

Ouvrir le code des algorithmes ? — oui, mais... (2/2)

Voici le deuxième volet (si vous avez raté le premier) de l'enquête approfondie d'Hubert Guillaud sur l'exploration des algorithmes, et de son analyse des enjeux qui en découlent.

Dans le code source de l'amplification algorithmique : que voulons-nous vraiment savoir ?

par Hubert GUILLAUD

Que voulons-nous vraiment savoir en enquêtant sur l'amplification algorithmique ? C'est justement l'enjeu du projet de recherche qu'Arvind Narayan mène au Knight Institute de l'université Columbia où il a ouvert un blog dédié et qui vient d'accueillir une grande conférence sur le sujet. Parler d'amplification permet de s'intéresser à toute la gamme des réponses qu'apportent les plateformes, allant de l'amélioration de la portée des discours à leur suppression, tout en se défiant d'une réduction binaire à la seule modération automatisée, entre ce qui doit être supprimé et ce qui ne doit pas l'être. Or, les phénomènes d'amplification ne sont pas sans effets de bord, qui vont bien au-delà de la seule désinformation, à l'image des effets très concrets qu'ont les influenceurs sur le commerce ou le tourisme. Le gros problème, pourtant, reste de pouvoir les étudier sans toujours y avoir accès.

Outre des analyses sur TikTok et les IA génératives, le blog recèle quelques trésors, notamment une monumentale synthèse qui fait le tour du sujet en expliquant les principes de fonctionnements des algorithmes (l'article est également très riche en liens et références, la synthèse que j'en propose y recourra assez peu).

Narayan rappelle que les plateformes disposent de très nombreux algorithmes

entremêlés, mais ceux qui l'intéressent particulièrement sont les algorithmes de recommandation, ceux qui génèrent les flux, les contenus qui nous sont mis à disposition. Alors que les algorithmes de recherche sont limités par le terme recherché, les algorithmes de recommandation sont bien plus larges et donnent aux plateformes un contrôle bien plus grand sur ce qu'elles recommandent à un utilisateur.

La souscription, le réseau et l'algorithme

Pour Narayan, il y a 3 grands types de leviers de propagation : la souscription (ou abonnement), le réseau et l'algorithme. Dans le modèle par abonnement, le message atteint les personnes qui se sont abonnées à l'auteur du message. Dans le modèle de réseau, il se propage en cascade à travers le réseau tant que les utilisateurs qui le voient choisissent de le propager. Dans le modèle algorithmique, les utilisateurs ayant des intérêts similaires (tels que définis par l'algorithme sur la base de leurs engagements passés) sont représentés plus près les uns des autres. Plus les intérêts d'un utilisateur sont similaires à ceux définis, plus il est probable que le contenu lui sera recommandé.

À l'origine, les réseaux sociaux comme Facebook ou Twitter ne fonctionnaient qu'à l'abonnement : vous ne voyiez que les contenus des personnes auxquelles vous étiez abonnés et vous ne pouviez pas republier les messages des autres ! Dans le modèle de réseau, un utilisateur voit non seulement les messages créés par les personnes auxquelles il s'est abonné, mais aussi les messages que ces utilisateurs choisissent d'amplifier, ce qui crée la possibilité de cascades d'informations et de contenus "viraux", comme c'était le cas de Twitter jusqu'en 2016, moment où le réseau introduisit le classement algorithmique. Dans le modèle algorithmique, la souscription est bien souvent minorée, le réseau amplifié mais surtout, le flux dépend principalement de ce que l'algorithme estime être le plus susceptible d'intéresser l'utilisateur. C'est ce que Cory Doctorow désigne comme « l'emmerdification » de nos flux, le fait de traiter la liste des personnes auxquelles nous sommes abonnés comme des suggestions et non comme des commandes.

Le passage aux recommandations algorithmiques a toujours généré des contestations, notamment parce que, si dans les modèles d'abonnement et de réseau, les créateurs peuvent se concentrer sur la construction de leur réseau, dans le « *modèle algorithmique, cela ne sert à rien, car le nombre d'abonnés n'a*

rien à voir avec la performance des messages » (mais comme nous sommes dans des mélanges entre les trois modèles, le nombre d'abonnés a encore un peu voire beaucoup d'influence dans l'amplification). Dans le modèle algorithmique, l'audience de chaque message est optimisée de manière indépendante en fonction du sujet, de la « qualité » du message et d'un certain nombre de paramètres pris en compte par le modèle.

Amplification et viralité

La question de l'amplification interroge la question de la viralité, c'est-à-dire le fait qu'un contenu soit amplifié par une cascade de reprises, et non pas seulement diffusé d'un émetteur à son public. Le problème de la viralité est que sa portée reste imprévisible. Pour Narayan, sur toutes les grandes plateformes, pour la plupart des créateurs, la majorité de l'engagement provient d'une petite fraction de contenu viral. Sur TikTok comme sur YouTube, 20 % des vidéos les plus vues d'un compte obtiennent plus de 70 % des vues. Plus le rôle de l'algorithme dans la propagation du contenu est important, par opposition aux abonnements ou au réseau, plus cette inégalité semble importante.

Parce qu'il est particulièrement repérable dans la masse des contenus, le contenu viral se prête assez bien à la *rétropropagation*, c'est-à-dire à son déclasserment ou à sa suppression. Le problème justement, c'est qu'il y a plein de manières de restreindre le contenu. Facebook classe les posts rétrogradés plus bas dans le fil d'actualité qu'ils ne le seraient s'ils ne l'avaient pas été, afin que les utilisateurs soient moins susceptibles de le rencontrer et de le propager. À son tour, l'effet de la rétrogradation sur la portée peut être imprévisible, non linéaire et parfois radical, puisque le contenu peut devenir parfaitement invisible. Cette rétrogradation est parfaitement opaque, notamment parce qu'une faible portée n'est pas automatiquement suspecte, étant donné qu'il existe une grande variation dans la portée naturelle du contenu.

Amplification et prédiction de l'engagement

Les plateformes ont plusieurs objectifs de haut niveau : améliorer leurs revenus publicitaires bien sûr et satisfaire suffisamment les utilisateurs pour qu'ils reviennent... Mais ces objectifs n'aident pas vraiment à décider ce qu'il faut donner à un utilisateur spécifique à un moment précis ni à mesurer comment ces décisions impactent à long terme la plateforme. D'où le fait que les plateformes

observent l'engagement, c'est-à-dire les actions instantanées des utilisateurs, comme le like, le commentaire ou le partage qui permettent de classer le contenu en fonction de la probabilité que l'utilisateur s'y intéresse. « *D'une certaine manière, l'engagement est une approximation des objectifs de haut niveau. Un utilisateur qui s'engage est plus susceptible de revenir et de générer des revenus publicitaires pour la plateforme.* »

Si l'engagement est vertueux, il a aussi de nombreuses limites qui expliquent que les algorithmes intègrent bien d'autres facteurs dans leur calcul. Ainsi, Facebook et Twitter optimisent les « *interactions sociales significatives* », c'est-à-dire une moyenne pondérée des likes, des partages et des commentaires. YouTube, lui, optimise en fonction de la durée de visionnage que l'algorithme prédit. TikTok utilise les interactions sociales et valorise les vidéos qui ont été regardées jusqu'au bout, comme un signal fort et qui explique certainement le caractère addictif de l'application et le fait que les vidéos courtes (qui ont donc tendance à obtenir un score élevé) continuent de dominer la plateforme.

En plus de ces logiques de base, il existe bien d'autres logiques secondaires, comme par exemple, pour que l'expérience utilisateur ne soit pas ralentie par le calcul, que les suggestions restent limitées, sélectionnées plus que classées, selon divers critères plus que selon des critères uniques (par exemple en proposant des nouveaux contenus et pas seulement des contenus similaires à ceux qu'on a apprécié, TikTok se distingue à nouveau par l'importance qu'il accorde à l'exploration de nouveaux contenus... c'est d'ailleurs la tactique suivie désormais par Instagram de Meta via les Reels, boostés sur le modèle de TikTok, qui ont le même effet que sur TikTok, à savoir une augmentation du temps passé sur l'application)...

« *Bien qu'il existe de nombreuses différences dans les détails, les similitudes entre les algorithmes de recommandation des différentes plateformes l'emportent sur leurs différences* », estime Narayan. Les différences sont surtout spécifiques, comme Youtube qui optimise selon la durée de visionnage, ou Spotify qui s'appuie davantage sur l'analyse de contenu que sur le comportement. Pour Narayan, ces différences montrent qu'il n'y a pas de risque concurrentiel à l'ouverture des algorithmes des plateformes, car leurs adaptations sont toujours très spécifiques. Ce qui varie, c'est la façon dont les plateformes ajustent l'engagement.

Comment apprécier la similarité ?

Mais la grande question à laquelle tous tentent de répondre est la même : « *Comment les utilisateurs similaires à cet utilisateur ont-ils réagi aux messages similaires à ce message ?* »

Si cette approche est populaire dans les traitements, c'est parce qu'elle s'est avérée efficace dans la pratique. Elle repose sur un double calcul de similarité. D'abord, celle entre utilisateurs. La similarité entre utilisateurs dépend du réseau (les gens que l'on suit ou ceux qu'on commente par exemple, que Twitter valorise fortement, mais peu TikTok), du comportement (qui est souvent plus critique, « *deux utilisateurs sont similaires s'ils se sont engagés dans un ensemble de messages similaires* ») et les données démographiques (du type âge, sexe, langue, géographie... qui sont en grande partie déduits des comportements).

Ensuite, il y a un calcul sur la similarité des messages qui repose principalement sur leur sujet et qui repose sur des algorithmes d'extraction des caractéristiques (comme la langue) intégrant des évaluations normatives, comme la caractérisation de discours haineux. L'autre signal de similarité des messages tient, là encore, au comportement : « *deux messages sont similaires si un ensemble similaire d'utilisateurs s'est engagé avec eux* ». Le plus important à retenir, insiste Narayan, c'est que « *l'enregistrement comportemental est le carburant du moteur de recommandation* ». La grande difficulté, dans ces appréciations algorithmiques, consiste à faire que le calcul reste traitable, face à des volumes d'enregistrements d'informations colossaux.

Une histoire des évolutions des algorithmes de recommandation

« *La première génération d'algorithmes de recommandation à grande échelle, comme ceux d'Amazon et de Netflix au début des années 2000, utilisait une technique simple appelée filtrage collaboratif : les clients qui ont acheté ceci ont également acheté cela* ». Le principe était de recommander des articles consultés ou achetés d'une manière rudimentaire, mais qui s'est révélé puissant dans le domaine du commerce électronique. En 2006, Netflix a organisé un concours en partageant les évaluations qu'il disposait sur les films pour améliorer son système de recommandation. Ce concours a donné naissance à la « *factorisation*

matricielle », une forme de deuxième génération d'algorithmes de recommandation, c'est-à-dire capables d'identifier des combinaisons d'attributs et de préférences croisées. Le système n'étiquette pas les films avec des termes interprétables facilement (comme "drôle" ou "thriller" ou "informatif" ...), mais avec un vaste ensemble d'étiquettes (de micro-genres obscurs comme « documentaires émouvants qui combattent le système ») qu'il associe aux préférences des utilisateurs. Le problème, c'est que cette factorisation matricielle n'est pas très lisible pour l'utilisateur et se voir dire qu'on va aimer tel film sans savoir pourquoi n'est pas très satisfaisant. Enfin, ce qui marche pour un catalogue de film limité n'est pas adapté aux médias sociaux où les messages sont infinis. La prédominance de la factorisation matricielle explique pourquoi les réseaux sociaux ont tardé à se lancer dans la recommandation, qui est longtemps restée inadaptée à leurs besoins.

Pourtant, les réseaux sociaux se sont tous convertis à l'optimisation basée sur l'apprentissage automatique. En 2010, Facebook utilisait un algorithme appelé EdgeRank pour construire le fil d'actualité des utilisateurs qui consistait à afficher les éléments par ordre de priorité décroissant selon un score d'affinité qui représente la prédiction de Facebook quant au degré d'intérêt de l'utilisateur pour les contenus affichés, valorisant les photos plus que le texte par exemple. À l'époque, ces pondérations étaient définies manuellement plutôt qu'apprises. En 2018, Facebook est passé à l'apprentissage automatique. La firme a introduit une métrique appelée « *interactions sociales significatives* » (MSI pour *meaningful social interactions*) dans le système d'apprentissage automatique. L'objectif affiché était de diminuer la présence des médias et des contenus de marque au profit des contenus d'amis et de famille. « *La formule calcule un score d'interaction sociale pour chaque élément susceptible d'être montré à un utilisateur donné* ». Le flux est généré en classant les messages disponibles selon leur score MSI décroissant, avec quelques ajustements, comme d'introduire de la diversité (avec peu d'indications sur la façon dont est calculée et ajoutée cette diversité). Le score MSI prédit la probabilité que l'utilisateur ait un type d'interaction spécifique (comme liker ou commenter) avec le contenu et affine le résultat en fonction de l'affinité de l'utilisateur avec ce qui lui est proposé. Il n'y a plus de pondération dédiée pour certains types de contenus, comme les photos ou les vidéos. Si elles subsistent, c'est uniquement parce que le système l'aura appris à partir des données de chaque utilisateur, et continuera à vous proposer des photos si vous les appréciez.

« Si l'on pousse cette logique jusqu'à sa conclusion naturelle, il ne devrait pas être nécessaire d'ajuster manuellement la formule en fonction des affinités. Si les utilisateurs préfèrent voir le contenu de leurs amis plutôt que celui des marques, l'algorithme devrait être en mesure de l'apprendre ». Ce n'est pourtant pas ce qu'il se passe. Certainement pour lutter contre la logique de l'optimisation de l'engagement, estime Narayan, dans le but d'augmenter la satisfaction à long terme, que l'algorithme ne peut pas mesurer, mais là encore sans que les modalités de ces ajustements ne soient clairement documentés.

Est-ce que tout cela est efficace ?

Reste à savoir si ces algorithmes sont efficaces ! « Il peut sembler évident qu'ils doivent bien fonctionner, étant donné qu'ils alimentent des plateformes technologiques qui valent des dizaines ou des centaines de milliards de dollars. Mais les chiffres racontent une autre histoire. Le taux d'engagement est une façon de quantifier le problème : il s'agit de la probabilité qu'un utilisateur s'intéresse à un message qui lui a été recommandé. Sur la plupart des plateformes, ce taux est inférieur à 1 %. TikTok est une exception, mais même là, ce taux dépasse à peine les 5 %. »

Le problème n'est pas que les algorithmes soient mauvais, mais surtout que les gens ne sont pas si prévisibles. Et qu'au final, les utilisateurs ne se soucient pas tant du manque de précision de la recommandation. *« Même s'ils sont imprécis au niveau individuel, ils sont précis dans l'ensemble. Par rapport aux plateformes basées sur les réseaux, les plateformes algorithmiques semblent être plus efficaces pour identifier les contenus viraux (qui trouveront un écho auprès d'un grand nombre de personnes). Elles sont également capables d'identifier des contenus de niche et de les faire correspondre au sous-ensemble d'utilisateurs susceptibles d'y être réceptifs. »* Si les algorithmes sont largement limités à la recherche de modèles dans les données comportementales, ils n'ont aucun sens commun. Quant au taux de clic publicitaire, il reste encore plus infinitésimal - même s'il est toujours considéré comme un succès !

Les ingénieurs contrôlent-ils encore les algorithmes ?

Les ingénieurs ont très peu d'espace pour contrôler les effets des algorithmes de

recommandation, estime Narayan, en prenant un exemple. En 2019, Facebook s'est rendu compte que les publications virales étaient beaucoup plus susceptibles de contenir des informations erronées ou d'autres types de contenus préjudiciables. En d'autres termes, ils se sont rendu compte que le passage à des interactions sociales significatives (MSI) a eu des effets de bords : les contenus qui suscitaient l'indignation et alimentaient les divisions gagnaient en portée, comme l'a expliqué l'ingénieure et lanceuse d'alerte Frances Haugen à l'origine des Facebook Files, dans ses témoignages. C'est ce que synthétise le tableau de pondération de la formule MSI publié par le Wall Street Journal, qui montrent que certains éléments ont des poids plus forts que d'autres : un commentaire vaut 15 fois plus qu'un like, mais un commentaire signifiant ou un repartage 30 fois plus, chez Facebook. Une pondération aussi élevée permet d'identifier les messages au potentiel viral et de les stimuler davantage. En 2020, Facebook a ramené la pondération des partages à 1,5, mais la pondération des commentaires est restée très élevée (15 à 20 fois plus qu'un like). Alors que les partages et les commentaires étaient regroupés dans une seule catégorie de pondération en 2018, ils ne le sont plus. Cette prime au commentaire demeure une prime aux contenus polémiques. Reste, on le comprend, que le jeu qui reste aux ingénieurs de Facebook consiste à ajuster le poids des paramètres. Pour Narayan : piloter un système d'une telle complexité en utilisant si peu de boutons ne peut qu'être difficile.

Le chercheur rappelle que le système est censé être neutre à l'égard de tous les contenus, à l'exception de certains qui enfreignent les règles de la plateforme. Utilisateurs et messages sont alors rétrogradés de manière algorithmique suite à signalement automatique ou non. Mais cette neutralité est en fait très difficile à atteindre. Les réseaux sociaux favorisent ceux qui ont déjà une grande portée, qu'elle soit méritée ou non, et sont récompensés par une plus grande portée encore. Par exemple, les 1 % d'auteurs les plus importants sur Twitter reçoivent 80 % des vues des tweets. Au final, cette conception de la neutralité finit par récompenser ceux qui sont capables de pirater l'engagement ou de tirer profit des biais sociaux.

Outre cette neutralité, un deuxième grand principe directeur est que *« l'algorithme sait mieux que quiconque »*. *« Ce principe et celui de la neutralité se renforcent mutuellement. Le fait de confier la politique (concernant le contenu à amplifier) aux données signifie que les ingénieurs n'ont pas besoin d'avoir un*

point de vue à ce sujet. Et cette neutralité fournit à l'algorithme des données plus propres à partir desquelles il peut apprendre. »

Le principe de l'algorithme qui sait le mieux signifie que la même optimisation est appliquée à tous les types de discours : divertissement, informations éducatives, informations sur la santé, actualités, discours politique, discours commercial, etc. En 2021, FB a fait une tentative de rétrograder tout le contenu politique, ce qui a eu pour effet de supprimer plus de sources d'information de haute qualité que de faible qualité, augmentant la désinformation. Cette neutralité affichée permet également une forme de désengagement des ingénieurs.

En 2021, encore, FB a entraîné des modèles d'apprentissage automatique pour classer les messages en deux catégories : bons ou mauvais pour le monde, en interrogeant les utilisateurs pour qu'ils apprécient des contenus qui leurs étaient proposés pour former les données. FB a constaté que les messages ayant une plus grande portée étaient considérés comme étant mauvais pour le monde. FB a donc rétrogradé ces contenus... mais en trouvant moins de contenus polémique, cette modification a entraîné une diminution de l'ouverture de l'application par les utilisateurs. L'entreprise a donc redéployé ce modèle en lui donnant bien moins de poids. Les corrections viennent directement en conflit avec le modèle d'affaires.

1111100101
0010101010
11110101
0101001
0101001
11110101
1001
1001
1010
1101
11

life



KEVIN SLAVIN

THOSE ALGORITHMS THAT GOVERN OUR



3.17.2013
FMA

first, a prologue:

a eulogy for lower manhattan

cities had to learn how to listen. (listen for incoming aircraft... RADAR)

STEALTH

The way to make something disappear is to break up a big thing and make it look like a lot of little things.

carries over to **stock market** (lots of little trades masking a BIG trade.)

shave 2ms off your trade

weird real estate spikes...

a **story** about DJing... in the late 1980s

algorithms determine what we hear

PLAYLIST

drive

eat

date

even who gets arrested

watch

how we get there

"NYC is being OPTIMIZED TO RUN LIKE A MOTHERBOARD"

- Neal Stephenson

1

OPACITY

bin-packing algorithm

elevator with only one button!!!

roomba doesn't behave like a HUMAN!

2

INSCRIPTABILITY

This is a PROBLEM when there's a flash crash

which people should you date? (Epaggix)

which movies should be made?!

Our capacity for that willful distortion is endless.

to be oblivious to facts & figures

we've seen algorithms go awry on the stock market... -10% in 11min!

3

DARK POOLS?

"there's no ASTRONOMY column in the newspaper (only astrology)"

STORIES

dark pools

these are things humans can write but no longer read!

genetic algorithms

Francis Galton (Darwin's cousin)

determine if good wine from data from grapes! before wine is made!!

Orley Ashenfelter

MATH determines affects

taste & preference

60% of NETFLIX rentals are recommended!

takes it further

production

consumption

mean?

What does the FLASH CRASH look like... for HOLLYWOOD? for the MUSIC INDUSTRY? for online dating? (maybe it's already happened?) for the CRIMINAL JUSTICE SYSTEM?

trading outside the algorithms!

«leaving town»

huge amounts of liquidity moving around outside algorithms.

So, what do the DARK POOLS look like... ???

Illustration par Jason Alderman « Those Algorithms That Govern Our Lives - Kevin Slavin » . (CC BY 2.0)

Pourquoi l'optimisation de l'engagement nous nuit-elle ?

« *Un grand nombre des pathologies familières des médias sociaux sont, à mon avis, des conséquences relativement directes de l'optimisation de l'engagement* », suggère encore le chercheur. Cela explique pourquoi les réformes sont difficiles et pourquoi l'amélioration de la transparence des algorithmes, de la modération, voire un meilleur contrôle par l'utilisateur de ce qu'il voit (comme le proposait Gobo mis en place par Ethan Zuckerman), ne sont pas des solutions magiques (même si elles sont nécessaires).

Les données comportementales, celles relatives à l'engagement passé, sont la matière première essentielle des moteurs de recommandations. Les systèmes privilégient la rétroaction implicite sur l'explicite, à la manière de YouTube qui a privilégié le temps passé sur les rétroactions explicites (les likes). Sur TikTok, il n'y a même plus de sélection, il suffit de swipper.

Le problème du *feedback* implicite est qu'il repose sur nos réactions inconscientes, automatiques et émotionnelles, sur nos pulsions, qui vont avoir tendance à privilégier une vidéo débile sur un contenu expert.

Pour les créateurs de contenu, cette optimisation par l'engagement favorise la variance et l'imprévisibilité, ce qui a pour conséquence d'alimenter une surproduction pour compenser cette variabilité. La production d'un grand volume de contenu, même s'il est de moindre qualité, peut augmenter les chances qu'au moins quelques-uns deviennent viraux chaque mois afin de lisser le flux de revenus. Le fait de récompenser les contenus viraux se fait au détriment de tous les autres types de contenus (d'où certainement le regain d'attraits pour des plateformes non algorithmiques, comme Substack voire dans une autre mesure, Mastodon).

Au niveau de la société, toutes les institutions sont impactées par les plateformes algorithmiques, du tourisme à la science, du journalisme à la santé publique. Or, chaque institution a des valeurs, comme l'équité dans le journalisme, la précision en science, la qualité dans nombre de domaines. Les algorithmes des médias

sociaux, eux, ne tiennent pas compte de ces valeurs et de ces signaux de qualité. *« Ils récompensent des facteurs sans rapport, sur la base d'une logique qui a du sens pour le divertissement, mais pas pour d'autres domaines »*. Pour Narayan, les plateformes de médias sociaux *« affaiblissent les institutions en sapant leurs normes de qualité et en les rendant moins dignes de confiance »*. C'est particulièrement actif dans le domaine de l'information, mais cela va bien au-delà, même si ce n'est pas au même degré. TikTok peut sembler ne pas représenter une menace pour la science, mais nous savons que les plateformes commencent par être un divertissement avant de s'étendre à d'autres sphères du discours, à l'image d'Instagram devenant un outil de communication politique ou de Twitter, où un tiers des tweets sont politiques.

La science des données en ses limites

Les plateformes sont bien conscientes de leurs limites, pourtant, elles n'ont pas fait beaucoup d'efforts pour résoudre les problèmes. Ces efforts restent occasionnels et rudimentaires, à l'image de la tentative de Facebook de comprendre la valeur des messages diffusés. La raison est bien sûr que ces aménagements nuisent aux résultats financiers de l'entreprise. *« Le recours à la prise de décision subconsciente et automatique est tout à fait intentionnelle ; c'est ce qu'on appelle la « conception sans friction ». Le fait que les utilisateurs puissent parfois faire preuve de discernement et résister à leurs impulsions est vu comme un problème à résoudre. »*

Pourtant, ces dernières années, la réputation des plateformes n'est plus au beau fixe. Narayan estime qu'il y a une autre limite. *« La plupart des inconvénients de l'optimisation de l'engagement ne sont pas visibles dans le cadre dominant de la conception des plateformes, qui accorde une importance considérable à la recherche d'une relation quantitative et causale entre les changements apportés à l'algorithme et leurs effets. »*

Si on observe les raisons qui poussent l'utilisateur à quitter une plateforme, la principale est qu'il ne parvient pas à obtenir des recommandations suffisamment intéressantes. Or, c'est exactement ce que l'optimisation par l'engagement est censée éviter. Les entreprises parviennent très bien à optimiser des recommandations qui plaisent à l'utilisateur sur l'instant, mais pas celles qui lui font dire, une fois qu'il a fermé l'application, que ce qu'il y a trouvé l'a enrichi. Elles n'arrivent pas à calculer et à intégrer le bénéfice à long terme, même si

elles restent très attentives aux taux de rétention ou aux taux de désabonnement. Pour y parvenir, il faudrait faire de l'A/B testing au long cours. Les plateformes savent le faire. Facebook a constaté que le fait d'afficher plus de notifications augmentait l'engagement à court terme mais avait un effet inverse sur un an. Reste que ce regard sur leurs effets à longs termes ne semble pas être une priorité par rapport à leurs effets de plus courts termes.

Une autre limite repose sur l'individualisme des plateformes. Si les applications sociales sont, globalement, assez satisfaisantes pour chacun, ni les utilisateurs ni les plateformes n'intériorisent leurs préjudices collectifs. Ces systèmes reposent sur l'hypothèse que le comportement de chaque utilisateur est indépendant et que l'effet sur la société (l'atteinte à la démocratie par exemple...) est très difficile à évaluer. Narayan le résume dans un tableau parlant, où la valeur sur la société n'a pas de métrique associée.

Level	Effect	Example metric	How to measure
Immediate	Engagement	CTR, MSI	Automated; optimized
Short term	User satisfaction	DAU	Manual (A/B test)
Long term	Value to user	Six-month retention	Too slow to measure
Global	Value to society	-	Impossible to measure

Tableau montrant les 4 niveaux sur lesquels les algorithmes des plateformes peuvent avoir des effets. CTR : Click Through Rate (taux de clic). MSI : Meaningful Social Interactions, interactions sociales significatives, la métrique d'engagement de Facebook. DAU : Daily active users, utilisateurs actifs quotidiens.

Les algorithmes ne sont pas l'ennemi (enfin si, quand même un peu)

Pour répondre à ces problèmes, beaucoup suggèrent de revenir à des flux plus chronologiques ou à des suivis plus stricts des personnes auxquelles nous sommes abonnés. Pas sûr que cela soit une solution très efficace pour gérer les volumes de flux, estime le chercheur. Les algorithmes de recommandation ont été la réponse à la surcharge d'information, rappelle-t-il : « *Il y a beaucoup plus*

d'informations en ligne en rapport avec les intérêts d'une personne qu'elle n'en a de temps disponible. » Les algorithmes de classement sont devenus une nécessité pratique. Même dans le cas d'un réseau longtemps basé sur l'abonnement, comme Instagram : en 2016, la société indiquait que les utilisateurs manquaient 70 % des publications auxquelles ils étaient abonnés. Aujourd'hui, Instagram compte 5 fois plus d'utilisateurs. En fait, les plateformes subissent d'énormes pressions pour que les algorithmes soient encore plus au cœur de leur fonctionnement que le contraire. Et les systèmes de recommandation font leur entrée dans d'autres domaines, comme l'éducation (avec Coursera) ou la finance (avec Robinhood).

Pour Narayan, l'enjeu reste de mieux comprendre ce qu'ils font. Pour cela, nous devons continuer d'exiger d'eux bien plus de transparence qu'ils n'en livrent. Pas plus que dans le monde des moteurs de recherche nous ne reviendrons aux annuaires, nous ne reviendrons pas aux flux chronologiques dans les moteurs de recommandation. Nous avons encore des efforts à faire pour contrecarrer activement les modèles les plus nuisibles des recommandations. L'enjeu, conclut-il, est peut-être d'esquisser plus d'alternatives que nous n'en disposons, comme par exemple, d'imaginer des algorithmes de recommandations qui n'optimisent pas l'engagement, ou pas seulement. Cela nécessite certainement aussi d'imaginer des réseaux sociaux avec des modèles économiques différents. Un autre internet. Les algorithmes ne sont peut-être pas l'ennemi comme il le dit, mais ceux qui ne sont ni transparents, ni loyaux, et qui optimisent leurs effets en dehors de toute autre considération, ne sont pas nos amis non plus !