux : clarification et cartographie. Pour une approche sociotechnique. *Décisions Marketing*, 70, 107-117.

Stewart, D. (2017). A comment on privacy. *Journal of the Academy of Marketing Science*, 45(2), 156-159.

Stewart, K. & Segars, A. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research*, 13(1), 36-49.

Stutzman, F. & Kramer-Duffield, J. (2010). Friends only: Examining a privacy-enhancing behavior in Facebook. *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 1553-1562.

Swaen, V. (2017). LLSMS2005 : *Méthodes avancées de recherche marketing*. Document non publié, Louvain School of Management, Louvain-la-Neuve.

Taddicken, M. (2014). The privacy paradox in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication*, 19(2), 248-273.

Trepte, S., Reinecke, L., Ellison, N., Quiring, O., Yao, M. & Ziegele, M. (2017). A cross-cultural perspective on the privacy calculus. *Social Media + Society*, 3(1), 1-13.

Van Rijmenam, M. (2014). *Tesco and Big Data analytics, a recipe for success*? Retrieved from https://datafloq.com/read/tesco-big-data-analytics-recipe-success/665.

Volle, P. (2011). Marketing : comprendre l'origine historique. In *MBA Marketing* (pp. 23-45). Paris: Eyrolles.

Vouters, C. & Lavielle, S. (2018). *Les Big Data au service de la santé : quoi de neuf, docteur ?* Retrieved from https://transformation-digitale.pwc.fr/data/les-big-data-au-service-de-la-sante-quoi-de-neuf-docteur.

Watson, H. (2014). Tutorial: Big Data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34, 1247-1268.

Warren, S. & Brandeis, L. (1890). The right to privacy. *Harvard Law Review*, 4(5), 193-220.

Weiss, R.& Townsend, R. (2005). Using Excel to clean and prepare data for analysis. *The Industrial-Organizational Psychologist*, 42(3), 89-97.

Westin, A. (1967). *Privacy and freedom*. New York: Atheneum.

Westin, A. (1990). *Consumer privacy issues in the nineties*. Atlanta: Equifax Inc.

Xu, F., Michael, K. & Chen, X. (2013). Factors affecting privacy disclosure on social network sites: An integrated model. *Electronic Commerce Research*, 13(2), 151-168.

Young, A. & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication and Society*, 16(4), 479-500.

Zhang, R., Chen, J. & Lee, C. (2013). Mobile commerce and consumer privacy concerns. *Journal of Computer Information Systems*, 53(4), 31-38.

Zlatolas, L., Welzer, T., Hericko, M., & Holbl, M. (2015). Privacy antecedents for SNS self-disclosure: The case of Facebook. *Computers in Human Behavior*, 45, 158-167.

ANNEXES

Annexe 1 : Questionnaire

Bonjour à tous,

Ce questionnaire vous parvient dans le cadre de mon mémoire de fin d'études en Ingénieur de Gestion à la Louvain School of Management (UCL).
Celui-ci ne prendra que quelques minutes de votre temps.
Il n'y a pas de bonne ou de mauvaise réponse, c'est votre avis qui m'intéresse.
Vos réponses seront traitées de manière anonyme et ne seront pas utilisées à d'autres fins que cette étude.
Merci d'avance pour votre collaboration.

Thibaud Aerts

Section 1 :

1. Conformément à la loi récemment entrée en vigueur (RGPD), j'ai compris que mes réponses seront enregistrées dans une base de données et seront analysées de façon anonyme, dans un but purement scientifique. Dans ce cadre, j'accepte de participer à cette enquête.

- Oui
- Non (fin du questionnaire)

Section 2 :

2. Avez-vous un compte Facebook ?

- Oui
- Non (fin du questionnaire)

Section 3 :

3. À quelle fréquence utilisez-vous Facebook ?

- Moins d'une fois par semaine
- Moins d'une fois par jour
- Moins de 10 minutes par jour
- Entre 10 et 30 minutes par jour
- Plus de 30 minutes par jour

Au cours des prochaines questions, il vous sera, à chaque fois, demandé de déterminer à quel degré vous êtes en accord ou pas avec les affirmations proposées.
1= « Pas du tout d'accord » ;
2 = « Pas d'accord » ;
3 = « Plutôt pas d'accord » ;
4 = « Ni d'accord, ni pas d'accord » ;
5 = « Plutôt d'accord » ;
6 = « D'accord » ;
7 = « Tout à fait d'accord ».

4. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ?
Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

DIV1	Mon profil Facebook en dit long sur moi							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
DIV2	Je tiens mes amis au courant de ce qui se passe dans ma vie à travers Facebook							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
DIV3	Je révèle beaucoup d'informations sur moi sur Facebook, notamment par mes publications							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
DIV4	Cela ne me dérange pas de mettre des informations personnelles sur Facebook							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

5. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ?
Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

PRO1	Je fais attention à ce que je publie sur les réseaux sociaux							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
PRO2	Je limite (au-delà des paramètres par défaut) l'accès à mes informations personnelles, à mes publications ou à mes pages likées sur Facebook							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
PRO3	J'évite d'utiliser des applications tierces, des jeux ou des tests sur Facebook							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
PRO4	Je fais attention avec qui je suis ami sur Facebook							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
PRO5	J'empêche Facebook d'avoir accès à ma localisation							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

Section 4 :

6. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ? Je crains que…

Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

INQ1	…les réseaux sociaux collectent trop d'informations personnelles sur moi							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
INQ2	…des personnes non autorisées puissent accéder à mes informations personnelles sur les réseaux sociaux							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
INQ3	…les informations que je partage sur les réseaux sociaux puissent être utilisées à mauvais escient							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
INQ4	…les réseaux sociaux utilisent mes informations personnelles à des fins qui ne me sont pas communiquées							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

Section 5 :

7. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ?

Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

AVA1	J'ai déjà entendu parler du Big Data							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
AVA2	Je suis conscient des avantages du Big Data pour les individus, notamment en ce qui concerne la personnalisation des services offerts							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
AVA3	Je suis conscient de la valeur du Big Data pour les entreprises, notamment en ce qui concerne l'exploitation plus rapide des données collectées							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
AVA4	Je connais la signification du terme Big Data							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

8. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ? Je suis conscient que mes données personnelles sur les réseaux sociaux…

Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

IMP1	…puissent être collectées							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
IMP2	…puissent être stockées pendant de nombreuses années							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
IMP3	…puissent être analysées de différentes façons							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
IMP4	…puissent être utilisées pour cibler les services ou les informations les plus susceptibles de m'intéresser							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
IMP5	…puissent être partagées avec différentes tierces parties							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

Section 6 :

9. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ? J'ai le droit (en respectant certaines formalités et conditions) de demander à toute entreprise ou administration…

Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

REG1	…d'avoir accès aux données personnelles que celle-ci possède sur moi							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
REG2	…la rectification de mes données personnelles si celles-ci sont inexactes							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
REG3	…que mes données personnelles soient effacées							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
REG4	…une copie de mes données personnelles afin de transférer celles-ci vers toute autre entreprise ou administration							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
REG5	…de ne pas traiter mes données personnelles							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

10. Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ?
Attribuez une note de 1 à 7 à l'expression proposée, avec 1 = « Pas du tout d'accord » et 7 = « Tout à fait d'accord » ; les notes intermédiaires vous permettent de nuancer votre jugement.

EXP1	Presque tous les jours, j'entends ou je lis des nouvelles sur la violation de la confidentialité des données personnelles							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
EXP2	Beaucoup d'informations sont publiées sur la façon dont nos données personnelles peuvent être utilisées ou détournées par les entreprises							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
EXP3	Le sujet du respect de la vie privée fait beaucoup parler de lui dans l'actualité							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord
EXP4	J'ai entendu ou lu beaucoup d'informations sur l'affaire Cambridge Analytica au cours de ces derniers mois							
Pas du tout d'accord	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Tout à fait d'accord

Section 7 :

11. Êtes-vous ?
- Un homme
- Une femme

12. Quel âge avez-vous ?
- -18 ans
- 18-24 ans
- 25-34 ans
- 35-44 ans
- 45-54 ans
- 55-65 ans
- +65 ans

13. Quel est le plus haut niveau d'étude que vous ayez achevé avec succès ?
- Enseignement primaire
- Enseignement secondaire
- Enseignement supérieur de type court
- Enseignement supérieur de type long

14. Quel est votre statut professionnel ?
- Étudiant(e)
- Stagiaire
- Ouvrier(ère)
- Fonctionnaire
- Employé(e)
- Cadre
- Indépendant(e)
- Sans emploi
- Retraité(e)
- Autre : …

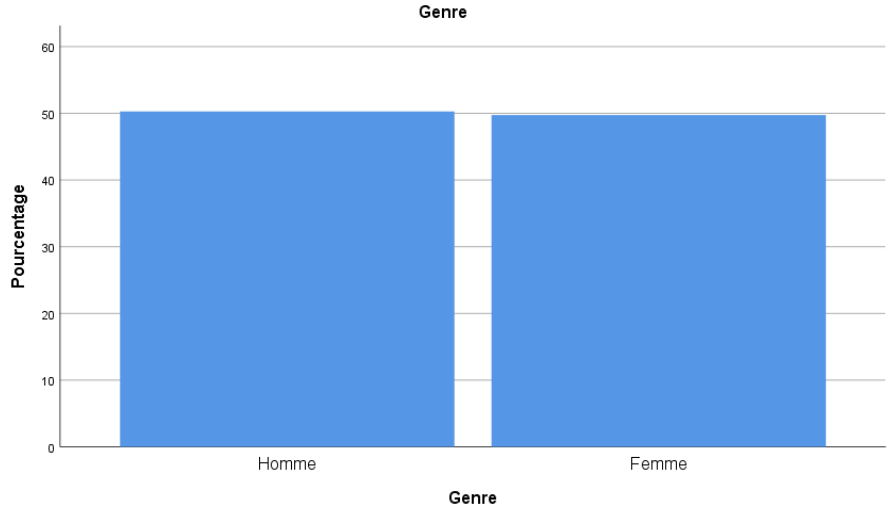
Annexe 2 : Plan de codage SPSS

Valeur dans SPSS	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Fréquence d'utilisation de Facebook	Moins d'une fois par semaine	Moins d'une fois par jour	Moins de 10 minutes par jour	Entre 10 et 30 minutes par jour	Plus de 30 minutes par jour		
Échelle de Likert	Pas du tout d'accord	Pas d'accord	Plutôt pas d'accord	Ni d'accord, ni pas d'accord	Plutôt d'accord	D'accord	Tout à fait d'accord
Genre	Homme	Femme					
Âge	-18 ans	18-24 ans	25-34 ans	35-44 ans	45-54 ans	55-65 ans	+65 ans
Niveau d'éducation	Primaire et secondaire	Supérieur de type court	Supérieur de type long				
Profession	Étudiant / Stagiaire	Employé	Cadre	Indépendant	Fonctionnaire	Retraité / Sans emploi	

Annexe 3 : Profil sociodémographique de l'échantillon

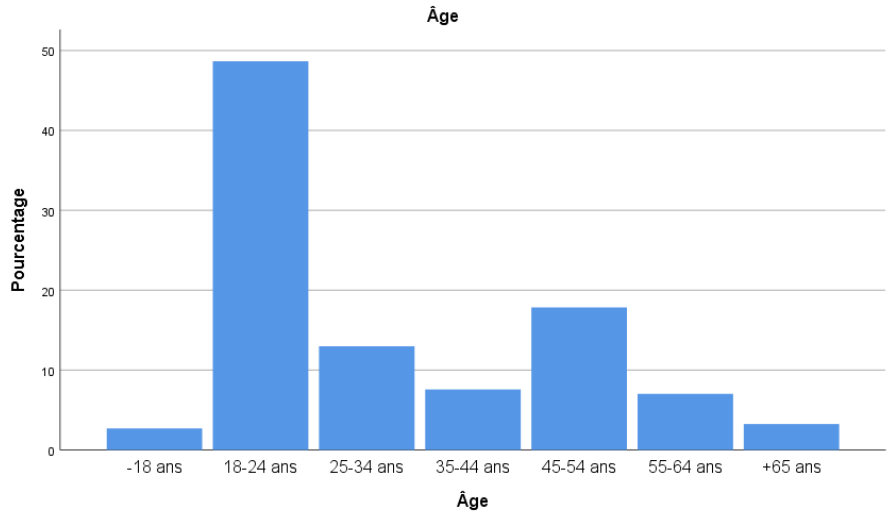
Genre

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Homme	93	50,3	50,3	50,3
	Femme	92	49,7	49,7	100,0
	Total	185	100,0	100,0	



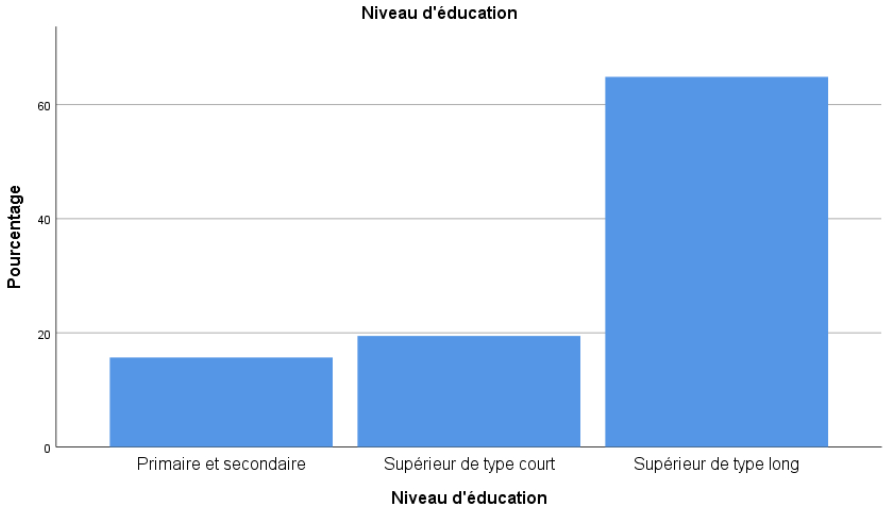
Âge

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	-18 ans	5	2,7	2,7	2,7
	18-24 ans	90	48,6	48,6	51,4
	25-34 ans	24	13,0	13,0	64,3
	35-44 ans	14	7,6	7,6	71,9
	45-54 ans	33	17,8	17,8	89,7
	55-64 ans	13	7,0	7,0	96,8
	+65 ans	6	3,2	3,2	100,0
	Total	185	100,0	100,0	



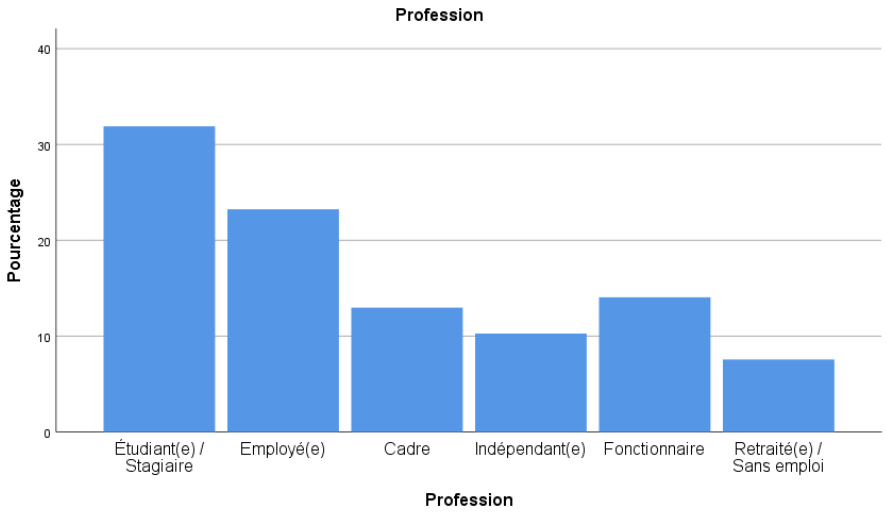
Niveau d'éducation

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Primaire et secondaire	29	15,7	15,7	15,7
	Supérieur de type court	36	19,5	19,5	35,1
	Supérieur de type long	120	64,9	64,9	100,0
	Total	185	100,0	100,0	



Profession

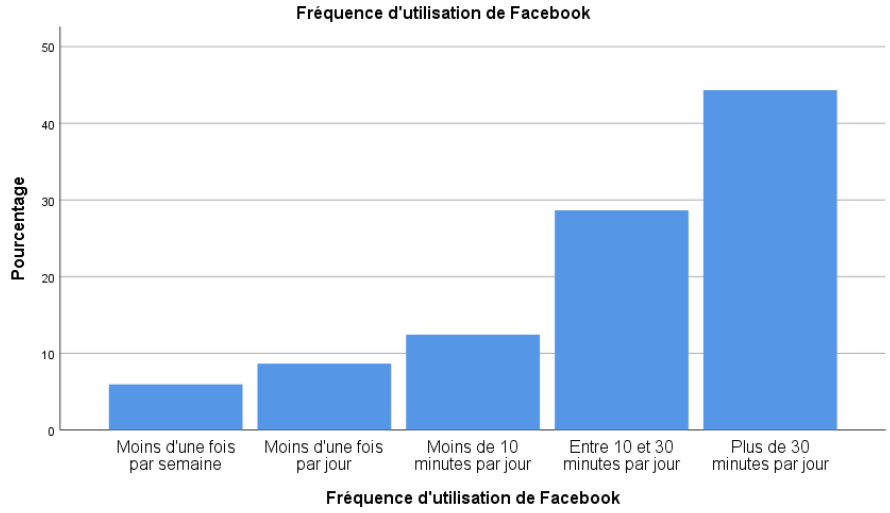
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Étudiant(e) / Stagiaire	59	31,9	31,9	31,9
	Employé(e)	43	23,2	23,2	55,1
	Cadre	24	13,0	13,0	68,1
	Indépendant(e)	19	10,3	10,3	78,4
	Fonctionnaire	26	14,1	14,1	92,4
	Retraité(e) / Sans emploi	14	7,6	7,6	100,0
	Total	185	100,0	100,0	



90.

Fréquence d'utilisation de Facebook

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Moins d'une fois par semaine	11	5,9	5,9	5,9
	Moins d'une fois par jour	16	8,6	8,6	14,6
	Moins de 10 minutes par jour	23	12,4	12,4	27,0
	Entre 10 et 30 minutes par jour	53	28,6	28,6	55,7
	Plus de 30 minutes par jour	82	44,3	44,3	100,0
	Total	185	100,0	100,0	



Annexe 4 : Répartition de l'âge et du genre des utilisateurs de Facebook en Belgique

Répartition en fonction de l'âge de la population cible

Âge	Nombre	Pourcentage
-18 ans	430.000	5,73%
18-24 ans	1.340.000	17,86%
25-34 ans	1.749.000	23,32%
35-44 ans	1.473.00	19,64%
45-54 ans	1.198.000	15,973%
54-65 ans	799.000	10,653%
+65 ans	511.000	6,813%
TOTAL	**7.500.000**	**100%**

Source : Hootsuite & WeAreSocial (2018)

Répartition en fonction du genre de la population cible

Genre	Nombre	Pourcentage
Homme	3.750.000	50%
Femme	3.750.000	50%
TOTAL	**7.500.000**	**100%**

Source : Hootsuite & WeAreSocial (2018)

Annexe 5 : Analyses de validité et de fiabilité

1. Modèle global

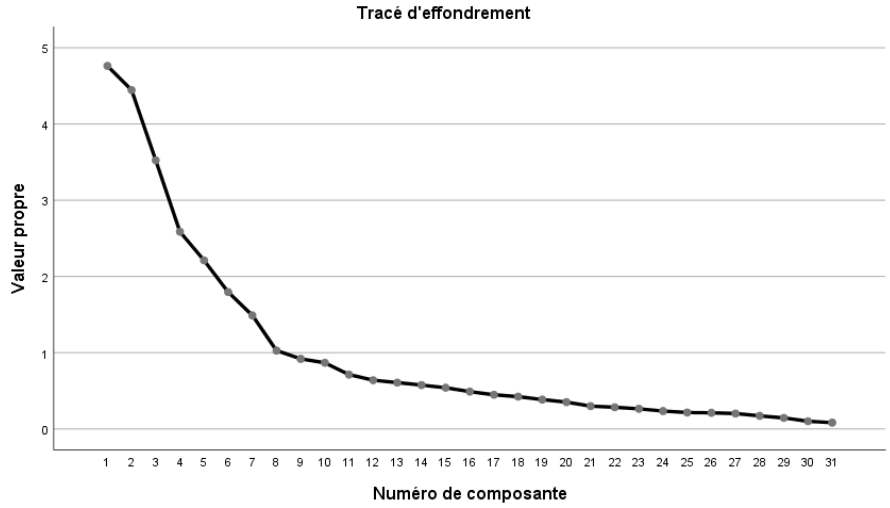
Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,776
Test de sphéricité de Bartlett	Khi-carré approx.	3186,972
	ddl	465
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,761	15,360	15,360	4,761	15,360	15,360	4,217	13,602	13,602
2	4,445	14,338	29,698	4,445	14,338	29,698	3,632	11,716	25,318
3	3,523	11,365	41,063	3,523	11,365	41,063	3,208	10,347	35,665
4	2,587	8,344	49,407	2,587	8,344	49,407	2,882	9,295	44,960
5	2,210	7,130	56,538	2,210	7,130	56,538	2,352	7,586	52,546
6	1,795	5,790	62,328	1,795	5,790	62,328	2,282	7,363	59,909
7	1,489	4,802	67,130	1,489	4,802	67,130	2,238	7,220	67,130
8	1,027	3,313	70,443						

Méthode d'extraction : Analyse en composantes principales.



Qualités de représentation

	Initiales	Extraction
DIV1	1,000	,596
DIV2	1,000	,598
DIV3	1,000	,780
DIV4	1,000	,464
PRO1	1,000	,465
PRO2	1,000	,407
PRO3	1,000	,429
PRO4	1,000	,469
PRO5	1,000	,537
INQ1	1,000	,700
INQ2	1,000	,739
INQ3	1,000	,704
INQ4	1,000	,643
AVA1	1,000	,852
AVA2	1,000	,798
AVA3	1,000	,847
AVA4	1,000	,847
IMP1	1,000	,814
IMP2	1,000	,860
IMP3	1,000	,895
IMP4	1,000	,801
IMP5	1,000	,781
RGP1	1,000	,734
RGP2	1,000	,640
RGP3	1,000	,679
RGP4	1,000	,650
RGP5	1,000	,601
MED1	1,000	,630
MED2	1,000	,670
MED3	1,000	,641
MED4	1,000	,541

Méthode d'extraction : Analyse en composantes principales.

Rotation de la matrice des composantes[a]

	Composante						
	1	2	3	4	5	6	7
IMP3	,944						
IMP2	,925						
IMP4	,889						
IMP1	,877						
IMP5	,877						
AVA1		,915					
AVA4		,909					
AVA3		,907					
AVA2		,837					
RGP1			,811				
RGP3			,791				
RGP4			,779				
RGP2			,778				
RGP5			,713				
INQ2				,835			
INQ3				,802			
INQ1				,776			
INQ4				,769			
PRO5					,704		
PRO4					,637		
PRO2					,620		
PRO1					,594		
PRO3					,571		
DIV3						,870	
DIV1						,766	
DIV2						,765	
DIV4					-,375	,481	
MED2							,783
MED3							,774
MED1				,307			,715
MED4		,440					,506

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

a. Convergence de la rotation dans 6 itérations.

2. Modèle global sans ‘DIV4’

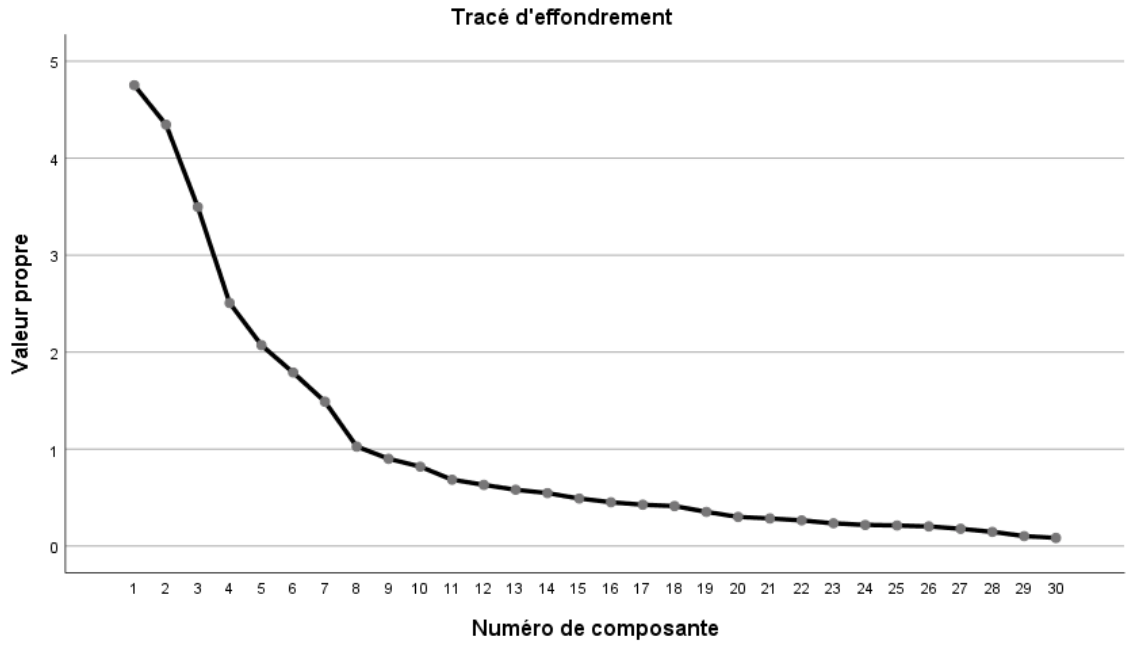
Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,779
Test de sphéricité de Bartlett	Khi-carré approx.	3111,109
	ddl	435
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,753	15,844	15,844	4,753	15,844	15,844	4,212	14,040	14,040
2	4,346	14,487	30,331	4,346	14,487	30,331	3,616	12,055	26,095
3	3,497	11,656	41,987	3,497	11,656	41,987	3,199	10,664	36,760
4	2,507	8,357	50,345	2,507	8,357	50,345	2,864	9,545	46,305
5	2,072	6,907	57,252	2,072	6,907	57,252	2,246	7,485	53,790
6	1,790	5,967	63,219	1,790	5,967	63,219	2,213	7,376	61,167
7	1,489	4,962	68,180	1,489	4,962	68,180	2,104	7,014	68,180
8	1,026	3,419	71,599						

Méthode d'extraction : Analyse en composantes principales.



Qualités de représentation

	Initiales	Extraction
DIV1	1,000	,637
DIV2	1,000	,594
DIV3	1,000	,805
PRO1	1,000	,470
PRO2	1,000	,430
PRO3	1,000	,440
PRO4	1,000	,478
PRO5	1,000	,519
INQ1	1,000	,699
INQ2	1,000	,738
INQ3	1,000	,704
INQ4	1,000	,643
AVA1	1,000	,852
AVA2	1,000	,797
AVA3	1,000	,849
AVA4	1,000	,846
IMP1	1,000	,811
IMP2	1,000	,861
IMP3	1,000	,895
IMP4	1,000	,801
IMP5	1,000	,781
RGP1	1,000	,734
RGP2	1,000	,645
RGP3	1,000	,679
RGP4	1,000	,649
RGP5	1,000	,600
MED1	1,000	,620
MED2	1,000	,687
MED3	1,000	,643
MED4	1,000	,544

Méthode d'extraction : Analyse en composantes principales.

Rotation de la matrice des composantes[a]

	Composante						
	1	2	3	4	5	6	7
IMP3	,944						
IMP2	,925						
IMP4	,889						
IMP1	,878						
IMP5	,877						
AVA1		,915					
AVA4		,908					
AVA3		,908					
AVA2		,838					
RGP1			,812				
RGP3			,791				
RGP2			,780				
RGP4			,778				
RGP5			,711				
INQ2				,835			
INQ3				,802			
INQ1				,777			
INQ4				,769			
PRO5					,695		
PRO4					,644		
PRO2					,640		
PRO1					,600		
PRO3					,580		
MED2						,794	
MED3						,774	
MED1				,306		,710	
MED4		,441				,511	
DIV3							,887
DIV1							,790
DIV2							,761

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

a. Convergence de la rotation dans 6 itérations.

3. Divulgation

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
DIV1	3,42	1,431	185
DIV2	2,02	1,236	185
DIV3	1,98	1,073	185
DIV4	2,50	1,665	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,682
Test de sphéricité de Bartlett	Khi-carré approx.	171,475
	ddl	6
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	2,214	55,362	55,362	2,214	55,362	55,362
2	,831	20,765	76,127			
3	,605	15,122	91,249			
4	,350	8,751	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
DIV1	1,000	,582
DIV2	1,000	,604
DIV3	1,000	,721
DIV4	1,000	,307

Méthode d'extraction : Analyse en composantes principales.

Matrice des composantes

	Composante 1
DIV1	,763
DIV2	,777
DIV3	,849
DIV4	,554

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,693	4

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
DIV1	6,49	9,262	,493	,617
DIV2	7,89	9,901	,537	,595
DIV3	7,93	10,207	,619	,567
DIV4	7,41	9,341	,343	,741

4. Divulgation sans ‘DIV4’

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
DIV1	3,42	1,431	185
DIV2	2,02	1,236	185
DIV3	1,98	1,073	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,639
Test de sphéricité de Bartlett	Khi-carré approx.	146,522
	ddl	3
	Signification	,000

Variance totale expliquée

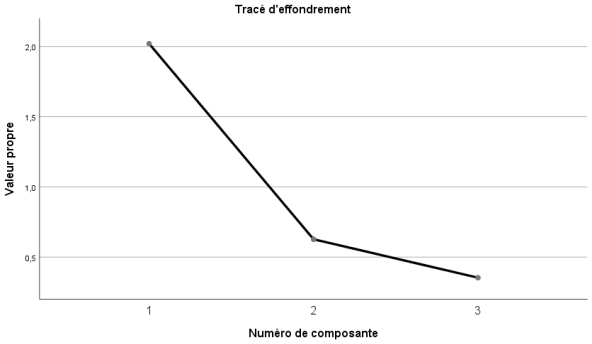
	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	2,020	67,326	67,326	2,020	67,326	67,326
2	,627	20,896	88,222			
3	,353	11,778	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
DIV1	1,000	,634
DIV2	1,000	,602
DIV3	1,000	,783

Méthode d'extraction : Analyse en composantes principales.



Matrice des composantes

	Composante 1
DIV1	,796
DIV2	,776
DIV3	,885

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,741	3

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
DIV1	3,99	4,158	,537	,711
DIV2	5,39	5,012	,507	,723
DIV3	5,43	4,899	,693	,540

5. Protection

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
PRO1	6,50	,835	185
PRO2	5,72	1,614	185
PRO3	5,42	1,749	185
PRO4	5,90	1,368	185
PRO5	4,55	2,136	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,755
Test de sphéricité de Bartlett	Khi-carré approx.	108,559
	ddl	10
	Signification	,000

Variance totale expliquée

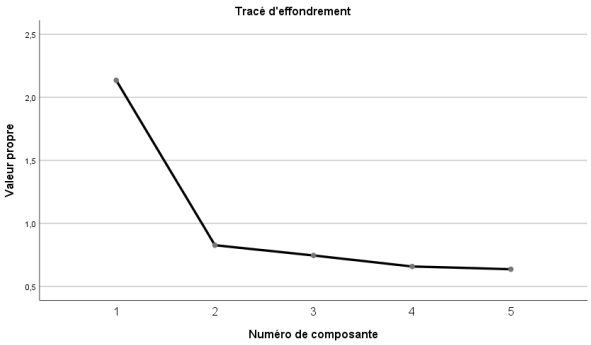
	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	2,135	42,696	42,696	2,135	42,696	42,696
2	,827	16,532	59,227			
3	,745	14,900	74,128			
4	,658	13,151	87,279			
5	,636	12,721	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
PRO1	1,000	,438
PRO2	1,000	,379
PRO3	1,000	,418
PRO4	1,000	,456
PRO5	1,000	,444

Méthode d'extraction : Analyse en composantes principales.



Matrice des composantes

	Composante 1
PRO1	,662
PRO2	,616
PRO3	,647
PRO4	,675
PRO5	,666

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,634	5

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
PRO1	21,59	21,992	,418	,600
PRO2	22,37	18,192	,375	,586
PRO3	22,67	17,081	,403	,573
PRO4	22,19	19,035	,424	,568
PRO5	23,54	14,391	,432	,570

6. Inquiétudes

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
INQ1	5,45	1,398	185
INQ2	5,46	1,605	185
INQ3	5,57	1,587	185
INQ4	5,74	1,426	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,753
Test de sphéricité de Bartlett	Khi-carré approx.	314,964
	ddl	6
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	2,748	68,703	68,703	2,748	68,703	68,703
2	,513	12,821	81,524			
3	,475	11,864	93,387			
4	,265	6,613	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
INQ1	1,000	,668
INQ2	1,000	,710
INQ3	1,000	,716
INQ4	1,000	,654

Méthode d'extraction : Analyse en composantes principales.

Matrice des composantes

	Composante 1
INQ1	,818
INQ2	,843
INQ3	,846
INQ4	,809

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,847	4

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
INQ1	16,77	15,571	,670	,813
INQ2	16,75	13,894	,706	,798
INQ3	16,65	13,947	,713	,794
INQ4	16,48	15,512	,657	,818

7. Avantages_BD

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
AVA1	5,22	2,230	185
AVA2	4,56	2,181	185
AVA3	5,12	2,146	185
AVA4	4,80	2,291	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,830
Test de sphéricité de Bartlett	Khi-carré approx.	618,072
	ddl	6
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	3,319	82,976	82,976	3,319	82,976	82,976
2	,329	8,227	91,202			
3	,195	4,884	96,086			
4	,157	3,914	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
AVA1	1,000	,843
AVA2	1,000	,790
AVA3	1,000	,851
AVA4	1,000	,835

Méthode d'extraction : Analyse en composantes principales.

Matrice des composantes

	Composante 1
AVA1	,918
AVA2	,889
AVA3	,922
AVA4	,914

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,931	4

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
AVA1	14,48	36,903	,852	,906
AVA2	15,14	38,465	,803	,922
AVA3	14,58	37,745	,858	,905
AVA4	14,90	36,397	,844	,909

8. Implications_BD

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
IMP1	6,00	1,493	185
IMP2	5,84	1,606	185
IMP3	5,92	1,534	185
IMP4	6,13	1,502	185
IMP5	5,56	1,756	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,885
Test de sphéricité de Bartlett	Khi-carré approx.	899,310
	ddl	10
	Signification	,000

Variance totale expliquée

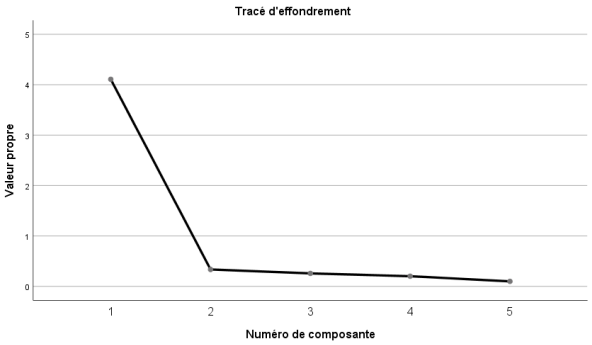
	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,107	82,149	82,149	4,107	82,149	82,149
2	,335	6,708	88,856			
3	,257	5,139	93,995			
4	,201	4,026	98,022			
5	,099	1,978	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
IMP1	1,000	,797
IMP2	1,000	,858
IMP3	1,000	,898
IMP4	1,000	,786
IMP5	1,000	,768

Méthode d'extraction : Analyse en composantes principales.



Matrice des composantes

	Composante 1
IMP1	,893
IMP2	,926
IMP3	,948
IMP4	,887
IMP5	,876

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,944	5

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
IMP1	23,45	34,336	,829	,934
IMP2	23,61	32,413	,879	,925
IMP3	23,53	32,718	,911	,920
IMP4	23,32	34,307	,824	,935
IMP5	23,90	31,919	,809	,940

9. Règles_RGPD

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
REG1	5,81	1,617	185
REG2	6,02	1,404	185
REG3	5,62	1,766	185
REG4	5,19	1,840	185
REG5	5,09	1,960	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,795
Test de sphéricité de Bartlett	Khi-carré approx.	387,409
	ddl	10
	Signification	,000

Variance totale expliquée

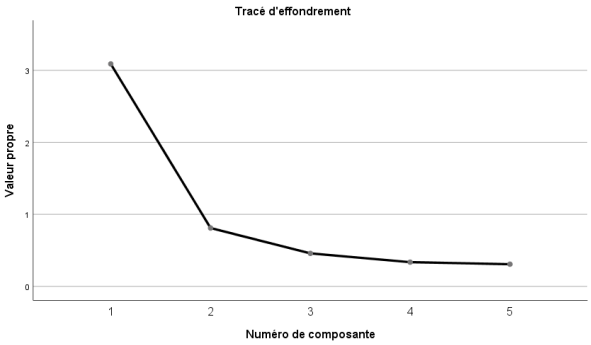
	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	3,089	61,773	61,773	3,089	61,773	61,773
2	,810	16,204	77,976			
3	,458	9,159	87,135			
4	,336	6,717	93,852			
5	,307	6,148	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
REG1	1,000	,663
REG2	1,000	,630
REG3	1,000	,637
REG4	1,000	,643
REG5	1,000	,516

Méthode d'extraction : Analyse en composantes principales.



Matrice des composantes

	Composante 1
REG1	,814
REG2	,794
REG3	,798
REG4	,802
REG5	,719

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,839	5

Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
REG1	21,91	30,753	,671	,799
REG2	21,70	32,938	,651	,809
REG3	22,10	29,234	,683	,795
REG4	22,52	28,936	,661	,801
REG5	22,63	29,266	,580	,828

10. Exposition_Médias

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
EXP1	4,56	1,560	185
EXP2	4,63	1,627	185
EXP3	5,39	1,486	185
EXP4	4,03	2,193	185

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,737
Test de sphéricité de Bartlett	Khi-carré approx.	150,067
	ddl	6
	Signification	,000

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	2,216	55,403	55,403	2,216	55,403	55,403
2	,754	18,840	74,243			
3	,570	14,256	88,499			
4	,460	11,501	100,000			

Méthode d'extraction : Analyse en composantes principales.

Qualités de représentation

	Initiales	Extraction
EXP1	1,000	,506
EXP2	1,000	,680
EXP3	1,000	,588
EXP4	1,000	,442

Méthode d'extraction : Analyse en composantes principales.

Matrice des composantes

	Composante 1
EXP1	,711
EXP2	,825
EXP3	,767
EXP4	,664

Méthode d'extraction : Analyse en composantes principales.

Statistiques de fiabilité

Alpha de Cronbach	Nombre d'éléments
,710	4

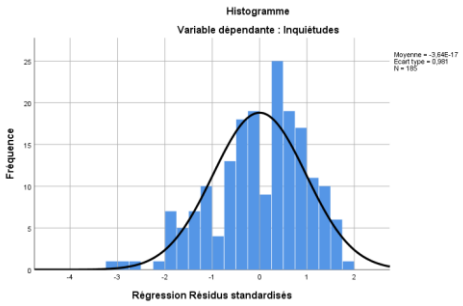
Statistiques de total des éléments

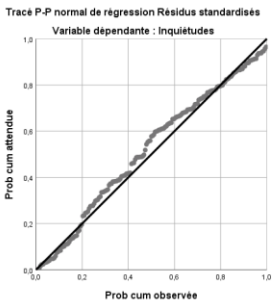
	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
EXP1	14,05	17,470	,459	,670
EXP2	13,98	15,353	,618	,577
EXP3	13,22	17,083	,536	,632
EXP4	14,58	13,951	,435	,716

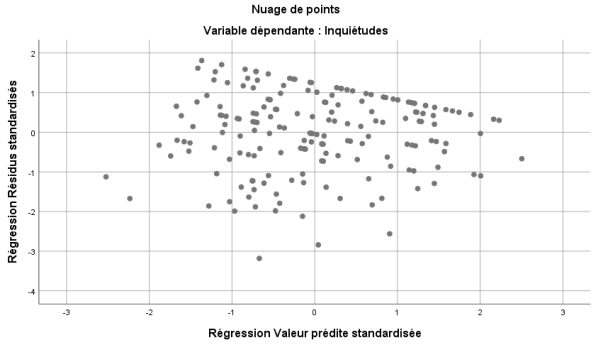
Annexe 6 : Validation des hypothèses

1. Antécédents → Inquiétudes

Régression multiple 1







Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,393[a]	,155	,121	1,17001	2,053

a. Prédicteurs : (Constante), Niveau d'éducation, Implications_BD, Genre, Règles_RGPD, Exposition_Médias, Âge, Avantages_BD

b. Variable dépendante : Inquiétudes

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	44,286	7	6,327	4,622	,000[b]
	de Student	242,298	177	1,369		
	Total	286,584	184			

a. Variable dépendante : Inquiétudes

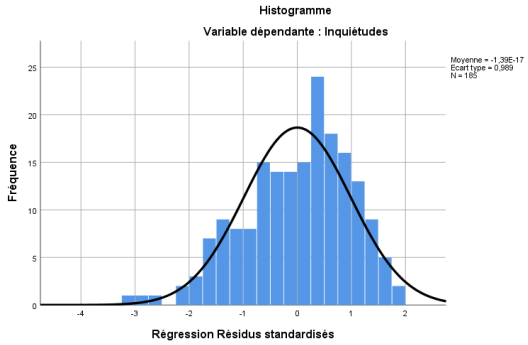
b. Prédicteurs : (Constante), Niveau d'éducation, Implications_BD, Genre, Règles_RGPD, Exposition_Médias, Âge, Avantages_BD

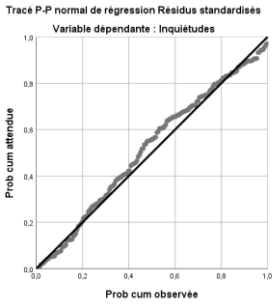
Coefficients[a]

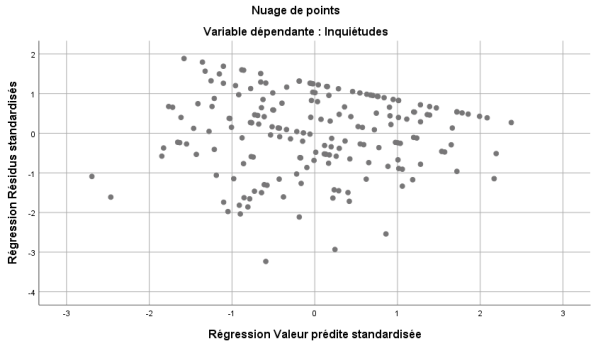
Modèle		Coefficients non standardisés		Coefficients standardisés			Statistiques de colinéarité	
		B	Erreur standard	Bêta	t	Sig.	Tolérance	VIF
1	(Constante)	3,363	,615		5,465	,000		
	Avantages_BD	-,094	,054	-,152	-1,743	,083	,632	1,582
	Implications_BD	,048	,062	,056	,783	,435	,949	1,053
	Exposition_Médias	,189	,076	,193	2,505	,013	,805	1,243
	Règles_RGPD	,173	,067	,187	2,596	,010	,923	1,083
	Âge	,064	,059	,081	1,093	,276	,870	1,150
	Genre	,463	,182	,186	2,549	,012	,897	1,115
	Niveau d'éducation	,036	,137	,022	,264	,792	,698	1,432

a. Variable dépendante : Inquiétudes

Régression multiple 2







Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,382[a]	,146	,127	1,16633	2,055

a. Prédicteurs : (Constante), Genre, Règles_RGPD, Exposition_Médias, Avantages_BD

b. Variable dépendante : Inquiétudes

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	41,724	4	10,431	7,668	,000[b]
	de Student	244,860	180	1,360		
	Total	286,584	184			

a. Variable dépendante : Inquiétudes

b. Prédicteurs : (Constante), Genre, Règles_RGPD, Exposition_Médias, Avantages_BD

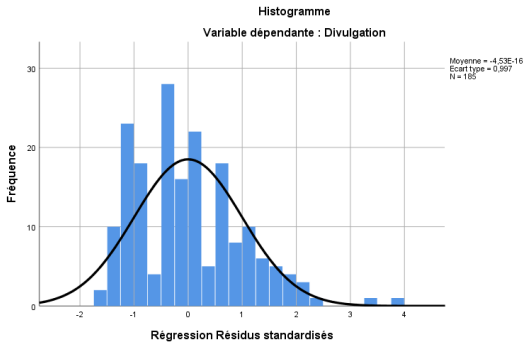
Coefficients[a]

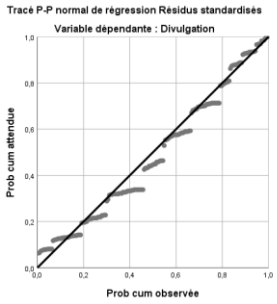
Modèle		Coefficients non standardisés		Coefficients standardisés			Statistiques de colinéarité	
		B	Erreur standard	Bêta	t	Sig.	Tolérance	VIF
1	(Constante)	3,762	,469		8,026	,000		
	Avantages_BD	-,087	,047	-,141	-1,844	,067	,817	1,224
	Exposition_Médias	,209	,074	,213	2,840	,005	,843	1,187
	Règles_RGPD	,182	,065	,196	2,784	,006	,957	1,045
	Genre	,484	,178	,194	2,719	,007	,930	1,076

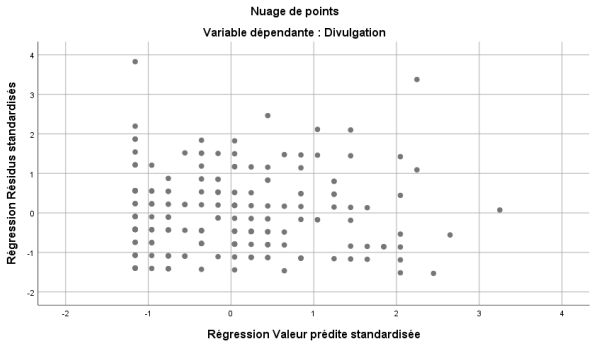
a. Variable dépendante : Inquiétudes

2. Inquiétudes → Conséquences

Hypothèse 7







Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,037[a]	,001	-,004	1,02086	1,958

a. Prédicteurs : (Constante), Inquiétudes

b. Variable dépendante : Divulgation

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	,259	1	,259	,249	,619[b]
	de Student	190,716	183	1,042		
	Total	190,975	184			

a. Variable dépendante : Divulgation

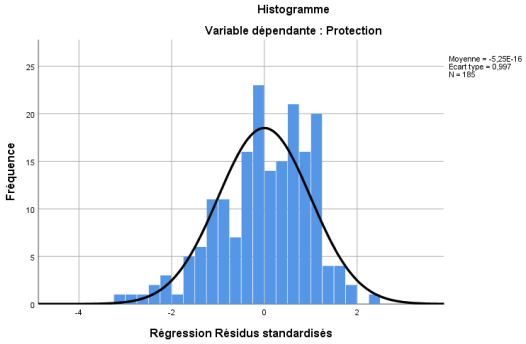
b. Prédicteurs : (Constante), Inquiétudes

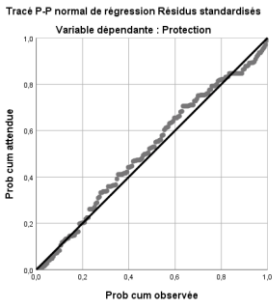
Coefficients[a]

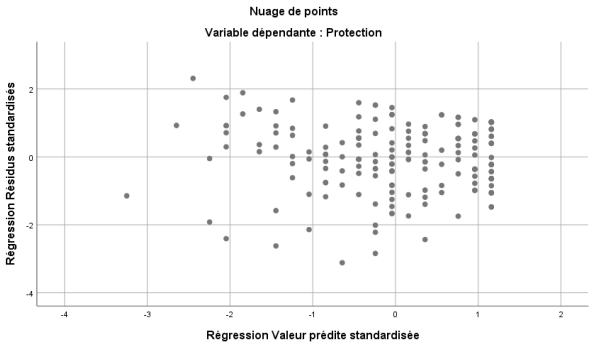
Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	2,637	,343		7,684	,000
	Inquiétudes	-,030	,060	-,037	-,499	,619

a. Variable dépendante : Divulgation

Hypothèse 8







Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,337[a]	,114	,109	,96203	2,105

a. Prédicteurs : (Constante), Inquiétudes

b. Variable dépendante : Protection

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	21,731	1	21,731	23,480	,000[b]
	de Student	169,367	183	,926		
	Total	191,098	184			

a. Variable dépendante : Protection

b. Prédicteurs : (Constante), Inquiétudes

Coefficients[a]

Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	4,089	,323		12,642	,000
	Inquiétudes	,275	,057	,337	4,846	,000

a. Variable dépendante : Protection

LOUVAIN-LA-NEUVE | BRUXELLES | MONS | TOURNAI | CHARLEROI | NAMUR
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/lsm**