

Netflix, une boîte noire dans une boîte noire

Netflix: A Black Box Within a Black Box

Zoé Buranello, Elsa Negre, Juliette Rouchier

Paris-Dauphine University, PSL Research Universities, UMR CNRS 7243, LAMSADE, Paris, France,
zoe.buranello@gmail.com

RÉSUMÉ. Dans cet article nous nous intéressons au système de recommandation du site de vidéo à la demande avec abonnement (SVOD) Netflix®, en prenant une double approche : le design, qui fonde notre questionnement sur la perception que l'utilisateur peut avoir sur les recommandations, et l'informatique, qui nous permet d'expliquer certains mécanismes génériques, déjà connus, des systèmes de recommandation. Pour cela nous nous concentrons sur l'influence que le choix des films et séries a sur les recommandations postérieures. Une de nos interrogations initiales porte sur le "démarrage à froid" du système et nous comparons, dans nos expériences, deux types de profil-utilisateur : l'un est ancien, alimenté depuis 7 ans, l'autre est créé pour l'occasion. Notre première expérience permet de mettre en avant le rôle crucial de la catégorisation des films - sur des critères de contenu, de popularité, ou d'exclusivité, par exemple. Ce travail nous a permis d'observer plusieurs éléments. Comme nous l'avions anticipé du fait de l'inertie des méthodes d'apprentissage, sur un profil récent l'expérience est très nettement plus visible que sur un profil ancien, et il est simple d'influencer par ses choix les recommandations. Ensuite, le nouveau profil reçoit beaucoup de recommandations génériques, puis très rapidement un peu contenu personnalisé qui peut être présenté dans différentes catégories. Enfin, la réversibilité du processus de recommandation est rapide sur le profil ancien, car quelques jours sans visionnage permettent de faire disparaître les traces de l'expérience. Ces expériences nous ont en outre permis de nous interroger sur les catégories qui sont finalement très nombreuses, plus ou moins précises, et non disjointes en contenu. Cette observation nous oriente vers une réflexion à venir sur la construction des catégories dans l'organisation pratique du capitalisme de surveillance.

ABSTRACT. This article focuses on the recommendation system of the subscription video-on-demand (SVOD) platform Netflix®, specifically exploring what can be understood from an end-user perspective. To this end, we investigate how viewing choices influence the recommendations displayed during the subsequent platform visit. To conduct this analysis, we designed two experiments comparing two user profiles: one long-standing profile active for 7 years, and another newly created profile. We concentrated on elements observable directly by the user: categories, recommended titles within personalized categories, and the « top banners » displayed prominently on the homepage. Our findings revealed the following: first, recommendations for a recent profile are more quickly and strongly influenced by the content viewed on that profile, whereas an older profile shows little change. Second, the new profile receives recommendations spanning a wide variety of genres, including popular content as well as some personalized suggestions. Finally, over time, while the older profile initially received increasing numbers of documentary recommendations based on its viewing history, only a few days of inactivity were enough for these recommendations to disappear entirely. Conversely, the recent profile continued to receive documentary suggestions. These experiments also allowed us to observe the evolution of suggested categories for each profile. The significant diversity of categories and the variability in how films are ranked within them emerged as important factors. This observation raises further questions about how the recommendation system creates and uses categories to encourage user engagement.

MOTS-CLÉS. Système de recommandation, Contexte, Diversité, Design, Genre cinématographique.

KEYWORDS. Recommender system, Context, Diversity, Design, Cinematic genre.

1. Introduction

Du protocinéma au streaming, la consommation de films s'est grandement transformée au cours des 130 dernières années. Les premiers films, diffusés dans les bars et les cafés ont rapidement trouvé leurs salles dédiées. Puis, l'arrivée de la télévision a resserré la consommation sur le cercle familial, privé, en proposant d'abord peu de programmes, puis un nombre de plus en plus important avec la multiplication des chaînes. Ensuite, la cassette vidéo (puis le DVD et le Blu-ray) a permis de visionner des films en choisissant individuellement ses horaires et de choisir si le film est regardé en entier d'une traite, par bout, ou en évitant certains passages. Finalement, avec l'avènement du streaming, la consommation est devenue individuelle et personnalisée. La Télé-internet se distingue de la télévision : au lieu de diffuser

des programmes selon une grille prévue à l'avance pour chaque chaîne, il s'agit de donner à l'utilisateur le choix du contenu qu'il veut regarder ainsi que le moment durant lequel il veut le regarder. Le succès des plateformes de streaming est notamment dû à ce système de recommandation, qui produit une représentation des goûts de l'utilisatrice et permet de proposer des films ou séries à regarder (FERRIER et al., 2024). Ce système de recommandation transforme la notion de public en individualisant l'expérience - tout en l'inscrivant dans des collectifs implicites de spectateurs, réunis par une forme de similarité par les algorithmes. En terme d'expérience individuelle, chacun se voit proposer un univers personnel, promesse d'un accord parfait avec ses goûts.

Nous nous sommes focalisées sur l'expérience faite par un utilisateur de Netflix®, inventeur de ce type de consommation en ligne qui s'est dorénavant généralisée. Netflix®¹ a été fondé en 1997 par Reed Hastings et Marc Randolph. À l'origine, il s'agissait d'un service en ligne d'achat et de location de DVD, livrés à domicile. Face aux coûts d'envoi des DVD, l'entreprise innove en proposant un nouveau modèle de consommation dès 1999 : un abonnement mensuel pour la location de DVD, sans pénalité en cas de retard.

En 2007, l'entreprise se transforme et propose alors un service de vidéo à la demande, disponible en ligne par abonnement. Entre 2010 et 2012, Netflix® s'exporte dans le monde, en commençant par le Canada et l'Amérique du Sud, puis l'Europe. En 2016, l'entreprise est présente dans la quasi-totalité des pays du monde. Netflix® utilise un système de recommandation hybride, c'est-à-dire qu'il allie filtrage basé sur le contenu et filtrage collaboratif (MADDODI and KARANI, 2019). Les recommandations sont ainsi basées à la fois sur l'historique de l'utilisateur et sur les profils qui lui sont similaires. On peut penser que cette analyse de données a permis à Netflix® d'augmenter la satisfaction de ses clients, qui sont devenus de plus en plus nombreux, de produire du contenu original et de devenir leader sur le marché du streaming. L'adéquation aux goûts de l'utilisateur grâce à la recommandation est un véritable enjeu économique au sein de ce qui s'appelle de nos jours "le capitalisme de surveillance" (ZUBOFF et al. (2022)) : il s'agit de s'assurer, grâce à la personnalisation la plus fine, que l'utilisateur reste abonné. Les systèmes de recommandation ont néanmoins des faiblesses qui créent des biais : démarrage à froid, dispersion des données, précision, diversité... (MADDODI and KARANI, 2019).

On peut noter que l'étude de Netflix® a un intérêt supplémentaire dans la mesure où, depuis 2013, Netflix® s'est lancé dans la production de contenus, qu'elle nomme « Netflix Originals ». Ceci implique que la sélection ciblée (création d'un univers personnel) mais aussi la production (en fonction de la demande perçue - fixant par exemple des cahiers des charges importants) sont en mesure de faire de Netflix® un acteur culturel responsable également de la production du goût du public. C'est ainsi que l'étude de la personnification interroge en retour sur le paysage culturel commun des utilisateurs et l'influence individuelle sur ce paysage.

L'algorithme de Netflix® est privé : par souci de protection du secret industriel, mais aussi des données utilisateurs suite à une faille révélée lors de la deuxième édition du Netflix Prize (HALLINAN and STRIPHAS, 2016; MADDODI and KARANI, 2019). Au cours de ce concours, certains participants ont prouvé qu'ils pouvaient retrouver les identités des utilisateurs dont les données avaient été partagées anonymement par Netflix®. Au vu de cette privatisation, à laquelle s'ajoute l'opacité du système de recommandation pour l'utilisateur (son fonctionnement n'est pas expliqué sur l'interface), nous choisis-

1. <https://www.netflix.com/fr/>

sons de nous placer du point de vue de l'utilisateur, au plus proche de ses habitudes. Notre méthodologie empruntera donc au design, cette discipline permettant une attention particulière à la perception de l'utilisateur. Nous nous concentrerons uniquement sur ce qui est accessible à l'utilisateur : l'interface. Nous étudierons donc en détail les changements observables sur la page d'accueil.

Dans cet article expérimental, nous cherchons à comprendre le fonctionnement de ces systèmes de recommandation, mais en nous plaçant du point de vue de l'utilisatrice qui veut comprendre l'influence de ses choix et éventuellement de ses changements de goût sur les propositions à venir. Cet article est ainsi principalement méthodologique, mais permet d'identifier des éléments qui confirment la compréhension théorique de ce système (issue de la connaissance informatique). Notre méthode répond à deux contraintes : nous souhaitons prendre le point de vue du design pour observer un système informatique complexe ; l'algorithme de Netflix® n'est pas public et on ne peut donc précisément comprendre le poids de chaque élément utilisé par le système de recommandation sans faire des tests où l'algorithme de recommandation de Netflix® est abordé comme une boîte noire.

Il est à noter que nous ne nous intéressons pas à l'expérience utilisateur (UX) dans le sens d'une recherche d'amélioration de la qualité de l'expérience vécue par l'utilisateur dans toute situation d'interaction, ni à l'ergonomie ou à l'utilisabilité d'un système informatique, ou encore à l'UX Design (méthode de conception centrée sur l'utilisateur). Au lieu de cela, nous tentons de rendre moins opaques les systèmes de recommandation sous l'angle abordé par les designers (souvent vus comme des boîtes noires et qui, de fait, inspirent la défiance des utilisateurs).

Nous nous posons la question suivante : un utilisateur de la plateforme peut-il percer la boîte noire du système de recommandation de Netflix® ? Nous cherchons à comprendre comment sont visibles, du point de vue de l'utilisateur uniquement en contact avec l'interface, les orientations proposées par les algorithmes de recommandation. Nous abordons cette question par les "genres" de films ou de séries, à travers deux expériences de visionnage. Ce travail est décrit dans cet article en cinq sections :

1. Un état de l'art sur les systèmes de recommandation, le mode de consommation proposé par Netflix®, le fonctionnement de son système de recommandation, la mesure de la diversité;
2. Nos hypothèses de départ ;
3. La description du protocole de nos expériences ainsi que de notre méthode d'analyse ;
4. Les résultats des 2 expériences menées ainsi qu'un retour sur nos hypothèses ;
5. Un ensemble de questions soulevées par notre étude, des observations que nous ne pouvons expliquer avec certitude et qui soulèvent un questionnement nouveau : celui de la catégorisation, c'est-à-dire le système de classement des films en diverses catégories, selon des thèmes, proposées à l'utilisateur sur l'interface.

2. État de l'art

2.1. Systèmes de recommandation

Le système de recommandation cherche à adapter l'accès à l'information à un utilisateur et ainsi l'aider à choisir des éléments de contenu dans un catalogue. En raison du déluge de données et d'informations, ce catalogue peut être très — voire trop — volumineux. Il s'agit donc de guider l'utilisateur lors de son exploration de la quantité d'informations à sa disposition, en identifiant pour lui les contenus qui

semblent les plus pertinents. C'est une forme particulière de filtrage, visant à présenter des éléments (films, musique, livres, images, pages web, ...) susceptibles de l'intéresser.

Généralement, le système de recommandation tente d'estimer l'« avis » que donnerait l'utilisateur à chaque élément, et met en avant ceux qui seraient les mieux évalués.

Par ailleurs, un utilisateur peut être influencé dans ses choix par le contexte dans lequel il se trouve au moment de la sélection. La définition du contexte la plus largement acceptée est celle de DEY (2001), où le contexte est défini comme toute information pouvant servir à caractériser la situation d'une entité (par exemple : le moment de la journée, les personnes présentes, etc.). Ainsi, les systèmes de recommandation dits « contextuels » ont été introduits pour tenir compte de ces éléments supplémentaires et proposer des suggestions encore plus adaptées (ADOMAVICIUS and TUZHILIN, 2011; KULKARNI and RODD, 2020). Il n'existe pas une méthode unique qui serait efficace dans tous les cas (PANNIELLO et al., 2009). Les résultats dépendent fortement du type de données disponibles et du domaine d'application. De plus, même lorsque les performances sont bonnes, les approches existantes restent limitées par la rareté des données (PANNIELLO et al., 2009; LE et al., 2021), ce qui peut freiner leur efficacité.

Un autre défi bien connu est celui du « démarrage à froid », rencontré lorsqu'il faut faire des recommandations pour des utilisateurs ou des contenus pour lesquels aucune information n'est encore disponible, qu'elle soit explicite ou implicite (SCHEIN et al., 2001). Plusieurs cas peuvent se présenter (KLOPOTEK, 2009), comme l'arrivée d'un nouvel utilisateur ou l'introduction d'un nouveau produit. Dans ces situations, le système manque de repères pour effectuer des choix pertinents.

Certains services en ligne ont développé des stratégies avancées pour répondre à ces enjeux. C'est le cas de Netflix®, dont le système de recommandation, Cinematch, s'appuie sur une grande variété de données pour construire une page d'accueil la plus personnalisée et pertinente possible. Il récolte notamment des informations sur la manière dont les membres regardent le contenu (le support utilisé, le moment de la journée, le jour de la semaine, le nombre d'épisodes visionnés, etc.), ce qui lui permet de proposer des recommandations tenant compte du contexte (NETFLIX, CENTRE D'AIDE, 2024; GOMEZ-URIBE and HUNT, 2016). Il enregistre également des données sur la position du contenu dans l'interface ainsi que sur les recommandations proposées qui n'ont pas été retenues.

L'objectif est de proposer une page d'accueil structurée autour de sections intuitives, avec des thématiques facilement compréhensibles pour l'utilisateur. Cette organisation est assurée par plusieurs algorithmes, dont le *Personalized video ranker* (classement personnalisé des vidéos) et le *Video-video similarity* (similarité entre vidéos). Ces algorithmes suivent une approche hybride, combinant filtrage basé sur le contenu et filtrage collaboratif.

Face au problème du démarrage à froid, Netflix® a mis en place une solution simple : lors de la création d'un nouveau profil, l'utilisateur est invité à sélectionner quelques titres qu'il apprécie (GOMEZ-URIBE and HUNT, 2016). Cela permet de disposer dès le départ d'éléments de préférence, pour proposer rapidement des recommandations adaptées et rendre l'expérience plus satisfaisante.

2.2. Transformation de la consommation audio-visuelle

Le système de recommandation de Netflix® a été élaboré en réponse à un problème : alors que la firme proposait ses DVD à la location, les films sortis récemment étaient très demandés par la clientèle, et bien

souvent indisponibles. En recommandant à ses abonnés des titres susceptibles de leur plaire, Netflix® a fait chuter le nombre de leur demande de 20% (MADDODI and KARANI, 2019). Désormais, au vu de la taille du catalogue, le système de recommandation est nécessaire pour éviter à l'utilisateur de passer en revue des milliers de titres à chaque session (SCHAFFNER et al., 2023) (6 842 en sur Netflix France® en mars 2023 (STOLL, 2023)).

Désormais, le système de recommandation a pour but d'inciter l'utilisateur à utiliser et à rester le plus longtemps possible sur la plateforme, pour qu'il prolonge son abonnement. L'entreprise formule deux objectifs précis : le premier c'est de « gagner des moments de vérité », c'est-à-dire proposer à l'utilisateur un contenu qui lui plait et dans lequel il s'investira en quelques secondes. C'est ce phénomène qui réduit les chances d'abandon du service. Ensuite, le second objectif est de proposer un contenu niche. En effet, certains titres ne seraient pas diffusables à la télévision, parce que leur audimat trop faible ne pourrait attirer les annonceurs nécessaires à la diffusion. Or, avec ce qu'il nomme la Télé-internet, Netflix® peut trouver le public niche qui aimera le contenu niche (GOMEZ-URIBE and HUNT, 2016).

En 2013, Netflix® commence à sortir des saisons entières d'un seul coup et popularise alors le *binge-watching*, une pratique née avec la sortie des séries en DVD. Celle-ci consiste à regarder plusieurs épisodes d'affilée, voire une saison entière d'une série en une session. Ce mode de consommation a d'abord gagné en popularité chez les adolescents, avant de devenir un passe-temps répandu. Les conséquences sont multiples : sanitaires, avec l'apparition de symptômes semblables à ceux de l'addiction, mais aussi sociales. En effet, certaines séries popularisées par le *binge-watching* traitent de sujets comme le changement climatique ou la diversité de genre et suscitent nombre de discussions d'actualité (SCHAFFNER et al., 2023). Compte tenu de ces conséquences, nous nous intéressons ensuite à l'impact sur l'utilisateur de la consommation sur les plateformes de SVOD.

2.3. Impact sur l'utilisateur et sur sa consommation de contenu audio-visuel

Le service que propose Netflix® est à l'intersection du *storytelling*, le fait de raconter des histoires à des fins de communication, et d'Internet. L'entreprise se réclame ainsi d'inventer la Télé-internet, avec un enjeu principal, le choix : que regarder, où et quand. Pour faciliter ce choix, elle a développé un système de recommandation complexe. L'entreprise met en place une stratégie de transparence, qui semble donc s'opposer à la qualité supposément opaque du système de recommandation que nous décrivions plus tôt. Cette stratégie de transparence consiste à présenter le système de recommandation comme une aide à l'utilisateur. Ce dernier est encouragé à faciliter son fonctionnement en laissant des évaluations sur les titres qu'il consomme pour profiter pleinement des services de la plateforme (GOMEZ-URIBE and HUNT, 2016).

Cependant, Netflix® est parfois accusé de nuire à l'agentivité de l'utilisateur, en portant atteinte à ses choix : l'algorithme reproduit et renforce des tendances, et il opère avant toute conscience de choix (SCHAFFNER et al., 2023) de la part de l'utilisateur. De fait, il agit comme une boîte noire qui sélectionne et met en forme des informations sans que l'on sache comment (DRUMOND et al., 2018). On passe d'une aide à la décision, à une contrainte (CHILDS, 2021). 80% des heures visionnées sur la plateforme proviennent de titres recommandés (GOMEZ-URIBE and HUNT, 2016). Selon un article

de CSI Magazine de 2025, 30% des 25-34 ans et 31% des foyers avec jeunes enfants se fient plus aux algorithmes de recommandation qu'aux conseils de leurs amis et familles².

Une étude portée sur 20 consommateurs réguliers (SCHAFFNER et al. (2023)) illustre ce paradoxe : d'un côté, Netflix® est une plateforme de divertissement, les utilisateurs sont peut-être moins méfiants vis-à-vis de ce qu'ils consomment que sur un site d'informations. Plusieurs personnes déclarent même utiliser Netflix® comme un divertissement d'évasion, cédant alors sciemment le choix du contenu à la plateforme. D'un autre côté, la même expérience révèle qu'ils ne sont pas nécessairement satisfaits, et préfèrent parfois des recommandations plus humaines, formulées par leur entourage par exemple. Par ailleurs, deux éléments du design de l'interface contribuent à réduire le sentiment d'agentivité (MOORE, 2016) et ébranle la confiance que l'utilisateur place en la plateforme. Tout d'abord, il y a le bouton *autoplay*, qui apparaît à la fin d'un épisode et lance automatiquement l'épisode suivant. Également, à la fin d'une saison ou d'un film, l'utilisateur fait face à l'absence de signal d'arrêt, puisqu'une bande-annonce est automatiquement lancée.

L'utilisateur est également encouragé à *binge-watcher*, pour accroître son investissement et son attachement. Cette pratique est encouragée par la proposition de séries originales acclamées par la critique (*The Crown*, *Narcos*, *Stranger Things*), par le bouton *autoplay* qui lance automatiquement l'épisode suivant, ainsi que par l'absence de publicité sur certaines plateformes. Ce mode de consommation présente des avantages : amusement et loisir immersif. Mais aussi des inconvénients : fatigue, problèmes de sommeil, problèmes d'attention, sentiment d'isolement, de solitude et de culpabilité vis-à-vis du temps « perdu ». . . (KAUR and ASHFAQ, 2023)

Ce phénomène s'inscrit dans l'économie de la jouissance (DRUMOND et al., 2018) : acquérir de grosses entrées de données, et en échange produire des listes de contenus. Plus l'utilisateur consomme, plus on lui proposera des contenus qu'il sera susceptible d'aimer, et plus il consommera. De fait, il sera plus enclin à poursuivre son abonnement. Le système de recommandation construit ainsi son « utilisateur idéal » (DRUMOND et al., 2018). Ce faisant, la recommandation peut cependant nuire à la diversité.

2.4. Mesure de la (non-)diversité

Le problème de (non-)diversité dans les systèmes de recommandation est un enjeu crucial dans les algorithmes utilisés par des plateformes comme Netflix®, Amazon®, Spotify®, et bien d'autres. En effet, les systèmes de recommandation ont tendance à suggérer des éléments très similaires à ceux que l'utilisateur a déjà consommés. Par exemple, si un utilisateur a regardé plusieurs films d'un même genre ou d'un même réalisateur, l'algorithme pourrait continuer à recommander des films très similaires, négligeant d'autres genres ou réalisateurs. Cela peut réduire les opportunités de découvrir de nouveaux intérêts ou perspectives, ou encore renforcer des biais culturels ou des stéréotypes en favorisant un type de contenu au détriment d'autres, limitant ainsi la représentation et la visibilité de diverses cultures et perspectives (NEGRE, 2015). Afin de pallier cette non-diversité, la plupart du temps, il s'agit d'intégrer explicitement la diversité comme un critère de recommandation, en plus de la pertinence (ZIEGLER et al., 2005). Il existe d'ailleurs dans la littérature, des techniques d'évaluation de la diversité (KUNAVAR and POŽRL, 2017).

2. CSI. (25/03/2025). Streaming algorithms overtake word-of-mouth for viewing choices. CSI. https://www.csimagazine.com/csi/Streaming-algorithms-overtake-word-of-mouth-for-viewing-choices.php?utm_source=chatgpt.com

La littérature distingue deux grandes familles de mesures pour évaluer la diversité dans les systèmes de recommandation : les mesures de diversité locale, centrées sur l'expérience individuelle de l'utilisateur, et les mesures de diversité globale, qui évaluent la capacité du système à exploiter l'ensemble du catalogue et à servir l'ensemble des utilisateurs.

- Mesures de diversité locale : elles quantifient la variété des éléments recommandés au sein d'une même liste proposée à un utilisateur donné.
 - Diversité intra-liste (*Intra-List Diversity*, ILD) : mesure la dissimilarité moyenne entre tous les couples d'éléments d'une liste de recommandations individuelle. Plus cette valeur est élevée, plus la liste est variée pour l'utilisateur (KUNAVAR and POŽRL, 2017).
 - nDCG (*Normalized Discounted Cumulative Gain*) : bien que principalement utilisée pour la pertinence, cette métrique peut être adaptée pour pondérer la diversité locale en tenant compte de la position des éléments diversifiés dans la liste (KUNAVAR and POŽRL, 2017).
- Mesures de diversité globale : elles évaluent la capacité du système à couvrir l'ensemble du catalogue et à fournir des recommandations pertinentes à un large éventail d'utilisateurs.
 - *Item-Coverage* : pourcentage d'éléments du catalogue qui sont recommandés à au moins un utilisateur sur une période donnée. Une couverture élevée indique que le système exploite une plus grande partie de son catalogue (KUNAVAR and POŽRL, 2017).
 - *User-Coverage* : pourcentage d'utilisateurs pour lesquels le système est capable de générer des recommandations pertinentes, reflétant la capacité à servir une population diversifiée (KUNAVAR and POŽRL, 2017).

Dans la suite de cet article, nous nous concentrerons exclusivement sur les mesures de diversité locale, c'est-à-dire celles qui évaluent la variété au sein des listes de recommandations individuelles, en particulier l'ILD. En effet, la diversité locale a un impact direct sur l'expérience utilisateur en particulier du fait du risque de « bulle de filtres »³ qui peut limiter la découverte de contenus (BE MY MEDIA, 2025), ce qui peut réduire l'engagement des utilisateurs (KUNAVAR and POŽRL, 2017).

On constate que Netflix® inclut de la diversité dans ses recommandations en recommandant du contenu populaire (GOMEZ-URIBE and HUNT, 2016). La plateforme pioche dans les contenus populaires, c'est-à-dire les contenus massivement regardés, ceux qui circulent énormément, pour produire de la diversité à l'échelle individuelle.

Dans le cadre des plateformes de streaming musical, TARISSAN (2022) propose d'analyser l'utilisation qui en est faite avant les recommandations (ce qu'il nommera « diversité organique ») puis après, en analysant les effets des algorithmes de recommandation. Il fait les observations suivantes : des profils similaires sont affectés différemment. Pour les profils peu diversifiés, plus il y a de suivi des recommandations, plus les profils se diversifient. Pour les profils déjà diversifiés, le suivi des recommandations déséquilibre la diversité. Pour TARISSAN (2022), la manière dont les recommandations s'inscrivent dans les habitudes d'écoute est plus importante que leur diversité.

Cet état de l'art nous a permis d'aborder le fonctionnement des systèmes de recommandation ainsi que les problématiques qui leur sont liées (contexte, démarrage à froid). Nous avons également abordé la transformation de la consommation de contenu audio-visuel, notamment à travers la question du choix que posent la Télé-internet et la recommandation, ainsi que la pratique du *binge-watching*. Cependant,

3. Situation où l'utilisateur est exposé à des contenus similaires, limitant la diversité de l'information.

nous avons abordé ces questions d'un point de vue informaticien. L'objectif de nos expériences est de comprendre comment l'utilisateur perçoit le système de recommandation, quels effets lui sont visibles et intelligibles. Nous proposons ici d'appliquer notre question à la plateforme de streaming Netflix®.

3. Protocole expérimental

3.1. Objectif

Nous voulons comprendre comment l'utilisateur perçoit le système de recommandation. Pour ce faire, notre objectif est de comprendre quel comportement de l'utilisateur induit quelle réponse du système de recommandation, c'est-à-dire comment ce que l'utilisateur visionne sur Netflix® impacte les recommandations qui lui sont faites.

Nous avons mis au point deux expériences afin d'observer les effets du système de recommandation visibles sur l'interface. Ces deux expériences fonctionnent par la comparaison d'un profil ancien et d'un profil récemment créé, sur lesquels les mêmes contenus sont visionnés. Ensuite, nous comparons les recommandations faites sur chacun des profils. La question du démarrage à froid sera adressée avec la comparaison des deux profils. L'information utilisée pour caractériser le contexte de visionnage sera le moment de la journée (matin, après-midi, soir).

Nous n'avons pas pu automatiser l'expérience, et chaque film a dû être visionné en entier, ce qui signale au système qu'il est apprécié.

3.2. Hypothèses de départ

Au cours de deux expériences, nous allons comparer deux-à-deux 3 profils utilisateurs. Nous utilisons le profil A, ancien (dit « historique »), que nous comparons à un profil récent. Leurs caractéristiques sont détaillées dans la Figure 1.

	Profil A	Profil B	Profil C
Ancienneté	Utilisé régulièrement depuis 7 ans Profil « historique »	Vient d'être créé	
Préférences enregistrées	Préférences issues de l'historique : <ul style="list-style-type: none"> - Habitudes de visionnage : beaucoup de séries comiques, de comédies romantiques et de titres primés, ainsi que de nombreux contenus déjà visionnés Informations indiquées à la création du profil : <ul style="list-style-type: none"> - Français indiqué comme langue - Prénom féminin 	Aucune préférence de contenu n'a été enregistrée : <ul style="list-style-type: none"> - L'étape qui consiste à sélectionner 3 titres appréciés au moment de la création du profil a été sautée Informations indiquées à la création du profil : <ul style="list-style-type: none"> - Français indiqué comme langue - Le prénom choisi pour le profil est neutre, ni masculin, ni féminin, pour ne pas constituer de variable supplémentaire dans les résultats 	
État des recommandations au démarrage des expériences	Correspondent aux habitudes de visionnage	Automatiquement générées par Netflix avant les premiers visionnage	
Utilisé pour les expériences...	Expérience 1 Expérience 2, partie 1 et partie 2 <i>À noter : à l'issue de l'expérience 1, il faut ajouter aux préférences enregistrées le visionnage de films de vampire</i>	Expérience 1	Expérience 2, partie 1 et partie 2

Figure 1. Caractéristiques des différents profils comparés. Ce tableau rassemble les caractéristiques des trois profils utilisés lors des deux expériences.

Avant l'exécution des expériences, nos hypothèses sont les suivantes :

1. **Les recommandations sont davantage marquées par le contenu visionné sur les profils nouveaux.** Les recommandations des profils B et C vont être plus rapidement et plus fortement marquées par le contenu que nous faisons voir à ces profils. « Regarde des films de vampires/ des documentaires » étant la seule information donnée au système de recommandation, il basera la plupart de ses recommandations là-dessus. Le profil A sera peu modifié au cours de l'expérience, le visionnage de films de vampires/ de documentaires/ de films d'animation jeunesse pouvant être considéré comme anecdotique. Le profil A n'ayant jamais regardé ces types de contenus, il est peu probable que ces recommandations soient nombreuses face aux poids d'un apprentissage de 7 ans nourri aux séries comiques et aux comédies romantiques.
2. **Au démarrage, les recommandations sont des choix populaires et non personnalisés.** Les profils B et C auront beaucoup de recommandations populaires, de genres différents, mais aucune recommandation personnalisée. En effet, lorsque le système de recommandation n'a aucune information sur l'utilisateur, il lui propose du contenu populaire (MADDODI and KARANI, 2019).
3. **Un profil récemment créé est plus durablement marqué par les contenus visionnés qu'un profil ancien.** Lors de la partie 2 de l'expérience 2, le profil A « oubliera » vite le documentaire, tandis que le profil C s'en verra toujours proposer. Puisque le visionnage de films de vampires n'aura duré que 5 jours, le système de recommandation pourra comprendre que le visionnage de documentaires sera lui aussi éphémère. Au contraire, pour le profil C, regarder des documentaires constituera son seul historique, aucun élément ne pourra encore relativiser son importance dans les habitudes de visionnage du profil C.

3.3. Expérience 1

L'expérience 1 nous a servi d'étalon pour définir les paramètres de notre expérience et les données significatives à relever. Nous comparons le profil A, historique, et le profil B, récent (voir Figure 1).

À chacun des deux profils, nous faisons regarder les mêmes films, des films de vampires, en entier. Nous lançons manuellement le visionnage de contenu, 1 à 3 programmes par jour. Ne disposant que d'un appareil, le visionnage d'un même titre se fait à un jour d'intervalle mais à des heures similaires (matinée, après-midi, soirée...). Par exemple, *Survivre au paradis*, est visionné le mercredi 8 mai 2024 à 23h41 sur le profil A, et jeudi 9 mai 2024 à 23h32 sur le profil B.

Ensuite, nous regardons, après visionnage sur chaque profil, les recommandations faites par les algorithmes de Netflix®. Nous récoltons les données de recommandation accessibles à l'utilisateur sur sa page d'accueil, rien de plus que ce qui ne lui est explicitement montré. Cette étude se fait également à un jour de décalage pour chaque profil, toujours à des moments similaires de la journée, et avant le visionnage du film suivant.

Pour une version schématisée du déroulé de l'expérience 1, voir Figure 2.

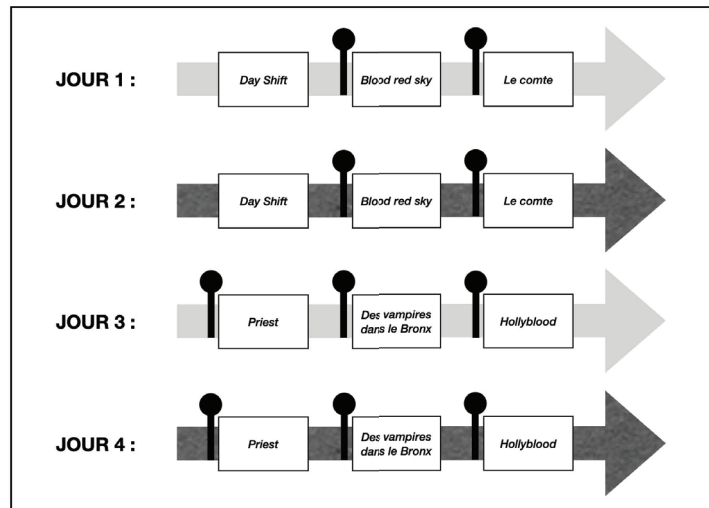


Figure 2. Déroulé schématique de l'expérience 1. Ce schéma montre le déroulé sur 4 jours de l'expérience 1, avec les visionnages de films suivis des relevés de recommandations sur les profils A et B. Les flèches représentent les profils : flèche claire = profil A, flèche foncée = profil B. Chaque rectangle indique le titre du film visionné, et chaque symbole en forme de repère vertical noir correspond à un relevé des recommandations effectué après le visionnage.

3.4. Expérience 2

Nous avons construit une seconde expérience, en deux parties. L'enjeu était toujours de comparer des profils, un ancien (le profil A) et un nouveau (le profil C), pour observer par comparaison les effets visibles du système de recommandation (voir Figure 1).

Cette expérience se déroule en deux parties. Dans la première partie, nous faisons regarder des documentaires animaliers aux deux profils. Dans la seconde partie, nous leur faisons regarder des films d'animation jeunesse.

Pour chaque partie de l'expérience 2, nous répétons le même protocole que dans l'expérience 1, c'est-à-dire que nous faisons regarder les mêmes films en entier aux deux profils, à un jour d'intervalle. Nous récoltons ensuite les données qui concernent les recommandations.

3.5. Compléments

Les genres cinématographiques choisis pour les deux expériences sont détaillés dans la Figure 3. Nous avons extrait les informations partagées par Netflix® à propos des films visionnés : âge conseillé pour le public, durée, genres, caractéristiques.

Pour définir les contenus visionnés sur chaque profil, nous avons choisis trois genres cinématographiques précis : les films de vampires, les documentaires animaliers et l'animation jeunesse. Ce choix s'explique pour trois raisons :

- Tout d'abord, le profil ancien (profil A) n'en a quasiment jamais visionnés, leur recommandation ne pouvait donc qu'être la conséquence de leur visionnage au cours de notre expérience.
- Ensuite, ces catégories précises nous permettent d'établir un échantillon de films à visionner le plus homogène possible.

- Enfin, ces catégories sont suffisamment précises pour que nous distinguions les recommandations les plus personnalisées mais aussi les tentatives de diversifications. Par exemple, recommander un film de vampire constituerait une recommandation extrêmement personnalisée, quand recommander un film d'horreur (sans vampires) ferait état d'une sensible diversification, tout en restant proche du contenu visionné. Ce choix nous permet ainsi de nuancer la personnalisation.

	Age	Durée	Genres rattachés	Caractéristiques
Film de vampires (9)	+16 (6) +13 (3)	Plus court : 78 minutes Plus long : 123 minutes Moyenne : 97 minutes	Film d'horreur (9) Comédie (5) Comédie d'horreur (4) Film pour ados (1) Horreur avec des ados (1) Horreur de série B (1) Film fantastique (1)	Effrayant (5) Macabre (3) Gore (3) Vampires (3) Sombre (2) Cérébral (1) Violent (1) Absurde (1) Insolite (1) Surnaturel (1) Dystopique (1)
Documentaire animalier (7)	+7 (6) Tous publics (1)	Plus court : 44 minutes Plus long : 75 minutes Moyenne : 64 minutes Hors épisodes de 50 minutes Moyenne avec : 62 minutes	Film documentaire (6) Film jeunesse et famille (4) Documentaire nature et écologie (3) Docu sciences et nature (1) Série à voir en famille (1) Émissions sciences et nature (1)	Captivant (5) Pour la famille (4) Documentaire (2) Les abysses (2) Nature et écologie (2) Relaxant (2) Nos amis les animaux (1) Enquête (1) Émotion (1)
Film d'animation jeunesse (11)	+7 (7) Tous publics (3) +10 (1)	Plus court : 85 minutes Plus long : 108 minutes Moyenne : 93 minutes	Film pour la famille (11) Film jeunesse et famille (11) Aventure en famille (5) Comédie (3) Grandir en musique (2) Comédie d'action (2) Film inspiré d'une pièce de théâtre (1) Film hollywoodien (1)	Jeunesse (7) Magique (3) Préhistoire (3) Exaltant (2) Onirique (2) Musique pour les enfants (2) Authentique (2) Nommé aux Annie Awards (2) Premier amour (1) Destin de stars (1) Rivalité (1) New York (1) Dinosaure (1) Changement climatique (1) Crise de la quarantaine (1) Liens familiaux (1)

Figure 3. Distinction et caractéristiques des genres cinématographiques. Ce tableau rassemble les informations partagées par Netflix® sur chacun des films visionnés, rassemblés par genre. Le chiffre qui suit une information indique combien de films elle concerne.

Pour voir le déroulé des deux expériences, voir les Figures 4 et 5.

	Genres cinématographiques	Dates	Nombres de films vus	Horaires des récoltes des recommandations	
Expérience 1	Films de vampires	25 avril au 1er mai 2024	9	Début de journée :	11h20 à 14h42
				Début de soirée :	19h01 à 21h37
				Fin de soirée :	22h à 23h50
Expérience 2	Partie 1 Documentaires animaliers	4 mai au 13 mai 2024	6 + 1 série de 12 épisodes	Matin :	9h16 à 12h33
	Partie 2 Films d'animation jeunesse	23 mai au 6 juin 2024	11	Après-midi :	16h20 à 18h27
				Soir :	20h23 à 00h31

Figure 4. Déroulé des deux expériences. Ce tableau récapitule les caractéristiques des 2 expériences : dates, nombres de titres visionnés, modalités de récoltes des recommandations.

Chronologie des deux expériences

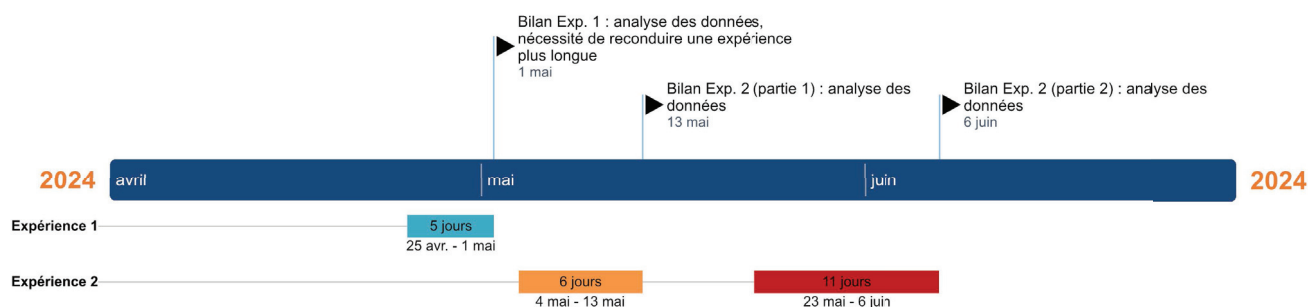


Figure 5. Chronologie des expériences. Ce diagramme illustre la chronologie des deux expériences, entre les mois d'avril et de juin 2024. En bleu, l'expérience 1, en orange la première partie de l'expérience 2, et en rouge la partie 2 de l'expérience 2.

3.6. Méthode d'analyse

Au cours de l'analyse des données, nous effectuons quatre mesures :

- **Mesure 1 :** Les rangs de catégories dans lesquels sont présentés horizontalement quelques titres du catalogue. Ces rangs sont comptés, mais aussi analysés pour leur formulation. Nous cherchons en particulier à savoir comment évolue le nombre de catégories horreur/documentaire/jeunesse proposées.
- **Mesure 2 :** Le nombre de films d'horreur/documentaires recommandés dans une catégorie personnalisée : « Notre sélection du jour pour vous » ou « Notre sélection pour *profil A/profil B* ». Dans la partie 2 de l'expérience 2, nous n'avons pu relever le nombre de titres d'animation jeunesse dans une catégorie personnalisée, puisque aucune n'est apparue sur la page d'accueil.
- **Mesure 3 :** La tête de gondole, le programme présenté en grand, tout en haut de la page d'accueil, et dont la bande-annonce se lance automatiquement à l'ouverture d'une session. Il s'agit de déterminer si le titre présenté en tête de gondole appartient à l'un des genres cinématographiques de notre expérience.
- **Mesure 4 :** La mesure de la (non-)diversité des recommandations. Nous nous focalisons sur l'ILD (Diversité Intra-Liste) qui s'appuie sur la Similarité Intra-Liste (ILS). Pour une liste de n éléments $\{i_1, i_2, \dots, i_n\}$, la formule de l'ILS avec une similarité binaire (nous faisons le choix d'une similarité binaire pour indiquer que les films sont exactement les mêmes ou pas) s'écrit :
$$ILS = \frac{2}{n(n-1)} \sum_{1 \leq k < l \leq n} \mathbb{1}_{\{i_k = i_l\}}$$
 où $\mathbb{1}_{\{i_k = i_l\}}$ vaut 1 si i_k et i_l sont identiques, 0 sinon. L'ILD correspond alors à : $ILD = 1 - ILS$.

Nous allons donc tout d'abord porter attention aux catégories proposées sur la page d'accueil de chaque profil. Netflix® détaille le fonctionnement de divers algorithmes qui constituent son système de recommandation (GOMEZ-URIBE and HUNT, 2016). L'un d'entre eux a pour mission de générer une page d'accueil spécifique à chaque profil. Il existe alors plusieurs catégories de classement :

- Celles de genre (« Drames », « Comédies sur l'amitié », « Comédies romantiques de Noël »...),
- Celles de similarité (« Parce que vous avez regardé [titre] »),
- Celles de popularité (« Les plus recherchés »).

Ces catégories ne sont pas exclusives : la série *Fiasco* apparaît dans « Nouveautés sur Netflix » ainsi que dans « Trouvez votre prochain coup de coeur ».

Ainsi, des milliers de combinaisons sont possibles, et l'algorithme a pour but de sélectionner les plus susceptibles de plaire à l'utilisateur, tout en favorisant une certaine diversité. Il permet également à l'utilisateur de sauter directement des catégories qui en théorie pourraient lui plaire, mais ne correspondent pas à ce qu'il veut regarder à cet instant. De plus, puisqu'il n'y a pas de modèle défini, les combinaisons de catégories sont infinies : un utilisateur peut se voir présenté trois catégories de genre, une catégorie de popularité et deux catégories de similarité, tandis qu'un autre n'aura que des catégories de genre.

De plus, ces catégories peuvent se combiner : « Séries saluées par la critique à regarder sans modération » ; « Drames européens inspirés de livres et salués par la critique » par exemple.

Pour simplifier la lecture des données, tout en conservant les spécificités des catégories qui sont apparues au cours de l'expérience, nous proposons la classification suivante :

- **Personnalisation** : Ces catégories, avec des formulations telle que « Notre sélection du jour pour vous », nous permettent de mesurer, de la façon la plus évidente, l'impact des visionnages de l'utilisateur sur ses recommandations.
- **Popularité** : La popularité devrait fluctuer selon le taux de personnalisation (GOMEZ-URIBE and HUNT, 2016).
- **Nouveauté/exclusivité/binge** : Les catégories d'exclusivité, de nouveauté et de binge-watching sont rassemblées. D'après DRUMOND et al. (2018), elles représentent à elles trois un usage idéal de la plateforme du point de vue de l'économie du système. Avec la catégorie nouveauté, l'utilisateur fait le constat du renouvellement régulier du catalogue, ce qui l'encourage à poursuivre son abonnement. Avec la catégorie exclusivité, il découvre des contenus qu'il ne retrouverait nulle part ailleurs, et Netflix® s'assure de la promotion de ces contenus autoproduits. Avec la catégorie binge, l'utilisateur est encouragé à consommer massivement des séries, pour éprouver l'intérêt de son abonnement et le poursuivre. De plus, les combiner est également un moyen de voir de façon plus significative l'évolution de ces données.
- **Primés/représentation** : Les catégories primées et représentation sont rassemblées parce qu'elles n'apparaissent que rarement au cours de l'expérience, les combiner nous permet alors de réduire la quantité de données présentées. Il s'agit des catégories mettant en valeur des films ou séries récompensés, ainsi que des productions volontairement mises en avant par la plateforme par souci de représentation, comme « Femme derrière la caméra » ou « Histoires et talents noirs ».
- **Genres** : Le détail des catégories de genre devrait nous permettre de mesurer l'influence des contenus visionnés par les profils.

Enfin, pour estimer la diversité des recommandations au cours de nos expériences, nous avons compté le nombre de titres différents proposés à l'ouverture d'une session. Nous comparons ce chiffre pour le profil A et pour les profils B et C afin de comparer la diversité d'un profil ancien avec celle d'un nouveau.

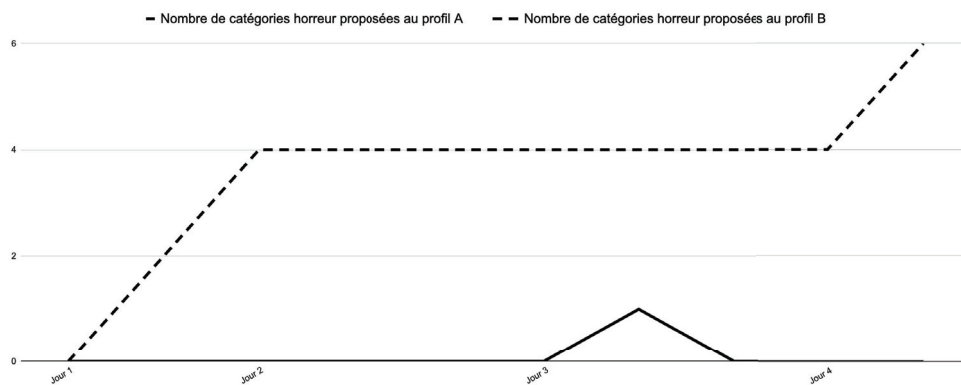


Figure 6. Évolution du nombre de catégories relatives au genre de l'horreur proposées sur la page d'accueil, au fil des jours. Ce graphique montre l'évolution du nombre de catégories horreur proposées à un utilisateur à l'ouverture de la page d'accueil. Ce nombre peut varier de 0 à 6, et notre étude se déploie sur 4 jours.

4. Résultats

4.1. Expérience 1 : films de vampires (Profils A et B)

Pour cette analyse, nous considérons comme recommandations liées aux films de vampires les catégories personnalisées relatives aux films visionnés (« Parce que vous avez regardé *The Priest* ») ainsi que toutes les catégories de genre liées à l'horreur, aux films de monstres, et aux thrillers fantastiques. Les films relevant de ces catégories ont été identifiés grâce aux mots clés « Genres rattachés » et « Caractéristiques » indiqués sur les pages de présentation des films sur la plateforme. Pour faciliter la lecture, nous nous y référerons en tant que films d'horreur.

Nombre de catégories horreur

Nous regardons combien de catégories relatives à l'horreur apparaissent (voir Figure 6). Pour le profil A, une seule catégorie horreur apparaît au jour 3. Pour le profil B, le nombre de catégories horreur augmente progressivement les deux premiers jours, puis stagne. Entre le jour 3 et le jour 4, ce nombre augmente à nouveau pour atteindre un pic de 6 catégories horreur au jour 4, dernier jour de cette première expérience. Finalement, il faut 3 jours de visionnage de films de vampires pour que le système recommande des films d'horreur au profil A (connu) tandis que le profil B (nouveau) s'en voit recommander dès le premier jour.

Nombre de titres horreur dans une catégorie personnalisée

Ensuite, nous regardons dans une catégorie personnalisée, « Notre sélection pour profil A/B », combien des 6 vignettes visibles à l'ouverture de la page d'accueil représentent des films d'horreur (voir Figure 7).

Différence à noter, au début de l'expérience 1, aucun des titres proposés dans la catégorie « Notre sélection pour profil A » ne correspond à un film d'horreur. Au contraire, le profil B compte 4 titres d'horreur sur les 6 présentés. Tout au long de l'expérience 1, un écart d'au moins 2 titres perdure entre les deux profils.

Les deux courbes suivent quasiment la même progression : une période de stagnation, puis une hausse suivie d'une baisse. C'est cette dernière qui est la plus suprenante, pourquoi diminuer le nombre de recommandations horreur alors que l'utilisateur est encore en train d'en regarder ?

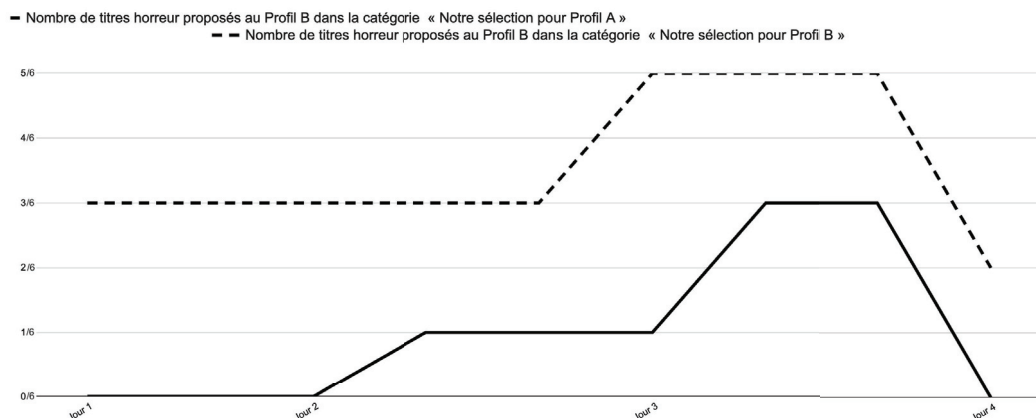


Figure 7. Évolution du nombre de titres horreur proposés dans la catégorie « Notre sélection pour profil A/profil B ». Sur les 6 vignettes visibles à l'ouverture de la page d'accueil, nous regardons lesquelles correspondent à des films ou des séries d'horreur. Nous regardons l'évolution sur 4 jours.

Occurrence de l'horreur en tête de gondole

Enfin, nous nous intéressons à la tête de gondole qui apparaît à l'ouverture d'une nouvelle session. Notre étude se porte sur le genre du titre proposé, avec une attention particulière portée aux titres horribles (voir Figure 8). Les recommandations proposées peuvent être distinguées en plusieurs catégories, selon que le titre soit :

- un film de vampires,
- un film de monstres (hors vampires), un film d'horreur ou un thriller fantastique, rassemblés en « films d'horreur »,
- un autre genre de film.

Le profil A n'a eu que des recommandations d'autres types de titre en début de journée, 25% de films de vampires en début de soirée, et 70% de films d'horreur en fin de soirée.

Le profil B s'est vu recommander 50% de films de vampires et 50% de films d'horreur en début de journée, 100% de titres autres en début de soirée, et 70% de films d'horreur en soirée.

Les titres de vampires et d'horreur sont plus présentés en tête de gondole au profil B, en particulier le matin.

Diversité

Nous nous intéressons à mesurer la (non-)diversité de nos recommandations. Comme déjà souligné, nous nous concentrons sur la diversité locale, i.e centrée sur l'expérience individuelle de l'utilisateur. De manière assez triviale, nous calculons dans un premier temps la proportion de titres uniques proposés à chaque utilisateur/profil.

- Pour le profil A, sur 234 titres proposés, 219 n'apparaissent qu'une seule fois, i.e. 93,5% de titres uniques.
- Pour le profil B, sur 228 titres proposés, 216 n'apparaissent qu'une seule fois, i.e. 94,7% de titres uniques.

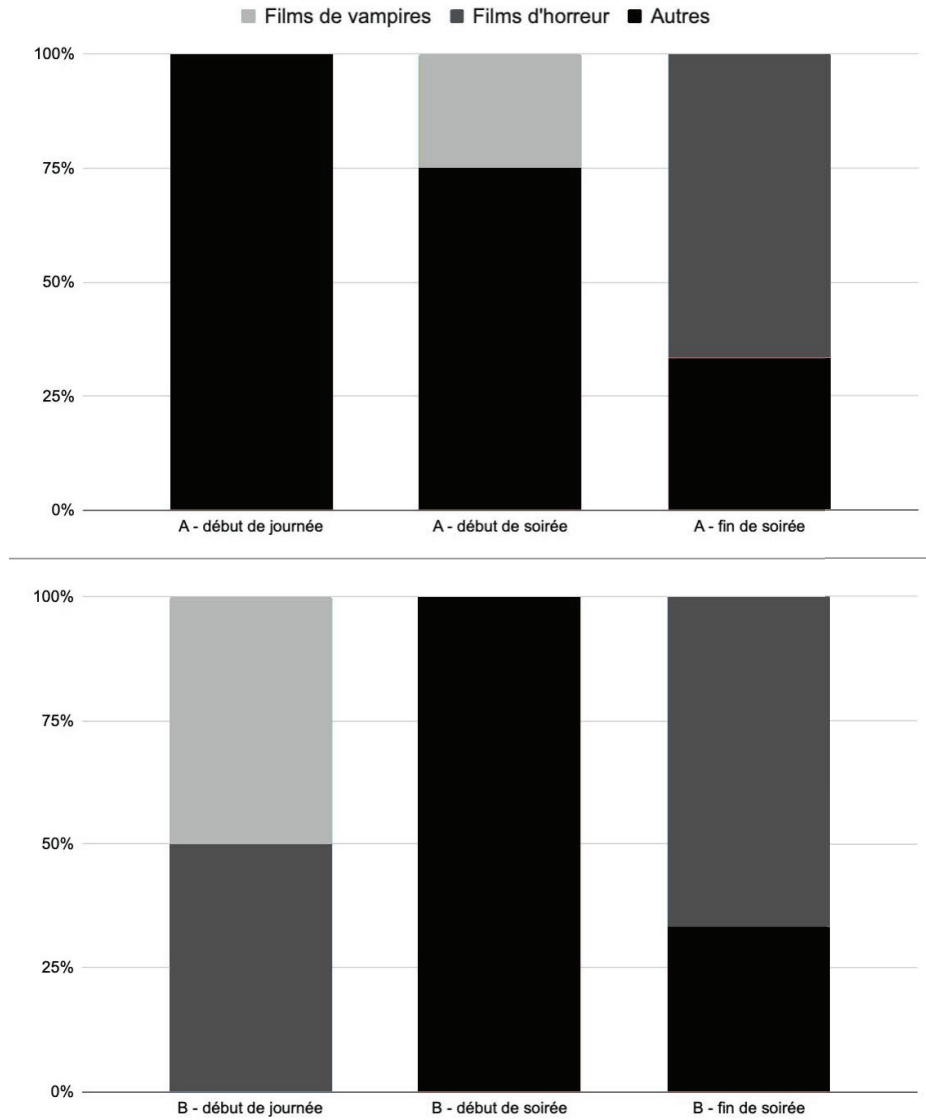


Figure 8. Variation de la part des programmes d'horreur et autres. Cet historiogramme détaille les parts de programmes d'horreur et d'autres genres proposés en tête de gondole. Il mesure leurs variations selon le moment de la journée (matin, début de soirée, soir), et entre les profils A et B.

Les films proposés plusieurs fois apparaissent 2 fois, parfois 3, et sont souvent des titres populaires et/ou des ajouts récents. La mesure de la proportion de titres uniques dépasse 93% pour les 2 profils, cela montre une diversité locale conséquente.

Pour le profil A, la liste de recommandations comporte $n^A = 234$ titres. On observe que 8 titres apparaissent plusieurs fois, avec les occurrences suivantes : 3 occurrences pour 2 titres, et 2 occurrences pour 6 titres. Le nombre total de paires d'éléments identiques est donc : $N_{\text{identiques}} = \sum_{t \in \text{titres}} \frac{k_t(k_t-1)}{2}$ où k_t est le nombre d'occurrences du titre t . Dans notre cas : $N_{\text{identiques}}^A = 12$. Le nombre total de paires possibles est : $N_{\text{paires}}^A = \frac{234 \times 233}{2} = 27\,261$. On obtient alors : $ILS^A = \frac{12}{27\,261} \approx 0.00044$ et $ILD^A = 1 - 0.00044 = 0.99956$

Pour le profil B, la liste de recommandations comporte $n^B = 228$ titres. On observe que 11 titres apparaissent plusieurs fois, avec les occurrences suivantes : 3 occurrences pour 1 titre, et 2 occurrences pour 10 titres. Ainsi $N_{\text{identiques}}^B = 13$ et $N_{\text{paires}}^B = 25\,878$. On obtient alors $ILS^B \approx 0.000502$ et $ILD^B = 0.999498$

Pour les profils A et B, la valeur proche de 1 de l'ILD indique une diversité locale quasi maximale dans la liste de recommandations de chacun des profils, au sens strict de l'absence de doublons.

Autres observations

Nous avons également fait les observations suivantes :

- Pour le profil A, le nombre de catégories de genre reste constant, mais les genres en question se diversifient.
- Pour le profil A, une seule catégorie horreur est apparue (au jour 3, voir Figure 6), mais des titres horribles font leur apparition dans d'autres catégories. Par exemple, *Dracula* apparaît dans « Séries internationales doublées en français ».
- Pour le profil A, le nombre de catégories de réalisations primées passent de 9 au début de l'expérience à 7 à la fin.
- Pour le profil B, on découvre au démarrage 23 catégories de genres, 4 catégories populaires, mais aussi 2 catégories personnalisées, ce qui nuance notre deuxième hypothèse qui postulait qu'il aurait une grande diversité de genre, de la popularité mais pas de la personnalisation.
- Pour le profil B, les catégories de popularité proposent aussi, au fur et à mesure, de plus en plus de titres horribles ou de science-fiction.

Cependant, aucune dynamique significative ne ressort de cette étude. Ayant seulement été déployée sur 4 jours, nous disposons de trop peu de données pour une analyse pertinente.

Cette première expérience nous a tout de même permis de bâtir notre protocole d'analyse et de fixer les catégories. Par exemple, pour l'étude des têtes de gondole (voir Figure 8), nous ajoutons pour l'expérience suivante la classification « Ajout récent ». Si un documentaire sorti récemment est présenté en tête de gondole, sa recommandation est peut-être due à sa nouveauté et non à notre protocole de visionnage, nous choisissons donc de les distinguer.

L'expérience 1 est finalement une expérience préliminaire nous permettant de mettre au point un protocole efficace pour la deuxième expérience.

4.2. Expérience 2 (Profils A et C)

Pour cette seconde expérience, qui se déroule en deux temps, nous comparons le profil A historique, déjà utilisé lors de l'expérience 1, avec un nouveau profil naïf, le profil C qui vient d'être créé.

Partie 1 : documentaires animaliers

Pour cette analyse, en plus des catégories documentaire, nous considérons comme recommandation liées aux documentaires animaliers les catégories personnalisées liées aux films visionnés (« Parce que vous avez regardé *Notre planète* ») ainsi que toutes les catégories proche documentaire, comme les « Émissions TV sur l'écologie ». Pour faciliter la lecture, nous nous y référerons en tant que catégorie documentaire.

Nombre de catégories documentaire

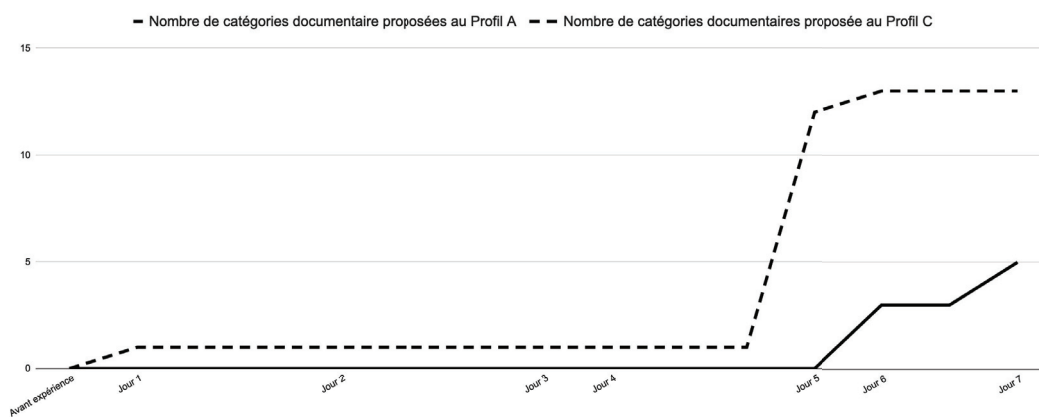


Figure 9. Évolution du nombre de catégories relatives au genre du documentaire proposées sur la page d'accueil, au fil des jours. Ce graphique montre l'évolution du nombre de catégories documentaire proposées à un utilisateur à l'ouverture de la page d'accueil. Ce nombre peut varier de 0 à 14, et notre étude se déploie sur 7 jours.

Nous regardons l'évolution du nombre de catégories documentaire (voir Figure 9).

Pour les deux profils, le nombre de catégories documentaire augmente quasiment au même moment, mais cette augmentation est plus importante pour le profil C (pour rappel, le profil C est un nouveau profil).

Pour le profil A, le nombre de catégories documentaire reste à 0 jusqu'au jour 5. Au jour 7, nous comptons 5 catégories documentaire.

Pour le profil C, s'il n'y a aucune catégorie documentaire au démarrage de l'expérience, on en dénombre 1 dès le premier jour. Nous dénombrons 13 catégories documentaire à la fin de l'expérience.

Le système met donc 5 jours pour recommander des documentaires au profil A tandis qu'il en recommande au profil C dès le premier jour.

Nombre de titres documentaire dans une catégorie personnalisée

En plus des catégories de genre spécialisées dans le documentaire, les recommandations de documentaires se font également par le biais des catégories exclusivité ou personnalisées, comme « On pense que vous allez adorer. . . », ou encore « Notre sélection du jour pour vous ».

Nous nous intéressons ici au nombre de documentaires sur les 6 titres présentés dans cette catégorie (voir Figure 10).

Pour les deux profils, le nombre de documentaires suggérés augmente progressivement jusqu'au jour 4, avant de diminuer au jour 5.

Nous observons une courbe similaire à celle de l'expérience 1, mais l'écart est moindre entre les profils A et C.

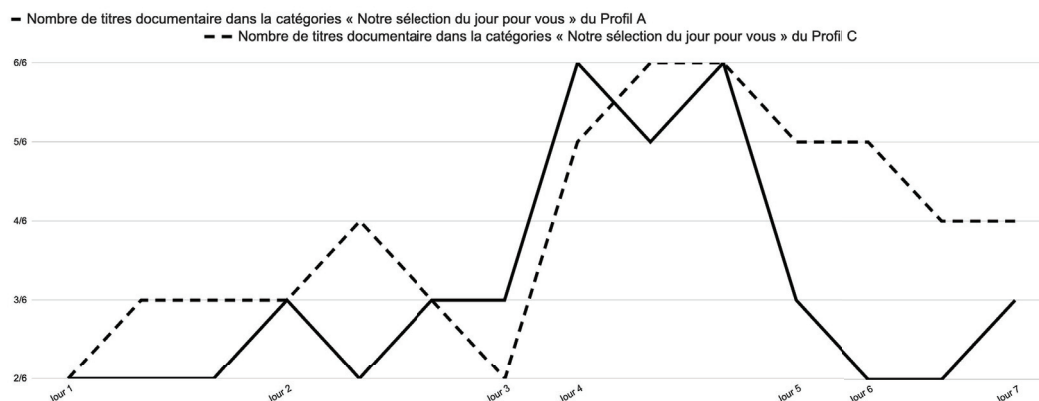


Figure 10. Évolution du nombre de titres documentaire proposés dans la catégorie « Notre sélection du jour pour vous ». Sur les 6 vignettes visibles à l'ouverture de la page d'accueil, nous regardons lesquelles correspondent à des documentaires. Nous regardons l'évolution sur 7 jours.

Occurrence du documentaire en tête de gondole

Enfin, nous nous intéressons aux occurrences de documentaires animaliers en tête de gondole pour chaque profil selon le moment de la journée (voir Figure 11). Les recommandations proposées peuvent être distinguées en plusieurs catégories :

- le titre est un documentaire **animalier**,
- le titre est un documentaire (hors documentaire animalier),
- le titre est un documentaire mais est également un ajout récent au catalogue Netflix®,
- le titre n'est pas un documentaire.

Pour le profil A, 75% des têtes de gondole proposées le matin sont des documentaires, dont 25% sont des documentaires animaliers. La part de documentaire tombe à 50% l'après-midi, avec autant de documentaires animaliers que d'autres types de contenus. Le soir, nous comptons 75% de programmes d'autres genres.

Pour le profil C, à chaque moment de la journée sont proposés 50% de documentaires animaliers. Le matin, toutes les têtes de gondole proposent des documentaires, et d'autres types de contenus sont proposés l'après-midi et le soir.

Pour le profil C, 50% des recommandations sont des documentaires animaliers quel que soit le moment de la journée, tandis que pour le profil A, la recommandation de documentaires animaliers peine à atteindre les 25%.

Diversité

Nous mesurons de nouveau ici la (non-)diversité locale des recommandations. Nous calculons dans un premier temps la proportion de titres uniques proposés à chaque utilisateur/profil.

- Pour le profil A, sur 228 titres proposés, 212 n'apparaissent qu'une seule fois, i.e. 92,9% de titres uniques.
- Pour le profil C, sur 219 titres proposés, 202 n'apparaissent qu'une seule fois, i.e. 92,23% de titres uniques.

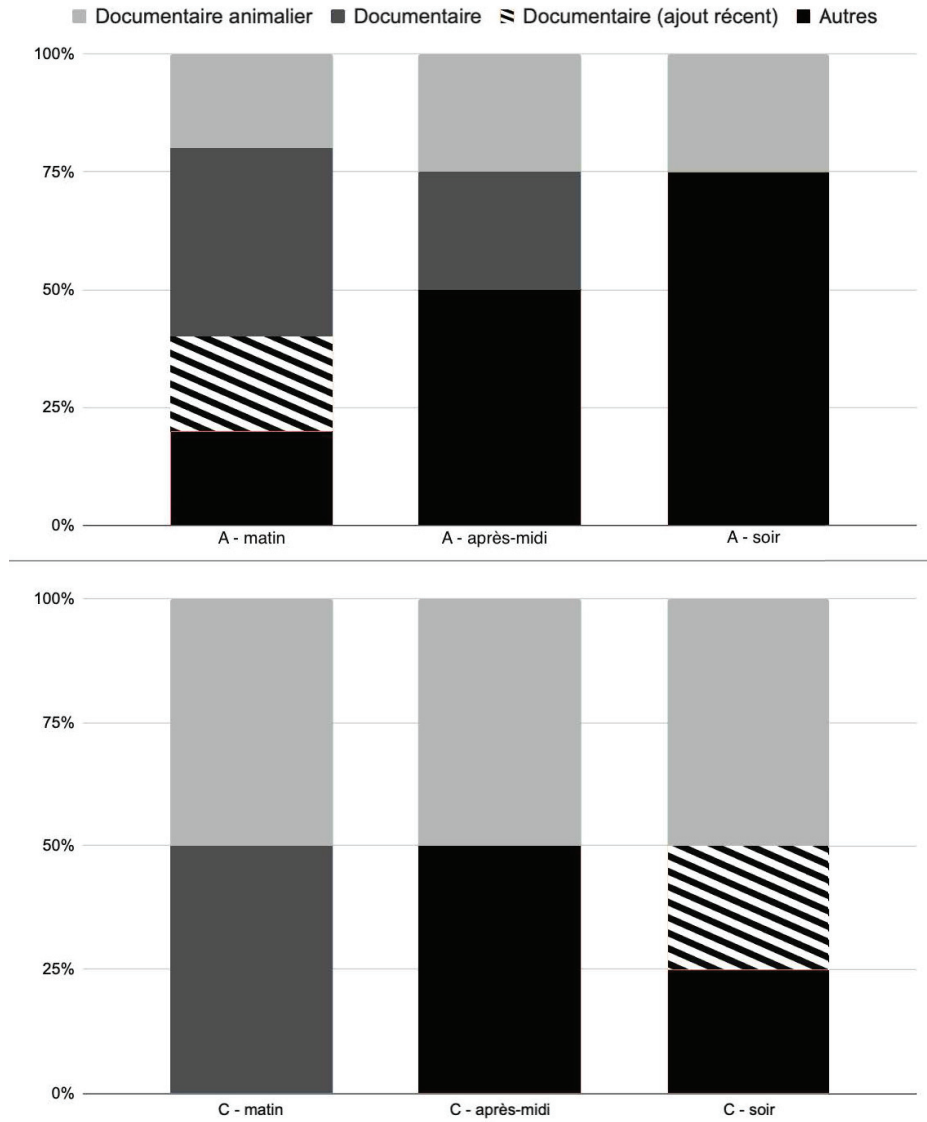


Figure 11. Variation de la part des programmes documentaire et autres. Cet historiogramme détaille les parts de programmes documentaire et autres proposés en tête de gondole. Il mesure leurs variations selon le moment de la journée (matin, début de soirée, soir), et entre les profils A et C.

La mesure de la proportion de titres uniques dépasse 92% pour les 2 profils, cela montre une diversité locale conséquente.

Nous calculons également l'ILD pour chaque profil (avec une similarité binaire).

Pour le profil A dans l'expérience 2, la liste de recommandations comporte $n^{A_2} = 228$ titres. $N_{\text{identiques}}^{A_2} = 11$ et $N_{\text{paires}}^{A_2} = 25\,878$. On obtient alors : $ILS^{A_2} = \frac{11}{25\,878} \approx 0.00043$ et $ILD^{A_2} = 1 - 0.00043 = 0.99957$

Pour le profil C, la liste de recommandations comporte $n^C = 219$ titres. $N_{\text{identiques}}^C = 10$ et $N_{\text{paires}}^C = 23\,871$. On obtient alors $ILS^C \approx 0.00042$ et $ILD^C = 0.99958$

Pour les profils A et C, la valeur proche de 1 de l'ILD indique une diversité locale quasi maximale dans la liste de recommandations de chacun des profils, au sens strict de l'absence de doublons.

Partie 2 - animation jeunesse

Pour cette analyse, nous considérons comme recommandation liées à l'animation jeunesse les catégories personnalisées liées au visionnage de titres jeunesse (« Parce que vous avez regardé *Leo* » par exemple). Nous comptabilisons les catégories proposant de l'animation jeunesse (films ou séries), mais aussi les catégories de séries et de films en live-action tant que ceux-ci sont adressés à la jeunesse.

Nombre de catégories jeunesse

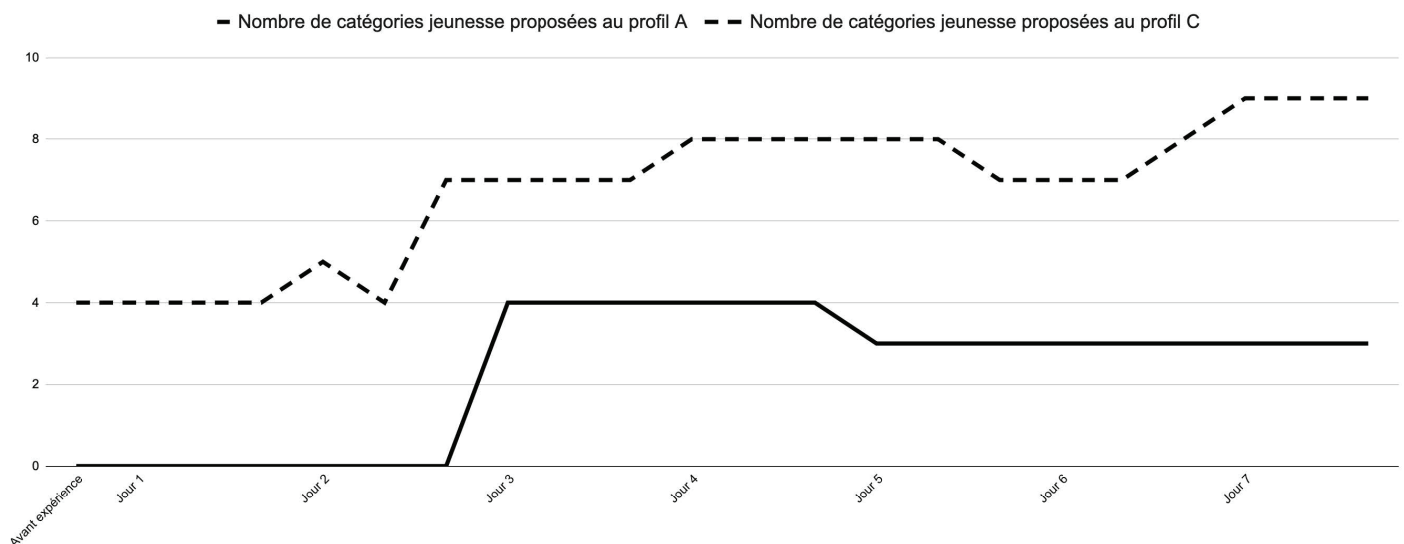


Figure 12. Évolution du nombre de catégories relatives au dessin-animé jeunesse proposées sur la page d'accueil, au fil des jours. Ce graphique montre l'évolution du nombre de catégories animation jeunesse proposées à un utilisateur à l'ouverture de la page d'accueil. Le trait continu correspond au nombre de catégories animation jeunesse proposées au profil A, tandis que le trait en pointillés correspond au profil C. Ce nombre peut varier de 0 à 9, et notre étude se déploie sur 7 jours.

Nous commençons par regarder le nombre de catégories jeunesse qui apparaissent au cours de l'expérience (voir Figure 12).

À nouveau, les courbes des profils A et C sont très similaires : le nombre de catégories jeunesse augmente autour du jour 3 puis varie légèrement jusqu'à la fin de l'expérience.

Une différence demeure entre les deux profils : aucune catégorie jeunesse n'est proposée au profil A jusqu'au jour 3, quand le profil C s'en voit proposé 4 dès le jour 1. Un écart d'au moins 3 titres se maintient tout au long de l'expérience.

Occurrence de l'animation jeunesse en tête de gondole

À nouveau, nous cherchons à savoir si le moment de la journée a une influence sur la tête de gondole, proposée tout en haut de la page d'accueil (voir Figure 13). Les recommandations proposées peuvent être distinguées en plusieurs catégories, selon que le titre soit :

- un film d'animation jeunesse,
- un programme jeunesse (hors animation),

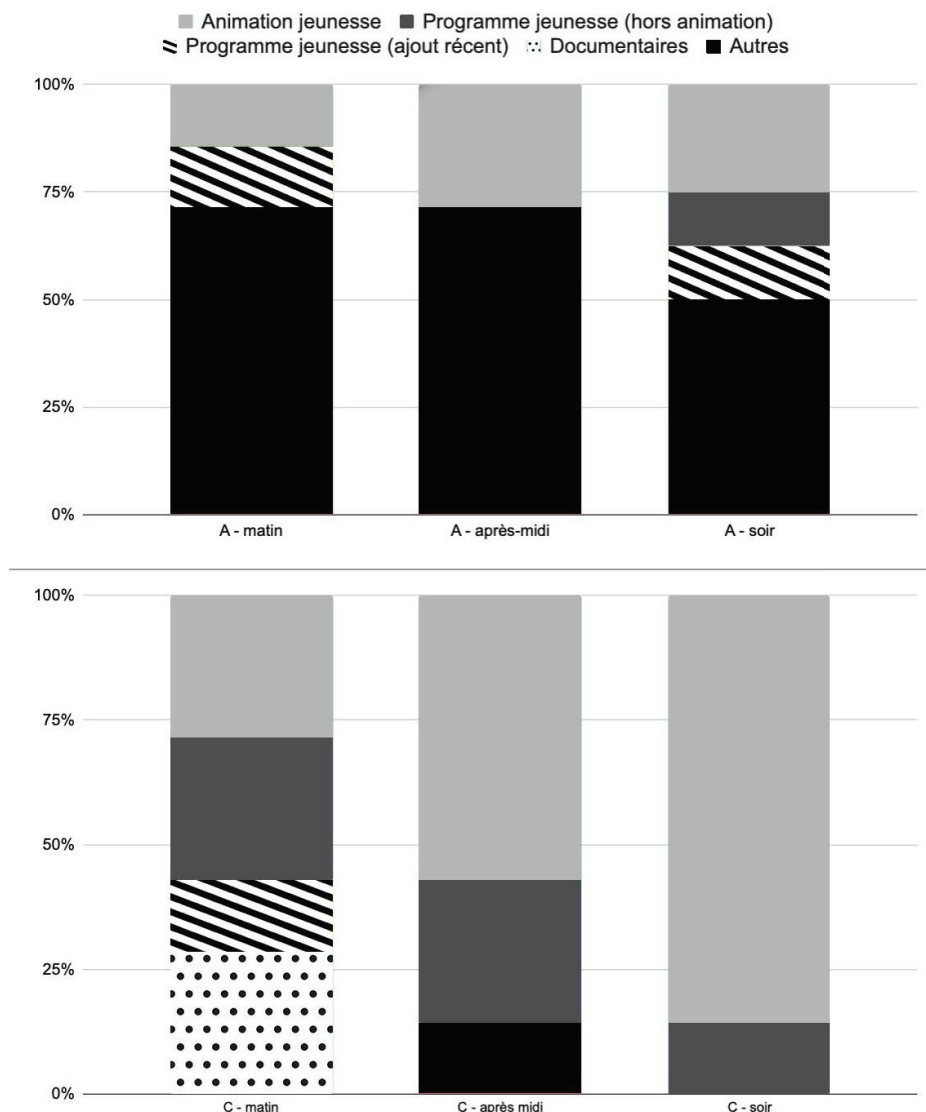


Figure 13. Variation de la part des programmes jeunesse et autres. Cet historiogramme détaille les parts de programmes jeunesse et d'autres genres proposés en tête de gondole. Il mesure leurs variations selon le moment de la journée (matin, début de soirée, soir) et entre les profils A et C.

- un programme jeunesse mais est également un ajout récent au catalogue Netflix®,
- un documentaire,
- un titre qui appartient à une autre catégorie.

Qu'importe le moment de la journée, au cours des 7 jours d'expérience, le profil A s'est vu recommander un peu plus de 20% d'animation jeunesse. La majorité des recommandations concernaient d'autres types de contenus : environ 65% le matin, presque 80% l'après-midi et un peu plus de 50% le soir. Si nous comptons les films d'animation, les programmes jeunesse et les programmes jeunesse ajoutés récents, la part de recommandation jeunesse s'élève à près de 50% le soir.

Pour le profil C, la part de recommandation jeunesse augmente tout au long de la journée. Le matin, nous comptons un total de 80% de programmes jeunesse, dont environ 50% d'animation. Les 20% restant sont occupés par le documentaire. L'après-midi, plus de 65% des recommandations sont de l'animation jeunesse, auxquels s'ajoutent environ 20% de programmes jeunesse. Le soir enfin, 80% des titres recom-

mandés correspondent à de l'animation jeunesse. Les 20% restant sont occupés à parts égales par les programmes jeunesse hors animation et les programmes jeunesse récents.

Pour cette deuxième partie, nous n'avons pas pu étudier le nombre de titres correspondant à de l'animation jeunesse dans une catégorie « Notre sélection du jour pour vous » ou « Notre sélection du jour pour profil A/ profil C » puisqu'aucune des deux n'est apparue.

Croisement des parties 1 et 2 de l'expérience 2

Points communs :

- Le profil C a à chaque fois été transformé plus rapidement que le profil A, c'est-à-dire qu'il proposait plus vite un plus grand nombre de catégories documentaire/jeunesse. De même, le profil C avait plus de têtes de gondoles tournées vers le documentaire puis vers le programme jeunesse que le profil A.
- En plus de l'augmentation variable du nombre de catégories documentaire/jeunesse, beaucoup de titres documentaire/jeunesse sont proposés dans d'autres catégories chez le profil C. Par exemple, au cours de la partie 1 de l'expérience 2, on retrouve des documentaires dans les catégories « Films primés » et « Nouveautés » au jour 4 sur le profil C. Également, dès le jour 2, les catégories « Hollywood » et « Incontournables » proposent des films d'animations jeunesse sur le profil C quand ces mêmes catégories proposent des titres plus variés sur le profil A. Cependant, ce phénomène est plus accentué pour les titres jeunesse que pour les documentaires.
- Au cours des deux parties de l'expérience 2, nous avons observé sur le profil A une augmentation et une diversification des catégories de genre. Ce phénomène avait également été observé dans l'expérience 1. Comme le profil A a visionné au cours de ces expériences des contenus inhabituels, il est possible que l'algorithme propose des catégories plus variées pour en apprendre plus sur les goûts de l'utilisateur.

Différences :

- Pour les profils A et C, le nombre de catégories jeunesse augmente dès le jour 3. C'est une hausse plus rapide que celle du nombre de catégories documentaire, qui n'a quasiment pas augmenté avant le jour 5. Elle est aussi plus graduelle, quand les catégories documentaire apparaissaient soudainement.
- Si les recommandations jeunesse sont plus nombreuses plus tôt, leur nombre maximal reste néanmoins inférieur à celui des recommandations documentaire. Le maximum atteint est ici de 9 catégories animation jeunesse contre 13 catégories documentaire pour le profil C, et 4 jeunesse contre 5 documentaire pour le profil A.
- Les têtes de gondole proposant des documentaires étaient plus fréquentes le matin, quand les têtes de gondole orientées jeunesse étaient plus fréquentes le soir.
- Pour le profil A, lors de la partie 2, aucun documentaire n'a été proposé en tête de gondole, malgré les 7 documentaires visionnés au cours de l'étape précédente. Pour le profil C, le matin, un quart de documentaire était toujours proposé. Pour ce profil nettement plus récent, si on regarde son historique de visionnage, le format documentaire représente la moitié des titres vus, même si aucun n'a été vu depuis la fin de l'expérience 2 partie 1.

5. Bilan des expériences 1 et 2

5.1. Retour sur nos hypothèses

Les deux parties de l'expérience 2 confirment notre première hypothèse : Les recommandations sont davantage marquées par le contenu visionné sur les profils nouveaux. Le profil récent est plus rapidement transformé par l'algorithme de recommandation que le profil historique.

- Les catégories documentaire/jeunesse apparaissent plus rapidement et plus nombreuses sur le profil C que sur le profil A.
- Il y a plus de titres documentaire dans la catégorie « Notre sélection du jour pour vous » du profil C que dans celle du profil A.
- Les têtes de gondole sont plus souvent des documentaires/titres jeunesse sur le profil C que sur le profil A, qu'importe le moment de la journée.

Notre deuxième hypothèse est en partie vérifiée : Au démarrage, les recommandations sont des choix populaires et non personnalisés. Les profils B et C auront beaucoup de recommandations populaires, beaucoup de recommandations de genres différents, et aucune recommandation personnalisée.

- Pendant l'expérience 1, au démarrage, le profil B s'est vu proposé 23 catégories de genre et 4 de popularité, mais aussi 2 catégories personnalisées.
- Pendant l'expérience 2, au démarrage, le profil C s'est vu proposé 23 catégories de genre et 2 de popularité, mais cette fois 4 catégories personnalisées.

Notre troisième hypothèse est confirmée : Un profil récemment créé est plus durablement marqué par les contenus visionnés qu'un profil ancien

- Les catégories documentaire ont continué d'apparaître sur le profil C une fois le visionnage de films d'animation jeunesse lancé, mais pas sur le profil A.
- Seul le profil C s'est vu proposer des documentaires en têtes de gondole (20% le matin).

Nous nous attendions à voir des changements dans les catégories proposées, mais nous avons été surprises par les changements des noms de catégories. Nous avons par ailleurs fait d'autres observations.

5.2. Autres observations

Prédominance du programme jeunesse :

Au cours de la partie 1 de l'expérience 2, le graphique de la Figure 12 montre que le profil A commence avec 0 catégorie jeunesse, quand le profil C en comptait déjà 4, sans que nous indiquions vouloir un profil Enfant. Depuis le démarrage de ce dernier, il n'a cessé d'en proposer, même quand nous ne regardions que des documentaires. Ensuite, dans la partie 2, beaucoup de titres jeunesse sont proposés dans d'autres catégories chez le profil C, comme expliqué précédemment. À partir du jour 5, les catégories « Trouvez votre prochain coup de coeur », « Notre sélection du jour pour vous », « Les plus recherchés », « Films et séries en français », « Films hollywoodiens doublés en français », « Nouveautés sur Netflix », « Comédies », « Seulement sur Netflix » et « Pour se détendre » comptent de nombreux films d'animation pour enfants, comme si ces derniers constituaient peu à peu le film par défaut du profil C. Notre hypothèse est que le format documentaire étant souvent présenté comme un programme fa-

milial, comme les films d'animation (voir Figure 3), le système de recommandation nous ait considérés favorables à l'animation jeunesse dès le début de l'expérience 2.

Documentaire anecdotique :

Au cours de la partie 2 de l'expérience 2, aucun documentaire n'a été proposé en tête de gondole sur le profil A, malgré les 7 documentaires visionnés au cours de l'étape précédente. Il semblerait que le système de recommandation ait considéré comme anecdotique (et à raison !) la semaine de visionnage de documentaires.

De la sélection du jour aux habitudes de visionnage :

Au cours de la partie 1 de l'expérience 2, le nombre de titres documentaire proposés dans la catégorie « Notre sélection du jour pour vous », et surtout sa diminution à partir du jour 5, pose question (voir Figure 10). À première vue, nous supposons que pour le profil A, le visionnage exclusif de documentaire constituait un moment anecdotique. Le système de recommandation choisissait dès le jour 5 de varier à nouveau les recommandations, tout en gardant un nombre certains de documentaires en tête. Pour ce qui est de la diminution de ces mêmes recommandations chez le profil C, nous pensions qu'étant un profil nouveau (cas du démarrage à froid), l'algorithme cherchait encore à être le plus exhaustif possible quant aux goûts de l'utilisateur, et lui proposait alors d'explorer d'autres types de contenus. Mais dans le graphique de la Figure 9, nous avons observé que le nombre de catégories relatives au documentaire augmentait fortement pour les deux profils au jour 5. De fait, si moins de titres documentaire sont proposés dans la catégorie « Notre sélection du jour pour vous », ces derniers ne disparaissent pas de la page d'accueil, ils ont au contraire leurs propres catégories dédiées.

6. Discussion

6.1. Étude des noms de catégories

Complexification des catégories :

Au démarrage du profil B, on retrouve des catégories de genres très simples : « Comédies », « Thrillers », « Drames »... tandis que le profil A se voit recommander des catégories plus complexes : « Drames européens inspirés de livre », « Films romantiques avec un triangle amoureux ». Le profil A étant alimenté depuis 7 ans, ses goûts ont pu être cernés avec précision par le système de recommandation qui est désormais à même d'être aussi précis. En captant avec autant de précision le registre d'un utilisateur, le système augmente ses chances de le satisfaire, et donc de le faire rester sur la plateforme.

Avec la diversification impliquée par l'apparition du documentaire dans le profil A, des catégories inédites sont apparues : « Séries politiques », « Drames cérébraux », « Histoire de cupidité et de corruption », « Divertissement, variété et talk-shows »... quand le profil C s'est vu recommandé des catégories orientées plus franchement vers le documentaire : « Films Émotion primés », « Nommés et primés aux News & Documentary Emmy Awards », « Émission TV sur l'écologie », « Films de 90 minutes à voir en famille ». Le profil C s'est orienté vers les genres proches du documentaire, mais il est plus difficile de comprendre le changement de recommandation faites au profil A. Des enjeux plus « sérieux » ? De la fiction moins légère ? Avec la partie 2 de l'expérience 2 et le visionnage de films d'animation pour enfants,

les catégories du profil A ont continué à se diversifier, notamment en revenant aux simples « Comédies » et « Blockbusters ».

L'animation pour enfant propose aussi un large choix de catégories spécifiques : « À regarder avec les enfants plus âgés », qui rappelle l'importance des réglementations liées à l'âge, « Films pour la famille inspirés de livres », mais aussi « Comédies d'action oniriques », « Films de chiens et chats », « Films magiques pour la famille », ou encore « Films père-fille jeunesse et famille » pour des thèmes plus précis.

Dénominations surprenantes :

Ensuite, des catégories à la formulation surprenante sont apparues, comme « Rire en pagaille, mais pas en boîte » à la suite des films de vampires. Il se trouve que cette expression fait référence aux rires préenregistrés typiques des sitcoms⁴. L'algorithme nous proposerait là un genre précis de la comédie, en s'appuyant pour ce faire sur sa particularité technique.

La catégorie « Grandes vacances » apparue sur le profil A le dernier jour de l'expérience animation jeunesse pourrait proposer des programmes pour les enfants en vacances, mais il n'en est rien. Elle propose des contenus très variés, des films et des séries, parfois pour un public plus vieux : *Fiasco*, *Le Monstre des mers*, *Mes meilleures amies*, *H* et *Grace and Frankie*. Entre un mockumentaire⁵, un film d'animation, une comédie décalée sur le mariage, une sitcom médicale française, et une série comique décalée, on se demande ce qui les rassemble sous la bannière « Grandes vacances ». Peut-être Netflix® tient-il ici la clé d'un nouveau genre cinématographique dont il garde le secret... à moins qu'il s'agisse d'une catégorie saisonnière ? Elle est apparue le mercredi 5 juin, c'est-à-dire à quelques jours des vacances d'été. Les élèves de seconde finissent généralement leur année en juin, peut-être que Netflix® propose pour eux une grande variété de contenus à *binge-watch* ?

Le visionnage de films de vampires (expérience 1) a fait apparaître une catégorie « Soirée filles ». Une hypothèse est qu'il s'agit d'un biais marketing, une catégorisation implicite « en film de filles ». La recommandation de catégories pourraient donc relever aussi de critères implicites, et ici Netflix® participe à une re-genrification du cinéma qui interroge. Mais là encore, les titres proposés surprennent : *Désenchantée*, *Umbrella Academy*, *Mes premières fois*, *Good Girls*, *Glow*, *Young Royals*... Puis, lorsque la catégorie apparaît à nouveau pendant la partie 2 de l'expérience 2 : *The Crown*, *Pièce montée*, *Ma vie avec les Walter boys*, *Toi chez moi et vice-versa*, *Mémoires d'une Geisha*, *L'école du bien et du mal*... Séries comiques d'action et d'aventure, séries de super-héros, séries dramatiques à suspense, séries romantique dramatique pour ados, séries historiques... là encore, on se demande ce qui les réunit.

Proximité avec l'utilisateur :

Les deux expériences nous ont par ailleurs permis d'observer le langage utilisé dans les catégories. Certaines contiennent des adresses à l'utilisateur : « Notre sélection pour profil A », « Trouvez votre prochain coup de coeur »... De même, on peut observer un processus de personnification : « **On** pense que vous allez adorer », « **Notre** sélection pour vous »... Est-ce pour Netflix® un moyen de dissimuler la dimension mécanique, automatique, programmée de la recommandation ? De rétablir une proximité avec l'utilisateur ?

4. Séries souvent humoristiques, caractérisées par un comique de situation et une unité de lieu

5. Faux documentaire

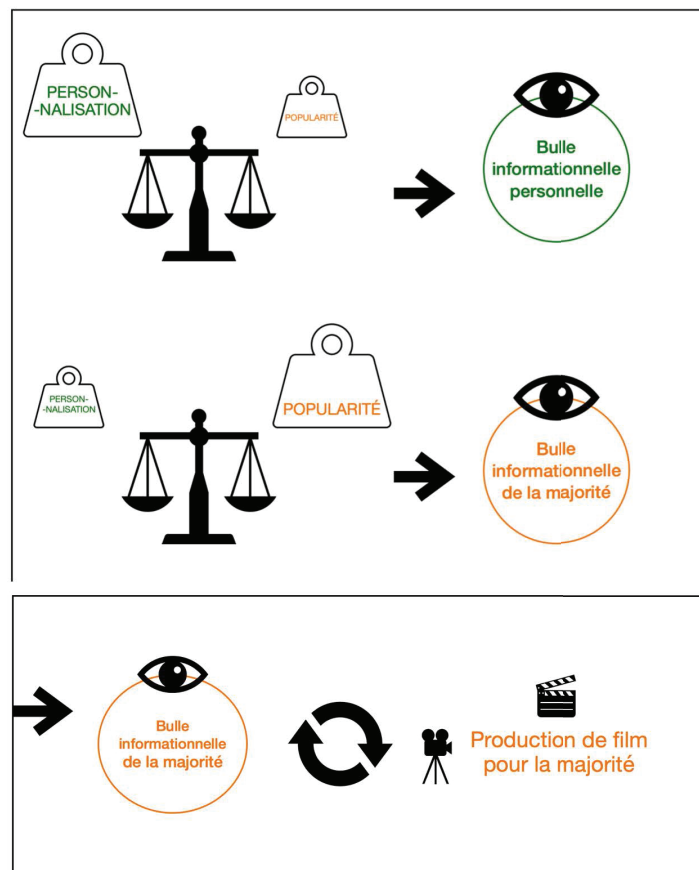


Figure 14. *Risque d'uniformisation dans la diversification par le contenu populaire.*

6.2. Questions

La personnalisation et la popularité sont présentées comme deux outils complémentaires (GOMEZ-URIBE and HUNT, 2016), permettant de construire des recommandations qui plaisent à l'utilisateur, mais aussi de varier ces dernières, d'y apporter de la diversité et de la nouveauté. Comme illustré par la Figure 14, si un profil se voit seulement recommander des titres susceptibles de lui plaire, il risque d'enfermer l'utilisateur dans une bulle informationnelle personnalisée, un paysage culturel seulement en accord avec ses goûts et ses opinions (en rapport avec les notions de Bulles de filtre et de Chambre d'écho définies par PARISER (2011)). Proposer du contenu populaire, c'est introduire de nouvelles idées. L'utilisateur peut alors découvrir les titres qui circulent beaucoup, qui sont appréciés et partagés. Cependant, se contenter de la popularité comme facteur de diversification c'est aussi risquer d'enfermer l'utilisateur dans le paysage culturel de la majorité. Si seuls les titres populaires sont recommandés, ce sont finalement les mêmes opinions et les mêmes goûts qui circulent. Les contenus niches sont oubliés, et les nouvelles productions s'accordent avec les goûts de la majorité pour avoir du succès. Si la popularité ne peut pas être le seul critère de diversification, lequel choisir ?

Dans le manga *Demokratia* de Motorò Mase (MASE, 2016), 3000 internautes sont sélectionnés au hasard pour décider à la majorité des actions d'un robot. Cependant, le système de votes ne se contente pas de mettre en valeur les propositions les plus suggérées : il synthétise les 3 actions les plus suggérées, mais aussi les 3 actions les moins proposées. Ainsi, les propositions regroupent à la fois majorité et créativité, et la plus originale peut être choisie face à la plus commune. Un fonctionnement similaire pourrait être imaginé pour accompagner le système de recommandation de Netflix®. Imaginons un algorithme chargé de recueillir les titres les moins regardés du catalogue, ou les plus éloignés des goûts d'un utilis-

teur pour lui proposer de découvrir autre chose, à l'instar de ce qui se fait en recommandation musicale (CELMA, 2010).

La plateforme de streaming MK2 Curiosity®⁶ déclare «Un algorithme ne vous propose que ce que vous aimez déjà. Nous vous proposons ce que vous n'aimez pas encore.» Elle se revendique ainsi d'ouvrir l'horizon culturel de ses abonnés en ne proposant que des films rares du cinéma mondial.

Une autre possibilité serait de penser d'autres catégories, de trier différemment les titres du catalogue. Au lieu de regarder des films romantiques avec un triangle amoureux, des comédies ironiques et des sitcoms américaines saluées par la critique, un abonné pourrait découvrir toutes les séries qui proposent au moins un épisode en huis-clos, des comédies romantiques qui s'ouvrent sur un travelling avant, ou encore la comédie la plus vue par les grenoblois en 2014 (FERRIER et al., 2024). Les possibilités sont infinies pour catégoriser les titres : selon des éléments précis, en jouant autrement avec son profil (goût, historique...), selon la structure du film (processus de réalisation et produit final), selon la sortie ou la réception par le public, ou relatif au titre, comme Netflix® s'amuse avec la formulation de ses catégories.

Le genre cinématographique, s'il est censé regrouper des œuvres partageant des caractéristiques communes, reste néanmoins variable, flottant, difficile à saisir et à fixer. Il nécessite reconnaissance à la fois par les critiques et par le public, et est ainsi pleinement historique, dépendant de l'évolution du cinéma (AUMONT and MARIE, 2016). Raphaëlle Moine (MOINE, 2009) ajoute qu'il nécessite reconnaissance, communauté mais aussi création d'un vocable commun par cette communauté. De plus, elle propose d'étudier le genre cinématographique non comme une catégorie de classement, mais comme une catégorie d'interprétation. Dès lors, il s'agit de comprendre les liens qui relient un film à un ou plusieurs genres, tout en s'interrogeant sur la façon dont on identifie ces liens (identités prêtées, ce qu'on voit de ces films, ce qu'on en dit lorsqu'ils sont regroupés sous un genre) (MOINE, 2009).

Le but du système de recommandation de Netflix® est d'augmenter la vitesse de la prise de décision chez l'utilisateur, ou du moins la permettre. C'est d'ailleurs le but de la plupart des algorithmes de recommandation : faire rapidement le tri dans un nombre incalculable de données. Peut-on questionner cet idéal de rapidité ? Les chiffres montrent qu'un utilisateur qui ne se décide pas se désintéresse (GOMEZ-URIBE and HUNT, 2016), mais est-il condamné à ce modèle ? N'est-il pas imputable à un environnement de réduction de l'attention et de la surstimulation ? Pourrions-nous imaginer une plateforme qui ne repose pas sur l'efficacité, mais qui contraint l'utilisateur à faire un choix, à décider consciencieusement, sans précipitation ?

7. Conclusion et Perspectives

Dans cet article, nous avons tenté de comprendre le fonctionnement du système de recommandation de Netflix®, du point de vue de l'utilisateur. Plus précisément, nous avons cherché à observer les effets de ses visionnages sur ses recommandations, et la perception qu'il peut en faire. Nous nous sommes ainsi concentrées, au cours de deux expériences de comparaison de profils, sur les éléments observables par l'utilisateur lui-même : les catégories, les titres recommandés dans les catégories personnalisées, ainsi que la tête de gondole figurant en haut de la page d'accueil.

6. <https://www.mk2curiosity.com/home>

L'expérience 1 nous a servi d'étalon. Notre échantillon test, les films de vampires, étant trop petit, nous n'avons pas pu mener l'expérience sur une durée suffisante pour recueillir assez de données. Des tendances se dégagent des résultats, mais rien que nous puissions affirmer avec certitude. Cependant, cette expérience nous a permis de définir le cadre d'analyse de l'expérience suivante.

L'expérience 2 s'est faite en deux parties : avec des documentaires animaliers, puis avec des films d'animation pour enfants. Cette fois-ci, nous avons pu répondre à nos hypothèses de départ :

1. Les recommandations du profil C vont être plus rapidement et plus fortement marquées par le contenu que nous faisons voir à ces profils tandis que le profil A sera peu modifié. **Cette hypothèse a été vérifiée** : au cours des deux parties de l'expérience 2, le profil C a été transformé plus rapidement et dans une plus grande mesure que le profil A, c'est-à-dire qu'il a reçu plus de recommandations centrées sur le documentaire animalier/l'animation jeunesse plus rapidement.
2. Les profils B et C (nouveaux) auront beaucoup de recommandations populaires, beaucoup de recommandations de genres différents, et aucune recommandation personnalisée. **Celle-ci est partiellement validée**, puisque le profil B avait à son premier démarrage (expérience 1) 2 catégories personnalisées, et le profil C (expérience 2, partie 1) 4 catégories personnalisées.
3. Lors de la partie 2 de l'expérience 2, le profil A « oubliera » vite le documentaire, tandis que le profil C s'en verra toujours proposer. **Cette hypothèse se vérifie**, des documentaires ont bel et bien continué à être suggérés au profil C.

Mais ces deux expériences nous ont aussi amenées à nous interroger sur les catégories en elles-mêmes. Les catégories qui composent la page d'accueil de la plateforme ont grandement changées au cours de notre expérience. Elles constituent ainsi la manifestation la plus explicite de l'influence du système de recommandation. En voulant étudier les biais possibles des algorithmes de recommandation de Netflix®, ou du moins leur influence sur le contenu présenté à l'utilisateur, nous en sommes finalement venues à nous questionner sur les biais de catégorisation. Dès lors, nous supposons que lorsqu'il navigue sur la plateforme de streaming, l'expertise de l'utilisateur se révèle être sa capacité à naviguer au sein de ces catégories.

Nous pouvons nous interroger sur les frontières entre film d'horreur, film de monstres et thriller fantastique, mais aussi interroger des choix qui peuvent sembler arbitraires : que signifie les catégories « Soirée filles » et « Grandes vacances » ? Comment sont choisis ces noms ? Sont-ils les mêmes partout dans le monde ? Comment sont choisis les titres présentés dedans ? Et comment affectent-ils la perception des films ? Alors que le genre est à considérer du point de vue historique et comme une catégorie d'interprétation, nous pouvons nous interroger sur la place qu'il occupe vraiment dans le choix et la réception du film par l'utilisateur.

À notre connaissance, ces questionnements ne font pas l'objet d'une application dans le domaine de la SVOD. Cette application sera donc au cœur de nos prochaines recherches. De plus, il serait pertinent, dans des travaux futurs, d'étendre cette analyse à d'autres plateformes de streaming afin de comparer les résultats obtenus et de vérifier si les tendances observées sur Netflix® se confirment dans différents contextes.

Bibliographie

- ADOMAVICIUS, G. and TUZHILIN, A. (2011). Context-aware recommender systems. pages 217–253.
- AUMONT, J. and MARIE, M. (2016). *Dictionnaire théorique et critique du cinéma*. Armand Colin.
- BE MY MEDIA (2025). Les bulles de filtre algorithmiques. Consulté en juillet 2025.
- CELMA, O. (2010). *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*.
- CHILDS, M. (2021). Runaway Recommendation Engine : Planet Money.
- DEY, A. (2001). Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7.
- DRUMOND, G., COUTANT, A., and MILLERAND, F. (2018). La production de l’usager par les algorithmes de Netflix. *Les Enjeux de l’information et de la communication*, 19/2:29–44.
- FERRIER, S., VEILLON, M., and BURANELLO, Z. (2024). Arleohtigm. Master’s thesis, ENS Paris-Saclay.
- GOMEZ-URIBE, C. and HUNT, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Trans. Manage. Inf. Syst.*, 6(4):13–31.
- HALLINAN, B. and STRIPHAS, T. (2016). Recommended for you: The netflix prize and the production of algorithmic culture. *New Media & Society*, 18(1):117–137.
- KAUR, H. and ASHFAQ, R. (2023). The impact of netflix on viewer behaviour and media consumption: An exploration of the effects of streaming services on audience engagement and entertainment preferences. *Journal of Media, Culture and Communication(JMCC)*, 3(04):9–23.
- KLOPOTEK, M. (2009). Approaches to cold-start in recommender systems. *Studia Informatica : systems and information technology*, Vol. 1(12):47–54.
- KULKARNI, S. and RODD, S. (2020). Context aware recommendation systems: A review of the state of the art techniques. *Computer Science Review*, 37:100255.
- KUNAVAR, M. and POŽRL, T. (2017). Diversity in recommender systems – a survey. *Knowledge-Based Systems*, 123:154–162.
- LE, Q., VU, S., NGUYEN, T., and LE, T. (2021). A state-of-the-art survey on context-aware recommender systems and applications. *Int. J. Knowl. Syst. Sci.*, 12:1–20.
- MADDODI, S. and KARANI, K. (2019). Netflix bigdata analytics- the emergence of data driven recommendation. *SSRN Electronic Journal*.
- MASE, M. (2015-2016). *Demokratia*. Kazé.
- MOINE, R. (2009). Film, genre et interprétation. *Le français aujourd’hui*, 165(2):9–16.
- MOORE, J. (2016). What is the sense of agency and why does it matter? *Frontiers in Psychology*, 7.

- NEGRE, E. (2015). *Recommender and Information Systems*. Wiley. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1848217544.html>.
- NETFLIX, CENTRE D'AIDE (2024). Fonctionnement du système de recommandations de netflix. <https://help.netflix.com/fr/node/100639> [Accessed: 07/01/2024].
- PANNIELLO, U., TUZHILIN, A., GORGLOGLIONE, M., PALMISANO, C., and PEDONE, A. (2009). Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of RecSys '09*, pages 265–268. ACM.
- PARISER, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group, The.
- SCHAFFNER, B., STEFANESCU, A., CAMPILI, O., and CHETTY, M. (2023). Don't let netflix drive the bus: User's sense of agency over time and content choice on netflix. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1):128–159.
- SCHEIN, A., POPESCU, A., UNGAR, L., and PENNCOCK, D. (2001). Generative models for cold-start recommendations. In *Proceedings of the 2001 SIGIR workshop on recommender systems*, volume 6.
- STOLL, J. (2023). Countries with the Biggest Netflix Library Worldwide.
- TARISSAN, F. (2022). Les systèmes de recommandation favorisent-ils la diversité ? *Interstices*.
- ZIEGLER, C., MCNEE, S., KONSTAN, J., and LAUSEN, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pages 22–32, New York, NY, USA. Association for Computing Machinery.
- ZUBOFF, S., FORMENTELLI, B., and HOMASSEL, A. (2022). *L'Âge du capitalisme de surveillance*. Zulma.