

Le phénomène des bulles de filtres sur Internet

Le moteur de recherche Google nous oriente-t-il à notre insu à cause de son
algorithme de personnalisation ?

Mémoire réalisé par
Gilles Guebels

Promoteur
Lionel Groetaers

Année académique 2017-2018
CORP2MS/G – Master [120] en communication, à finalité spécialisée

Tables des matières

Partie I - Introduction	4
1. Avant-propos	4
1.1. <i>Une pertinence orientée ?</i>	9
1.2. <i>Problématique</i>	10
1.3. <i>Apports aux recherches académiques et scientifiques</i>	11
1.4. <i>Méthodologie</i>	13
1.5. <i>Structure mémoire</i>	14
Partie II – Ancrage théorique	16
1. État de l’art des bulles de filtres	16
1.1. <i>Le concept de la bulle de filtres</i>	16
1.2. <i>Des bulles aux formes multiples</i>	17
1.3. <i>Économie de l’attention et capital attentionnel</i>	19
1.4. <i>Synthèse des études empiriques : la sociologie à la rescousse ?</i>	24
1.5. <i>Le fonctionnement du PageRank</i>	33
Partie III – Analyse d’une étude quantitative	39
1. Question de recherche et hypothèse	39
2. Corpus et méthodologie	39
3. Résultats et interprétations	50
3.1. <i>Présentation des données</i>	50
3.2. <i>Traitements des données</i>	51
3.3. <i>Interprétations des résultats</i>	58
Partie IV - Conclusion	61
1. Synthèse des résultats et perspective	61
2. Limites et retour critique sur l’étude	64
3. Mot de conclusion	65
Partie V - Bibliographie	68
1. Monographie :	68
2. Articles scientifiques	68
3. Webographie :	72

Remerciement

En préambule de ce mémoire, je souhaite adresser mes remerciements aux personnes qui m'ont apporté leur aide et qui ont contribué, chacune à leur façon, à l'élaboration de ce travail de fin d'étude.

En premier lieu, je tiens à remercier Monsieur Lionel Groetaers, mon promoteur, pour sa supervision éclairée et sa disponibilité durant ces deux dernières années. Ses conseils avisés, ses directives, son recul critique et son expérience professionnelle m'ont permis de dessiner avec précision les contours de mon mémoire, de résoudre certaines problématiques inhérentes à mon sujet et surtout, de mettre en place une méthodologie qui tienne la route, ce qui n'était pas gagné d'avance. En dehors du strict cadre académique, je tiens également à le remercier pour son soutien et sa patience.

Je tiens également à remercier vivement le professeur Damien Renard, qui m'a apporté une aide inestimable dans l'élaboration de la méthodologie sur Excel et qui m'a prodigué des conseils judicieux dans la manière de valider scientifiquement mes résultats, alors même qu'il n'était lié d'aucune façon à ce mémoire.

J'aimerais également remercier tous les internautes qui ont participé aux entretiens pour leur patience et confiance. Leur contribution m'a permis de récolter des données précises et authentiques, ciment essentiel de ma recherche.

Je souhaite également remercier chaleureusement mes amis et tout particulièrement Loïc, Coralie et Déborah, pour leur amitié, leur soutien, leur bienveillance, leurs conseils et surtout, pour tous les souvenirs inoubliables que nous avons partagés ensemble.

Enfin, ce mémoire ne serait pas ce qu'il est sans l'aide précieuse de mes parents, de mon frère Kévin, de ma belle-sœur Sophie et de ma famille. De près ou de loin, chacun à participer à la réalisation de ce projet. Merci !

Partie I - Introduction

1. Avant-propos

C'est le premier réflexe des internautes du monde entier lorsqu'ils ouvrent leur navigateur Internet pour obtenir un renseignement : se connecter au moteur de recherche Google, quand celui-ci n'est pas déjà configuré comme moteur par défaut. Quelle que soit l'information recherchée sur la toile, le géant américain a la préséance sur tout autre outil. À tel point que la plupart des utilisateurs ne se posent plus la question du service qu'il convient d'utiliser, Google faisant partie de leurs pratiques web les plus enracinées.

Ce constat s'appuie sur des statistiques¹ qui donnent le vertige et font pâlir plus d'un concurrent. Au sein de la bataille stratégique que se livrent les moteurs de recherche, Google atomise la concurrence. Actuellement², le géant américain représente 90,14% de la recherche en ligne au niveau mondial³ (*figure 1*) et possède une position hégémonique sur l'ensemble des continents, sans exception, puisque la majorité des régions du monde affiche en effet une part de marché aux alentours de 90%, y compris en Asie où la Chine et ses deux milliards d'habitants n'ont pas accès aux services de Google suite à la censure imposée par le gouvernement⁴. Au niveau européen, la part de marché moyenne s'élève à 91,63% et la Belgique (*figure 2*), quant à elle, ne déroge pas à la règle : 93,51% des recherches Internet sont effectuées à travers ce moteur⁵. On notera toutefois – non sans une certaine pointe d'ironie – que l'endroit de la planète où le taux d'utilisation du moteur de recherche américain est le plus *faible* est justement ... l'Amérique du Nord, emplacement historique de la firme de la Silicon Valley.

¹ Statistiques compilées par le Stat Counter : <http://gs.statcounter.com/> (consulté le 03/02/2018)

² C'est-à-dire à la date du dernier relevé disponible, à savoir juin 2018.

³ <http://gs.statcounter.com/search-engine-market-share> (consulté le 30/06/2018)

⁴ https://lexpansion.lexpress.fr/high-tech/la-chine-face-a-google_1492142.html (consulté le 04/04/2018)

⁵ <http://gs.statcounter.com/search-engine-market-share/all/belgium> (consulté le 30/06/2018)

À la lumière de ces chiffres, la situation monopolistique du moteur de recherche est indiscutable. Les statistiques de terrain confirment d'ailleurs, s'il y avait encore le moindre doute, le rôle central de Google comme source du trafic Internet. L'étude de Shareaholic⁶ apporte une mise en perspective supplémentaire à ce sujet, démontrant qu'au second semestre 2017, le moteur de recherche était à l'origine de 36,82% du trafic des sites Internet, en augmentation de plus de 20% en un an (*figure 3*). C'est mieux que le mastodonte Facebook, qui ne représentait que 18,16% à la même époque.

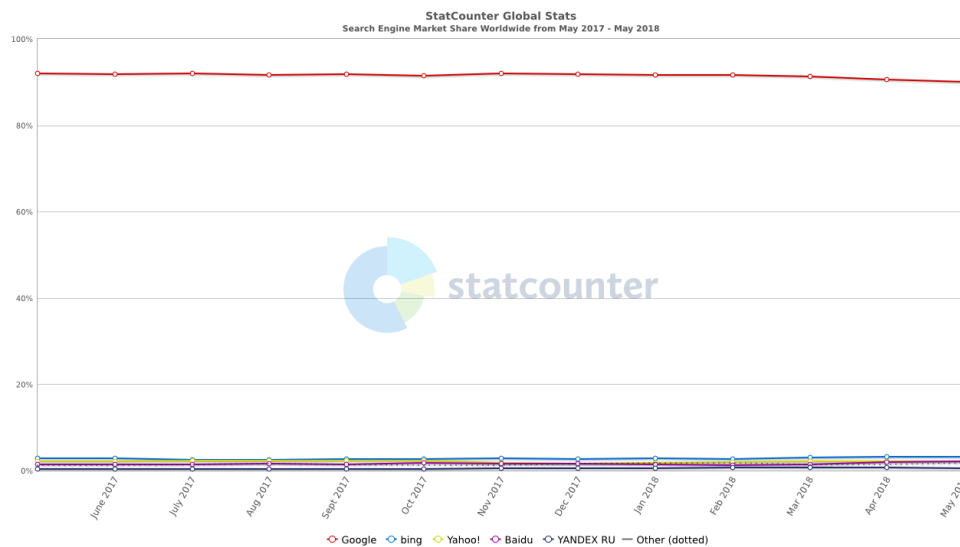


Figure 1 - Part de marché des moteurs de recherche au niveau mondial entre juin 2017 et juin 2018

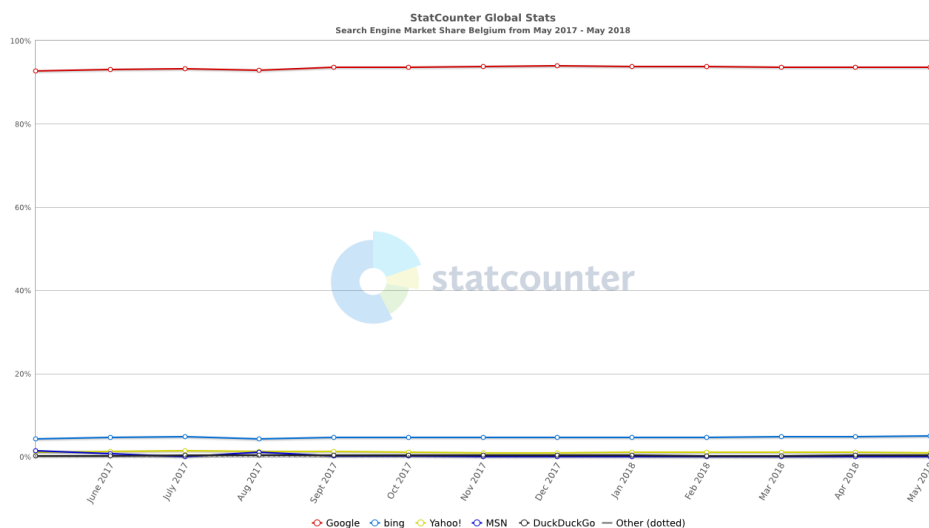


Figure 2 - Part de marché des moteurs de recherche en Belgique entre juin 2017 et juin 2018

⁶ <https://blog.shareaholic.com/search-engine-social-media-traffic-trends-report-2017/> (consulté le 28/02/2018)

Search Engine	1H '16	2H '16	1H '17	2H '17	YoY	
Google	27.91%	29.20%	32.50%	36.82%	20.94%	5.98% PP
Bing	0.66%	0.67%	0.73%	0.82%	16.40%	0.11% PP
Yahoo!	0.15%	0.15%	0.17%	0.19%	17.37%	0.03% PP
DuckDuckGo	0.03%	0.03%	0.04%	0.04%	21.83%	0.01% PP
AOL	0.02%	0.02%	0.02%	0.03%	7.65%	0.002% PP

Figure 3 – Origine du trafic des sites Internet par moteurs de recherche, par trimestre et total à l'année

Depuis sa création en 1998 dans un garage de Menlo Park en Californie, Google s'est transformé en un véritable empire du web, positionnant ses pions au sein de tous les domaines stratégiques d'Internet. Car au-delà de son produit phare, le moteur de recherche, la galaxie Google est composée de services annexes tout aussi puissants⁷ : Gmail (900 millions d'utilisateurs), Android (1 milliard de terminaux équipés), YouTube (1,5 milliard d'utilisateurs actifs), ainsi que Google Maps, Google Drive ou encore les régies publicitaires Adwords, Adsense et Analytics. Grâce à cette myriade de services, Google s'est petit à petit imposé dans la vie numérique des internautes. Que ce soit pour connaître l'horaire d'ouverture de la pizzeria d'en face, consulter ses e-mails, déterminer l'itinéraire le plus court, regarder les vidéos de son *YouTubeur* favori ou s'informer des dernières actualités, ses services sont pratiquement devenus incontournables.

La position de Google est si importante qu'elle est désormais dans le viseur de la Commission européenne. En juin 2017, l'Union européenne a condamné Google pour abus de position dominante à cause de son service Google Shopping, un comparateur de prix, qui aurait été favorisé dans les résultats de recherche par rapport aux comparateurs de prix concurrents. Le groupe est par ailleurs sous le coup de deux autres enquêtes de la Commission : l'une concernant « AdSense », son service publicitaire, l'autre à l'encontre d'Android, le système d'exploitation pour smartphone.⁸

⁷ https://abc.xyz/investor/news/earnings/2016/Q3_alphabet_earnings/ (consulté le 24/01/2018)

⁸ http://www.lemonde.fr/economie/article/2017/06/27/bruxelles-punit-google-d-une-amende-record_5151710_3234.html (consulté le 23/01/2018)

Face à cette omniprésence, d'autres acteurs essayent tant bien que mal d'exister et, parfois, de se distinguer en proposant un modèle de fonctionnement radicalement différent. Si Bing (propriété de Microsoft- 2,7% de part de marché mondial) ou Yahoo (2,2%)⁹ suivent peu ou prou les mêmes méthodes que Google – même si en pratique, nous pourrions le lire, les différences sont notables - certains moteurs de recherche se distinguent par des fonctionnalités uniques ou une hiérarchisation de l'information différente. Ainsi, le français **Qwant** assure respecter la vie privée en ne stockant et en n'utilisant pas les données personnelles des utilisateurs. Les résultats obtenus après une requête ne sont donc pas personnalisés, mais influencés par l'actualité et les tendances. L'américain **DuckDuckGo** propose quant à lui de ne pas enregistrer l'adresse IP de l'internaute et de puiser auprès de 400 sites de références pour proposer des résultats de recherche pertinents. Dans une autre optique encore, **Xaphir** fait le pari de remettre en cause le principe de base du PageRank, le tout puissant et secret algorithme de Google qui sera au cœur de cet objet d'étude. Au lieu de prendre en compte la popularité d'un site web pour juger de sa pertinence, le moteur de recherche applique une égalité dans le classement des pages et part du principe qu'un site populaire n'est pas nécessairement celui qui répondra le mieux à la requête de l'internaute. L'autre innovation de ce petit nouveau réside dans la manière même d'effectuer sa recherche. Xilopix, la société derrière Xaphir, indiquait à cet égard : « *Xhapir permet simultanément et pour la première fois, ce qu'aucun autre moteur n'a réussi à faire à ce jour : combiner des formats hétérogènes qu'il s'agisse de textes, d'images ou de vidéos dans la barre de recherche, affiner sa page de résultats par des itérations successives, en choisissant ou excluant les résultats proposés par un simple clic et explorer visuellement et tactilement pour chercher et trouver de manière intuitive et ludique* ». ¹⁰

⁹ <http://gs.statcounter.com/search-engine-market-share> (consulté le 30/06/2018)

¹⁰ <https://www.lemondeinformatique.fr/actualites/lire-xaphir-dernier-ne-des-moteurs-de-recherche-francais-68161.html> (consulté le 11/12/2017)

Si ces solutions ont le mérite d'offrir une réelle alternative à Google et de répondre à des attentes légitimes, la protection de la vie privée en tête, elles ne parviennent pas à casser la suprématie du géant américain. Leurs statistiques d'utilisation¹¹ sont pourtant loin d'être ridicules : Qwant a généré 1,6 milliard de requêtes l'an dernier et DuckDuckGo, près de 5 milliards. Mais cela reste tout de même à des années-lumière de Google et ses 3,3 milliards de requêtes ... par jour.

L'une des clefs pour comprendre la domination de Google sur le marché des moteurs de recherche réside en grande partie dans la terrible efficacité de son algorithme, le PageRank, à trier les pages web pour ne faire ressortir que celles qui rentrent en concordance avec l'objet de nos requêtes. Le mode de fonctionnement de l'algorithme consiste à mesurer quantitativement la popularité des pages web et à les hiérarchiser selon cette popularité. Dans le brevet destiné à protéger l'algorithme, Larry Page, l'un des cofondateurs de Google, décrit de façon limpide le mécanisme qui le régit : « *Un document doit être important (sans considération pour son contenu) s'il est beaucoup cité par d'autres documents* »¹². Le PageRank considère donc que plus une page est citée par d'autres sur Internet (à l'instar, par exemple, d'un article académique qui serait de nombreuses fois cité par des scientifiques), plus ce document fait preuve d'*autorité* et peut être considéré comme un document de *qualité*. L'algorithme lui accorde alors une place de choix au sein des résultats de recherche, généralement dans les premiers.

Cette explication est volontairement simplifiée à l'extrême, car nous y reviendrons en détails plus tard. Elle permet néanmoins de comprendre techniquement comment Google parvient à trouver si rapidement la réponse à une requête, et surtout à fournir des résultats qui ne nous déçoivent que très rarement. La prouesse est notable : grâce au PageRank, Google est capable de fournir les résultats les plus pertinents parmi les milliards de pages indexées sur le World Wide Web.

¹¹ <http://www.eskimo.fr/saviez-vous-qu'il-existe-des-moteurs-de-recherche-alternatifs-a-google/> (consulté le 03/02/2018)

¹² Page Larry, « *Method for node ranking in a linked database* », Patent 6285999, 1998

Chez Google, cette notion de pertinence est le nerf de la guerre. Car comme ils aiment le répéter, Google ne fournit pas des résultats. Il fournit des résultats *pertinents*. Sur un site dédié au fonctionnement du moteur de recherche, l'entreprise explique que la pertinence est la condition *sine qua non* pour atteindre un haut degré de réussite : « *Larry Page a dit un jour que le moteur de recherche parfait doit comprendre exactement ce que vous voulez et vous donner exactement ce que vous cherchez* »¹³. Et force est de constater que c'est sur base de ce critère qu'ils parviennent à se distinguer nettement de la concurrence, qui éprouve toutes les peines du monde à afficher des résultats en phase avec la demande initiale. Le journaliste Marc Zaffagni expliquait d'ailleurs que l'absence de hiérarchisation chez les concurrents (notamment Xaphir), « *pouvait dérouter* » et que pour utiliser d'autres moteurs de recherche « *il faut apprendre à se défaire du réflexe qui nous fait considérer les résultats placés en tête comme étant les plus pertinents* »¹⁴.

1.1. Une pertinence orientée ?

Des voix se sont rapidement élevées pour dénoncer une supposée influence de la part du géant américain. Car, en devenant le premier moteur de recherche en taux d'utilisation, Google a acquis le statut de premier relayeur de contenu sur Internet et a profondément modifié la façon dont les informations atteignent le public. Lors de son étude sur l'algorithme PageRank, Dominique Cardon expliquait que via « *ses procédures, ses décisions, ses interdits, et tout l'appareil de codification du comportement des internautes qu'il a fabriqué, [...] Google a imposé au web un ordre qui lui est propre* »¹⁵.

D'un point de vue éditorial, le moteur de recherche n'a rien à se reprocher dans la mesure où il ne fait que véhiculer l'information et n'est pas l'auteur du contenu des pages qu'il relaye. Mais à une époque où la visibilité est la clef de voûte de la réussite d'une page Internet, les sites et autres

¹³ https://www.google.com/search/howsearchworks/responses/#?modal_active=none (consulté le 14/07/2017)

¹⁴ <https://www.futura-sciences.com/tech/actualites/internet-xaphir-moteur-recherche-francais-veut-faire-ombre-google-67270/> (consulté le 11/11/2017)

¹⁵ Cardon Dominique, « *Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google* », Réseaux n° 177, p. 65, 2013

documents qui apparaissent dans les tout premiers résultats de recherche font figure de parole d'évangile. De par sa stature de premier moteur de recherche au monde, Google fournirait mécaniquement les meilleurs résultats, qui pourraient être pris comme une vérité incontestée. Une inquiétude qui est partagée par l'entreprise elle-même. En 2005, l'ex-PDG de Google, Eric Schmidt, s'inquiétait de voir le moteur de recherche devenir un vecteur de désinformation et avouait être impuissant face aux fausses informations et au phénomène de diffamation qui apparaît parfois dans les premiers résultats de Google¹⁶. Selon lui, la technique n'est - aujourd'hui - pas capable de juger l'authenticité de l'information présentée.

Dans une veine d'idée plus ou moins similaire, le militant d'Internet Eli Pariser a développé depuis 2011 une théorie selon laquelle Google, mais aussi Facebook et Yahoo, enfermeraient les internautes dans une « **bulle de filtres** ». La *bulle de filtres* est un concept qui désigne à la fois le filtrage arbitraire de l'information opéré par les géants d'Internet et l'état d'isolement, d'enfermement, dans lequel serait plongé l'internaute quand les résultats qu'il obtient après une requête résultent d'une personnalisation mise en place à son insu.

C'est précisément cela qu'explique Eli Pariser dans « The Filter Bubble » :

« Nous pensons tous que quand nous googlons un même mot, tout le monde a les mêmes résultats [...]. Avec la recherche personnalisée, vous obtenez en réalité le résultat que l'algorithme de Google pense être le plus adapté à vous en particulier – mais quelqu'un d'autre verra apparaître d'autres résultats. »¹⁷

1.2. Problématique

C'est précisément ce phénomène qui va nous intéresser, car les enjeux sociétaux et démocratiques semblent évidents. Au fond, si la multitude d'informations qui nous parviennent chaque jour n'est en réalité que le reflet

¹⁶ Interview d'Éric Schmidt, documentaire télévisé « *Defining Google* » diffusée par CBSNews (janvier 2005) (consulté le 12/12/2017)

¹⁷ Pariser Eli, « *The Filter Bubble : What The Internet Is Hiding From You* », Editions Penguin, New-York, 2012

de nos propres idées, nous risquons à terme de nous conforter dans nos propres convictions et croire que le monde pense comme nous. Indéniablement, cela détériore les débats entre citoyens, chacun étant persuadé de posséder les bonnes informations et d'avoir une vision éclairée du monde. Par exemple, un internaute d'une orientation politique X pourrait voir les algorithmes Google totalement supprimer les contenus concernant l'orientation politique Y, rendant le débat d'idée atone. Une phrase de Pariser reflète assez bien cette idée : « *La démocratie requiert du citoyen qu'il voit le problème du point de vue de l'autre, et nous, nous sommes de plus en plus enfermés dans notre bulle* ».

La problématique de ce mémoire se concentrera donc sur les effets des algorithmes de Google sur la personnalisation et la hiérarchisation des informations, et cherchera plus précisément à savoir si ces effets favorisent, ou non, l'effet bulle de filtres. La question de recherche qui sera le fil conducteur de ce mémoire sera donc celle-ci :

« Dans quelle mesure l'algorithme mis en place par Google pour personnaliser et hiérarchiser les contenus sur son moteur de recherche favorisent-ils l'effet « bulle de filtre » ? »

Dans une société de l'information et de l'immédiateté, où le digital prend une place considérable, et dans un écosystème numérique où le moteur de recherche Google dispose d'une force de frappe sans commune mesure avec celle de ses concurrents, le phénomène des bulles de filtres mérite que l'on s'y attarde.

1.3. Apports aux recherches académiques et scientifiques

Médiatisé depuis la sortie du livre d'Eli Pariser, le concept des bulles de filtres a connu un sursaut d'intérêt dans l'opinion publique suite à l'élection présidentielle américaine de 2016. La victoire inattendue de Donald Trump a en effet poussé les journalistes et certains analystes d'Internet à mener des investigations pour comprendre comment un tel résultat n'avait pas pu être prédit. Très rapidement, les réseaux sociaux, Facebook en tête, ont été pointés du doigt. Reprenant l'hypothèse des bulles de filtre, de nombreux quotidiens

ont reproché à Facebook d'avoir apposé des œillères à ses utilisateurs et donc d'avoir influencé le résultat du scrutin. *The New York Magazine* a ainsi affirmé que « *Facebook a permis la victoire de Trump* »¹⁸, tandis que le site spécialisé TechCrunch a accusé le réseau social d'avoir « *amplifié et déformé les idées de ses utilisateurs* »¹⁹ et le Guardian d'avoir « *divisé la narration politique en deux bulles de filtrage distinctes : une pour les conservateurs, et une pour les libéraux* »²⁰. Dans sa défense, le patron de Facebook, Mark Zuckerberg, n'a pas nié l'existence d'une bulle, mais a expliqué que celle-ci n'était pas due au réseau social : « *Le plus grand filtre du système, ce n'est pas que le contenu ne soit pas là, ou que vous n'ayez pas d'amis qui soutiennent l'autre candidat (...), mais que vous le rejetez quand vous le voyez. Nous ne cliquons tout simplement pas dessus [...]* »²¹.

Décriées de toutes parts et cible de nombreuses critiques, les bulles de filtrage ne font pourtant pas l'objet d'un intérêt scientifique important : très peu d'études empiriques ont été menées pour confirmer ou infirmer le bien-fondé de ce concept. Bien entendu, les médias généralistes et spécialisés ont largement évoqué la thématique, mais la plupart des enquêtes se sont limitées à une description basique du phénomène et n'ont pas suivi les principes d'une recherche scientifique, avec un corpus représentatif et une méthodologie rigoureuse. Une absence qui pose d'autant plus question lorsque l'on remarque que les rares recherches effectuées sur le sujet tendent à démontrer que les bulles de filtres n'auraient pas l'impact négatif escompté et que leurs effets seraient plus limités que l'on pourrait penser. C'est tout du moins la conclusion de l'étude menée par Flaxman, Goel et Rao (2016)²², qui s'est appuyée sur les habitudes en ligne de 50.000 utilisateurs américains. En 2015, Bakshy, Messing et Adamic²³ en étaient arrivés à des conclusions similaires

¹⁸ <http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html> (consulté le 29/04/2017)

¹⁹ <https://techcrunch.com/2016/11/09/rigged/?guccounter=1> (consulté le 29/04/2017)

²⁰ <https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories> (consulté le 29/04/2017)

²¹ https://www.lemonde.fr/pixels/article/2016/11/11/accuse-d-avoir-influence-l-election-americaine-facebook-se-defend_5029769_4408996.html (consulté le 05/06/2017)

²² Flaxman S., Goel S., Rao J., « *Filter Bubbles, Echo Chambers, and Online News Consumption* », in *Public Opinion Quarterly*, Vol. 80, Issue S1, p. 298–320, 2016

²³ Bakshy E., Messing S., & Adamic L. A., « *Exposure to ideologically diverse news and opinion on Facebook* », in *Science*, N°348, Issue 6239, p. 1130–1132, 2015

à propos de la polarisation des opinions politiques sur Facebook, ceux-ci se basant sur un corpus extrêmement large : 7 millions articles partagés par les utilisateurs. Mais la crédibilité et l'indépendance de l'étude menée par Bakshy, Messing et Adamic fut rapidement remise en question dans la mesure où les trois auteurs travaillent pour le ... Facebook Labs, le centre de recherche du réseau social. Enfin, une étude de 2012, menée par Messing et Westwood²⁴, suggérait même que les individus étaient plus susceptibles de réagir avec un contenu contraire à leur opinion lorsqu'il est présenté avec des informations sociales.

Enfin, et c'est peut-être le plus important, il faut noter que la littérature scientifique aborde essentiellement la question des bulles de filtres sous le prisme des réseaux sociaux. Or, la présente étude se concentre uniquement sur Google et son moteur de recherche. Et force est de constater qu'il n'existe tout simplement pas de recherches scientifiques sur les bulles de filtres appliquées à Google. L'intérêt de ce mémoire sera donc double : à la fois, enrichir de connaissances supplémentaires le champ de recherche sur les bulles de filtre et l'impact des algorithmes de recommandation ; mais aussi approfondir ce sujet par rapport au premier moteur de recherche au monde, et pas uniquement via les réseaux sociaux.

1.4. Méthodologie

La difficulté principale de cette recherche résidait dans la méthodologie à mettre en place. D'abord, parce que la petite taille de la littérature scientifique sur le sujet n'a pas permis de s'inspirer de méthodes ayant déjà fait leurs preuves. Ensuite, parce qu'il était nécessaire de réaliser des tests grandeur nature pour confirmer ou infirmer l'hypothèse selon laquelle les systèmes de personnalisation proposent des contenus différents selon les internautes. Un tel procédé sous-entendait – comme on le verra plus tard – avoir accès à leurs ordinateurs et à leurs comptes Google.

²⁴ Messing S., Westwood S.J., « *Selective Exposure in the Age of Social Media* », in *Communication Research*, Vol. 41, Issue 8, p. 1042, 2012

En l'absence d'outils préconçus, une méthodologie inédite a donc été mise en place. Dans les grands traits, celle-ci consistait à effectuer une même recherche Google sur les ordinateurs d'internautes ayant *des degrés d'utilisation Google* différents et de comparer ensuite les résultats obtenus avec une recherche dite « *neutre* », c'est-à-dire ayant été effectuée avec un degré d'utilisation égal à zéro, pour déterminer le nombre de résultats différents entre les deux recherches. Dans un second temps, grâce à un tableau croisant le degré d'utilisation de l'internaute et le nombre de résultats différents, il fut possible d'analyser, selon le principe de corrélation bivarié, si un lien pouvait être établi entre le taux d'utilisation de Google et la personnalisation des résultats.

Cette notion de *degré d'utilisation Google* est une notion créée spécifiquement pour ce mémoire. Elle peut être définie comme une échelle ou un score allant de 0 à 35 qui permet de mesurer quantitativement la proportion, l'intensité, avec laquelle un internaute utilise les services de Google. Cette échelle repose sur le principe selon lequel plus le score est élevé, plus l'internaute utilise les services du géant américain, et plus celui-ci est censé disposer d'informations à son égard pour personnaliser les résultats du moteur de recherche.

Ces notions seront amplement développées dans la section réservée à la méthodologie. Mais elles permettent néanmoins d'obtenir un premier aperçu de la manière avec laquelle une problématique relativement abstraite va pouvoir être décryptée scientifiquement.

1.5. Structure mémoire

Après cette introduction, ce mémoire sera divisé en cinq sections distinctes. La première offrira un état de l'art le plus exhaustif possible concernant le fonctionnement des bulles de filtres et de l'algorithme de Google, mais offrira aussi un aperçu détaillé des recherches qui ont été menées sur le sujet, notamment en sociologie, et qui apportent un éclairage indispensable pour comprendre la problématique. La seconde section reviendra en détails sur la question de recherche qui guide ce mémoire ainsi que sur l'hypothèse qui en

découle. La troisième partie se concentrera, elle, sur la méthodologie et le corpus. L'avant-dernière section sera consacrée à la présentation, le traitement et l'interprétation des résultats. Enfin, le présent travail se terminera sur une conclusion en trois étapes : une synthèse finale des résultats, un retour critique sur les limites de l'analyse, et un mot de conclusion.

Partie II – Ancrage théorique

1. État de l’art des bulles de filtres

1.1. *Le concept de la bulle de filtres*

On désigne par bulle de filtres à la fois le phénomène par lequel les internautes se retrouvent confrontés uniquement à des informations correspondantes à leurs opinions, mais aussi la *bulle intellectuelle* qui en résulterait et dans laquelle s’enfermerait les internautes en pensant détenir des sources d’informations variées.

Développé depuis 2011 par l’activiste américain Eli Pariser dans son livre « *The Filter Bubble: What the Internet Is Hiding from You* », le concept de bulle de filtres repose sur l’idée que les algorithmes au cœur des moteurs de recherche et des réseaux sociaux ne trient pas les informations présentes sur Internet de la même manière pour tout le monde, mais tiennent compte des données personnelles et des habitudes de navigation des internautes pour leur offrir des contenus en adéquation avec leurs avis et leurs goûts : c’est la personnalisation.

Les résultats de recherche, ou les contenus qui s’affichent dans le fil d’actualité d’un média social, seraient donc tout sauf neutres et résulteraient d’une personnalisation mise en place à l’insu des internautes. Grâce à une multitude de critères et de données à caractère personnel, les algorithmes opéreraient un filtrage de l’information pour ne faire ressortir que des contenus optimisés pour l’internaute à l’origine de la requête, à l’image de l’allégorie de la caverne de Platon. Pour Eli Pariser, « *nous pensons tous que quand nous googlons un même mot, tout le monde a les mêmes résultats [...]. Avec la recherche personnalisée, vous obtenez en réalité le résultat que l’algorithme de Google pense être le plus adapté à vous en particulier – mais quelqu’un d’autre verra apparaître d’autres résultats.* »²⁵

²⁵ PARISER Eli, « *The Filter Bubble : What The Internet Is Hiding From You* », Editions Penguin, New-York, 2012

Ces systèmes de filtrage de l'information, bien qu'indispensables étant donné le flot incessant de données, créent des effets pervers. En n'ayant accès uniquement aux informations conformes à leurs propres opinions, et pas aux avis contraires puisqu'ils auront été effacés des résultats de recherche, les internautes risquent de se polariser en pensant que le reste du monde partage les mêmes points de vue et qu'ils détiennent la vérité. Autrement dit, l'effacement dans les résultats de recherche de toutes les idées, théories, documents, articles qui pourraient perturber leur vision du monde, risque de les conforter dans leurs propres convictions et réduire leur champ informationnel en leur renvoyant l'extension de leurs propres idées.

Eli Pariser propose un exemple frappant pour illustrer sa théorie : « *si un citoyen n'ayant pas de sympathie pour le mouvement écologiste effectue une recherche Google sur la société pétrolière BP, il risque d'obtenir des liens sur les possibilités d'investissement dans BP, alors qu'un autre obtiendrait probablement des informations sur la marée noire provoquée par l'entreprise* »²⁶.

1.2. Des bulles aux formes multiples

Le phénomène des bulles de filtres apparaît essentiellement sur les réseaux sociaux et sur les moteurs de recherche. Dans les deux cas, les algorithmes ont été développés de manière à ne pas afficher l'entièreté des informations disponibles, mais seulement celles susceptibles de rencontrer les intérêts de l'utilisateur. Mais les critères sur lesquels s'appuie l'algorithme pour personnaliser les contenus ne sont pas comparables. En 2015, le sociologue Dominique Cardon proposait une grille de lecture permettant de distinguer quatre grandes familles d'algorithmes selon la métrique de mesure qu'ils utilisent pour opérer une hiérarchisation de l'information : la popularité, l'autorité, l'affinité et la prévision.²⁷ Ce sont précisément les deux dernières métriques qui sont au cœur des bulles de filtres, respectivement, sur les médias sociaux et les moteurs de recherche. Si ce mémoire se concentre sur

²⁶ <https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427/> (consulté le 12/07/2018)

²⁷ Cardon Dominique, « *À quoi rêvent les algorithmes : nos vies à l'heure des big data* », Éditions Le Seuil, 112p, 2015

l'univers Google, les explications de Dominique Cardon sur le fonctionnement des mécanismes à l'œuvre à l'intérieur du plus grand réseau social au monde sont révélatrices du mode de pensée diamétralement opposé entre les deux géants du Web.

Dans le cas des moteurs de recherche, qu'il s'agisse de Google ou d'un moteur de recommandation intégré à un site, la personnalisation des contenus se base sur la **prévision**. À partir des comportements passés, des habitudes de navigation, ou encore des données enregistrées par les cookies, les algorithmes dits « prédictifs » fournissent un ensemble de contenu qui épouse ce que nous sommes, notre personnalité. Le futur de l'internaute est prédit par le passé de ceux qui lui ressemblent (Cardon, p.34, 2015). C'est grâce à cette méthode que Netflix se targue de savoir ce que vous allez regarder ou Amazon de connaître par avance votre prochain achat.

C'est à travers cette conception de la personnalisation que Eli Pariser a imaginé le concept des bulles de filtres. Nous nous limiterons à cette brève description pour l'instant, et nous reviendrons en profondeur sur les bulles de filtres relatifs aux moteurs de recherche dans la section consacrée au PageRank.

Dans le cas des réseaux sociaux, les systèmes de personnalisation se basent sur un principe radicalement différent : **l'affinité**. Alors que les méthodes traditionnelles de classement de l'information, à l'image du PageRank, se basaient sur la notion d'autorité et de dépersonnalisation pour établir une hiérarchie, les réseaux sociaux « *classent les documents en fonction des jugements subjectifs que s'échangent des personnes liées par une relation d'affinité* »²⁸. Les contenus qui apparaissent sur notre écran lorsqu'on parcourt Facebook ou Twitter sont donc le fruit des liens que chaque individu a tissés avec ses amis. Par exemple, l'EdgeRank, l'algorithme de Facebook, « *hiérarchise les informations en fonction de la proximité relationnelle entre les personnes* ». Dans la pratique, l'algorithme analyse cette proximité vis-à-vis des interactions qui peuvent exister entre deux amis Facebook. Plus deux

²⁸ Cardon Dominique, « *Du lien au like sur Internet. Deux mesures de la réputation* », Communications, 2/2013 (n° 93), p. 180.

amis Facebook sont régulièrement en contact via des publications, des commentaires ou des « likes », plus l'algorithme favorisera la visibilité de leurs publications respectives dans le fil d'actualité de l'autre.²⁹

Cette manière de mesurer l'intérêt pour un document n'est pourtant pas nouvelle. En adoptant une métrique basée sur la proximité relationnelle, les médias sociaux ne font que mettre en application les théories énoncées en sociologie, qui indiquent que « *cette proximité est un bon outil d'approximation des goûts partagés* »³⁰. Toutefois, le véritable tour de force des réseaux sociaux a été cette capacité à mettre en place des outils permettant d'exprimer son appréciation pour un contenu (incarnés à travers les boutons « J'aime » sur Facebook ou « Retweet » sur Twitter) et surtout, d'utiliser les données produites par ces outils pour déterminer ce qui plait et ce qu'il ne plait pas à l'internaute. Mais aussi évident que cela puisse paraître, un tel procédé nécessitait d'élargir « *les droits de l'internaute à participer au classement des documents* », ce qui n'était absolument pas le cas avant l'apparition du web 2.0, dans la mesure où Internet était classé selon le principe, on l'a vu, de l'autorité.

1.3. *Économie de l'attention et capital attentionnel*

Indéniablement, les effets négatifs mis en avant dans la théorie de la bulle de filtres amènent avec eux plusieurs questions inquiétantes. D'un point de vue purement informationnel, l'avènement d'Internet devait initialement ouvrir les portes du savoir en rendant l'information accessible et sans limites, à l'image d'un grand débat public. Les concepteurs et les pionniers du net partageaient l'intime conviction que cette nouvelle technologie, encore naissante, permettrait de s'affranchir des limites sociales, économiques et culturelles inhérentes à la société. Dominique Cardon, reprenant Pierre Lévy, expliquait ainsi qu'au cœur même de la conception du web se situait « *une vision idéalisée d'un monde d'idées dialoguant entre elles dans un rapport d'argumentations et de raison débarrassé du poids des intérêts, de la*

²⁹ Ibidem

³⁰ Ibidem

personnalité ou de la psychologie de ceux qui les ont émises ».³¹ Les pionniers ne cesseront d'ailleurs de critiquer les évolutions successives d'Internet lorsque celles-ci s'éloigneront trop de la philosophie originale de leur invention. Mais vingt-huit ans après l'invention du World Wide Web par Tim Berners-Lee, ces idéaux d'un Internet neutre semblent appartenir à une époque révolue. La théorie des bulles de filtres aurait fait voler en éclat l'idée selon laquelle le web permettrait à tout un chacun de découvrir de nombreux points de vue différents et de faire jouer la *sérendipité*. Ce concept, d'origine anglaise et qui s'est répandu dans les sciences humaines, est l'art de tirer profit de circonstances imprévues (Van Andel & Bourcier, 2008) ou l'art de prêter attention à ce qui surprend et d'en imaginer une interprétation pertinente (Catellin, 2014).

Ce que dénonce Eli Pariser à travers les bulles de filtres, ce sont les pratiques malsaines adoptées par les géants d'Internet. Selon l'activiste américain, les services Internet mis à la disposition du grand public seraient développés dans le seul et unique but de capter l'attention des consommateurs dans un marché numérique en surabondance et ultra-concurrentiel, où ladite attention fait désormais office de valeur marchande.

Cette démarche s'inscrit résolument dans les pratiques de **l'économie de l'attention**. Cette économie, qui peut être décrite comme une économie dont l'attention constituerait la première rareté et la plus précieuse source de valeur (Citton, 2014, p.3), n'est pas nouvelle. En 1969 déjà, Herbert Simon explicitait le concept : « *Dans un monde riche en information, l'abondance d'information entraîne la pénurie d'une autre ressource [...] c'est l'attention de ses receveurs. Donc une abondance d'information crée une rareté de l'attention et le besoin de répartir efficacement cette attention parmi la surabondance des sources d'informations qui peuvent la consommer* »³². Plus récemment, en 2004, l'ex-PDG de TF1 Patrick La Lay avait défrayé la chronique en expliquant, de manière aussi simple que révélatrice, que le

³¹ Cardon Dominique, « *Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google* », Réseaux, n° 177 (« Politique des algorithmes. Les métriques du web »), p. 75, 2013

³² Simon, H. A., « *Designing Organizations for an Information-Rich World* », in Martin Greenberger, Computers, Communication, and the Public Interest, Baltimore, 1971

modèle d'affaire de la chaîne de télévision était basé sur l'économie de l'attention : « *Ce que nous vendons à Coca-Cola, c'est du temps de cerveau humain disponible* »³³.

Cette économie de l'attention est aujourd'hui enracinée dans les pratiques des GAFAs et constitue même la base de leur modèle économique et publicitaire. Cette recherche effrénée de l'attention est la conséquence de la diversification et la multiplication des acteurs sur Internet. Alors qu'elle était auparavant focalisée sur un nombre limité de *produits*, l'attention s'est désormais diluée au travers des milliers de propositions, de stimuli, de tentations qu'offrent les concurrents. Évoluant au sein d'un marché éminemment concurrentiel, les géants d'Internet ont donc rapidement compris que l'attention des internautes constituait la clef de voûte du système et que celui qui parviendrait à la capter aurait toutes les chances de tirer son épingle du jeu. Là où Yves Citton indique que les médias sont de véritables banques de l'attention, « *leurs modèles reposant sur leur capacité à accumuler un capital attentionnel* »³⁴, il est possible de percevoir les Google, Amazon, Facebook, Microsoft et autre Netflix comme des entreprises dont la valeur dépend en grande partie de l'*attention* qu'elles suscitent auprès de l'audience Internet.

Dans ce contexte, les algorithmes de personnalisation ont pour mission première d'attirer toujours plus l'œil des utilisateurs en proposant des éléments susceptibles de les intéresser et de susciter de l'engagement. Ainsi sollicités avec des contenus qu'ils apprécient, comme une carotte devant un âne, les internautes sont disponibles et l'entreprise peut valoriser cette attention – toujours momentanée – en commercialisant des espaces publicitaires. Car c'est bien là que réside tout l'enjeu pour les entreprises : les algorithmes ont avant tout un objectif commercial. En investissant et en développant des systèmes capables d'*analyser* et de *comprendre* les désirs de ceux qui les utilisent, les géants d'Internet parviennent à mieux cibler les publicités selon les centres d'intérêt et à attirer des annonceurs désireux

³³ <https://www.nouvelobs.com/culture/20040710.OBS2633/le-lay-nous-vendons-du-temps-de-cerveau.html> (consulté le 15/04/2018)

³⁴ <http://maisouvaleweb.fr/pour-une-ecologie-de-lattention-lantichambre-du-temps-de-cerveau-disponible-essai-yves-citton/> (consulté le 14/07/2017)

d'investir leur budget publicitaire sur une plateforme leur garantissant une audience intéressée par leurs activités. De ce point de vue, ils ne font que répondre à une logique de marché et d'acquisition de parts d'audience. Comme l'indiquait l'article « *La société des calculs sous la loupe de la sociologie* », le fonctionnement des algorithmes n'est jamais neutre et reflète les choix des individus qui les ont conçus. Ils sont par nature éminemment politiques et relèvent de choix de société³⁵.

Mais en agissant de la sorte, les techniques de personnalisation enfermeraient dans le même temps les internautes dans une bulle cognitive. L'auteur met en garde contre l'isolement que ces bulles de filtres peuvent engendrer et le trompe-œil dangereux qu'elles représentent. En ne percevant sur Internet que des informations qui vont dans son sens, l'internaute risque à terme de s'enfermer dans une *bulle cognitive*³⁶ en parfaite concordance avec sa personnalité, et dans laquelle il pourrait se complaire, puisqu'il évoluerait dans une réalité alternative où les opinions contraires aux siennes n'existent tout simplement pas. Eli Pariser constatait ainsi avec effroi que son profil Facebook s'était transformé en un univers lisse et sans relief. En ne commentant pas, en likant pas et en ne réagissant pas aux contenus de ses amis conservateurs, le réseau social avait pris la décision de les supprimer de son champ de vision. Or, s'il n'adhérait pas à leurs opinions, il ne souhaitait pas pour autant être privé de ces points de vue contraires.

Frédéric Joignot, dans un papier du Monde, expliquait que « *nous vivons dans un cocon à notre convenance, où plus personne ni aucune nouvelle désagréable ne nous dérange [...] nous faisons circuler les messages, les posts, les vidéos et les hashtags sans les discuter* »³⁷. En 2001, le philosophe et juriste américain Cass Sunstein s'inquiétait déjà des effets de filtre des algorithmes. Celui-ci expliquait que si Internet offrait bel et bien les conditions pour être un eldorado où se confronterait des opinions diverses et contraires, puisque le web possède cette faculté de faire cohabiter tous les

³⁵ <http://maisouvaleweb.fr/la-societe-des-calculs-sous-la-loupe-de-la-sociologie-a-quoi-revent-les-algorithmes-dominique-cardon-lecture/> (consulté le 14/07/2017)

³⁶ Il s'agit du terme exact utilisé par Eli Pariser.

³⁷ <http://fredericjoignot.blog.lemonde.fr/2016/08/30/comment-les-algorithmes-nous-enferment-dans-une-bulle-intellectuelle/> (consulté le 12/07/2018)

points de vue sur d'innombrables plateformes³⁸, les filtres coupent le citoyen d'une nécessaire diversité au maintien de la vie en société.³⁹ En pensant le futur comme une simple reproduction du passé, les algorithmes favoriseraient donc le conformisme, distordraient la et ne prendraient pas en considération le besoin d'avoir accès aux opinions contraires dans la construction d'un débat démocratique sain.

Si cette problématique est revenue sur le devant de la scène avec Internet, elle n'est pourtant pas nouvelle dans le champ des médias et de la communication. Elle évoque le cas des chambres d'écho, cette métaphore définissant « *une situation dans laquelle l'information, les idées ou les croyances sont amplifiées ou renforcées par la communication et la répétition dans un système défini* »⁴⁰. Ainsi répétées de manière continue, les croyances sont prises pour argent comptant et les idées déviantes sont n'ont pas lieu d'être. Dans le cas d'Internet, des chambres d'écho se produisent parce que les algorithmes de personnalisation ne renvoient que des informations qui font *écho* aux opinions de l'internaute. Un tel scénario est évidemment préjudiciable pour l'esprit critique, la réalité étant déformée par les points de vue qui nous parviennent et dont nous pensons que ce sont les seuls qui sont légitimes, mais qui en réalité, ne représente qu'une partie très étroite des perspectives possibles⁴¹.

Le phénomène est d'autant plus vicieux qu'il est invisible. Lors d'une recherche sur Google ou Facebook, rien n'indique à l'utilisateur qu'un puissant outil est en train de procéder à une filtration des contenus et qu'une grande majorité de ceux-ci seront rejetés en dehors de la bulle. Dans sa navigation quotidienne, l'internaute ne se rend donc pas nécessairement compte qu'il est en train de passer à côté d'un tas important de données que les moteurs de recherche auront décidé de passer sous silence.

³⁸ Ibidem

³⁹ Sunstein Cass, « *Republic.com* », Princeton University Press, 224 p, 2001

⁴⁰ [https://fr.wikipedia.org/wiki/Chambre_d%27%C3%A9cho_\(m%C3%A9dias\)](https://fr.wikipedia.org/wiki/Chambre_d%27%C3%A9cho_(m%C3%A9dias)) (consulté le 23/04/2018)

⁴¹ Parrish, Shane, "*How Filter Bubbles Distort Reality: Everything You Need to Know*", Farnam Street, 2017

La préoccupation envers les bulles de filtres est d'autant plus légitime lorsqu'on constate qu'environ la moitié⁴² des internautes accèdent désormais à l'information à travers les réseaux sociaux (Facebook en tête avec 44%), et que 36% des individus sont d'accord de se fier aux suggestions personnalisées des systèmes de recommandation pour consommer l'actualité.

1.4. Synthèse des études empiriques : la sociologie à la rescousse ?

Une grande partie de la presse et quelques spécialistes – on l'a vu ci-dessus - présument qu'Internet exacerbe la fragmentation des citoyens. Pourtant, alors que les thèses autour des bulles de filtres n'ont jamais eu autant de popularité, plusieurs recherches empiriques tendent à démontrer que le phénomène aurait moins d'impact qu'initialement prévu et que la polarisation des opinions sur le web, initialement imputé aux bulles de filtre, pourrait davantage s'expliquer au travers de concepts sociologiques et psychologiques.

L'une des premières théories à avoir abondé dans ce sens est celle de la dissonance cognitive exposée par Leon Festinger en 1957, et qui tente de comprendre comment les individus parviennent à gérer les situations où ils se retrouvent face à des éléments incompatibles. Selon le psychosociologue américain, la dissonance cognitive peut être définie comme la tension psychologique que ressent une personne « *lorsqu'elle est confrontée à une information, une pensée, un comportement, qui entre en contradiction avec ses propres idées ou croyances* ». ⁴³ Au cours de sa recherche, Festinger mit en lumière différents types de dissonance qui peuvent apparaître chez les individus, mais l'une d'entre elles résonne particulièrement avec l'effet présumé des bulles de filtres : l'exposition sélective. Selon ce paradigme, « *les individus ont tendance à rechercher des informations consistantes avec leurs cognitions et éviter les informations inconsistantes* » ⁴⁴. Au moment de la confrontation avec une information incompatible, les individus se

⁴² https://www.lemonde.fr/actualite-medias/article/2016/06/15/les-reseaux-sociaux-prennent-une-place-croissante-dans-l-acces-a-l-information_4950771_3236.html (consulté le 03/09/2017)

⁴³ https://fr.wikipedia.org/wiki/Dissonance_cognitive (consulté le 29/11/2017)

⁴⁴ Bardin B., Perrissol S., Fos Y., Py J., Dagot L., « *Exposition sélective aux informations de prévention et attitude implicite envers le tabac* », dans Psychologie Française, Volume 62, Issue 1, p.1, 2017

convaincraient du bien-fondé de leurs propres croyances en rejetant ladite information et en recherchant le soutien d'autres personnes qui partagent les mêmes avis (Harmon-Jones, p.101, 2002).

Cette théorie a été mise en application dans de nombreux domaines. Ainsi, des politologues américains avaient découvert que les partisans démocrates refusaient de s'exposer aux messages qui présentaient le point de vue des républicains⁴⁵. Dans le même ordre d'idées, les êtres humains auraient également recours au principe de l'exposition sélective quand ils choisissent leurs amis, en ayant tendance à se regrouper avec ceux qui ont des attitudes et des opinions semblables aux leurs.⁴⁶ A la lumière de cette théorie, les bulles de filtre ne seraient alors qu'une transposition d'un phénomène humain et social sur les sites web et réseaux sociaux numériques. C'est notamment ce que déclarait Dominique Cardon dans une interview au quotidien Le Monde⁴⁷ : « *La bulle, c'est nous qui la créons, par un mécanisme typique de reproduction sociale. Le vrai filtre, c'est le choix de nos amis, plus que l'algorithme de Facebook* ».

Cette hypothèse de « transposition » entre en résonnance avec un autre travail de Leon Festinger, celui sur la théorie des processus de comparaison sociale. Cette recherche fondatrice en psychologie indique que les êtres humains tendent à comparer leurs propres opinions avec celles des autres afin de voir s'ils sont semblables ou différents. Ce n'est rien d'autre que cela qu'indique Festinger en disant qu'« *en l'absence de moyens objectifs non sociaux, on évalue ses opinions et ses aptitudes en les comparant avec les opinions et les aptitudes des autres* »⁴⁸. Ainsi, les internautes auraient une tendance naturelle à comparer tous les contenus qui apparaissent sur le moteur de recherche ou le fil d'actualité pour déterminer s'ils sont en adéquation avec leurs opinions.

⁴⁵ John V. Petrof, « Comportement du consommateur et marketing », Presses Université Laval, p.225, 1993

⁴⁶ Ibidem

⁴⁷ https://www.lemonde.fr/pixels/article/2016/11/14/facebook-faiseur-de-rois-de-l-election-americaine_5030959_4408996.html (consulté le 12/05/2017)

⁴⁸ Festinger Leon, « *A Theory of Social Comparison Processes* », Human Relations, vol. 7, n°2, 1954, p. 117-140 (consulté en ligne le 04/06/2017 : <https://goo.gl/qYaGWo> et sa version traduite : <https://goo.gl/bMVF1s>)

Un phénomène qui rappelle celui de l'homophilie sociale : les individus ont tendance à se connecter, à créer des liens, à établir des interactions relationnelles, avec des personnes qui sont socialement proches d'eux. Autrement dit, on choisit de devenir amis avec ceux qui nous ressemblent le plus.

D'autres biais cognitifs ont été avancés pour nuancer l'effet bulle de filtres. C'est notamment le cas du **biais de confirmation**. En psychologie, ce biais consiste à accorder un poids et une légitimité supplémentaire aux informations qui rentrent en concordance avec ses idées ou ses hypothèses, y compris si celles-ci sont fausses, et en corolaire, à considérer les informations jouant en défaveur de ses conceptions comme peu pertinentes. Ce biais aurait un impact considérable dans la recherche et la sélection de l'information, car les individus auraient tendance à ne lire que les sources qui confirment leur position actuelle et à les analyser de manière à ne tirer parti que des éléments qui alimentent leur thèse, y compris lorsque des preuves contraires sont présentes. En ne rassemblant et en ne se souvenant que des contenus qui renforcent leurs attentes, via ce que l'on appelle communément la **mémoire sélective**, les personnes contribueraient à renforcer leurs croyances personnelles et à adopter un comportement biaisé vis-à-vis de la réalité. De ce point de vue, le biais de confirmation serait davantage la cause des bulles de filtres que sa conséquence et nuancerait fortement l'impact de ces bulles sur Internet.

Plus de cinquante ans après les travaux de Festinger, deux chercheurs de l'Université de Stanford, Messing et Westwood, se sont demandés si la théorie de l'exposition sélective s'appliquait encore à l'âge des réseaux sociaux. Alors que les études menées à ce sujet ont tendance à conclure qu'Internet accrut la polarisation des opinions, Messing et Westwood ont pourtant mis en évidence que l'utilisation généralisée des médias sociaux modifiait la consommation d'informations. Sans remettre en cause les travaux sur la dissonance cognitive, les auteurs indiquent toutefois que les réseaux sociaux n'auraient pas plus d'effets de polarisation ou d'exposition sélective que les autres médias, et seraient même parfois plus sains. En effet, ceux-ci

offrent aux lecteurs une multitude de sources différentes pour s'informer, dans un contexte qui met l'accent sur la valeur sociale plutôt que sur l'affiliation partisane (Messing & Westwood, p.1, 2012). Et dans deux expériences distinctes, les auteurs démontrent que les individus seraient plus susceptibles de s'engager avec un contenu contraire à leurs propres opinions quand celui-ci est présenté dans un contexte social, comme c'est le cas sur les réseaux sociaux.

En mai 2015, le laboratoire de recherche de Facebook a publié une étude dans la revue Science afin de contrecarrer les critiques selon lesquelles Facebook serait lieu par excellence pour observer le phénomène des bulles de filtres. En s'appuyant sur un corpus d'environ 10 millions d'utilisateurs américains de Facebook ayant indiqué leur orientation politique sur leur profil (conservateur Vs. libéral), Bakshy, Adamic et Messing ont souhaité observer les effets de l'algorithme EdgeRank sur la sélection des informations politiquement opposées dans le fil d'actualité des internautes.

Premièrement, il en est ressorti qu'une certaine diversité était présente dès le départ puisqu'en moyenne, 23% des amis Facebook des personnes sondées font partie du camp politique opposé.

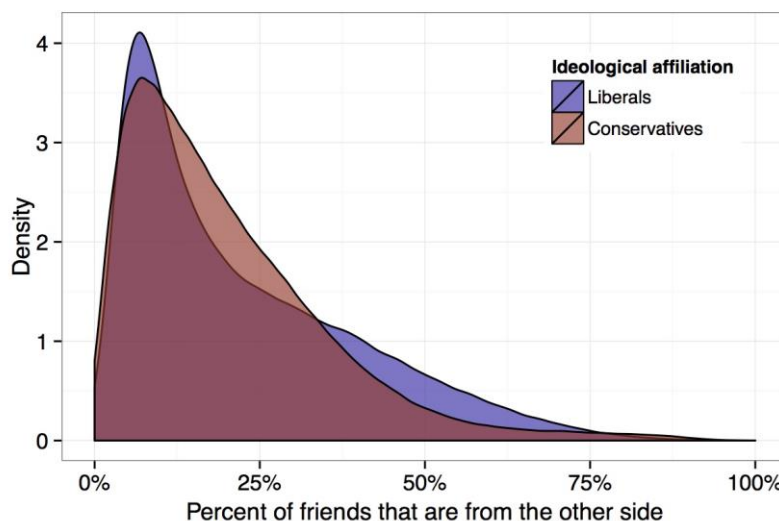


Figure 4 - Pourcentage d'amis qui sont de l'autre bord politique – Bakshy, Adamic, Messing (2015)

Deuxièmement, les auteurs indiquent que le fil d'actualité de Facebook montre bel et bien l'ensemble des contenus possibles. Cependant, l'algorithme considère que les individus n'ont tout simplement pas le temps

de tout passer en revue et trie donc les informations afin de montrer en premier les contenus les plus pertinents pour eux. Dans ce contexte, les tests montrent que 23% des contenus partagés par les libéraux concernent une information opposée à leur point de vue, tandis que ce chiffre grimpe à 34% pour les conservateurs. Mais une fois que l'algorithme a fait son œuvre, ces contenus opposés ne représentent plus que, respectivement, 22% et 33% des contenus visibles dans le fil d'actualité. Pour Facebook, cela signifie que l'EdgeRank crée bien un effet bulle de filtre, mais que celui est marginal, en ne diminuant que de 5 à 8% l'exposition aux informations inconsistantes.

Le fil d'actualité reflèterait donc plutôt fidèlement la diversité des opinions. En montrant que les gens sont exposés à une quantité substantielle de contenu provenant d'amis ayant des points de vue opposés (Bashky, Adamic, Messing, 2015), les résultats contredisent les préoccupations selon lesquelles les individus ne parleraient uniquement aux personnes partageant les mêmes idées. Les auteurs ajoutent que ce sont les choix individuels, à travers « *la composition de nos réseaux sociaux* », et le choix de nos amis, qui restreignent le plus l'exposition aux contenus opposés. Autrement dit, « le classement des informations a un impact moindre sur la diversité » que le principe social de l'homophilie.

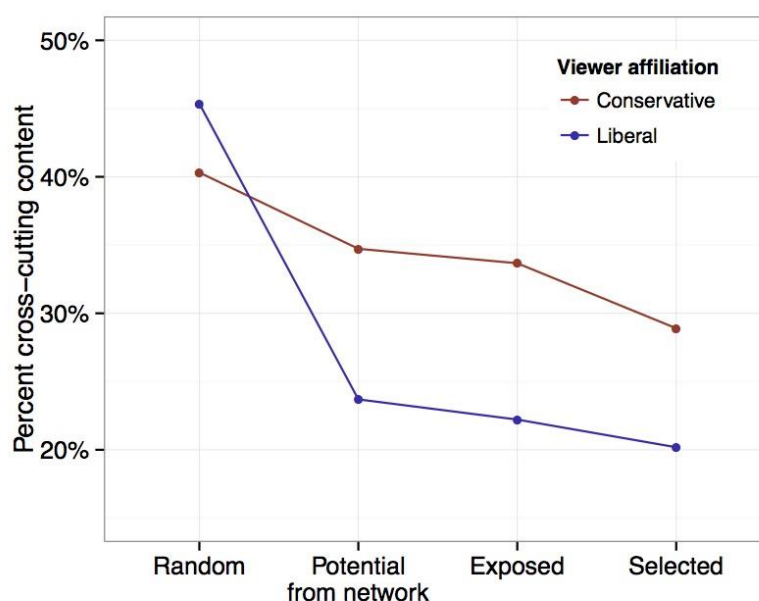


Figure 5 - Le pourcentage de contenu diversifié, partagé par hasard (random), partagé par les amis (potential from network), qui apparaissent effectivement dans le fil d'actualité (exposed), sur lequel on a cliqué (selected) - Bashky, Adamic, Messing (2015)

Il faut toutefois reconnaître que l'étude a été vivement critiquée. D'abord, pour son manque flagrant d'indépendance, les trois scientifiques qui en sont à l'origine travaillant précisément pour Facebook. Ensuite, parce Eli Pariser lui-même s'est montré médusé face aux résultats et à l'interprétation de ceux-ci par Facebook⁴⁹. Selon lui, une modification de 5 à 8% des contenus ne peut pas être considérée comme « *marginal* » et 23% de publications *opposées* n'est franchement pas élevé. Pariser s'étonne également que l'étude ne concerne que 9% des usagers de Facebook – ceux qui ont indiqué une orientation politique – et se demande si les mêmes conclusions pourraient être tirées en prenant en considération tous les autres usagers. Enfin, et c'est le point le plus inquiétant, Facebook indique que seuls 7% des articles partagés sur le réseau peuvent être considérés comme des *hard news*, c'est-à-dire des actualités sérieuses. Pour Eli Pariser, c'est bien trop peu, dans la mesure où Internet devrait constituer une agora propice au débat d'idées.

L'enquête du journaliste Pablo Barbera tend également à prouver que l'effet des bulles de filtres ne serait pas si important que prévu. Selon lui, les médias sociaux encourageraient les relations entre des personnes ayant des liens dits faibles, c'est-à-dire des personnes plus éloignées de notre cercle d'amis habituel, comme des collègues, des personnes rencontrées en voyage ou un parfait inconnu découvert sur Twitter. Pour Barbera, ces personnes ont tendance à être plus hétérogènes que notre cercle de proche, ce qui nous exposerait à davantage de diversité.

Pour prouver ses dires, le journaliste a appliqué une méthode pour mesurer les positions idéologiques de millions d'individus en Allemagne, en Espagne et aux USA. Les résultats de son panel montrent que la plupart des internautes sont baignés sur un environnement diversifié et que contrairement aux idées reçues, l'utilisation de Facebook et Cie. réduirait la polarisation politique et que l'effet bulle de filtres serait bien plus complexe qu'il n'y paraît. Pablo Barbera ajoute : « *Contrairement à un nombre croissant de travaux qui suggèrent qu'Internet fonctionne comme une «chambre d'écho», où les citoyens sont principalement exposés à des opinions politiques similaires,*

⁴⁹ <https://www.wired.com/2015/05/did-facebooks-big-study-kill-my-filter-bubble-thesis/> (consulté le 23/06/2018)

mes résultats démontrent que la plupart des utilisateurs de médias sociaux reçoivent des informations diversifiées [...] »⁵⁰.

Enfin, la dernière enquête scientifique immanquable lorsqu'on aborde la question des bulles de filtres est celle conduite par l'Université d'Oxford. Seth Flaxman, Sharad Goel et Justin Rao se sont basés sur les habitudes de lecture en ligne de 50.000 utilisateurs pour déterminer l'existence, ou non, du phénomène popularisé par Eli Pariser. Selon les conclusions de l'étude, les réseaux sociaux augmenteraient de façon sensible l'effet de polarisation idéologique, notamment lorsqu'il est question d'articles dits d'opinion. Les articles trouvés sur les réseaux sociaux ou via un moteur de recherche seraient en effet associé à une plus grande ségrégation idéologique que ceux trouvés directement sur un site d'actualité. Mais l'étude met également en évidence que ces canaux numériques ont tendance à exposer davantage aux perspectives opposées que les médias traditionnels. Par ailleurs, les auteurs indiquent aussi que 94% des internautes s'informent uniquement auprès de deux sources et que la grande majorité de la consommation d'information en ligne ne fait qu'imiter les habitudes de lecture hors ligne traditionnelles, les personnes se rendant directement sur les pages d'accueil de leurs médias préférés pour renforcer leurs croyances.

Ils en viennent donc à la conclusion que si l'effet bulle de filtre est avéré, les algorithmes ne contribueraient que marginalement à renforcer ces dernières dans la mesure où le comportement habituel des internautes fait que ceux-ci vivent déjà dans une bulle de filtre.⁵¹

Par rigueur intellectuelle, nous pourrions également citer deux autres recherches qui parviennent plus ou moins aux mêmes conclusions : l'étude conjointe entre l'Université de Brown et de Stanford, qui obtient « *des chiffres qui vont à l'encontre de l'hypothèse selon laquelle Internet serait l'un des principaux responsables de l'augmentation de la polarisation*

⁵⁰ http://www.niemanlab.org/2014/10/this-study-finds-that-social-media-use-reduces-political-polarization-instead-of-increasing-it/?utm_content=buffer04916&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer (consulté le 20/11/2017)

⁵¹ Flaxman S., Goel S., Rao J.M., « *Filter Bubbles, Echo Chambers, and Online News Consumption* », in *Public Opinion Quarterly*, Vol. 80, Issue S1, p. 298–320, 2016.

politique»⁵² ; et l'enquête de Gentzkow & Shapiro, qui rapporte que les gens étaient plus enclins à rencontrer des opinions divergentes sur les médias en ligne que dans leurs interactions de tous les jours⁵³.

Nous l'avons vu : certains auteurs soutiennent que les bulles de filtres mènent à des « chambres d'écho » et à des « bulles de filtration », tandis que d'autres se veulent plus optimistes en indiquant que les réseaux sociaux et les moteurs de recherche offrent un accès aux opinions divergentes. Interrogé par Slate sur les conclusions de l'étude de Gentzkow et Shapiro, qui démontent son hypothèse, Eli Pariser indique que leur travail s'arrête en 2012, trois ans seulement après l'introduction du bouton « Like » sur Facebook et le début de la personnalisation des recherches sur Google : « *Nous n'avons pas encore atteint le jour où les réseaux sociaux constitueront notre première manière de consommer les choses, explique Pariser. Mais je pense qu'en attendant ce jour, l'avenir nous réserve encore quelques surprises étranges et inquiétantes* »⁵⁴

Dans le même article, Slate a également demandé à Jesse Shapiro, l'un des auteurs de l'étude précédemment citée, d'expliquer comment il pouvait expliquer ce phénomène de polarisation si les réseaux sociaux ne sont pas aussi responsables que nous le pensions : « *Je pense que ça s'explique en partie parce que les médias traditionnels restent très importants. Et en partie aussi parce que, si Internet permet en effet d'avoir accès à des informations très tranchées, elles ne sont peut-être pas aussi tranchées que les gens se l'imaginent. Cette notion d'un public qui ne lirait que des sources extrêmes sans s'intéresser aux sources "mainstream" ne colle pas avec les données que nous avons consultées en 2011. Cela ne veut pas dire que le risque n'est pas là. Juste que ce n'est peut-être pas un facteur aussi puissant qu'il semble l'être* »⁵⁵

⁵² Boxell L., Gentzkow M., Shapiro J., « *Is the internet causing political polarization ? Evidence from demographics* », Stanford University, Brown University, 25p., 2017

⁵³ Gentzkow M., Shapiro J., « *Ideological Segregation Online and Offline* » The Quarterly Journal of Economics, Volume 126, Issue 4, Pages 1799–1839, 2011

⁵⁴ <http://www.slate.fr/story/142910/bulle-filtres-polarisation> (consulté le 13/11/2017)

⁵⁵ Ibidem.

De cet état de l'art, nous pouvons donc en conclure que :

- Les acteurs de la consommation médiatique en ligne tiennent indéniablement compte de nos préférences personnelles pour nous fournir leurs services. Facebook et Google poursuivent un modèle économique essentiellement basé sur l'économie de l'attention et cela les oblige à mettre en place des mécanismes permettant de capter cette attention, comme les algorithmes de personnalisation. Cela relève d'un choix politique et volontaire, qui reflète une certaine vision d'Internet et dont les conséquences sur la polarisation sont réelles. Malgré qu'Internet nous permettent d'avoir accès à un catalogue extrêmement varié d'idées, les auteurs s'accordent tous à dire que la bulle de filtre est un phénomène qui existe bel et bien et qu'une consommation médiatique unilatérale peut en aggraver les effets.
- Cependant, n'oublions pas que l'homme a depuis toujours créé des "bulles de filtres : nous nous entourons de personnes qui partagent les mêmes opinions, nos proches disposent des mêmes centres d'intérêt, nous nous regroupons à travers associations ou partis qui défendent les combats qui nous importent et nous choisissons des amis qui nous ressemblent. Ainsi, les bulles de filtres sur Internet ne seraient peut-être qu'une transposition sur le web d'une réalité sociale observée depuis le début des recherches en psychosociologie. En naviguant sur Internet, les internautes ne feraient donc que reproduire un comportement proprement humain qu'ils adoptent naturellement dans la vie quotidienne.

Le passage en revue des précédents travaux scientifiques sur le sujet, nécessaire pour obtenir des propos nuancés et un recul critique, a surtout mis en lumière que le phénomène des bulles de filtres est bien plus complexe que l'on pourrait le penser. Il confirme par ailleurs le besoin d'approfondir ce champ de recherche par une nouvelle étude empirique. Si une grande majorité des enquêtes se sont attardées sur le cas des réseaux sociaux, le présent mémoire se concentrera uniquement sur le moteur de recherche Google, et ce, pour toutes les raisons évoquées en introduction.

À noter que la recherche sera précédée par une explication synthétique, mais détaillée du fonctionnement du PageRank, l'algorithme de Google. S'il n'est pas nécessaire de rentrer dans les détails techniques – ils sont de toute manière, on va le voir, cachés – pour saisir l'étude qui va suivre, un minimum de connaissance est la bienvenue pour comprendre l'efficacité redoutable de Google, pour apprécier les subtilités de la recherche, ainsi que certaines terminologies propres au géant américain.

1.5. Le fonctionnement du PageRank

Le fonctionnement des mécanismes à l'œuvre au sein des algorithmes de Google est relativement complexe. Dans une longue enquête publiée dans la revue *Réseaux*, Dominique Cardon, spécialiste du domaine, explique de manière précise les procédures de l'algorithme, son mode de fonctionnement et les limites d'un tel modèle. Cardon détaille ainsi que « *l'intuition fondatrice dont Google est l'enfant est de considérer le lien entrant (et non sortant, comme pensé précédemment) comme support de toute opération visant à classer les sites de l'Internet* »⁵⁶. Pour déterminer quelle(s) page(s) mérite(nt) de se retrouver tout en haut des résultats de recherche, Google perçoit ces fameux liens hypertextes comme étant un « **vote** » d'une page pour une autre. En effet, plus une page Internet reçoit des liens hypertextes d'autres sites web, « *des votes* » selon les deux fondateurs, plus une page Internet peut être considérée comme importante.

Dans un article fondateur de Jon Kleinberg sur le sujet, celui-ci expliquait que « *le créateur de la page p, en incluant un lien vers la page q, a dans une certaine mesure, conféré de l'autorité à q* ». Autrement dit, comme l'expliquait Larry Page, « *de façon intuitive, un document [une page] doit être important s'il est beaucoup cité par d'autres documents* ». Si on reprend la métaphore du vote, pour connaître l'importance d'une page Internet, il suffit de « compter » le nombre de liens reçus par d'autres, comme on dépouillerait les votes lors d'une élection. C'est cela que Google explique en

⁵⁶ Cardon Dominique, « *Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google* », *Réseaux* n° 177, p.4, 2013

disant que « *tout lien pointant de la page A à la page B est considéré comme un vote de A en faveur de la page B* ».

Bien entendu, on peut légitimement se demander en quoi un lien hypertexte est une source fiable pour déterminer l'importance d'une page. Mais Kleinberg répond à cette problématique en détaillant que le « *lien hypertexte est à la fois un acte de reconnaissance et un signe d'autorité. « Les liens hypertextes », écrit-il, « encodent une somme considérable de jugements humains latents et nous prétendons que c'est exactement ce type de jugement qui est requis pour formuler la notion d'autorité.* »

Par ailleurs, pour déterminer l'importance d'une page sur le World Wide Web, Google ne se limite pas au principe du vote, il « *procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par Google ont plus de « poids », et contribuent ainsi à « élire » d'autres pages* ». Cette technique, nommée *influence weight*⁵⁷, propose de considérer que tous les liens n'ont pas le même poids et qu'il faut leur appliquer un « *attribut récursif permettant de calculer l'autorité du citeur dans le réseau en fonction du nombre de citations qu'il a lui-même reçues des autres* ». Ce principe permet d'éviter que certains sites web reçoivent beaucoup d'attention pour la seule raison qu'ils publient beaucoup.

Si on reprend une nouvelle fois la métaphore du vote, cela reviendrait à dire que l'on se trouve dans une « *méritocratie censitaire, qui n'accorde pas le même pouvoir à chaque vote* ». Comme le précise Cardon, une égalité des votes viendrait installer un principe de popularité et non d'autorité. Pour déterminer dès lors la prééminence de certains liens, Google se base sur la métaphore de la recommandation sociale : « *Ce que doit désormais faire l'algorithme n'est pas de faire voter une communauté d'égaux préalablement sélectionnés, mais de dégager parmi tous, ceux qui sont reconnus par les autres comme des experts pour que leur vote ait davantage de poids* ». En réalité, Google considère que les internautes sont égaux. « *Mais leurs pages ne le sont pas et il fait de cette séparation entre la personne et la page une*

⁵⁷ Selon la terminologie utilisée par les fondateurs de ce mécanisme, Gabriel Pinski et Francis Narin, en 1976.

manière de préserver le principe d'autorité lorsque le droit de publier est ouvert à tous ». Et c'est là que réside l'une des spécificités du système : l'autorité qui est mesurée n'accorde aucune n'espère d'importance aux statuts des individus qui sont à l'origine du contenu. Pour conserver une nécessaire neutralité, Google dépersonnalise son algorithme.

1.5.1. Et concrètement ?

Concrètement, l'algorithme opère en trois phases. La première étape consiste à explorer le web afin de repérer les nouvelles pages et les ajouter à la liste des pages connues. Ce processus, qui s'appelle l'**exploration**, se réalise lorsque les robots d'exploration de Google découvrent un nouveau lien sur une page ou lorsque l'administrateur d'un site Internet transmet aux équipes de Google une liste des différentes pages de son site.

La seconde étape est celle de l'**indexation**. Après avoir découvert une nouvelle page, les robots tentent de déterminer le sujet dont elle parle en analysant le texte, les photos, les vidéos ou toutes autres informations susceptibles d'apporter des informations pertinentes sur la page. Ces informations sont collectées, triées et rangées dans l'index Google, une base de données répartie sur une multitude de serveurs.

Enfin, la dernière étape consiste à proposer à l'internaute un ensemble de résultats pertinents lorsqu'il effectue une requête : c'est la **diffusion**. Durant cette phase, deux aspects rentrent en jeu : l'importance et la pertinence. Pour cette dernière, Google explique se baser sur *« de nombreux critères pour essayer de déterminer la réponse la plus appropriée à partir de notre index. Nous tenons compte d'autres éléments, tels que la zone géographique de l'internaute, sa langue ou son appareil (ordinateur ou téléphone), pour optimiser l'expérience utilisateur et fournir les réponses les plus pertinentes »*.⁵⁸ Pour l'importance, le PageRank essaye de déterminer les pages, parmi les plus pertinentes, qui font autorité et celles qui ne font pas autorité en établissant une cote de 1 à 10 sur échelle *« logarithmique qui mesure le nombre de liens reçus par la page venant d'autres pages en*

⁵⁸ <https://support.google.com/webmasters/answer/70897?hl=fr> (consulté le 23/03/2018)

considérant que les sites s'envoient les uns aux autres une force ». Plus la note obtenue sera élevée, plus la page web sera considérée comme faisant autorité et plus celle-ci apparaîtra en haut des résultats de recherche.

Dans ce domaine de la pertinence, Gérald Bronner expliquait ainsi que les informations qui nous parviennent sur Google se basent sur 57 critères, parmi lesquels notre historique de recherche, notre localisation, le type d'ordinateur que nous utilisons, notre langue, etc (Bronner, p.59, 2013). Nous y reviendrons en détails dans la méthodologie.

1.5.2. Google est-il un gatekeeper ?

S'il s'agit *grosso modo* de ce que l'on peut dire du PageRank sans entrer dans des explications techniques et mathématiques qui n'auraient que peu d'intérêts pour cette recherche, force est de constater que Google met un point d'orgue à maintenir secret les détails du fonctionnement de son moteur de recherche. Bernhard Rieder expliquait à cet égard que cette conception de l'autorité (le pilier central du fonctionnement de PageRank) avait au moins une conséquence : elle invite à dissimuler la métrique qui somme les liens afin qu'elle ne fasse pas l'objet d'agissements stratégiques ». ⁵⁹ Dans son article, Dominique Cardon ne disait pas le contraire en expliquant que « *la présupposition de neutralité implique-t-elle une revendication complémentaire d'invisibilité, afin de ne pas troubler le monde dont elle capture les agissements* ».

En effet, l'invisibilité et la neutralité sont les deux mamelles du moteur de recherche. Google ne cesse de répéter que pour neutraliser les aléas du jugement humain, « *il est préférable de faire confiance aux algorithmes et de s'en tenir là* » (Cardon, p.82, 2013). Dans son principe, le PageRank ne fait que prendre en considération les jugements que les internautes se sont portés les uns les autres à travers leurs liens. Mais en adoptant une telle position, la firme de Mountain View refuse d'être tenue pour responsable des résultats de

⁵⁹ Rieder Bernhard, « *What Is in Pagerank ? A Historical and Conceptual Investigation of a Recursive Status Index* », Computational Culture. A Journal of software studies, n° 2, 28 septembre 2012

la recherche et considère que ceux-ci ne sont que le reflet de ce que la communauté d'Internet considère comme important. L'utilisation d'un algorithme informatique déplace sa responsabilité : c'est le PageRank qui procure les résultats, pas Google. *« Ce souci de déléguer à une règle de calcul la responsabilité des classements préserve l'entreprise d'avoir à se justifier contre les mille et une accusations de partialité dont elle fait l'objet »* expliquera Cardon.

Pourtant, Google va rapidement se voir obligé d'intervenir. Malgré les multiples recommandations du géant américain, notamment celle selon laquelle il ne faut pas agir en tenant compte des algorithmes, les internautes vont faire preuve de malice pour tromper l'algorithme et se faire une place au soleil dans les résultats organiques, notamment en créant de faux sites liés entre eux pour se donner du poids mutuellement. Pour Google, ces pratiques remettent en cause le principe même de son fonctionnement : dès lors que les liens présents sur les sites Internet ont été produits de façon à tromper l'algorithme et à accaparer une autorité factice, *« ils apportent une information biaisée qui érode la pertinence du résultat global de la recherche »*⁶⁰

Ne pouvant accepter que les résultats soient biaisés, Google n'a pas d'autres solutions que celle d'intervenir pour corriger les failles, d'une manière ou d'une autre, perdant au passage le statut de neutralité qui le caractérisait jusque-là. Aujourd'hui, à travers des révisions toujours plus nombreuses du PageRank, Google tente de distinguer de mieux en mieux les liens qui transportent réellement de l'importance et ceux qui en transportent moins, et punit fermement les sites Internet qui ne respectent pas les règles édictées pour les webmasters en déclassant tout simplement le site en question. Google se retrouve donc dans une situation paradoxale : à la fois dénombrer les liens sans porter attention à leur contenu, mais en intervenant tout de même pour

⁶⁰ Cardon Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », Réseaux n° 177, p.89, 2013

distinguer vrais et faux liens. « *En devenant législateur et policier des écritures du web, Google ne cesse de perdre sa position d'extériorité* »⁶¹

Le dernier paradigme mis en avant pour critiquer le classement opéré par Google est celui de l'inégalité. Selon Pandurangan et al., 90% du PageRank du web est possédé par 10% des sites. Un très petit nombre de pages attire un nombre considérable de liens, pendant que la très grande majorité des sites sont liés à très peu de sites et ne sont souvent cités par aucun (Broder et al., 2000 ; Adamic et Huberman, 2001). Cela s'expliquerait par le fait que les grands acteurs du web sont en réalité ceux qui disposent déjà d'une centralité accrue dehors du web, comme les entreprises, les médias ou les institutions. En conséquence, la visibilité qu'ils acquièrent sur le réseau « *attire mécaniquement à eux de nouveaux liens surnuméraires. Les noeuds qui reçoivent le plus de liens exercent un effet de halo qui conduit les autres nœuds à rechercher activement leur proximité pour emprunter un peu de leur force ; phénomène qui conduit à donner une autorité imméritée à certains [...] tout en laissant dans l'ombre de nombreux méritants* »⁶². La distribution des liens sur Internet serait donc fortement inégale et répondrait à d'autres logiques, comme la centralité que les acteurs possèdent dans la vie sociale et leur capital économique (Diaz, 2005).

Enfin, nous concluons cette partie sur une réflexion proposée par Cardon par rapport à l'impact des systèmes de personnalisation – et donc des bulles de filtres – sur le PageRank : « *Au principe d'autorité qui a fait la force du PageRank, Google substitue de plus en plus un principe d'efficacité qui renvoie de manière toujours plus appropriée vers l'internaute les choix que l'algorithme a appris de ses comportements. De fait, la machine inventée par Google est devenue si complexe, si sensible aux tests statistiques qui ne cessent de la reparamétrer, si dévoreuse de variables et de traces, si auto-apprenante, que ses comportements ne peuvent plus désormais être compris et interprétés, pas même par ses géniteurs* ».

⁶¹ Ibidem.

⁶² Ibidem.

Partie III – Analyse d’une étude quantitative

1. Question de recherche et hypothèse

Après avoir établi un état de l’art le plus exhaustif possible sur le phénomène des bulles de filtres, il est temps de passer à la recherche à proprement dite. Pour rappel, la question de recherche qui guide l’ensemble de ce mémoire est la suivante :

« Dans quelle mesure l’algorithme mis en place par Google pour personnaliser hiérarchiser les contenus sur son moteur de recherche favorisent-ils l’effet « bulle de filtre ? »

De cette question de recherche découle une hypothèse, dont l’objectif sera de la confirmer ou de l’infirmar à l’issue de l’enquête :

- **« L’algorithme de personnalisation et hiérarchisation de l’information présente sur le moteur de recherche de Google créent une bulle informationnelle qui renforcent de manière substantielle les opinions et la vision du monde des internautes qui l’utilisent »**

2. Corpus et méthodologie

Comme évoqué en introduction, la difficulté majeure de cette recherche résidait dans la méthodologie à mettre en place. D’abord parce que, comme on l’a vu, les critères de fonctionnement des algorithmes sont volontairement cachés. Difficile donc de se baser sur des éléments tangibles ou ayant fait leurs preuves pour fonder la méthodologie.

Ensuite, parce que les données ne sont pas facilement accessibles. Les recherches empiriques menées sur le sujet disposaient pour la plupart d’un accès à une masse colossale de données (comme les chercheurs de Facebook qui pouvaient récupérer des données ... Facebook) et possédaient les technologies informatiques leur permettant d’analyser et de traiter les

résultats obtenus. Mais à son échelle, ce mémoire n'a pas pu bénéficier de tels avantages. Les outils à notre disposition étant donc pratiquement inexistant, une telle situation nous a poussé à mettre en place une méthodologie tout à fait inédite.

L'enquête a pris la forme d'une étude quantitative et s'est déroulée en deux temps. La première étape consistait à effectuer des recherches Google identiques sur plusieurs ordinateurs différents, appartenant à des internautes présentant *des degrés d'utilisation Google* différents. Dans un second temps, il était question de comparer les résultats obtenus avec une *recherche dite « neutre »*, c'est-à-dire ayant été effectuée avec un degré d'utilisation égal à zéro, pour déterminer le *nombre de liens différents* entre les deux recherches et ainsi savoir si une corrélation peut être établie entre le nombre d'hyperliens différents et le degré avec lequel on utilise Google.

Cette méthodologie repose sur trois concepts clefs créés spécifiquement pour ce mémoire : celui de *degré d'utilisation Google*, celui de *recherche neutre* et celui de *nombres de liens différents*.

Le *degré d'utilisation Google*, aussi nommé « Score Google dans la suite du travail », est une échelle graduée de 0 à 35 qui permet de mesurer quantitativement l'intensité avec laquelle un internaute est lié à Google. Selon un système de cotation, en deux parties, basé sur 29 critères, il est ainsi possible de déterminer si un utilisateur est fortement dépendant à Google ou non. Cette échelle repose sur le principe selon lequel plus un individu utilise les services de Google, plus celui-ci dispose d'informations à caractère personnel pour personnaliser les résultats de recherche et favoriser un effet bulle de filtres. Ainsi, plus le degré d'utilisation calculé sera élevé, plus cela signifiera que l'utilisateur a recours aux services du géant américain, et plus celui-ci est censé disposer d'informations à son égard pour faire fonctionner son algorithme.

La *recherche neutre* est la recherche qui sert d'étalon de comparaison. Afin d'observer une éventuelle personnalisation dans les résultats Google, il est nécessaire de comparer les résultats obtenus sur les ordinateurs des

internauts avec exactement la même recherche, mais à travers laquelle l'algorithme n'a pas pu faire son œuvre. La recherche neutre est donc une recherche effectuée sur un ordinateur n'ayant jamais été utilisé, avec un navigateur n'ayant jamais servi et qui n'est pas connecté à un compte Google. En principe, les recherches effectuées dans ce cadre doivent être neutres puisqu'elles ne peuvent pas avoir été influencées par Google.

Enfin, le ***nombre de liens différents*** se réfère quant à lui simplement au nombre de liens différents entre la recherche neutre et la recherche effectuée sur un autre ordinateur. Le chiffre obtenu offre un aperçu quantitatif de l'importance avec laquelle les deux recherches diffèrent.

De manière précise, la méthodologie s'est décomposée en cinq étapes :

1. Établir le ***degré d'utilisation Google (Score Google)*** des internautes :

Lors des tests réalisés auprès des internautes, la première étape consiste à établir leur profil Google et à calculer un score d'utilisation grâce à un questionnaire basé sur une grille d'analyse en deux parties⁶³.

La première partie de la grille, intitulée « **Service Google** », concerne les services proposés Google et cherche à savoir quels services l'internaute utilise régulièrement. En effet, plus l'utilisateur est actif sur ces différents services, plus Google dispose d'informations personnelles pour personnaliser les recherches et les annonces. Ce sont les dix services les plus célèbres du géant américain qui ont été retenus :

- **Le moteur de recherche Google** : bien entendu, comme l'étude porte précisément sur son utilisation, les internautes sélectionnés l'utilisaient forcément.
- **Gmail** : le service de messagerie électronique de Google, qui est connu pour espionner les correspondances afin d'obtenir des renseignements sur les internautes.⁶⁴

⁶³ Sous format Excel, par facilité dans la suite du traitement. Voir annexes.

⁶⁴ <https://www.nouvelobs.com/rue89/rue89-explicateur/20150929.RUE0753/la-loi-numerique-ne-veut-pas-que-google-lise-vos-mails-et-les-espions.html> (consulté le 26/07/2018)

- **Un compte YouTube** : la plateforme de partage de vidéos de Google, dont les historiques peuvent être enregistrés pour proposer des contenus similaires par la suite.
- **Les services de Cloud** : Google Drive, Slide, Cloud.
- **Google Map** : le service de localisation et d'itinéraire, qui, si activé, profite de la localisation de l'individu pour mieux cibler les contenus
- **Google Chrome** : le navigateur maison de Google
- **Blogger** : la plateforme permettant de créer son blog en ligne.
- **Google Livres** : le site offrant un accès à un catalogue numérisé d'ouvrages.
- **Google Photos** : le service photos de Google (anciennement Picasa)
- **L'application mobile Google** : la déclinaison pour smartphone des services Google, et dont l'utilisation est enregistrée si l'option n'a pas été désactivée.

Afin de transformer ses données qualitatives en données quantitatives, il a été convenu arbitrairement que chaque item coché dans la grille équivaldrait à un point. Pour cette partie du tableau, le score minimum est de **0** (s'il la personne n'utilise aucun service), et le score maximum est de **10** (s'il utilise l'entièreté des services). À noter que pour être considéré comme un « service utilisé régulièrement », on entend une utilisation d'au moins 1 fois par semaine.

Dans le même ordre d'idée, l'objectif de la deuxième partie de la grille d'analyse concerne le « **Score de confidentialité** » et cherche à déterminer dans quelle mesure l'internaute à paramétrer ses options Google de confidentialité et de partages de données⁶⁵, puisque ces paramétrages influent sur la quantité de données personnelles que Google peut collecter et donc sur la personnalisation des résultats. Dans ce cadre, il fut demandé à la personne interrogée de se rendre dans les paramètres de son compte Google et de vérifier, pour chacune des options des six catégories (voir tableau ci-dessus) si le paramètre est activé ou désactivé. Dans les cas où les paramètres sont

⁶⁵ Ces options de paramétrages sont accessibles via le profil personnel que chaque utilisateur Google possède : <https://myaccount.google.com/privacy>

des champs textuels à remplir (comme pour les données personnelles), on considéra le paramètre comme activé si le champ est rempli et vice-versa. Ces options, triées par catégories, sont les suivantes :

- **Vos informations personnelles** : nom et prénom, adresse e-mail, téléphone, anniversaire, sexe, recommandations partagées⁶⁶ et partage de position en direct.
- **Paramètre de recherche** : SafeSearch (option permettant de bloquer les images inappropriées ou explicites dans les résultats de recherche) et les résultats privés (options permettant d'afficher des résultats issus d'autres produits Google).
- **Commandes relatives à l'activité** : ces options concernent l'enregistrement des activités que nous menons sur Google. Cela comprend les activités sur le web et les applications, l'historique des positions, les informations provenant des appareils (via l'application mobile Google), l'activité vocale et audio, l'historique des recherches YouTube et l'historique des vidéos regardées sur YouTube. Dans notre grille d'analyse, ces options valent deux points, car ce sont celles qui influent le plus la personnalisation des résultats. Google le dit lui-même, *« les informations enregistrées dans votre compte nous permettent de vous offrir une expérience plus personnalisée des services Google. Mais aussi d'accélérer les recherches, améliorer les recommandations et personnaliser l'expérience utilisateur sur les services Google tels que la recherche ou Map. Dans votre compte Google, choisissez les paramètres pour lesquels les données sont enregistrées »*⁶⁷.
- **Examiner l'activité** : les trajets dans Google Maps, l'historique des positions, l'adresse du domicile renseigné et l'adresse du lieu de travail renseigné.

⁶⁶ Google définit cette option comme suit : « Grâce aux recommandations partagées, vous pouvez recevoir de précieuses recommandations de la part d'autres personnes ».

⁶⁷ <https://myaccount.google.com/activitycontrols> (consulté le 13/05/2018)

- **Paramètre des annonces :** personnalisation des annonces (« Définissez vos préférences de sorte que les annonces qui vous sont proposées présentent un réel intérêt pour vous »).

À l'instar de la première partie de la grille, il a été convenu arbitrairement que chaque item coché dans la grille équivaudrait à un point, sauf pour les options de la partie « *Commandes relatives à l'activité* » qui équivalent à deux points, car il s'agit de celle avec le plus d'influence sur la personnalisation. Pour cette partie du tableau, le score minimum est de **0** (si tous les paramètres sont désactivés et si aucune donnée personnelle n'est spécifiée), et le score maximum est de **25** (si toutes les options sont activées et si toutes les données personnelles spécifiées).

1.1 Calcul du degré d'utilisation final :

Une fois les deux parties de la grille remplies, il suffit alors de simplement additionner les deux scores pour obtenir le **degré d'utilisation de Google/Score Google** de l'internaute. Comme le premier questionnaire est sur 10 et le second sur 24, le score maximum est de 35.

Score Google - 10 points	
Service Google	Service utilisé ? (1pt ou 0pt)
Utilisateur du moteur de recherche	
Compte Gmail	
Compte YouTube	
Google Drive, Slide, Cloud	
Google Map	
Google Chrome	
Blogger	
Google Livres	
Google Photo	
Application mobile Google	
Score total :	

Figure 6 - Première partie de la grille d'analyse

Score de confidentialité - 25 points	
Etat des paramètres de confidentialité	Paramètre activé ou non
Données et informations personnelles - 1pt/item	
Nom et prénom	
Adresse e-mail	
Téléphone	
Anniversaire	
Sexe	
Recommandations partagées	
Partage de position	
Paramètre de recherche - 1pt/item	
Safe Search	
Résultats privés	
Commandes relatives à l'activité - 2pt/item	
Activité sur le Web et les applications	
Informations provenant des appareils	
Activité vocale et audio	
Historique des recherches YouTube	
Historique des vidéos regardées sur YouTube	
Examiner l'activité - 1pt/item	
Vos trajets dans Google Maps	
L'historique des positions	
Adresse du domicile renseigné	
Adresse du lieu de travail renseigné	
Paramètre des annonces - 2pt/item	
Personnalisation des annonces	
Score total :	

Figure 7 - Deuxième partie de la grille d'analyse

2. Création des recherches neutres

Afin de disposer d'un point de référence, d'un gabarit, qui n'a pas été influencé par l'algorithme de Google et avec lequel on peut comparer les résultats de recherche obtenus auprès des internautes sélectionnés, des recherches neutres ont été réalisées. Pour rappel, la recherche neutre nécessite d'effectuer exactement les mêmes requêtes, sur les mêmes thématiques et les mêmes mots-clefs, que les internautes, mais via un ordinateur et un moteur de recherche où aucun service ou compte Google n'est utilisé. Autrement dit, effectuer les recherches dans une situation où le degré d'utilisation Google est égal à 0.

Par ailleurs, il est primordial de noter qu'une recherche neutre a été effectuée chaque jour où une enquête était menée avec un internaute afin prendre en considération le **facteur temps**. En effet, le temps est, dans ce mémoire, un **biais potentiellement problématique**. Une même recherche, menée à quelques jours ou quelques semaines d'intervalle, peut donner des résultats totalement différents, notamment si les thématiques utilisées pour la recherche rentrent en résonance avec un événement dans l'actualité. Dès lors, pour éviter tout biais relatif à la temporalité, les résultats de recherche obtenus sur les ordinateurs des internautes ont été systématiquement comparés avec une recherche neutre effectuée le même jour.

3. Réaliser les tests grandeur nature :

L'avant dernière étape de la méthodologie est celle des tests à proprement dit auprès des internautes. Lors de chaque entretien, il fut demandé aux personnes de se connecter à leur compte Google (s'il en possédait un) et d'effectuer douze requêtes sur Google.be. Les thématiques et les mots-clefs pour les requêtes étaient les suivants :

➤ **La peine de mort :**

- Peine de mort
- Efficacité de la peine de mort
- Peine de mort arguments
- Faut-il rétablir la peine de mort

➤ **Crise financière 2008 :**

- Crise financière 2008
- Explication crise financière 2008
- Responsable crise financière 2008
- Crise financière conséquences

➤ **Révolutions arabes :**

- Révolutions arabes
- Printemps arabe
- Les causes du printemps arabe
- Les révolutions arabes résumé

Ces thèmes et ces mots-clefs n'ont pas été choisis par hasard. Pour pouvoir observer les effets de l'algorithme de personnalisation, il était nécessaire d'avoir recours à sujets clivants ou qui peuvent faire l'objet d'un positionnement idéologique. Le choix d'une thématique neutre (par exemple, la météo) n'aurait probablement pas généré beaucoup de résultats personnalisés.

Les mots-clefs s'inscrivent par ailleurs dans une stratégie visant à réduire au maximum le biais temporel. Il s'agit en effet de thématiques, soit universelles (la peine de mort), soit faisant référence à un ou des événements terminés (la crise financière de 2008 et les révolutions arabes). Dès lors, les probabilités qu'un fait d'actualité lié à ces sujets survienne et influence les résultats de recherches sont fortement réduites. Au cas où cela surviendrait tout de même, le biais est contrebalancé par les recherches neutres créées au même moment et prenant en compte ces changements.

Par ailleurs, il a été décidé de mener les requêtes avec quatre variantes par thématiques, mêlant courte et longue traine. Le but était double : à la fois d'obtenir un corpus relativement conséquent, et à la fois prendre en compte les variations potentielles qui peuvent survenir entre des recherches menées à l'aide de petits mots-clefs ou d'une phrase complète.

D'un point de vue technique, la comparaison entre les recherches et les liens ne pouvait s'opérer que grâce à une solution informatique et semi-

automatisée, compte tenu de la quantité de données à traiter. Ainsi, après chaque requête, les URL des pages apparaissant dans les résultats ont été extraites grâce à un plug-in installé au sein du navigateur de l'internaute. Il s'agit de **Link Klipper** sur Google Chrome, et de **Text Link** sur Firefox. Ces extensions permettent de récupérer, sous format Excel, la liste des URL des pages qui sont apparues dans les résultats de recherche⁶⁸. Comme une page Google renvoie en moyenne 10 résultats, et que douze recherches ont été effectuées par internaute, le fichier Excel final de chaque internaute comprenait plus ou moins 120 hyperliens⁶⁹. Et ce sont précisément grâce à ces URL que l'on a pu comparer des résultats entre eux.

Le choix des URL comme méthode de comparaison résulte d'une astuce informatique. En effet, il faut savoir qu'au sein d'une page de résultats Google, les URL sont l'unique variable qui reste identique, quel que soit l'ordinateur ou le navigateur. À l'inverse, les titres, par exemple, peuvent être différents selon le nombre de caractères affichés par l'écran. Pour s'assurer de pouvoir comparer des éléments analogues, les URL représentaient la solution idéale.

La comparaison en elle-même s'est déroulée sur Excel. Dans la mesure où nous étions face à une quantité très grande de données (+- 120 hyperliens par fichier), il aurait été chronophage de comparer manuellement les hyperliens entre deux fichiers. Mais grâce à la fonction **RECHERCHEV**, Excel offre la possibilité de comparer deux feuilles d'un classeur et d'indiquer si un élément présent dans la feuille A est aussi présent dans la feuille B. Chaque fichier récupéré après un entretien (feuille B) a ainsi été *copié/collé* dans le fichier de la recherche neutre correspondante (feuille A) pour déterminer le nombre de liens différents. Une fois la fonction appliquée au fichier, il en résulte un résultat positif ou négatif pour chaque URL : si le lien réapparaît dans la colonne de comparaison, c'est qu'il est également présent dans la liste des URL de l'internaute. Par contre, si la mention **N/A** s'inscrit, cela signifie que

⁶⁸ En réalité, ces programmes récupèrent également les titres et les meta descriptions des pages web. Pour ne conserver que les URL, les fichiers Excel ont été « nettoyés » grâce à la fonction « Remplacer » qui permet de supprimer les éléments qui ne nous intéressent pas.

⁶⁹ Les termes liens, URL et hyperliens sont synonymes dans le cas de ce mémoire.

l'URL n'est pas présente dans la liste de l'internaute, mais uniquement dans celle de la recherche neutre⁷⁰.

Il suffit alors de compter le nombre de N/A pour connaître le nombre de liens différents entre les deux recherches. Plus le pourcentage de liens *différents* est haut, plus cela signifie que les recherches varient sensiblement.

À ce stade, un fichier Excel fut créé pour centraliser l'ensemble des données précédemment récoltées. Ce fichier comprend :

- Une colonne avec le listing des internautes
- Une colonne avec le score cumulé de l'internaute, nommé « **Score Google** » (ou degré d'utilisation).
- Une colonne avec le nombre de liens différents (calculé selon la méthode expliquée précédemment), ainsi que cette valeur transformée en pourcentage
- Trois colonnes qui détaillent le nombre de liens différents par thématiques (par rapport au nombre total de liens récoltés pour cette thématique). Dans le tableau, PDN = Peine de mort ; CF2008 = Crise financière 2008 ; RA = Révolutions arabes
- Une colonne avec le navigateur utilisé pour la recherche
- Une colonne avec le numéro de la recherche neutre utilisée pour la comparaison.

Utilisateurs - Score Google		Nombre d'hyperliens différents par rapport :								Navigateur	N° de la recherche neutre utilisée
N° utilisateur	Score Google	Recherche neutre (R1)		PDN	Par thématique : RA, RA		CF2008	RA			
		En nombre	En %		En %	En %		En %	En %		
N°1	10	22	20,16%	5/38	13,15%	8/37	21,62%	9/35	25,71%	Chrome	1
N°2											
N°3											
N°4											
N°5											
N°6											
N°7											
N°8											
N°9											
N°10											
N°11											
N°12											
N°13											
N°14											
N°15											
N°16											
N°17											
N°18											
N°19											
N°20											
N°21											
N°22											
N°23											
N°24											
N°25											
N°26											
N°27											
N°28											
N°29											
N°30											

Figure 8 - Tableau Excel centralisateur, avec un exemple pour l'utilisateur n°1

⁷⁰ Par facilité, une vidéo de démonstration de la méthodologie a été incluse dans les annexes numériques. Des captures d'écran ne rendaient pas suffisamment compte du processus à l'œuvre.

3.1. Justification du corpus

Les internautes sélectionnés pour les entretiens ne devaient pas respecter de critères démographiques particuliers, comme cela peut être le cas lors de certaines études empiriques. Dans le cadre de cette enquête, de tels critères n'avaient en effet aucun intérêt puisque ce sont uniquement les résultats de leurs recherches qui importaient. Cependant, le panel n'a pas été choisi au hasard.

Pour parvenir à observer si le degré d'utilisation influençait les recherches et favorisait l'effet bulle de filtre, il était tout de même nécessaire d'obtenir un corpus qui comprenait à la fois des personnes qui n'utilisaient pas ou très peu Google (faible degré d'utilisation), des personnes qui utilisaient moyennement Google (degré moyen d'utilisation) et des personnes qui utilisaient fortement Google (haut degré d'utilisation). La seule contrainte était donc d'obtenir le même nombre d'entretiens dans chaque catégorie.

4. Analyse des résultats :

Finalement, pour tirer une conclusion scientifique des résultats, c'est-à-dire démontrer qu'il existe un lien, ou non, entre le degré d'utilisation de Google et le pourcentage d'hyperliens différents, et donc la mise en place d'un effet bulle de filtres, un **test de corrélation bivarié** fut effectué sur le fichier Excel centralisateur. Ce calcul a été réalisé grâce aux deux variables : les données de la colonne reprenant le « score Google », c'est-à-dire le degré d'utilisation des internautes, et la colonne reprenant tous les pourcentages de liens différents. Ensuite, grâce à la formule de corrélation, le logiciel nous indique immédiatement si les deux données sont corrélées. Le score de corrélation va de 0 à 1 : plus il est près de 1, plus les deux données sont corrélées. Pour rendre les conclusions plus visuelles, un graphique sera dressé dans la partie sur l'interprétation des résultats.

3. Résultats et interprétations

3.1. Présentation des données

Les données de l'étude – anonymisées - ont été collectées auprès de trente internautes lors d'entretiens qui se sont déroulés en trois sessions : le 01 mars, le 13 juillet et le 01 août 2018.

Ces entretiens ont permis d'exécuter **360 requêtes Google** et de générer une base de données composée de **3358 hyperliens**.

En se penchant sur les résultats en eux-mêmes, il est possible de classer les utilisateurs dans trois catégories selon l'importance de leur **Score Google**. On considèrera en effet que l'internaute à un :

- **Faible degré d'utilisation** lorsque son score est inférieur ou égal à 11 (en bleu dans le fichier Excel centralisateur) ;
- **Degré moyen d'utilisation** lorsque son score est compris entre 12 et 24 inclus (en orange) ;
- **Fort degré d'utilisation** lorsque son score est supérieur à 25 jusqu'à 35 (en rouge).

Dans notre étude, les internautes présentent des résultats variés, mais se répartissent équitablement au sein des trois catégories : 10 internautes par degré. Cette équité est volontaire : pour être capable de comparer des différences selon le degré d'utilisation et de personnalisation Google, il était primordial de disposer d'un corpus comprenant tous les cas de figure.

On observe que **le score le plus bas est de 1**, tandis que **le score le plus haut observé est de 35**. En **moyenne**, les internautes atteignent un **score de 17,73**. Comme des différences très importantes existent entre les internautes, il est intéressant de connaître la médiane, qui n'est pas sensible aux valeurs extrêmes. **La médiane se situe à 19,5** : la moitié des internautes atteignent un score en-dessous de 19,5 quand l'autre moitié le dépasse.

Par ailleurs, il est intéressant de noter qu'une seule personne a atteint le score maximum. Cela s'explique par la relative rareté de certains services, comme

Blogger ou Google Photos, qui ne sont pas les applications les plus utilisés dans leur domaine. Cela n'a toutefois pas d'incidence sur la pertinence des résultats, car ce ne sont pas les services jouant le plus grand rôle dans la personnalisation.

À l'inverse, quatre personnes ne dépassent pas le score minimum de 1 point⁷¹. Si cette constatation n'est évidemment pas représentative de la réalité des usagers d'Internet – ce n'est pas l'objectif - elle est tout de même notable dans la mesure où un tel score signifie que la personne n'a recours à aucun service du géant américain, hormis le moteur de recherche. Lorsqu'on se remémore les statistiques de Google et sa position monopolistique évoquée en début de mémoire, parvenir à se passer de tous les services relève presque d'une position militante.

Enfin, au niveau des navigateurs Internet dans notre panel, c'est une parfaite égalité. Les sondés privilégient autant **Mozilla Firefox** que Google Chrome. Nous verrons dans le traitement des données si le choix du navigateur est un facteur d'influence dans l'apparition des bulles de filtres.

3.2. Traitements des données

Afin de procéder à un traitement rigoureux des données, les analyses ont été divisées en cinq catégories :

- Une analyse du nombre de liens différents, afin de déterminer la moyenne de différence par recherche et mettre en lumière d'éventuels résultats extrêmes.
- Une **comparaison par thématiques**, pour détecter si un sujet est davantage propice à la personnalisation qu'un autre.
- Une **comparaison par catégories de scores** (faible, moyen, fort), pour identifier s'il existe des variations dans les sous-catégories.
- Une **comparaison par navigateur Internet**, pour chercher à savoir si le navigateur choisi peut avoir une influence quelconque.

⁷¹ Pour rappel, le score minimum est de 1 point – et non de zéro – puisque les internautes sélectionnés utilisent nécessairement le moteur de recherche Google et que ce critère vaut 1 point.

- Une **synthèse globale**, grâce au coefficient de corrélation, entre le Score Google et le nombre d'hyperliens différents par rapport la recherche neutre.

Ci-dessous, les termes « les recherches » ou « les résultats » feront référence, sauf mention contraire, aux 12 requêtes effectuées par entretien. Ainsi, s'il est indiqué qu'il y a, par exemple, 10 liens différents, cela signifie qu'il existe une différence de 10 liens pour *l'ensemble* des 12 requêtes.

Utilisateur : Score Google		Nombre d'hyperliens différents par rapport :								Navigateur :	N° de la recherche neutre : utilité
N° utilisateur	Score Google	Recherche neutre (RN)		Par l'internaute (Io, RN)		Par l'internaute (Io, RN)		RA			
		En nombre	En %	PDN	En %	CF2008	En %		En %		
N°1	16	22	20,16%	5/38	13,15%	8/37	21,62%	9/35	25,71%	Chrome	1
N°2	17	39	36,40%	11/37	29,72%	16/34	47,05%	12/37	32,43%	Firefox	1
N°3	11	14	11,77%	8/39	20,51%	3/41	7,31%	3/40	7,50%	Chrome	1
N°4	22	28	25,87%	12/37	32,43%	11/34	32,35%	5/39	12,82%	Firefox	1
N°5	26	12	9,63%	6/43	13,95%	2/41	4,87%	4/40	10%	Chrome	1
N°6	26	13	11,57%	2/36	5,56%	8/38	21,05%	3/37	8,10%	Chrome	1
N°7	1	20	16,73%	10/40	25%	7/40	17,50%	3/39	7,69%	Firefox	1
N°8	17	17	13,83%	8/42	19,04%	5/41	12,19%	4/39	10,25%	Chrome	1
N°9	18	24	20,60%	9/38	23,68%	7/41	17,07%	8/38	21,05%	Firefox	1
N°10	1	30	23,76%	11/42	26,19%	11/43	25,58%	8/41	19,51%	Firefox	1
N°11	1	14	11,99%	6/41	14,63%	4/38	10,52%	4/37	10,81%	Firefox	2
N°12	25	18	16,41%	4/37	10,81%	8/36	22,22%	6/37	16,21%	Firefox	2
N°13	21	23	20,79%	7/39	17,94%	11/36	30,55%	5/36	13,88%	Firefox	2
N°14	23	15	13,76%	5/36	13,88%	5/37	13,51%	5/36	13,88%	Firefox	2
N°15	29	7	6,45%	2/38	5,71%	3/37	8,10%	2/36	5,55%	Chrome	3
N°16	31	13	12,40%	2/41	4,87%	7/33	21,21%	4/36	11,11%	Chrome	3
N°17	12	23	20,79%	7/41	17,07%	10/33	30,30%	6/40	15%	Firefox	3
N°18	13	19	17,20%	7/38	18,42%	7/37	18,91%	5/35	14,28%	Chrome	3
N°19	7	15	13,54%	6/37	16,21%	5/36	13,88%	4/38	10,52%	Firefox	3
N°20	2	20	16,63%	6/45	13,36%	10/37	27,02%	4/42	9,52%	Chrome	3
N°21	26	16	13,42%	8/40	20%	4/40	10%	4/39	10,25%	Chrome	3
N°22	32	12	10,10%	7/39	17,94%	2/41	4,87%	3/40	7,50%	Chrome	3
N°23	15	19	17,04%	8/36	22,22%	6/39	15,38%	5/37	13,51%	Chrome	3
N°24	28	21	16,66%	8/42	19,04%	8/42	19,04%	5/42	11,90%	Chrome	3
N°25	26	21	17,96%	9/37	24,32%	7/41	17,07%	5/40	12,50%	Firefox	3
N°26	35	9	11,05%	4/41	9,75%	5/38	13,15%	4/39	10,25%	Chrome	3
N°27	11	30	27,40%	20/36	55,55%	5/38	13,15%	5/37	13,51%	Firefox	3
N°28	24	15	12,92%	6/40	15%	5/37	13,51%	4/39	10,25%	Firefox	3
N°29	24	14	18,33%	10/40	25%	15/40	37,50%	9/40	22,50%	Firefox	3
N°30	4	19	17,25%	7/38	18,42%	8/36	22,22%	4/36	11,11%	Chrome	3
Nombre de liens analysés : 3358											
MOYENNE TOTALE : 17,7333333											
MEDIANE TOTALE : 19,5											
Par catégorie : Faible 17,69% Moyen 19,59% Fort 14,30%											
Par navigateur : Chrome 13,64% Firefox 20,48%											
Corrélation : -0,2970511											

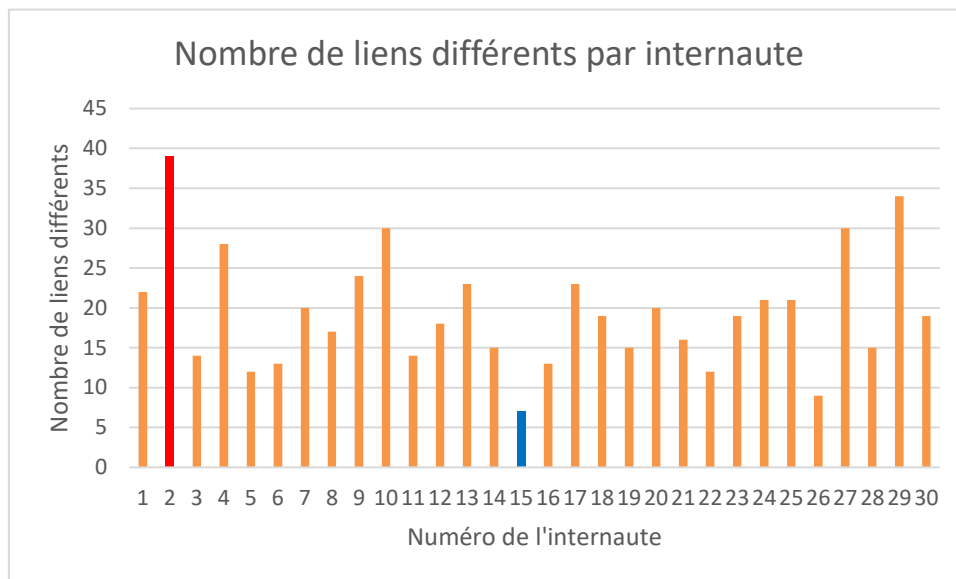
Figure 9 - Fichier récapitulatif des données récoltées

Voici les conclusions qui peuvent être tirées après avoir passé au crible **30 internautes** aux profils différents et analysé **3358 URL**.

I. Quel que soit le Score Google, une différence a été observée entre les résultats obtenus par les internautes et ceux obtenus à la suite de la recherche neutre.

En s'attardant sur la colonne réservée au nombre de liens différents, nous remarquons en effet qu'aucune recherche réalisée lors des entretiens ne correspond exactement à celle de la recherche neutre correspondante. Sans exception, les recherches des internautes présentent au moins une différence. Celle-ci peut être minime, comme avec le cas n°15, où seuls 7 liens différents

ont été comptabilisés (6,45% des liens récoltés différents) ; mais peut-être bien plus grande, comme avec l'utilisateur n°2 et ses 39 liens divergents (36,40%).

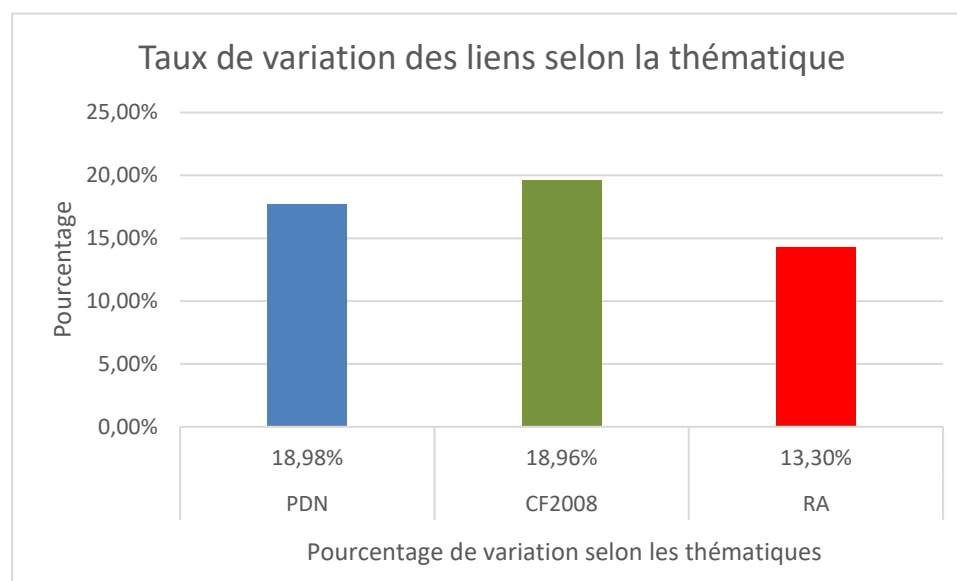


En **moyenne**, les recherches des internautes ont **19,4 liens différents** par rapport aux recherches neutres auxquelles elles ont été comparées. Autrement dit, cela signifie qu'après avoir effectué les douze recherches sur Google, les internautes ont été exposés, toujours en moyenne, à des **contenus 17,08% différents**. Un calcul de la médiane révèle par ailleurs que les résultats obtenus pour cette variable ne comprennent pas de valeurs extrêmes à même d'influencer négativement l'analyse. En effet, **la médiane se situe à 19**, ce qui équivaut à un **pourcentage de 16,65%**, soit une valeur très proche de la moyenne calculée.

II. Les thématiques ne semblent pas avoir d'influence sur le taux de variation des liens récoltés.

Si la variation entre les ordinateurs des internautes et les recherches neutres est un phénomène désormais avéré, il est intéressant de noter que les thématiques choisies n'ont pas d'incidence sur le taux de variation. Les hyperliens varient dans les mêmes proportions, quel que soit le sujet abordé. Aucun sujet, expression ou mot-clef ne semble polariser les résultats plus qu'un autre. En **moyenne**, les requêtes relatives à la **peine de mort (PDN)** varient de **7,36 liens (18,98%)**, celles concernant la **crise financière de 2008 (CF2008)** de **7,10 liens (18,96%)** et celles par rapport **aux révolutions**

arabes (RA) de 5 liens (13,30%). Au total des trois thématiques, la variation s'établit, en moyenne, à 6,48 liens.

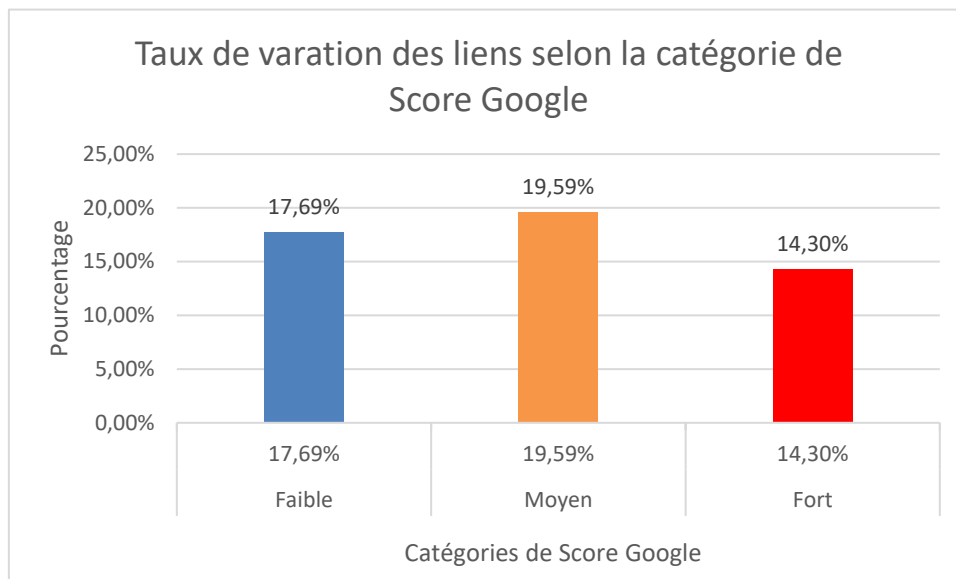


En indiquant que les thématiques n'influencent pas le taux de variation, nous ne ferons bien entendu pas les yeux sur l'écart de pourcentage qui existe entre les deux premiers sujets et celui sur les révolutions arabes, qui de l'ordre de 5,68% Cependant, cette disparité n'a pas de réelles explications. L'une des hypothèses consiste à dire qu'il s'agit d'un sujet plus récent que les deux autres, et que par conséquent, les sites Internet traitant de celui-ci sont plus rares. Une seconde hypothèse plausible considère que les révolutions arabes relèvent d'un champ historique plus complexe à traiter, notamment à cause de sa relative nouveauté, et qu'en conséquence, les sites web spécialisés ont tendance à réapparaître plus régulièrement. En tout état de cause, les statistiques obtenues pour la peine de mort et la crise financière 2008 nous confortent dans l'idée qu'il n'existe pas de réelle différence selon le sujet qui fait l'objet de la recherche.

III. La sous-catégorie du Score Google dans lequel se trouve l'internaute n'a pas d'impact sur le nombre de liens différents.

Lorsqu'on effectue une analyse par catégorie de scores, nous remarquons que celui-ci n'a pas d'importance. Les internautes appartenant à la catégorie « **faible** » (bleu dans fichier Excel) atteignent **17,69%** de liens différents,

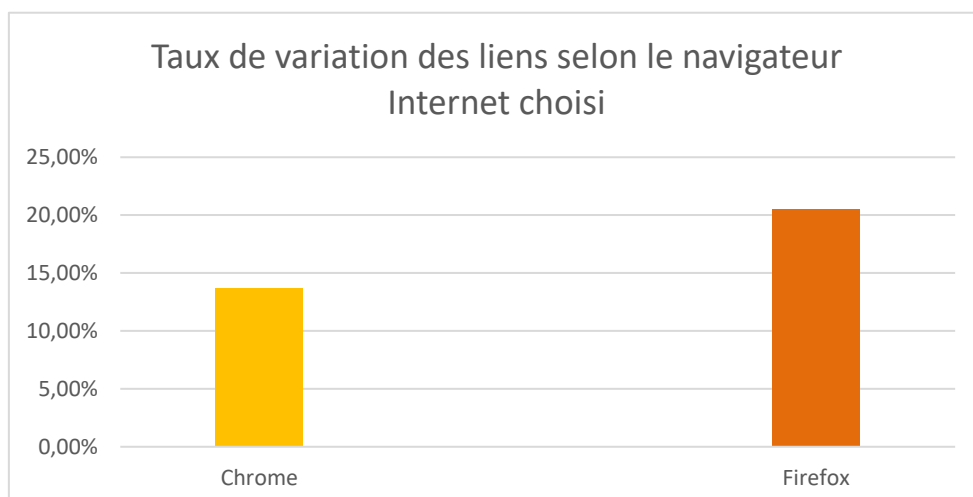
alors que ceux qui disposent d'un **score moyen** (orange) présentent **19,59%** de différence et les **forts** (rouge) un **taux de variation de 14,30%**.



De tels résultats vont même à l'encontre de l'hypothèse de départ, puisque les internautes ayant un fort degré d'utilisation Google, c'est-à-dire appartenant à la catégorie des forts, présentent dans le même temps le taux de variation le plus bas.

IV. Le navigateur Internet a tendance à influencer le taux de variation.

Le choix du navigateur Internet tend à influencer le taux de variation des liens. Les internautes ayant recours à **Mozilla Firefox** observent un taux s'élevant à **20,48%**, alors que les aficionados de **Google Chrome** n'atteignent *que* **13,68%** de différence.



Cette divergence peut s'expliquer par divers facteurs. Comme son nom l'indique, Google Chrome appartient au géant américain. Dans ce cadre, l'on pourrait imaginer que Google favorise certains de ces produits ou sites Internet dans les résultats de recherche effectuée depuis son propre navigateur. Par exemple, pour reprendre un site qui est apparu à plusieurs reprises, le site Google Scholar. Étant donné que les recherches neutres ont été créées depuis ce navigateur, cela pourrait être l'une des raisons pour lesquelles les recherches des internautes effectuées avec Chrome correspondent mieux.

Cette disparité pourrait également s'expliquer par la conception même du navigateur. Selon les configurations techniques choisies par l'internaute dans l'utilisation quotidienne de son navigateur, il est possible que celui-ci affiche plus ou moins de liens par rapport à Google Chrome et influence donc les résultats.

Ceci étant dit, les recherches préliminaires à l'étude n'ont pas fait mention d'une quelconque influence de la part des navigateurs dans le développement de l'effet bulle de filtres. Ces résultats sont donc à considérer avec les précautions d'usage.

V. Il n'existe pas de corrélation entre le degré d'utilisation Google et la personnalisation des résultats.

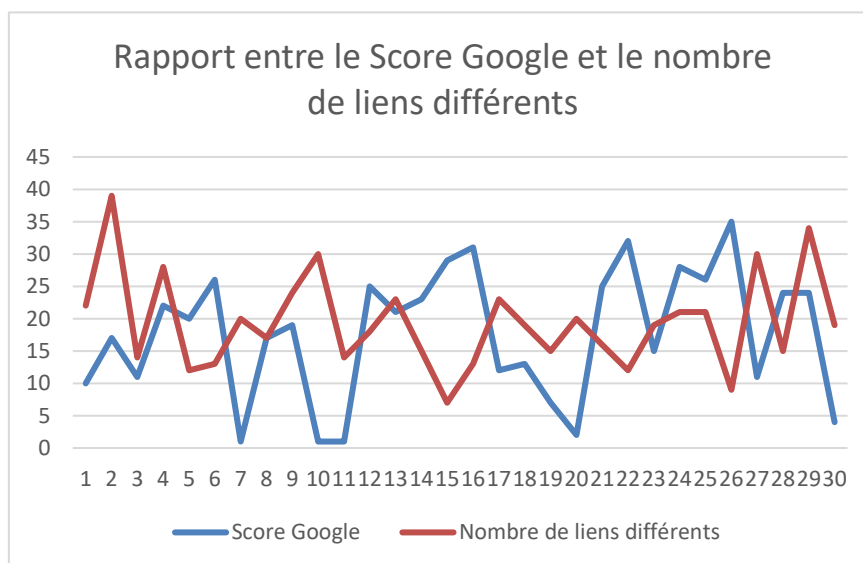
Les résultats explicités ci-dessus tendent à indiquer que des différences importantes peuvent être observées au sein des résultats de recherche entre des internautes différents. Mais cette constatation n'indique pas, pour l'instant, si ces différences sont liées au Score Google des internautes ou si elles sont le fruit d'une autre variable.

Mais grâce au tableau Excel croisant le Score Google et le nombre de résultats différents, il fut toutefois possible de vérifier si une corrélation pouvait être établie entre le taux d'utilisation de Google et la personnalisation des résultats. Pour y parvenir, nous avons calculé un coefficient de corrélation bivarié. On entend par statistique bivariée l'étude de(s) relation(s) entre deux variables.

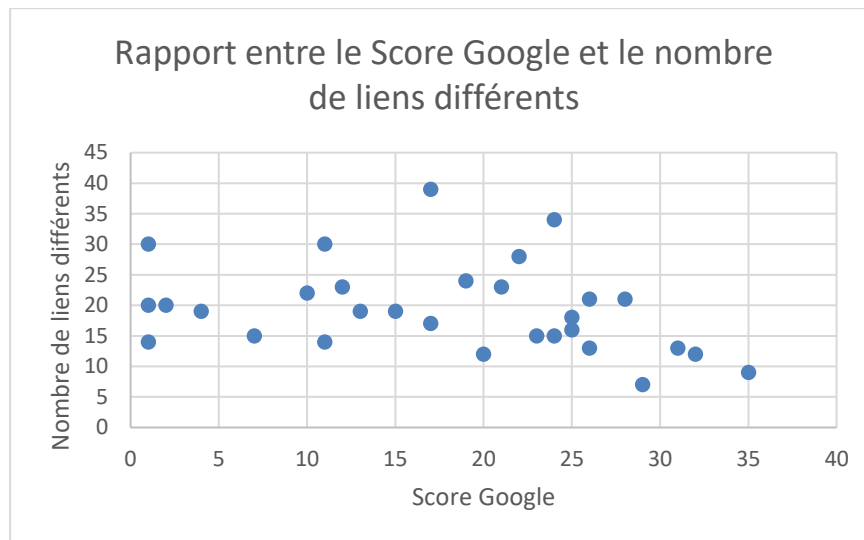
Après analyse entre les deux variables en question, il en résulte que le **coefficient de corrélation** entre le Score Google et le nombre de liens différents est égal à **- 0,2970511**. Cela signifie que ces variables sont corrélées négativement et à des valeurs trop basses pour établir un lien solide entre les deux.

Dans la mesure où ce chiffre brute n'est pas nécessairement parlant au premier abord, nous avons dressé deux graphiques qui démontrent visuellement qu'il n'y a pas de relation formelle entre le Score Google de l'internaute et le nombre de liens différents qu'il voit s'afficher sur son ordinateur.

Le premier graphique montre que la courbe bleue, indiquant le Score Google, ne progresse pas dans la même direction que la courbe rouge, indiquant le nombre de liens différents, ce qui devrait être le cas s'il y avait une corrélation. Au contraire, les deux courbes poursuivent des trajectoires parfois opposées.



Ce deuxième graphique, en nuage de points, explicite d'une autre façon l'absence de relation entre les deux variables. Si une corrélation avait été avérée, les points devraient former une droite rectiligne s'étalant de (0,0) à (40,45). Dans le cas présent, les points sont répartis de manière irrégulière et sans logique, prouvant que les deux variables n'obéissent pas ensemble.



3.3. Interprétations des résultats

Après avoir passé au crible les résultats des trente internautes et les 3358 hyperliens, plusieurs interprétations peuvent être tirées.

Le calcul de corrélation bivarié nous apporte un premier élément de réponse. Pour rappel, un coefficient égal à 1 ou à -1 signifie que les deux variables sont parfaitement corrélées. À l'inverse, un coefficient égal à 0 indique que les variables ne sont pas corrélées. Quant aux valeurs intermédiaires, elles informent de l'intensité de la dépendance entre les deux variables. Dans notre cas, le coefficient de corrélation tend à indiquer que la relation entre la variable « Score Google » et la variable « Nombre de liens différents » **n'est pas** corrélée à un niveau suffisamment élevé pour en conclure que l'algorithme de personnalisation et de hiérarchisation de l'information sur le moteur de recherche Google crée une bulle de filtre informationnelle. En effet, avec un coefficient de -0,2970511, la corrélation peut être considérée comme très faible.

Ceci étant dit, l'interprétation d'un coefficient de corrélation ne peut pas être aussi tranchée. Selon le domaine de recherche, les données qui sont traitées et les objectifs poursuivis par l'étude, un coefficient peut être plus ou moins révélateur d'un phénomène, même s'il n'est pas élevé dans l'absolu. C'est pourquoi le coefficient obtenu à la suite de cette étude nécessite d'être fortement nuancé.

Indéniablement, l'hypothèse de départ, selon laquelle l'algorithme de Google personnalise à l'excès les résultats de son moteur de recherche et crée par la même occasion un effet bulle de filtres, semble avoir du plomb dans l'aile. En effet, le niveau de dépendance à Google n'influence pas les résultats des internautes qui composent le panel de l'étude : avoir fourni l'ensemble de ses données personnelles et avoir activé les options de personnalisation ne permet pas de générer davantage de liens différents, et vice-versa. Toutefois, cette conclusion ne signifie pas qu'un phénomène de personnalisation n'existe pas et que les résultats de recherche sont les mêmes pour tout le monde. Elle indique uniquement que les deux variables ne sont pas corrélées et qu'elles ne peuvent expliquer l'effet bulle de filtres.

Même si elles avaient recours à d'autres dispositifs méthodologiques et analysaient d'autres supports (essentiellement les réseaux sociaux), les études scientifiques citées dans l'état de l'art en étaient venues aux mêmes conclusions. Tant Pablo Barbera, que les universitaires Flaxman, Goel et Rao, ou encore l'étude menée conjointement entre Stanford et Brown, ont indiqué que si l'effet bulle de filtres et les systèmes de personnalisation ne pouvaient être complètement niés, la polarisation des contenus en ligne ne trouvait pas sa cause principale sur Internet ou à l'intérieur des algorithmes.

Mais force est de constater que le phénomène de personnalisation est un phénomène bel et bien existant. En se penchant sur les autres statistiques issues de l'enquête, on remarque qu'en moyenne, les résultats de recherche diffèrent de **17,08%** au total. Remis à l'échelle, cela équivaut à obtenir **1,67 lien différent**⁷² pour chaque recherche effectuée. Dans certains cas de figure, le pourcentage de différence dépasse les 50% sur un sujet en particulier après quatre recherches⁷³. Vu d'une autre manière, cela signifierait que pratiquement 1/5^e des contenus que nous voyons après une requête Google ne sont pas les mêmes que ceux de notre voisin.

⁷² Il y a en moyenne 6,68% de différence par thématique pour 4 recherches. Pour ramener la donnée sur une seule recherche, il suffit de faire 6,68%/4.

⁷³ C'est le cas de l'utilisateur n°2 lors des recherches sur la crise financière de 2008.

Dans le contexte des bulles de filtres, et de manière plus globale, dans celui de l'accès démocratique à l'information, un tel pourcentage pose question. Cette situation rappelle les critiques qu'avait formulées Eli Pariser sur l'étude menée par les chercheurs de Facebook en 2015. En lisant la conclusion qui indiquait que seuls 5 à 8% des contenus pouvaient être modifiés, Pariser avait rétorqué qu'un tel nombre était déjà inquiétant et ne pouvait, en aucun cas, être considéré comme marginal.

Au-delà d'un effet bulle de filtres à proprement parlé, nous retiendrons que les résultats de recherche sur Google doivent être pris pour ce qu'ils valent et que le moteur de recherche ne représente qu'un accès à l'information parmi d'autres, aussi puissant et pertinent qu'il peut l'être. En surfant sur Google, l'internaute peut être rassuré : il n'est pas totalement isolé du monde et des contenus opposés à ses opinions. Mais il doit garder en tête que des différences, parfois importantes, peuvent survenir au sein des pages de résultats et qu'à cet égard, il est de son ressort d'agir pour accéder à des sources d'informations plus diversifiées.

Partie IV- Conclusion

1. Synthèse des résultats et perspective

Finalement, nous pouvons infirmer l'hypothèse de départ et en conclure que l'algorithme implémenté à l'intérieur du moteur de recherche Google ne contribue pas à la mise en place d'un effet bulle de filtres aussi puissant qu'annoncé. L'état de l'art développé en introduction avait mis en évidence, notamment grâce aux témoignages de Pariser et de Bronner, que Google se basait sur un ensemble de 57 critères pour personnaliser les résultats et, indirectement, créer une bulle informationnelle autour de ses utilisateurs. En théorie, le nombre de critères renseignés par l'internaute au sein de son compte Google est proportionnel à la capacité de l'algorithme à dresser son profil et lui proposer des contenus adaptés. Pourtant, l'analyse de **30 internautes** aux profils divergents et de **3558 liens** n'a pas permis de mettre à jour une corrélation solide entre le degré d'utilisation Google d'un individu et la polarisation de ses contenus.

Si les deux variables ne sont liées l'une à l'autre, il n'en demeure pas moins qu'un phénomène de personnalisation relativement important se produit bel et bien lors de requêtes Google. Les statistiques de l'enquête ont mis en lumière que le contenu issu de douze recherches Google variait en moyenne de **18,11% différent** par rapport à celui d'une recherche neutre. Remis à l'échelle d'une seule requête, cela signifie qu'un utilisateur est exposé, en moyenne, à 1,62 lien différent à chaque fois qu'il tape un mot-clef.⁷⁴

Nous sommes donc loin, très loin, du scénario catastrophe où les internautes seraient des pions aveuglés et mis à l'écart des contenus ne correspondant pas à leurs opinions. Toutefois, ces résultats doivent tout de même les usagers lors de leur utilisation quotidienne du moteur de recherche. Car si un ou deux liens différents peut sembler marginal, et peu à même d'influencer une

⁷⁴ En effet, comme il y a en moyenne 6,48 liens différents par thématique (voir ci-dessus), il suffit de diviser ce chiffre par 4 (puisque'il y a quatre recherches par thématiques) pour obtenir le nombre moyen de différences pour une seule requête.

personne, cela pose tout de même question lorsqu'il s'agit d'un contenu placé en haut de la page des résultats et bénéficiant d'une plus grande visibilité. Il suffit d'imaginer, par exemple, que le lien menant vers un article journalistique de qualité soit remplacé, selon les desideratas de l'algorithme, par un lien vers un site faisant la part belle à la désinformation. Dans la pratique, un seul lien aura été remplacé, mais avec un impact parfois considérable selon la place qu'il occupe dans la hiérarchie des résultats.

Cette constatation, qui n'est valable que pour Google, s'inscrit dans la lignée des recherches empiriques menées à ce sujet durant les dernières années. Si aucune n'avait pour objet d'étude un moteur de recherche, elles partageaient toute la conclusion si l'effet bulle de filtres était un phénomène avéré, les algorithmes ne renforçaient que marginalement ces effets. L'étude conduite par Jesse Shapiro et Gentzkow, et celle de l'Université d'Oxford, en avaient même conclu que le comportement des internautes était probablement une cause plus crédible pour expliquer la polarisation qui peut se produire sur Internet, à l'image des travaux menés un demi-siècle auparavant par les psychosociologues.

Face à cette situation, plusieurs initiatives ont été pensées pour aider les internautes à atténuer au maximum les effets de la personnalisation et des bulles de filtres.

La première solution, aussi simple qu'elle puisse paraître, est fondamentale puisqu'elle concerne l'éducation. L'enseignement aux médias dès le plus jeune âge pourrait permettre aux futurs adultes de mieux comprendre les rouages et les enjeux des outils numériques, et de disposer d'un recul critique nécessaire. Ainsi formés, les internautes seraient au courant qu'il est indispensable de sortir proactivement de sa bulle filtre et de se confronter aux points de vue contraires.

La deuxième piste est d'ordre politico-juridique. La Cour européenne des droits de l'homme rappelle que « *le principe général de tout traitement juridique et éthique de l'information doit se fonder sur une distinction claire entre opinion et information. La première concerne l'expression de pensées,*

*de croyances et de jugements de valeur, et la seconde a trait à des données et à des faits, faits qui peuvent faire l'objet d'une information «ou qui peuvent être d'une importance publique »*⁷⁵. Les acteurs du web pourraient ainsi se calquer sur cette recommandation. Mais disons-le d'emblée : il s'agit d'un vœu pieux.

Plusieurs sites Internet ont également été développés afin de créer des espaces numériques où se rassemblent toutes les informations de tous les points de vue différents. Se présentant comme des médiums non biaisés et ne jouant pas le jeu des bulles de filtres, ils sont censés montrer les opinions qui ont volontairement été omises par les algorithmes. C'est le cas notamment des sites américains : allsides.com³⁵ ou de hifromtheotherside.com. Ces médias sont toutefois à prendre avec des pincettes, car sous couvert de neutralité, *« ils tendent parfois à privilégier les extrêmes au détriment des points de vue nuancés et/ou du pluralisme des opinions »*⁷⁶. On notera également la présence de <http://unfiltered.news/>, un site web qui prend la forme d'une mappemonde où il est possible de voir les sujets les plus tendance par pays ou par région du monde.

Enfin, certains moteurs de recherche concurrents à Google, comme Qwant, suppriment la fonction de personnalisation et font *« valoir que leur absence de tri idéologique des résultats de recherche permettait de lutter contre les bulles de filtres »*⁷⁷

⁷⁵ https://assembly.coe.int/nw/xml/XRef/X2H-XrefViewHTML.asp?FileID=7207&lang=fr#P258_48077 (consulté le 03/08/2018)

⁷⁶ https://fr.wikipedia.org/wiki/Bulle_de_filtres#cite_note-32 (consulté le 03/08/2018)

⁷⁷ Ibidem.

2. Limites et retour critique sur l'étude

Notre étude présente plusieurs limitations qu'il convient d'expliciter pour en saisir la portée réelle.

Premièrement, la taille du corpus ne permet pas de tirer des conclusions définitives. Les bulles de filtres et les systèmes de personnalisation sont des sujets relativement complexes qui nécessitent bien souvent une masse importante de données pour en comprendre toutes les ramifications. Contrairement aux études empiriques citées plus haut, nous n'avions pas accès à des outils spécialisés nous permettant de constituer une base de données plus large. Malgré un processus semi-automatisé grâce à Excel, la récolte et le traitement des données exigeaient plus ou moins 1h de temps par internaute. L'établissement du profil Google des personnes interrogées et l'analyse des XXXX liens ont donc nécessité environ 30h de travail. Cet aspect pour le moins audacieux de l'enquête avait bien entendu été pris en considération en amont du mémoire, mais il permet de comprendre la difficulté de constituer un corpus exhaustif.

Deuxièmement, une enquête quantitative ne rend toujours pas compte de la complexité d'une problématique. Si les résultats obtenus offrent un aperçu pertinent du phénomène à l'œuvre sur les moteurs de recherche, ils ne reflètent pas d'autres dimensions abstraites. Durant notre étude, par exemple, le contenu en lui-même des résultats de recherche n'a pas été pris en compte, pour des raisons tant méthodologiques que de faisabilité. Mais en agissant de la sorte, la méthode quantitative nous indique uniquement la présence de liens différents par rapport à une recherche neutre, pas que ces liens dirigent vers des sites affichant des opinions opposées. Par exemple, si la recherche neutre affiche un lien qui mène à un article du journal Le Monde, mais que la recherche de l'internaute affiche un autre article du journal Le Monde, la méthodologie le comptera comme différent, y compris si les deux articles parlent *in fine* du même sujet. Un individu pourrait donc réalistement afficher un taux de différence élevé à la suite de notre test, mais être exposé, dans la pratique, à des contenus tout à fait similaires à la recherche neutre.

C'est pourquoi le développement d'une étude qualitative serait la bienvenue pour approfondir le sujet.

Troisièmement et dernièrement, le facteur temps est un biais potentiellement important. Même si celui-ci a été contrôlé par deux astuces méthodologiques (des recherches neutres effectuées le même jour que les recherches des internautes et des thématiques/mots-clefs recouvrant des sujets passés), il peut néanmoins influencer les résultats. Si de futures études scientifiques devaient être conduites sur le sujet, la gestion du facteur temps devrait se trouver en bonne place dans l'élaboration de la méthodologie.

Néanmoins, malgré ces limitations, nous sommes convaincus que notre recherche peut apporter quelques données scientifiques supplémentaires dans la compréhension d'un phénomène de plus en plus commenté médiatiquement.

3. Mot de conclusion

Si les analyses auxquelles nous avons procédées présentent des limites qu'il convient de prendre en compte dans l'interprétation des résultats, nous faisons le pari qu'elles apportent néanmoins une mise en perspective pertinente dans le cadre des bulles de filtres appliquées au moteur de recherche Google.

Contrairement à certains propos catastrophés dans l'opinion publique, cette étude émet des critiques circonstanciées qui ne cherchent pas à condamner coûte que coûte les algorithmes, consciente que ceux-ci sont nécessaires et utiles dans ce qui constitue aujourd'hui Internet. Au contraire, l'enquête empirique tend à prouver que ces systèmes de personnalisation et de recommandation sur les moteurs de recherche ne jouent peut-être un rôle aussi important que l'on pourrait le croire dans l'apparition du phénomène des bulles de filtres.

À travers une enquête et une méthodologie originale, qui a permis de passer en revue plusieurs milliers d'hyperliens, ce mémoire veut avant tout susciter une prise de conscience, tant du côté des citoyens que des acteurs du web. Si les internautes doivent garder en tête que leur comportement humain est tout

aussi présent sur Internet que dans la vie réelle, les GAFA et autres géants ne sont pas exempts de tous soupçons puisqu'ils détiennent le pouvoir d'atténuer ou d'aggraver les bulles de filtres. En réalité, pour paraphraser Dominique Cardon, les citoyens pourraient guider les rêves des ordinateurs afin qu'ils ne deviennent pas nos cauchemars de demain.

Derrière ce phénomène des bulles de filtres se cache deux réalités. D'un côté, celle des êtres humains, qui ont une tendance naturelle à privilégier ce qui les rassure et à adopter des comportements qui leur permettent d'éviter des informations inconsistantes. Cette théorie psychosociologique n'a rien de nouveau, mais elle est encore trop souvent laissée pour compte lorsqu'il s'agit de trouver ses éclaircissements aux phénomènes de polarisation. Si les informations que l'on obtient après une recherche sur le web résultent indéniablement d'une certaine forme de personnalisation, elles ne sont peut-être que le reflet de la personnalité de celui qui les a cherchées. Car études scientifiques à l'appui, Internet ne semble pas fragmenter davantage l'opinion publique que ces ancêtres, les médias de masse en tête. À l'instar d'un individu qui choisissait jadis de lire le journal correspondant à son idéologie politique, l'internaute du XXI^e siècle effectue des requêtes de manière à trouver l'information qui lui plait.

Mais cette situation est conditionnée au respect d'une ligne rouge à ne pas franchir. Au fil des évolutions d'Internet, les géants du web ont fondé des empires dont les modèles économiques reposent sur l'économie de l'attention et sur les données personnelles de leurs usagers. Devenues des entreprises dont la valeur dépend en partie de l'attention qu'elles suscitent, elles pourraient être tentées d'accroître d'un cran leurs systèmes de recommandations et de personnalisation pour s'assurer de maintenir l'intérêt de leur clientèle. Mais agissant de la sorte, le phénomène de bulle de filtres serait *gonflé* artificiellement et engendrerait une fragmentation et une polarisation à des niveaux supérieurs à ceux connus actuellement.

Les propos spectaculaires prononcés par Eli Pariser en 2011 avaient peut-être pour objectif, non pas tellement de dénoncer une situation existante, mais de prévenir des conséquences désastreuses des bulles de filtres si aucune limite

n'était fixée, et d'empêcher Internet de devenir un sanctuaire où se rencontrerait des camps convaincus de détenir le bon savoir. En réalité, l'ouvrage de l'activiste est un rappel brutal, mais salvateur qu'Internet, et en corolaire les algorithmes, sont avant tout des créations humaines et qu'à cet égard, elles relèvent de considérations éminemment politiques et économiques. Loin d'être une fatalité, l'avenir des bulles de filtres dépendra surtout des choix moraux qui seront opérés par les acteurs numériques de nos vies.

Partie V- Bibliographie

1. Monographie :

Bronner G., « *La démocratie des crédules* », Presses Universitaires de France, p.34, 2013

Catellin Sylvie, « *Sérendipité : Du conte au concept* », Éditions Le Seuil, Coll.: Science ouverte, 2014

Van Andel P. et **Bourcier D.**, « *De la sérendipité : dans la science, la technique, l'art, et le droit ; Leçons de l'inattendu* », Editions Hermann, 2008

Sadin Éric, « *La vie algorithmique : critique de la raison numérique* », Éditions L'Échappée, 288p., 2015

2. Articles scientifiques

Abbate Janet, « *L'histoire de l'Internet au prisme des STS* », dans *Le Temps des médias* (n° 18), p. 170-180, 2012. En ligne : <http://www.cairn.info.proxy.bib.ucl.ac.be:8888/revue-le-temps-des-medias-2012-1-page-170.htm> (consulté le 30/03/2017)

Bakshy E., Messing S., & Adamic L. A., « *Exposure to ideologically diverse news and opinion on Facebook* », in *Science*, N°348, Issue 6239, p. 1130–1132, 2015. En ligne : <https://research.fb.com/exposure-to-diverse-information-on-facebook-2/> (consulté le 21/05/2018)

Barbera Pablo, « *How social media reduces mass political polarization. Evidence from Germany, Spain, and the U.S.* », New York University Press, 46p., 2015

Bardin B., Perrissol S., Fos Y., Py J., Dagot L., « *Exposition sélective aux informations de prévention et attitude implicite envers le tabac* », dans *Psychologie Française*, Volume 62, Issue 1, p.1-14, 2017

Boxell L., Gentzkow M., Shapiro J., « *Is the internet causing political polarization ? Evidence from demographics* », Stanford University, Brown University, 25p, 2017. En ligne : <https://www.brown.edu/Research/Shapiro/pdfs/age-polars.pdf> (consulté le 09/07/2018)

Borgatti Stephen P., Martin G. Everett, and Jeffrey C. Johnson, “*Analyzing social networks*”, SAGE Publications, 2013

Bozdag E., « *Bias in Algorithmic Filtering and Personalization* », in *Ethics and Information Technology*, vol. 15, n° 3, p. 209-227, 2003 En ligne : <https://link-springer-com.proxy.bib.ucl.ac.be:2443/article/10.1007%2Fs10676-013-9321-6> (consulté le 23/07/2017)

Cardon Dominique, « *Du lien au like sur Internet. Deux mesures de la réputation* », *Communications* (n° 93), p. 173-186, 2013

Cardon Dominique, « *Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google* », *Réseaux* (n° 177), p. 63-95, 2013

Cardon Dominique, « *À quoi rêvent les algorithmes : nos vies à l'heure des big data* », Éditions Le Seuil, 112p, 2015

Eisermanne M., « *Comment fonctionne Google ?* », Grenoble, Université Joseph Fourier, s. e., 2009. En ligne sur : <www.igt.uni-stuttgart.de/eiserm/enseignement/google.pdf>, consulté le 09/03/2017

Eloire Fabien, « *Qui se ressemble s'assemble ? Homophilie sociale et effet multiplicateur : les mécanismes du capital social* », dans *Actes de la recherche en sciences sociales* (N°205), p.104-119, 2014

Festinger Leon, « *A Theory of Social Comparison Processes* », in *Human Relations*, vol. 7, n°2, p. 117-140, 1954. En ligne : <https://goo.gl/qYaGWO> et sa version traduite : <https://goo.gl/bMVF1s> (consultés le 07/01/2017)

Festinger Leon, « *A Theory of Cognitive Dissonance* », California : Stanford University Press, 1957

Flaxman S., Goel S., Rao J.M., « *Filter Bubbles, Echo Chambers, and Online News Consumption* », in *Public Opinion Quarterly*, Vol. 80, Issue S1, p. 298–320, 2016. En ligne : <https://academic.oup.com/poq/article-abstract/80/S1/298/2223402> (consulté le 15/04/2018)

Forse Michel, « *Les réseaux sociaux d'aujourd'hui. Un monde décidément bien petit* », dans *Revue de l'OFCE* (n° 126), p. 155-169, 2012

Gentzkow M., Shapiro J., « *Ideological Segregation Online and Offline* », *The Quarterly Journal of Economics*, Volume 126, Issue 4, Pages 1799–1839, 2011. En ligne : <https://doi.org/10.1093/qje/qjr044> (consulté le 23/11/2017)

Harmon-Jones E., « *A Cognitive Dissonance Theory Perspective on Persuasion* », in *The Persuasion Handbook: Developments in Theory and Practice*, Sage Publications, p. 101, 2002

Langville Amy N., Meyer Carl D., « *Google's pagerank and beyond : the science of search engine rankings* », Editions Princeton University Press, Princeton, 2006

Lazarfeld P. F., Berelson B. & Gaudet H., « *The people's choice: How the voter makes up his mind in a presidential election* ». New York: Duell, Sloan and Pearce, 1944

Mercanti-Guerin Maria, « *Facebook, un nouvel outil de campagne : Analyse des réseaux sociaux et marketing politique* », dans *La Revue des Sciences de Gestion* (n°242), p. 17-28, 2010

Messing S., Westwood S.J., « *Selective Exposure in the Age of Social Media* », in *Communication Research*, Vol. 41, Issue 8, p. 1042-1063, 2012. En ligne : <http://journals.sagepub.com/doi/abs/10.1177/0093650212466406>, consulté le 12/12/2017

Page Larry, « *Method for node ranking in a linked database* », Patent 6285999, 1998. En ligne : <https://patents.google.com/patent/US6285999B1/en> (consulté le 27/02/2018)

Pariser Eli, « *The Filter Bubble : What The Internet Is Hiding From You* », Editions Penguin, New-York, 2012

Parrish S., "How Filter Bubbles Distort Reality: Everything You Need to Know", Farnam Street, 2017. En ligne : <https://fs.blog/2017/07/filter-bubbles/> (consulté le 03/02/2018)

Lévy P., « *L'hypertexte, instrument et métaphore de la communication* », Réseaux, n° 46-47, pp. 59-68, 1991

McPherson M., **Smith-Lovin** L., **Cook** J.M., « *Birds of a feather: Homophily in social networks* », in Annual review of sociology, 27(1), 415–444, 2001.

Nickerson R.S., « *Confirmation bias: A ubiquitous phenomenon in many guises* » Review of general psychology, 2(2), 175-220, 1998

Rieder Bernhard, « *What Is in PageRank ? A Historical and Conceptual Investigation of a Recursive Status Index* », in Computational Culture. A Journal of software studies, n° 2, 2012

Robins Garry, « *Doing Social Network Research : Network-based Research Design for Social Scientists* », Sage Publications, London, 2015

Simon H. A. (1971), « *Designing Organizations for an Information-Rich World* », in Martin Greenberger, Computers, Communication, and the Public Interest, Baltimore, MD: The Johns Hopkins Press

Sire Guillaume, « *VI. Un traitement de l'information controversé* », dans *Les moteurs de recherche*. Paris, La Découverte, « Repères », p. 85-96, 2016. En ligne : <http://www.cairn.info.proxy.bib.ucl.ac.be:8888/les-moteurs-de-recherche--9782707184955-page-85.htm> (consulté le 02/02/2018)

Sire Guillaume, « *Cinq questions auxquelles Google n'aura jamais fini de répondre* », dans *Hermès, La Revue* (n° 73), p. 201-208, 2015. En ligne : <http://www.cairn.info.proxy.bib.ucl.ac.be:8888/revue-hermes-la-revue-2015-3-page-201.htm> (consulté le 02/02/2018)

Sire G. et Rieder B., « *Dans les ramifications de l'arbre hypertexte. Analyse des incitations générées par l'opacité du moteur Google* », in French Journal for Media Research, n° 3, 2015. En ligne :

<frenchjournalformediaresearch.com/index.php?id=468>, consulté le 09/09/2017

Sunstein Cass, « *Republic.com* », Princeton University Press, 224 p, 2001

Von Schoutz D., Van Nierkerk J., « *You According to Google : the Effects of Filter Bubbling* », Proceedings of the 14th annual conference on World Wide Web applications, Cape Town, Cape Peninsula University of Technology, 2012

3. Webographie :

Bensaid Andrea, « *Saviez-vous qu'il existe des moteurs de recherche alternatifs à Google* », Eskimoz, <http://www.eskimo.fr/saviez-vous-quil-existe-des-moteurs-de-recherche-alternatifs-a-google/> (consulté le 03/02/2018)

Benton Joshua, « *This study finds that social media use reduces political polarization instead of increasing it* », in Nieman Lab, http://www.niemanlab.org/2014/10/this-study-finds-that-social-media-use-reduces-political-polarization-instead-of-increasing-it/?utm_content=buffer04916&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer (consulté le 20/11/2017)

Coëffé Thomas, « *Chiffres Google – 2018* », <https://www.blogdumoderateur.com/chiffres-google/> (consulté le 19/11/2018)

Delcambre Alexis, « *Les réseaux sociaux prennent une place croissante dans l'accès à l'information* », dans Le Monde, https://www.lemonde.fr/actualite-medias/article/2016/06/15/les-reseaux-sociaux-prennent-une-place-croissante-dans-l-acces-a-l-information_4950771_3236.html (consulté le 03/09/2017)

Ducourtieux Cécile, « *L'Union européenne punit Google d'une amende record de 2,42 milliards d'euros* », Le Monde, http://www.lemonde.fr/economie/article/2017/06/27/bruxelles-punit-google-d-une-amende-record_5151710_3234.html (consulté le 23/01/2018)

Fradin Andréa, « *La loi numérique ne veut pas que Google lise vos mails. Et les espions ?* », dans Le Nouvel Observateur, <https://www.nouvelobs.com/rue89/rue89-explicateur/20150929.RUE0753/la-loi-numerique-ne-veut-pas-que-google-lise-vos-mails-et-les-espions.html> (consulté le 26/07/2018)

Joignot Frédéric, « *Comment les algorithmes nous enferment dans une bulle intellectuelle* », dans Journalisme pensif, <http://fredericjoignot.blog.lemonde.fr/2016/08/30/comment-les-algorithmes-nous-enferment-dans-une-bulle-intellectuelle/> (consulté le 12/07/2018)

Oremus Will, « *Dans leur bulle (de filtres) les internautes ? Pas si simple ...* », dans Slate, <http://www.slate.fr/story/142910/bulle-filtres-polarisation> (consulté le 13/11/2017)

Pariser Eli, « *Did Facebook's Big Study Kill My Filter Bubble Thesis ?* », dans Wired, <https://www.wired.com/2015/05/did-facebooks-big-study-kill-my-filter-bubble-thesis/> (consulté le 23/06/2018)

Pariser Eli, « *Beware online « filter bubbles »* », in TED, <https://www.youtube.com/watch?v=B8ofWfx525s&> (consulté le 21/05/2017)

Perez Sarah, « *Rigged* », Tech Crunch, <https://techcrunch.com/2016/11/09/rigged/?guccounter=1> (consulté le 29/04/2017)

Read Max, « *Donald Trump Won Because of Facebook* », New-York Times Magazine, <http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html> (consulté le 29/04/2017)

Régnauld Irénée, « *Pour une écologie de l'attention, l'antichambre du temps de cerveau disponible* », dans Mais où va le web ?, <http://maisouvaleweb.fr/pour-une-ecologie-de-lattention-lantichambre-du-temps-de-cerveau-disponible-essai-yves-citton/> (consulté le 14/07/2017)

Régnauld Irénée, « *La société des calculs sous la loupe de la sociologie : à quoi rêvent les algorithmes – Dominique Cardon [lecture]* », dans Mais où va le web ?, <http://maisouvaleweb.fr/la-societe-des-calculs-sous-la-loupe-de-la-sociologie-a-quoi-revent-les-algorithmes-dominique-cardon-lecture/> (le 14/07/2017)

Solon Olivia, « *Facebook's failure: did fake news and polarized politics get Trump elected ?* », The Guardian,

<https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories> (consulté le 29/04/2017)

Szadkowski Michäel, « Facebook, faiseur de rois de l'élection américaine », dans *Le Monde*, https://www.lemonde.fr/pixels/article/2016/11/14/facebook-faiseur-de-rois-de-l-election-americaine_5030959_4408996.html (consulté le 12/05/2017)

Tual Morgane, « *Accusé d'avoir influencé l'élection américaine, Facebook se défend* », *Le Monde*, https://www.lemonde.fr/pixels/article/2016/11/11/accuse-d-avoir-influence-l-election-americaine-facebook-se-defend_5029769_4408996.html (consulté le 05/06/2017)

Zaffagni Marc, « *Xaphir, le moteur de recherche français qui veut faire de l'ombre à Google* », *Futura Sciences*, <https://www.futura-sciences.com/tech/actualites/internet-xaphir-moteur-recherche-francais-veut-faire-ombre-google-67270/> (consulté le 11/11/2017)

Alphabet Investor Relations, « *Alphabet Announces Thrid Quarter 2016 Results* », *Press Release*, https://abc.xyz/investor/news/earnings/2016/Q3_alphabet_earnings/ (consulté le 24/01/2018)

The Atlantic, « *The Filter Bubble* », <https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427/> (consulté le 12/07/2018)

L'Express/L'Expansion, « *La bataille contre la censure du web chinois ouvre un nouveau front commercial* » https://lexpansion.lexpress.fr/high-tech/la-chine-face-a-google_1492142.html (consulté le 04/04/2018)

GlobalStats Statcounter, <http://gs.statcounter.com/> (consulté le 03/02/2018)

GlobalStats Statcounter, « *Search Engine Market Share Worldwide* », <http://gs.statcounter.com/search-engine-market-share> (consulté le 30/06/2018)

GlobalStats Statcounter, « *Search Engine Market Share Belgium* », <http://gs.statcounter.com/search-engine-market-share/all/belgium> (consulté le 30/06/2018)

Google, « *Comment fonctionne Google* », <https://support.google.com/webmasters/answer/70897?hl=fr> (consulté le 23/03/2018)

Google, <https://myaccount.google.com/activitycontrols> (consulté le 13/05/2018)

Google Search, « *Des réponses pertinentes et des formats qui évoluent* », https://www.google.com/search/howsearchworks/responses/#?modal_active=none (consulté le 14/07/2017)

Le Monde Informatique, « *Xaphir, dernier des moteurs de recherche français* », <https://www.lemondeinformatique.fr/actualites/lire-xaphir-dernier-ne-des-moteurs-de-recherche-francais-68161.html> (consulté le 11/12/2017)

Le Nouvel Observateur, « *Le Lay : "nous vendons du temps de cerveau"* », <https://www.nouvelobs.com/culture/20040710.OBS2633/le-lay-nous-vendons-du-temps-de-cerveau.html> (consulté le 15/04/2018)

Parlement Européen, « *Projet de recommandation adopté à l'unanimité par la commission le 8 juin 1993 ; commission de la culture et de l'éducation* », https://assembly.coe.int/nw/xml/XRef/X2H-XrefViewHTML.asp?FileID=7207&lang=fr#P258_48077 (consulté le 03/08/2018)

Pew Research Center, « *Differences by Medium* », http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/pi_17-08-23_socialmediaupdate_0-01/ (consulté le 08/02/2018)

Wikipedia, « *Chambre d'écho (médias)* », [https://fr.wikipedia.org/wiki/Chambre_d%27%C3%A9cho_\(m%C3%A9dias\)](https://fr.wikipedia.org/wiki/Chambre_d%27%C3%A9cho_(m%C3%A9dias)) (consulté le 23/04/2018)

Wikipedia, « *Dissonance cognitive* », https://fr.wikipedia.org/wiki/Dissonance_cognitive (consulté le 29/11/2017)

Wikipedia, « *Bulle de filtres* », https://fr.wikipedia.org/wiki/Bulle_de_filtres#cite_note-32 (consulté le 03/08/2018)

La question des bulles de filtres sur Internet apparait dès le début des années 2000, lorsque certains spécialistes remettent en question l'idée selon laquelle le web serait totalement neutre. Mais ce n'est qu'en 2011 que le phénomène se popularise, lorsque le militant d'Internet Eli Pariser publie « *The Filter Bubble* ». Il y indique que les algorithmes présents sur les sites web ont la fâcheuse tendance à filtrer l'information qui parvient aux internautes, et en conséquence, à créer un état d'isolement informationnel, intellectuel et culturel. Ce mémoire s'interroge spécifiquement sur l'effet réel des bulles de filtres au cœur du moteur de recherche Google et cherche à comprendre si celles-ci peuvent véritablement influencer sur l'opinion des citoyens. Grâce une enquête quantitative où nous avons pu passer en revue plusieurs milliers d'hyperliens, cette étude émet des critiques circonstanciées qui ne cherchent pas à condamner coûte que coûte les algorithmes, consciente que ceux-ci sont nécessaires et utiles dans ce qui constitue aujourd'hui Internet. Au contraire, l'enquête empirique tend à prouver que ces systèmes de personnalisation et de recommandation sur les moteurs de recherche ne jouent peut-être un rôle aussi important que l'on pourrait le croire dans l'apparition du phénomène des bulles de filtres.

Mots-clefs : bulle de filtres, Google, algorithme, personnalisation, chambre d'écho

