

Université Paris 1 Panthéon – Sorbonne

Mémoire de Master 2 Méthodes Informatiques Appliquées à la Gestion des
Entreprises, Systèmes d'Information et Innovation (MIAGE S2I)

Matthieu Porembski

Année Universitaire : 2020 - 2021

Date de soutenance : 27/09/2021

Directeur de mémoire : Nicolas HERBAUT

Membre du jury : Irina RYCHKOVA

Remerciement

Dans un premier temps, je tiens à remercier mon directeur de mémoire Monsieur Nicolas HERBAUT pour le temps qu'il m'a accordé, ses précieux conseils, sa pédagogie et son soutien tout au long de la réalisation de ce mémoire.

Je tiens aussi à présenter mes sincères remerciements à mon maître d'apprentissage Madame Pascale MONDETEGUY, qui durant mon année d'alternance chez Orange Events m'a donné l'opportunité d'apprendre, d'évoluer du point de vue professionnel et personnel tout en contribuant à l'organisation d'un évènement unique en son genre.

Mes remerciements vont également à tous les étudiants de Licence 3 et Master 1 qui participent au projet Streaming 4 Good, à Floriane OWCZAREK, et Ramona EL ALLY pour leurs conseils.

Je tiens aussi à remercier tous les enseignants, les intervenants et l'administration de la MIAGE, du CFA AFIA, de l'UFR 27 et de la Sorbonne que j'ai pu côtoyer durant mes cinq années d'étude à l'Université Paris 1 Panthéon-Sorbonne. Merci beaucoup pour vos efforts déployés afin d'assurer une formation complète, pluridisciplinaire et continue, durant des années qui ne furent pas de tout repos.

Enfin, afin de m'assurer de n'avoir oublié personne, je souhaite remercier tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce mémoire ainsi qu'à la réussite de mon parcours universitaire.

Glossaire

ACREP¹ : l'Autorité de régulation des communications électroniques est une autorité administrative Française chargée d'assurer la régulation des secteurs des communications électroniques.

OTT² : L'Over The Top (OTT) est un modèle de distribution de contenus par internet qui n'a pas besoin d'un fournisseur d'accès à internet comme intermédiaire, au-delà de l'acheminement des données.

CSA² : Le Conseil Supérieur de l'Audiovisuel est une autorité publique française en charge de la régulation de l'audiovisuel. Il veille sur le respect et la protection des droits et libertés individuels, la régulation économique et technologique du marché ainsi que la responsabilité sociale.

Directive SMA⁶ : La directive des Services de médias audiovisuels (SMA) vise à créer un marché unique des SMA dans l'Union européenne et à s'assurer du bon fonctionnement de celle-ci, tout en contribuant à la mise en avant de la diversité culturelle et garantissant un niveau de sécurité des consommateurs et des enfants

SVOD : Une plateforme de vidéo à la demande (subscription video on demand) est une plateforme qui propose d'accéder à un catalogue de contenu en échange d'un abonnement ou versement mensuel ou annuel.

Contrairement à la VOD (Video on Demand), qui est un service qui permet lui de commander et regarder un contenu sur internet. Il n'y a plus ici une notion d'abonnement mais plus de location ou achat.

Introduction

Pendant de nombreuses années la télévision, le cinéma, la location ou l'achat des DVD étaient des moyens de consommations et divertissements audiovisuel. Régi par des lois à l'échelle de la Nation (par exemple le Conseil Supérieur de l'Audiovisuel) ou de l'Union Européenne, ces sources de divertissements encadrées font aujourd'hui toujours face à de nouveaux acteurs internationaux, comme Netflix ou Amazon Prime Vidéo pour ne citer qu'eux. Avec la digitalisation du contenu, l'évolution des réseaux, l'accessibilité accrue des supports de diffusion et les changements des habitudes de consommation, le marché du divertissement a depuis bien évolué.

D'autant plus que ces changements ont été accentués par la pandémie mondiale que nous traversons encore actuellement, en transformant de nombreux aspects de la vie quotidienne. Elle a en effet bousculé nos routines et nos habitudes, comme par exemple le fait de sortir, profiter d'une terrasse ou des salles obscure (cinéma). En situation de confinement, la population fut forcée à rester à la maison et avec beaucoup de temps libre mais aucun nouveau divertissement sous une forme traditionnel disponible. Les plateformes OTT et les émissions sur le web sont apparues comme le meilleur outil pour remédier à cela. Avec le télétravail à distance et la virtualisation des différents éléments, nous avons vécu une transformation numérique sans précédent. Elle a profondément bouleversé le mode de consommation du contenu, en particulier celui du secteur des médias et du divertissement. En 2020, le leader du contenu en streaming (Netflix) fut obligé par la Commission européenne de réduire ses débits³ afin de ne pas saturer le réseau lors d'une période où le télétravail était omniprésent. Le géant américain mobilisait à lui seul un quart des capacités d'Internet⁴ en France selon l'ACREP.

L'entrée au fur et à mesure des différentes plateformes de service par contournement (OTT) comme Netflix, Amazon Prime Vidéo ou encore Disney+ a bousculé les services plus classiques. Avec un éventail de contenu large accessible à travers différents appareils connectés à Internet comme un ordinateur, une console de jeu, un smartphone ou une Smart TV, ces plateformes se sont aussi démarquées grâce à des moteurs de recommandations qui s'éloignent grandement d'une programmation télévisuelle linéaire. Le contenu proposé étant personnalisé en fonction des goûts de l'utilisateur.

En plus de cela, ces plateformes du numérique étrangères n'étaient pas liées par des accords⁵ avec la filière audiovisuelle (comme le sont les diffuseurs français) qui leur fixent des obligations et certaines règles. Etant donné que les membres de l'Union Européenne se réfèrent tous à la directive européenne SMA (Services médias audiovisuels - 2010/13/UE) pour ce qui est des services médias audiovisuels, cela a créé une concurrence déloyale entre les plateformes du numérique et les diffuseurs des différents pays. Mais dans l'optique d'une régularisation afin de

proposer des conditions de concurrence équitables entre les différents opérateurs, l'Europe est intervenue.

Le 10 octobre 2018, les députés ont voté (452 voix pour, 132 contre et 65 abstentions) la directive 2018/1808 qui vise à actualiser les règles qui régissent les services de médias audiovisuels. Outre des règles relatives à la protection des mineurs et au contrôle plus stricte des publicités, le Parlement Européen étend l'application de la directive SMA (2010/13/UE) aux plateformes numériques. Publié ensuite le 18 novembre 2018, voici une section⁶ de cette directive qui fut ajoutée au texte de 2010.

Article 13

1. Les États membres veillent à ce que les fournisseurs de services de médias relevant de leur compétence qui fournissent des services de médias audiovisuels à la demande proposent une part d'au moins 30 % d'œuvres européennes dans leurs catalogues et mettent ces œuvres en valeur.

Est-ce que Netflix respecte cet article ? Qu'est-ce qu'un contenu en valeur sur une grille infini de contenu où le choix final appartient à l'utilisateur ? Cela nous amène à la question de recherche étudié dans ce mémoire :

Peut-on grâce à la data analyse et la fouille de processus montrer la conformité de Netflix à la directive Européenne SMA ?

Afin de répondre à cette question, nous reviendrons dans un premier temps sur l'histoire de la plateforme et son expansion progressive. Nous verrons quels sont les outils que la plateforme met en place afin de se distinguer de la concurrence et l'impact sur la population. Ensuite nous verrons l'initiative universitaire Streaming for Good qui regroupe des étudiants de plusieurs années. Dans la seconde grande partie je détaillerai la méthodologie et le traitement des données avant de passer à l'analyse de celles-ci grâce à la fouille de processus ainsi que de la data analyse. L'objectif étant de savoir si l'organisation en vignette de Netflix (qui est à l'opposée d'une programmation télévisuelle) est en accord avec la directive Européenne.

I. Background

Dans cette section, nous allons revenir sur différents éléments importants qui sont nécessaires à la compréhension du sujet. Dans un premier temps nous allons revenir sur l'histoire de la plateforme numérique Netflix, voir comment celle-ci s'est adaptée et a changé les habitudes des consommateurs des médias numériques. Ensuite nous parlons de son expansion à l'international (notamment en France) et de son impact sur certaines habitudes des consommateurs. Enfin, nous verrons la présentation du projet Streaming for Good, une initiative d'étude des plateformes de distribution de contenus numériques.

1. L'histoire de Netflix

Après une année marquée par une crise sanitaire sans précédent, 2020 fut pour Netflix la meilleure année de son histoire en termes d'abonnement⁷. A lui seul le dernier trimestre a rapporté pas moins de 8 millions de nouveaux abonnements, ce qui a permis de dépasser la barre symbolique des 200 Millions de comptes (on utilise la notion de compte, étant donné que sur qu'un abonnement peut regrouper plusieurs comptes et/ou utilisateurs). Ce succès, la plateforme le doit à des années de persévérances, d'études du marchés et l'utilisation de la Big Data.

L'histoire de la plateforme débuta en 1997 aux États-Unis. A cette époque, Reed Hastings (actuel PDG) et Marc Randolph se sont associés afin de fonder une plateforme de location de DVD, Netflix. Mais contrairement à Blockbuster qui était le leader de la location et la vente de DVD aux USA avec 2800 boutiques physique à travers le monde, ils décidèrent d'innover en utilisant un catalogue accessible en ligne alors que le commerce numérique n'en est qu'à ses débuts⁸. Ainsi, un consommateur en quête de divertissement devait se rendre sur le site de Netflix, choisir le film ou la série qu'il voulait et il n'avait plus qu'à attendre à la maison que sa location arrive par les services postaux. Toutefois le prix de la location ne couvrait pas les frais de gestion et d'expédition aller-retour. C'est alors que les deux fondateurs décidèrent d'instaurer un abonnement à prix fixe qui changea le business model de l'entreprise. Dorénavant, un abonné payait un prix mensuel fixe et pouvait louer « gratuitement » un DVD. Ainsi, un simple abonnement de 19,95\$ par mois permettait de louer l'une des 925 contenus disponibles.

Un autre fait marquant du début de l'entreprise concerne la gestion de la demande grandissante des utilisateurs en termes de « Super Hits ». Toute l'industrie de la location des DVD devait gérer des utilisateurs qui pouvaient être mécontents lorsqu'ils voulaient impérativement voir les dernières productions populaires mais que celles-ci n'étaient pas disponibles. Afin de remédier à cela, Netflix a intégré un système de recommandation à son catalogue afin de faire

découvrir à ses abonnés certains contenus qu'ils n'auraient pas vu et ainsi proposer une alternative au contenu populaire du moment. L'implémentation de cette solution a fait diminuer de 20% la demande des nouveaux « Supers Hits » du moment comparée aux services de location DVD comme Blockbuster⁹. Ainsi, de par ses idées nouvelles et ce système de recommandation Netflix s'est basé sur le fait de considérer les films et les séries comme un divertissement régulier qu'il faut alimenter et renouveler⁹. Contrairement à Blockbuster qui est resté sur une idée plus classique de la location de DVD, où la majorité des locations se font sur la base d'une décision impulsive du consommateur qui veut regarder un film de manière immédiate, une tendance du moment ou une nouveauté. Avec les années et la démocratisation du Web dans de nombreux foyers, le principal concurrent de Netflix n'a pas su répondre à temps aux différents changements de la consommation et des nouvelles habitudes. Malgré une tentative d'adopter un modèle similaire à Netflix en 2004, Blockbuster est resté sur le modèle de distribution et modèle commercial initial, ce qui mena malheureusement l'entreprise à mettre la clé sous la porte en 2013. Avant cela, l'année 2007 fut pour Netflix le début du streaming du contenu depuis un ordinateur, ce qui a permis à l'entreprise de continuer son extension et l'acquisition de nouveaux utilisateurs. Après cette date, la plateforme commença à arriver progressivement sur différents supports comme les consoles de jeux, les décodeurs, les smartphones et les tablettes.

Comme on a pu le voir l'objectif de la plateforme est de considérer la consommation de film et série comme un divertissement continu. Il est donc nécessaire de renouveler de manière perpétuelle le contenu proposé afin de rentrer dans cette « continuité ». Par le biais d'algorithmes qui régissent l'entièreté du contenu que l'on retrouve sur la page d'accueil, Netflix a toujours su personnaliser l'expérience de l'utilisateur. Comme l'a dit Ted Sarandos (co-PDG de Netflix) en 2014 : « *Les algorithmes pilotent l'ensemble de notre site web - il n'y a pas de pouce d'espace éditorial non calculé* ». Le système de recommandation¹⁰ utilise l'ensemble des données produites par un utilisateur (contenu visionné, contenu recherché, contenu aimé, contenu observé, moment de la journée...) afin de proposer une expérience unique et individuelle (filtrage basé sur le contenu). A cela s'ajoute aussi l'étude faite sur des profils similaires à l'utilisateur (filtrage collaboratif), ce qui permet de perfectionner encore plus le contenu sélectionné pour la personne. Par son implication dans le développement d'outils de recommandation, Netflix fut listé comme la 30ème entreprise la plus innovante par le journal FastCompany dédié à la technologie et le business. Une position qui vient entre autres récompenser l'utilisation d'algorithmes afin de proposer un contenu attendu par le spectateur.

Les outils de recommandation mis en place par Netflix sont le fruit de longues années de travaux. Avec un investissement conséquent dans la Data Analyse, qui a débuté en 2006, l'objectif de la plateforme est d'offrir le contenu le plus personnalisé pour l'utilisateur. Afin de toujours proposer une recommandation plus précise, le service a promis d'offrir 1 Millions de dollars à celui qui proposera un algorithme de recommandation (basé sur le contenu visionné

par un utilisateur) 10% plus efficace que leur algorithme de recommandation Cinematch. En 2008, l'équipe BellKo's Pragmatic Chaos gagna le prix. Les nombreuses autres solutions furent réutilisées dans l'e-commerce.

En plus de proposer un contenu sur lequel l'entreprise a acquit un droit de diffusion, Netflix se lança dans la production de contenus originaux d'une manière bien spéciale. Une manière que certains estiment être un coup de poker très bien exécuté¹¹. Après avoir récolté les données générées par ses 27 millions d'abonnés aux USA et 33 millions dans le monde en 2011, Netflix a pu analyser le contenu visionné, les notes attribués et les recherches. Cette analyse de la Big Data a aidé Netflix à identifier la tendance de demain. Comme le montre la *Figure 1* publiée par le New York Times en 2013, Netflix a combiné grâce à la Big Data, les clés du succès.



Figure 1 « Les cercles de la réussite » de la création Original Netflix House of Cards – Source : The New York Times

En alliant un réalisateur (David Fincher) dont la dernière production (The Social Network) fut fortement regardée et appréciée à un acteur populaire lauréat des Oscars (Kevin Spacey) dans le « remake » (nouvelle version) d'une ancienne série Britannique des années 90, Netflix créa House of Cards U.S. Sans avoir vue un seul épisode, la plateforme a investi 100 millions de dollars pour produire 26 épisodes sur 2 saisons.

Cet investissement d'argent fut d'autant plus motivé par le succès d'un concurrent de la plateforme avec une production télévisuel sans précédents. En effet, le groupe Warner Media et sa filiale HBO proposa en 2011 la première saison de la série Game of Thrones. Une série qui bouleversa le « lore de l'industrie télévisuel »ⁱ de par sa qualité et son ambition. C'est pour

ⁱ En 2008 (*The evolution of industry lore in African American television trade*. Montreal, QC: International Communication Association), Timothy Havens (Professeur à l'Université de l'Iowa) met en avant la notion du lore de l'industrie qu'il définit comme les connaissances conventionnelles des initiés de l'industrie sur les types de culture médiatique qui sont et ne sont pas possibles, et sur ce qui va oui ou non attirer

cela, que pour rentrer dans les nouvelles normes de production et une cinématographie élevée Netflix s'appuie sur une telle équipe. A l'époque le Chief Content Officer de Netflix, Ted Sarandos déclara :

"L'objectif est de devenir HBO plus rapidement que HBO ne devienne comme nous"

Entre Stranger Things, Orange is the new black, Altered Carbon, Narcos, The Crown... les productions originales baptisées « Netflix Originals » se sont enchaînées (une notion qui regroupe les films ou séries commandé(e)s par Netflix ou des films et séries dont les droits de diffusion ont été exclusivement achetés par Netflix). En 2019, selon Thibault Henneton ces contenus originaux représentent 8%¹² du catalogue composé d'environ 15000 titres toutes bibliothèques confondues (chiffre selon la base de données uNoGS.com).

Toujours en termes d'évolution, depuis la fin de l'année 2020, Netflix a décidé de proposer une fonctionnalité (uniquement testée en France) qui s'approche de la télévision traditionnelle¹³. Selon Netflix, la France reste un pays où la consommation de TV traditionnelle reste très populaire et de nombreux abonnés aiment l'idée de ne pas avoir à choisir ce qu'ils vont regarder. C'est pourquoi, depuis le 5 novembre 2020, le service déploie progressivement la fonction : le Direct. Une fonctionnalité qui propose un flux en temps réel identique pour tous les abonnés afin de « découvrir les meilleurs contenus français, européens et internationaux, disponible sur le service. »

Ainsi, comme on a pu le voir, Netflix a toujours su utiliser les nouvelles formes des traditions industrielles comme un outil pour contrôler et réinventer les pratiques d'attraction des utilisateurs en utilisant toutes les informations dont elle dispose. Mais son ambition ne s'arrête pas là. Depuis le début de l'année 2021 de nombreuses informations circulent sur le fait que Netflix veut se lancer dans le Cloud Gamingⁱ. Ce marché en forte expansion depuis la pénurie des composants informatiques¹⁴ a fait naître de nombreux services propulsés par des géants comme Amazon (Luna), Google (Stadia), Nvidia (GeForce Now) ou encore Sony (PlayStation Now). Reed Hasting précise¹⁵ que ce nouveau cap a pour but de renforcer et améliorer le service de la plateforme.

l'audience. – La qualité des productions d'HBO a permis d'augmenter le nombre d'abonnés à la chaîne de télévision. HBO est passé du simple recyclage des films d'Hollywood à la production d'un contenu propre. L'objectif fut donc pour Netflix de reproduire ce modèle de production de qualité, mais avec une distribution numérique. -

ⁱ Le Cloud Gaming est un moyen de jouer à des jeux vidéo depuis un appareil connecté à internet qui n'est qu'un support d'image. Peu importe l'endroit, peu importe le moment, grâce à une connexion réseau, l'écran du joueur (télévision, smartphone, PC...) se connecte à un serveur qui va s'occuper de l'ensemble du traitement de l'information (graphique notamment). Cela permet de jouer par exemple à des jeux gourmands en ressources depuis un appareil peu puissant.

2. L'expansion de Netflix en France

Netflix est aujourd'hui disponible sur tout le globe à quelques exceptions près (Chine, Crimée, Corée du Nord et Syrie - Figure 2). L'expansion progressive à l'international de Netflix commença en 2010 avec le Canada. Ensuite elle continua en 2011 avec l'Amérique du Sud, les Caraïbes et l'Europe Anglophone en 2011, en 2013 aux Pays-Bas, puis dans l'ouest de l'Europe en 2014, pour enfin être disponible dans le monde entier (sauf 4 pays) en 2016.

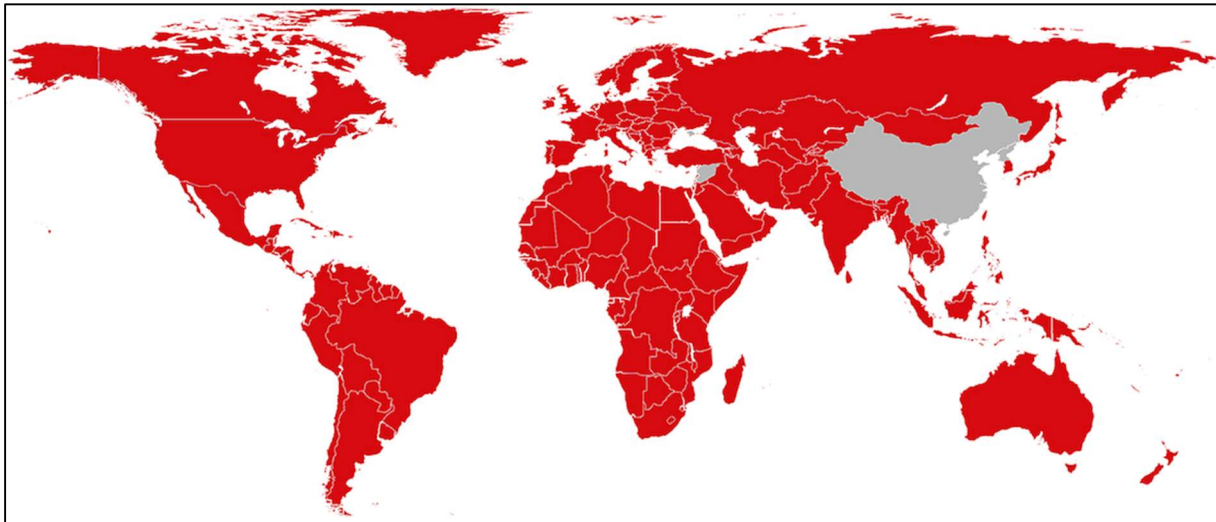


Figure 2 Pays où Netflix est disponible – Source Netflix.com

Cette expansion progressive fut agréablement adoptée par de nouveaux utilisateurs dans le monde entier qui voyaient en Netflix un nouveau moyen de consommer un éventail large de contenu sur différents supports. Cependant, cette arrivée ne fut pas si bien accueillie par certains acteurs du marché.

Lorsque la plateforme américaine fut lancée dans l'Hexagone, le secteur de la SVOD était principalement dominé par Canal Play lancé en 2011 par le groupe Canal +. Ce service comptait environ 800 000 membres et un catalogue composé de 1 700 films et 4 000 épisodes de séries (chiffres du webzine Puremédiast¹⁶). Mais le service a dû mettre la clé sous la porte car selon les dires du président du groupe Canal+, après quatre ans sur le marché face à Netflix, Canal Play est passé de 800 000 à 200 000 abonnés. « En deux ans, on a été rayés de la carte sur ce marché qui est en train de se substituer à la télévision. C'est terminé pour CanalPlay ». Au moment de cette déclaration (2018), le Centre national du cinéma et de l'image animée estime que Netflix regroupe 3,5 millions d'abonnés, ce qui représente 70% du marché de la SVOD en France. L'effet de l'arrivée de la plateforme américaine a aussi bousculé FilmoTV. En 2016, dans une interview¹⁷ pour CB News, le patron de ce service de vidéo à la demande créé en 2008 parle d'une concurrence déloyale. Celui-ci pointe du doigt le fait que les deux services ne

proposent pas le même registre. Netflix c'est en grande partie sur du contenu américains (80% selon lui), alors que FilmoTV c'est 60% de films européens, 40% de films français. En plus de cela, les opérateurs nationaux sont soumis à des règles (Chronologie des médias, financement des productions...) auxquelles la concurrence étrangère n'est pas soumise. C'est pour cela, qu'en février 2014, les présidents des trois grands groupes français (TF1, Canal+ et M6) ont adressé une lettre conjointe afin de demander une révision des obligations. Une demande qui s'est révélée être partagée par d'autres services membres de l'Union¹⁸.

Au début de l'année 2016, le vice-président de la commission chargé du marché numérique qui s'occupe de réviser la directive sur les services de médias audiovisuels (SMA) avait déclaré qu'il fallait « créer un terrain de jeu plus juste entre ces acteurs et les chaînes de télévision »¹⁹. C'est alors que le 10 octobre 2018, les députés ont voté (452 voix pour, 132 contre et 65 abstentions) la directive 2018/1808 qui vise à actualiser les règles qui régissent les services de médias audiovisuels. En plus d'imposer aux services numériques un quota de 20% de mise en avant du contenu européen, la commission a donné la possibilité aux Etats membres d'obliger les services numériques étrangers à contribuer au financement de la production locale.

Aujourd'hui, en cette mi-année 2021, Netflix avoisine les 8 millions d'abonnés en France. Un chiffre qui selon Médiamétrie (société qui mesure l'audience et les usages des médias audiovisuels et numériques en France) ne représente pas la totalité des utilisateurs, puisque la plateforme serait consultée par 19 millions de personnes²⁰.

3. L'impact de Netflix

Outre un impact sur le marché que l'on vient d'évoquer dans la partie précédente, le contenu que propose Netflix (mais aussi les autres plateformes) a une forte influence dans d'autres domaines. Par exemple c'est le cas de la mini-série de sept épisodes, *The Queen's Gambit* (Le jeu de la dame), qui en octobre 2020 a bouleversé le monde des échecs. Regardée par 62 millions de foyers dans les 28 premiers jours suivant sa sortie, cette adaptation du roman de Walter Tevis a fait bondir de 1100% la vente des échecs de l'entreprise de production de jouet Goliath Games. De même, le site lichess.org (qui permet de jouer aux échecs en ligne) a enregistré des millions de nouveaux membres²¹. Il en va de même pour le documentaire sur la Formule 1 : *Drive to Survive*, qui a fait augmenter la popularité du sport automobile aux Etats-Unis selon le CEO de McLaren Racing Zak Brown²².

Plus récemment, la série Française *Lupin* qui fut un véritable succès dans le monde entier a profité d'une certaine manière à la France. A sa sortie le 8 janvier 2021, la série de cinq épisodes coproduits par Gaumont a su se hisser au sommet. Elle a en effet occupé la première place dans de nombreux pays comme la France, le Mexique, les Pays-Bas et est même entrée dans le TOP 10 des contenus les plus regardés aux Etats-Unis. En France, afin de répondre à la demande,

les romans qui relatent les aventures du gentleman cambrioleur ont été tirés à 60 000 exemplaires selon Hachette²³. En pleine pandémie, la série a permis de faire voyager plus d'un spectateur en plein cœur de Paris. Lors d'un échange avec France Inter, Omar Sy dit être conscient d'être sur Netflix, « une plateforme qui va nous porter [la série] à l'internationale, on a voulu faire une série française qui claque, qui porte la France, qui porte Paris. [...] Je compare toujours Netflix aux Jeux Olympiques des séries, et je crois qu'on a pris une médaille, [...] de bronze, d'or ou d'argent je ne sais pas, mais en tout cas on est sur le podium ».

4. Découvrabilité du contenu

Etant donné que l'on va s'intéresser à la mise en avant d'un certain type de contenu sur la page d'accueil de Netflix, il est nécessaire de définir la notion de découvrabilité. Dans une étude sur la conception et la gestion du choix dans les interfaces des plateformes menée par Fenwick McKelvey et Robert Hun, des chercheurs de l'Université de Concordia au Canada, on retrouve une définition de la découvrabilité comme étant un « media power » (pouvoir médiatique) donné aux plateformes de découverte de contenu (Netflix, Spotify...) qui coordonne les utilisateurs, les créateurs de contenus et les logiciels afin de rendre un contenu plus ou moins engageant (attractif). En 2016, Karen Yeung qualifia cette force médiatique comme étant une régulation par le design. Afin de préciser cette notion de découvrabilité, les chercheurs abordent la notion de « surround » et de « vecteur ».

Proposé pour la première fois en 1955 le concept de surround fut utilisé par Fred Turner afin de décrire l'exposition des photographies « The Family of Man » au Musée d'art moderne de New York. Cette exposition qui avait réuni 503 photographies de 273 photographes de 68 pays différents était selon Turner une exhibition qui marqua un tournant par rapport aux « médias de masse à source unique au profit d'environnements médiatiques multi-images et multi-sources », qui fut en quelque sorte le prototype du système médiatique d'aujourd'hui. Avec comme thème le portrait de l'humanité, l'exposition mettait en valeur les différences entre les Hommes et les différentes communautés. Organisé autour de 37 thèmes (naissance, travail, joie, éducation...) Turner a analysé l'agencement de l'exhibition afin de comprendre les choix de l'organisation et le contrôle qui en découle. Si on applique alors cette notion de surround dans un monde composé de plateforme digitale, cela revient à étudier la gestion de l'espace sur un écran composé de contenu de différente source (Netflix par exemple).

Cependant avec l'augmentation du nombre de support numérique le surround décrit par Turner nécessite un certain ajustement. Etant donné que les utilisateurs possèdent différents appareils et que les services sont disponibles sur différents supports il est nécessaire de prendre cette variable en considération. C'est pourquoi, les chercheurs de Concordia alimentent cette notion avec le travail de Kathleen Oswald et Jeremy Packer (2012) et la notion de flow 1.0 et 2.0. La version 1.0 du flow définit pour les nouveaux médias, l'ensemble matériel des pratiques,

techniques et technologiques qui intègrent les individus dans une dynamique temporelle et spatiale d'une économie et des attentes. Par la suite, ils proposent le flow 2.0 où « l'environnement médiatique n'est plus consacré à garder les spectateurs fixés sur une transmission, mais plutôt des transmissions à travers de multiples écrans qui guident les consommateurs à travers tout le temps et l'espace ». Les chercheurs précisent alors que la nouvelle version de ce flow s'appuie sur les logs générés par les utilisateurs afin de suivre les utilisateurs, identifier les moments, les lieux de connexions afin de consommer. Le flow 2.0 est donc étroitement la conséquence des services à la demande accessibles partout et tout le temps. Combiné alors avec la notion de flow, le surround répond à des contraintes dynamiques afin de s'adapter aux supports numériques des utilisateurs. Ainsi le surround que propose Nieborg & Poel est un service numérique accessible sur différents supports et répond à des contraintes dynamiques qui s'adapte aux individus de manière individuelle en fonction de différents paramètres comme la taille de l'écran et propose un contenu divers de sources différentes à des fins de consommation.

En étudiant le surround de Netflix en 2017, McKelvey et Hun décrivent la page d'accueil du service comme étant à deux tiers un espace promotionnel de son propre contenu. Et ce n'est que bien plus bas qu'apparaît le contenu recommandé en fonction de ce que l'utilisateur a déjà regardé. Même si le discours de Netflix met en avant continuellement un système de recommandation basé uniquement sur le contenu regardé par l'individu ou le groupe, la prééminence des Netflix Originals suggère un modèle d'intégration vertical²⁴.

Comme on a pu le voir, le surround rassemble en un seul lieu différents choix. Le vecteur quant à lui coordonne les interactions des utilisateurs avec des algorithmes qui peuplent le surround. Ainsi un vecteur interagit avec les algorithmes afin de remplir l'espace disponible sur l'écran avec un contenu basé sur des suppositions probabilistes issues de l'interaction des utilisateurs. Le fait d'aimer une catégorie spécifique sur Netflix déclenche un vecteur qui alimentera ensuite le surround. Les chercheurs soulignent que cette notion de vecteur est aujourd'hui très importante au sein des plateformes numériques. Selon l'étude de Solsman en 2018, les propres données de YouTube indiquent que les vidéos recommandées représentent 70 % du temps de visionnage sur le site²⁵. Les vecteurs permettent donc d'alimenter continuellement en contenu l'espace que va voir l'utilisateur mais aussi le contenu qui lui sera proposé à la fin du visionnage d'un élément sans avoir à chercher.

La découvrabilité a toujours fait partie d'un point crucial pour les plateformes. Afin de garantir l'engagement des utilisateurs il est essentiel de créer un sentiment de personnalisation individuelle en constante évolution. Une philosophie diamétralement opposée à la télévision classique à diffusion continue.

5. Projet Streaming for Good

Réunissant aujourd'hui des étudiants de la L3, M1 et M2 MIAGE, le projet Stream4Good débuta avec l'initiative de recherche nommé Prime-Space, menée par des étudiants de l'école des médias et du numérique de la Sorbonne avec le soutien du maître de conférences à Paris 1, Nicolas Herbaut et le lead développeur Sylvain Vocale. Ce projet démarra avec la création d'une extension de navigateur qui vise à récolter la position de chaque élément que l'on trouve sur la page d'accueil de Netflix.

Tout en respectant la norme RGPD ce plugin fonctionne en arrière-plan. Lorsqu'un utilisateur se rend sur la plateforme de streaming le code derrière l'extension fait en sorte de récupérer l'ensemble des vignettes (thumbnails) affichées sur l'écran de l'utilisateur, le temps passé et le contenu sélectionné (ou non) entre autres. Par la suite, afin de récupérer différentes informations sur le contenu observé, un processus d'identification va permettre de faire une correspondance avec les données observées et l'Internet Movie Database (IMDB) où l'on trouve des métadonnées utiles pour les analyses. La récolte de ces données a permis par la suite de débiter une analyse afin de comprendre les mécanismes de mise en avant du contenu dans un environnement dynamique, régi par des algorithmes de recommandation et basé sur l'interaction de l'utilisateur.

Grégoire Bideau (étudiant de l'EMNS) a alors travaillé sur la définition de la notion de Prime-Space²⁶ que l'on peut comparer au Prime-Time (notion que l'on retrouve dans une programmation télévisuelle) mais dans un environnement bien différent comparé à la télévision traditionnelle. De par ses recherches, il en a conclu que le Prime-Space est somme toute une notion complexe à définir étant donné les nombreux paramètres qui rentrent en jeu, et souligne le fait qu'elle dépend du périmètre que l'on fixe pour l'étude. Une période de temps donnée ou à tout moment ? Pour un utilisateur unique ou un groupe ? En plus de cela il s'accorde à dire que la chose la plus importante pour cette plateforme de vidéo à la demande est le positionnement du contenu, étant donné la mise en page en forme de grille, où le choix appartient à l'utilisateur. Il dit même que la page d'accueil de Netflix est un « algorithmically-generated nudge »ⁱ (coup de coude généré par des algorithmes) dont l'objectif est d'attirer l'attention sur une centaine de titres et non le catalogue entier. Une petite sélection précise affinée de manière dynamique qui répond aux changements des comportements (et/ou

ⁱ Théorie du Nudge est un concept des sciences du comportement mis en avant par Richard Thaler et Cass Sunstein en 2008, qui est issu des pratiques de design industriel et qui montre que des suggestions indirectes peuvent, influencer la finalité, les choix et la prise de décision des individus et des groupes avec une efficacité similaire voir supérieur comparée à une instruction directe, explicite. – Théorie du nudge, Wikipédia

consommations) de l'utilisateur et du groupe qui met en valeur des tendances à l'échelle de la population.

Par la suite cette initiative a évolué en devenant le projet Streaming 4 Good (Figure 3).

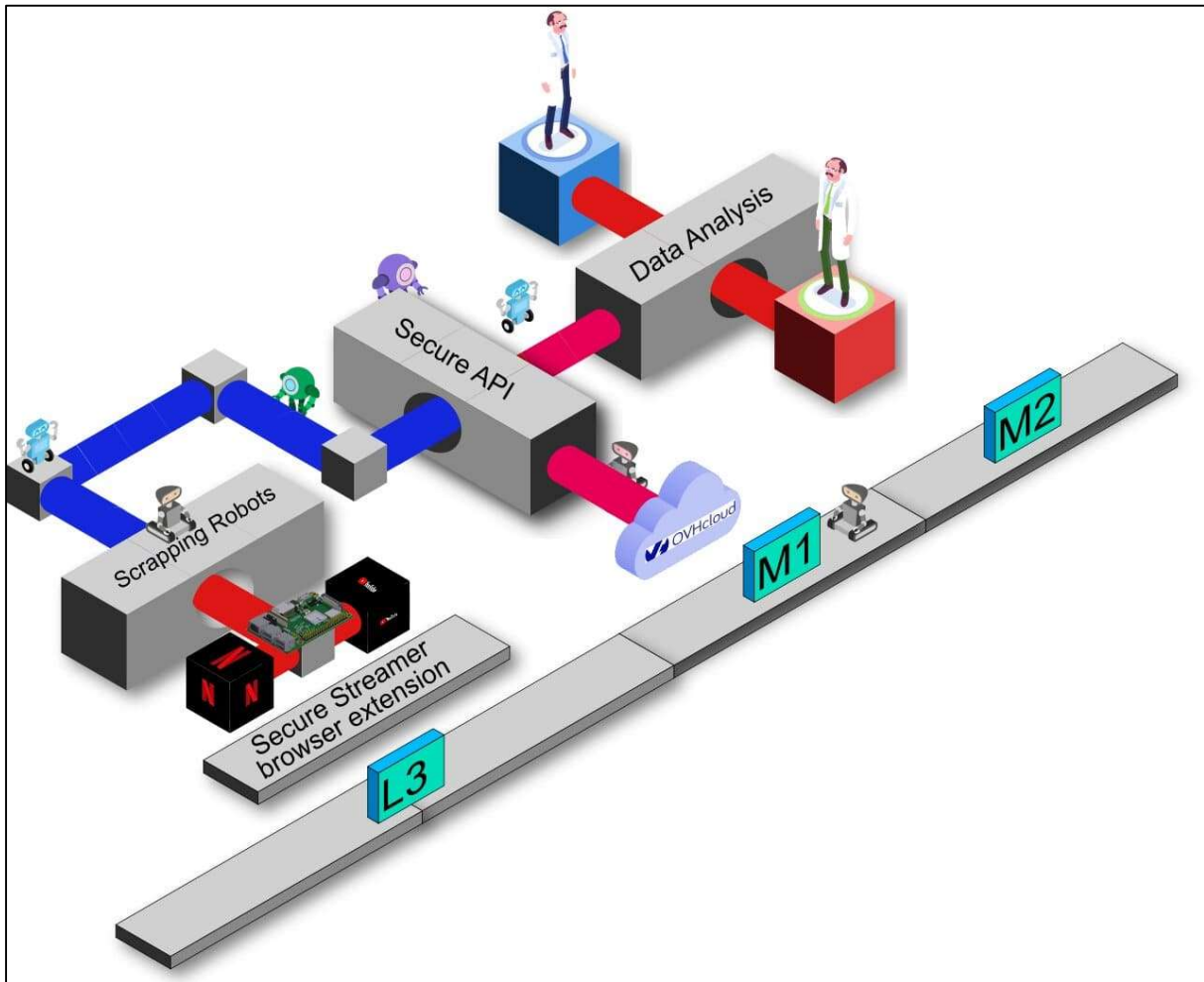


Figure 3 Streaming 4 Good schéma

Streaming 4 Good²⁷ regroupe aujourd'hui des étudiants de 3 années universitaires différents. Comme on peut le voir sur la figure 3, au début de cette « chaîne » on retrouve les étudiants de Licence 3. A ce niveau, ils travaillent sur la mise en place des robots de scrapping déployés sur des systèmes électroniques et informatiques autonomes (systèmes embarqués), qui enregistrent des données issues d'un environnement de laboratoire contrôlé. Cette source de donnée complémentaire vient alors s'ajouter aux données humaines issues de l'extension Prime-Space, pour alimenter les données utilisables par les étudiants de plus haut niveau. La mise en place des systèmes embarqués, étend par la même occasion le projet d'étude des plateformes de vidéo à la demande à un autre géant de l'industrie du divertissement : YouTube.

Ensuite, on trouve des étudiants de Master 1 qui s'occupent du déploiement ainsi que de l'audit de sécurité et fiabilité de l'ensemble des micro-services (architecture technique du projet) dans un cloud public, à travers un module de veille et projet commun. Les données récupérées sont

mises en ligne de façon sécurisée et anonyme sur un serveur Cloud situé en France. De cette façon les étudiants qui souhaitent les analyser, peuvent le faire de manière simple. D'autant plus, qu'elles sont proposées par le biais d'une API RESTⁱ qui respecte le paradigme HATEOSⁱⁱ qui met en évidence des liens hypermédia pour naviguer au sein des données.

Enfin, au bout de la chaîne des étudiants de Master 2 exploitent les différentes données disponibles à des fins d'analyses et recherches. En cette année universitaire 2021, on retrouve Léna Albert qui s'intéresse à YouTube et la promotion du contenu conspirationniste et complotiste, Marie-Françoise Edroux qui se penche sur le rôle du système de recommandation de Netflix dans la diversité des contenus et enfin moi, qui m'interroge sur Netflix et l'application d'une directive qui impose aux plateformes de SVOD de diffuser 30% de contenus issus de l'Union Européenne.

ⁱ Une API REST est un ensemble de règles qui définissent un style architectural qui permet à un programme sur un serveur d'échanger avec des applications clientes grâce à un protocole de transfert REST.

ⁱⁱ « Hypermedia as the Engine of Application State » définit le fait qu'un client REST doit se déplacer au sein d'une application Web seulement en utilisant des URI.

II. Méthodologie

Dans cette partie nous allons voir le processus de récupération des données du projet Streaming for Good ainsi que les différentes méthodes utilisées afin de répondre à la question de recherche.

1. Données Stream for Good

Comme nous avons pu le voir dans la présentation du projet Streaming for Good, le contenu qui alimente la base de données (concernant Netflix) provient de deux sources différentes.

Dans un premier temps on retrouve un abonné Netflix qui utilise son navigateur Chrome où se trouve l'extension (plugin) Prime-Space. Lorsque celui-ci se rend sur la plateforme de streaming, le code derrière l'extension va récupérer les différentes informations qui se trouvent sur la page sur laquelle il navigue. Comme le montre la Figure 4, ce processus est composé de plusieurs éléments.

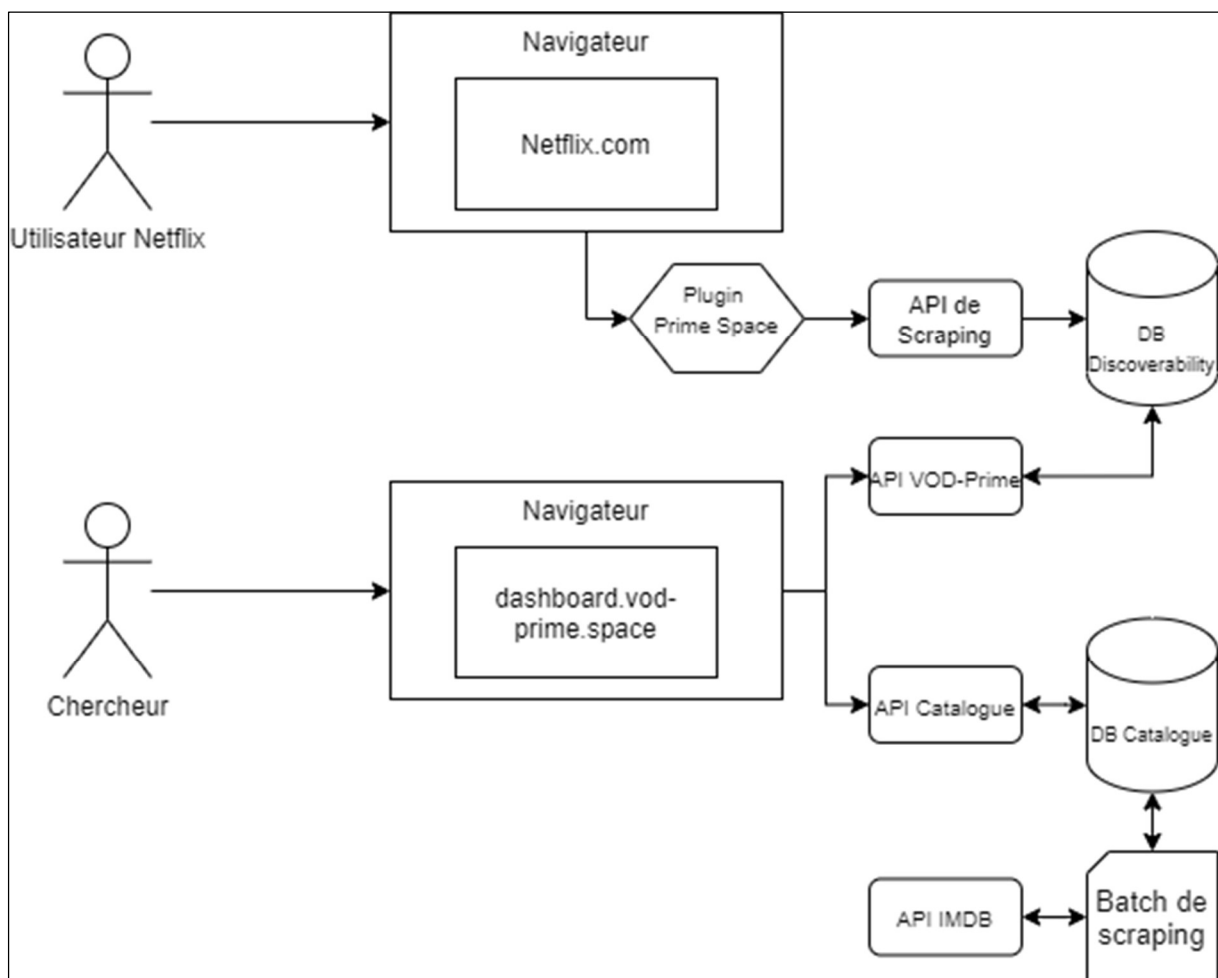


Figure 4 - Schéma architecture technique Stream4Good avec Humains

Si l'on suit le processus depuis le navigateur de l'utilisateur Netflix, on voit que l'extension Prime-Space est connecté à une API de scraping (collecte de donnée) qui envoie les données collectées à la base de données DB Discoverability. Les données sont ensuite restituées au(x) chercheur(s) par le biais de l'API Vod-Prime. Afin de faire correspondre les données et récupérer des métadonnées sur le contenu obtenu, une API Catalogue va récupérer l'ensemble des données stockées dans une base de données (alimenté par un batch de scraping - traitement par lot de donné) qui contient les informations du catalogue Netflix.

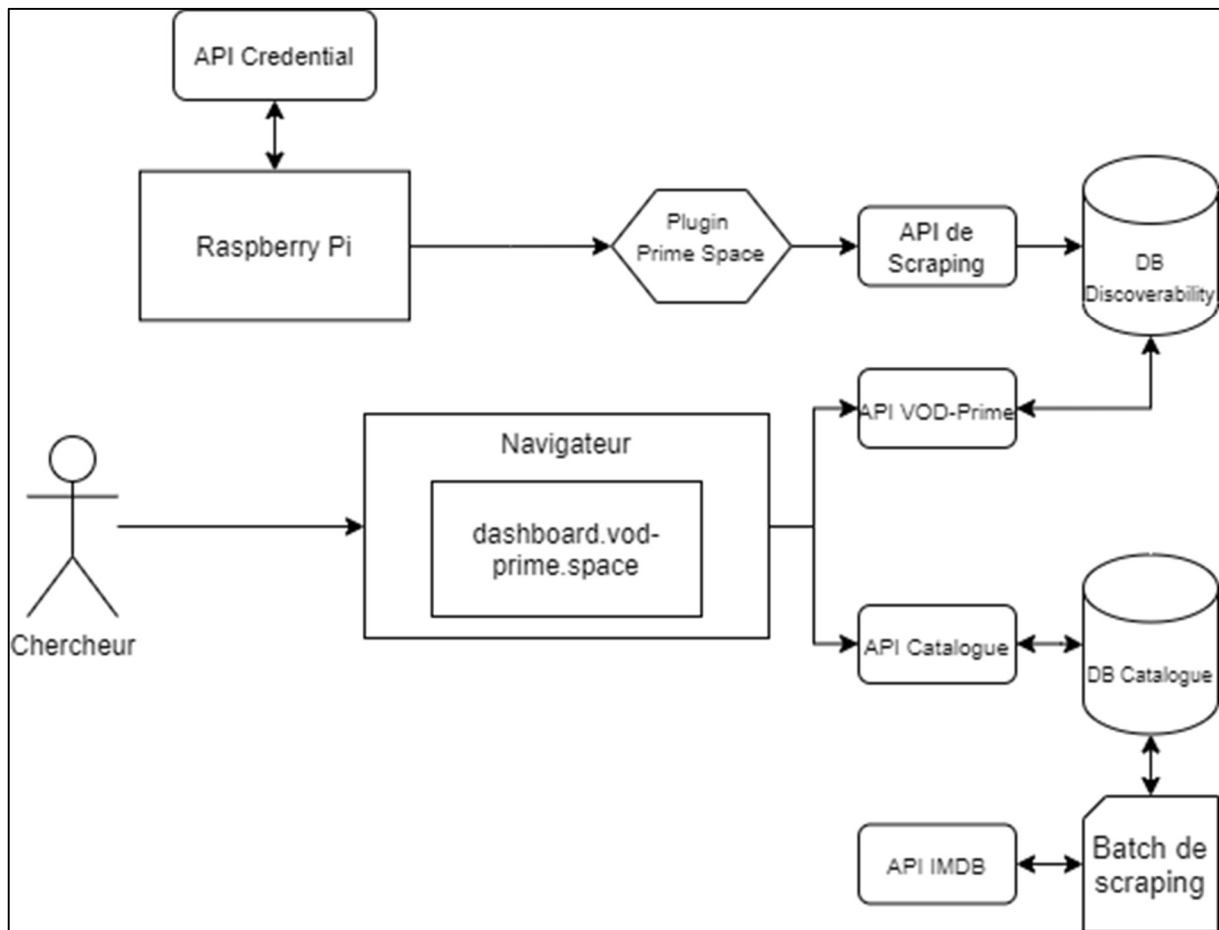


Figure 5 - Schéma architecture technique Stream4Good avec Robots

Parce que les utilisateurs peuvent être rare, ou consomme le divertissement proposé par Netflix d'une autre manière, des étudiants de la Licence 3 MIAGE se sont penchés sur l'instauration de robots dans le processus. Comme le montre la Figure 5, l'utilisateur Netflix est ici remplacé par un programme contenu sur un Raspberry Pi qui va pouvoir exécuter des actions définies par avance mais de manière automatique et continue, ce qui va permettre d'alimenter les données disponibles sans interruption. Comme l'explique des étudiants du projet dans un rapport publié en 2021 intitulé « Project prime space distributed scraping »²⁸ on retrouve 3 différents robots : Robot Watcher, Robot Thumbnail et Robot Direct. Le premier est programmé pour regarder du

contenu (série ou film) selon un personaⁱ. Ensuite le Robot Thumbnail viendra récupérer l'ensemble des contenus qui sont recommandés sur la page d'accueil du service à la suite des actions du Robot Watcher. Enfin, on retrouve le Robot Direct qui récolte le planning d'une des dernières fonctionnalités ajoutées à Netflix : Direct. Une fois que les données sont agrégées, celles-ci sont disponibles depuis un tableau de bord (Dashboard) accessible depuis le réseau.

Il est alors possible d'accéder depuis le tableau de bord à différents outils qui peuvent faciliter l'analyse du contenu. On retrouve par exemple (Figure 6) une représentation graphique de données statistiques (dataviz) sur laquelle il est possible de choisir une session utilisateur (anonymisée) et effectuer une recherche spécifique ; Titre, type, acteur, genre, année... les possibilités sont nombreuses. On obtient alors une grille qui montre selon un code couleur la correspondance de la recherche effectuée. Vert pour signaler une correspondance avec le(s) critère(s) de recherche, rouge pour un résultat négatif et noir pour signaler que le contenu n'a pas de métadonnée. Il est aussi possible d'accéder à un outil de recherche visuelle qui permet de trouver différentes informations sur certains contenus trouvés sur Netflix.

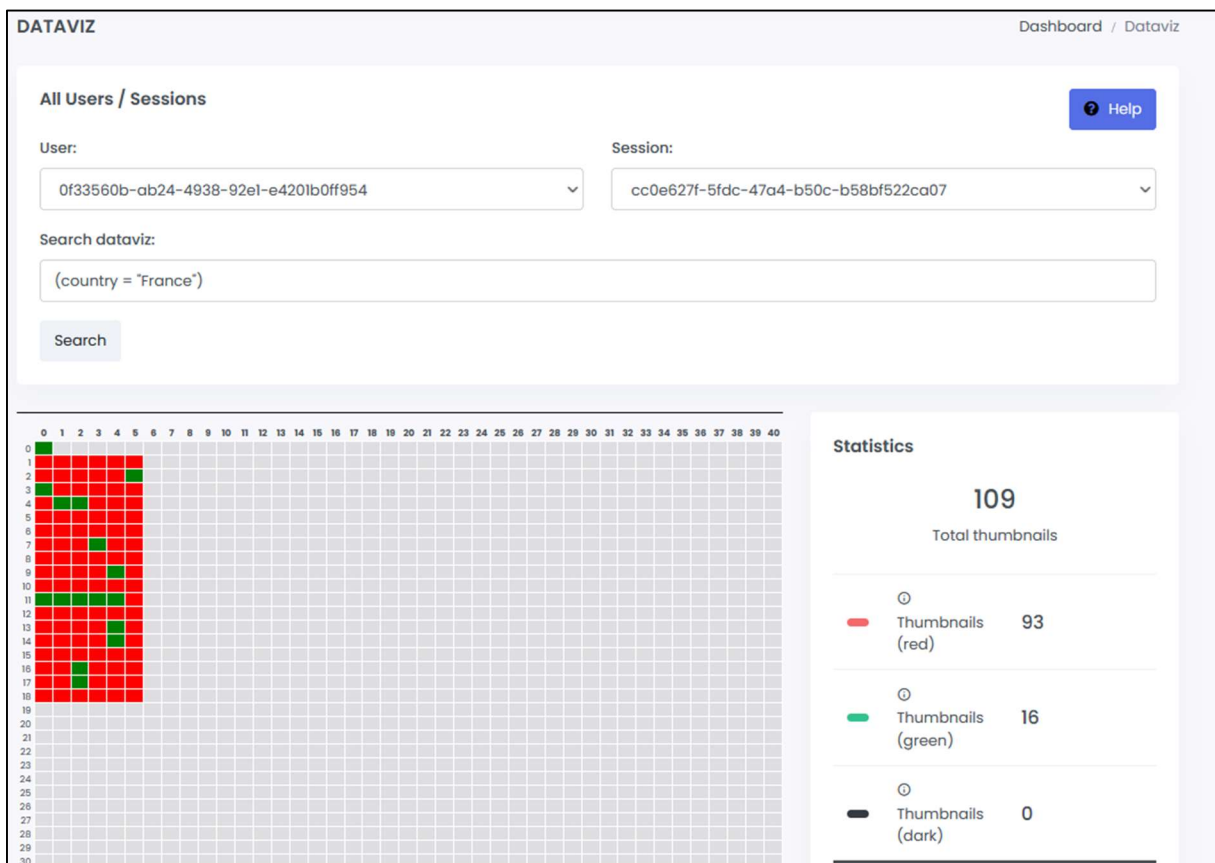
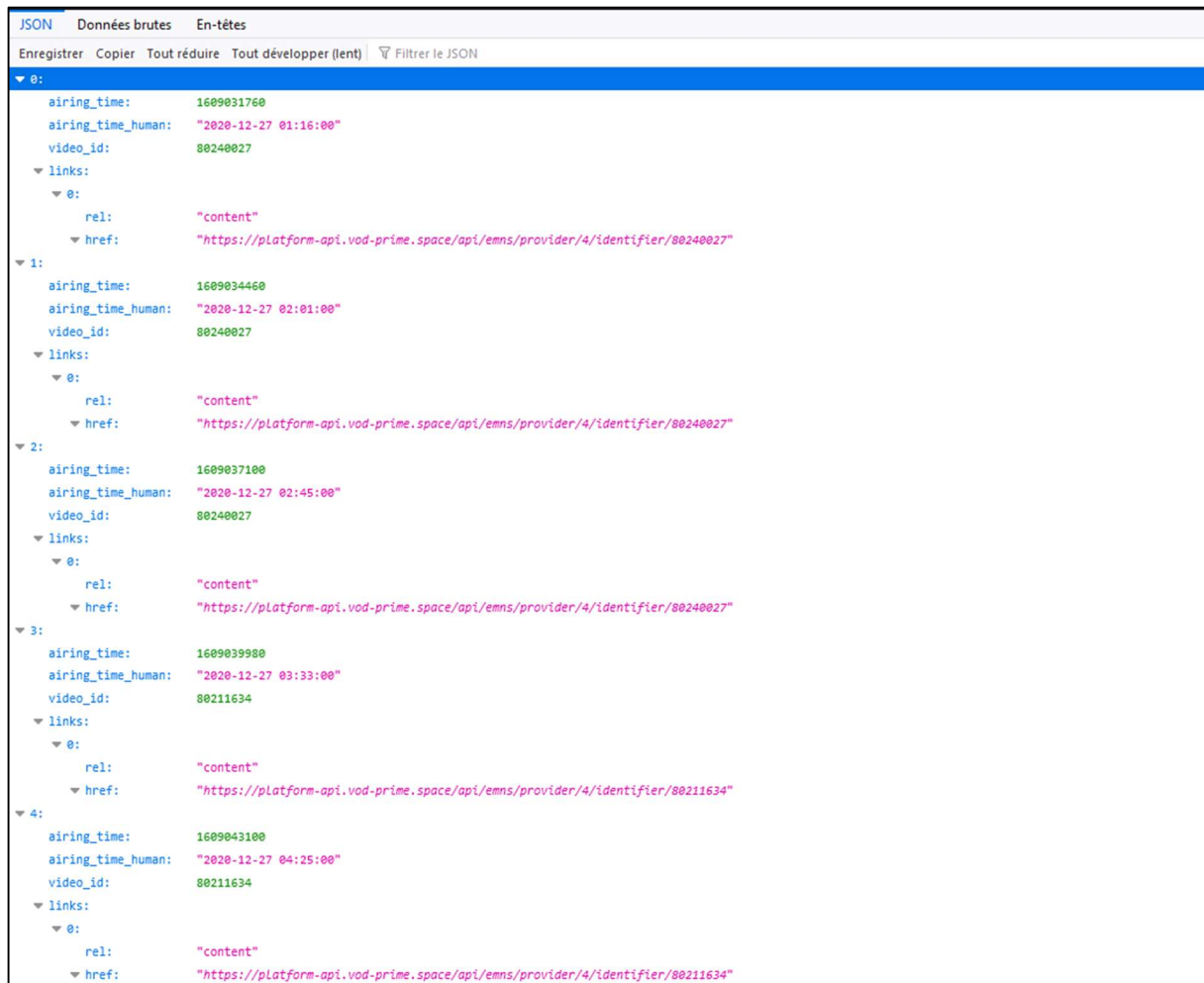


Figure 6 Dataviz Dashboard VOD Prime Space

ⁱ Un persona est un personnage semi-fictif qui représente une population qui ont des habitudes de consommation similaire.

En plus de ces outils visuels, il est possible d'accéder à l'ensemble des données par le biais d'une API. Représenté de manière structurée (Figure 7) à l'aide du format JSON (JavaScript Objet Notation) il est possible de naviguer grâce à des liens HyperText entre les différents éléments enregistrés dans la base de données.



```
JSON  Données brutes  En-têtes
Enregistrer Copier Tout réduire Tout développer (lent) Filtre le JSON
0:
  airing_time: 1609031760
  airing_time_human: "2020-12-27 01:16:00"
  video_id: 80240027
  links:
    0:
      rel: "content"
      href: "https://platform-api.vod-prime.space/api/emms/provider/4/identifieur/80240027"
1:
  airing_time: 1609034460
  airing_time_human: "2020-12-27 02:01:00"
  video_id: 80240027
  links:
    0:
      rel: "content"
      href: "https://platform-api.vod-prime.space/api/emms/provider/4/identifieur/80240027"
2:
  airing_time: 1609037100
  airing_time_human: "2020-12-27 02:45:00"
  video_id: 80240027
  links:
    0:
      rel: "content"
      href: "https://platform-api.vod-prime.space/api/emms/provider/4/identifieur/80240027"
3:
  airing_time: 1609039980
  airing_time_human: "2020-12-27 03:33:00"
  video_id: 80211634
  links:
    0:
      rel: "content"
      href: "https://platform-api.vod-prime.space/api/emms/provider/4/identifieur/80211634"
4:
  airing_time: 1609043100
  airing_time_human: "2020-12-27 04:25:00"
  video_id: 80211634
  links:
    0:
      rel: "content"
      href: "https://platform-api.vod-prime.space/api/emms/provider/4/identifieur/80211634"
```

Figure 7 API Direct - Stream for Good

Afin de faciliter encore plus l'exploitation des données par le biais de code, le projet met à disposition une bibliothèque python nommée S4Gpy²⁹. Celle-ci permet d'appeler des fonctions qui récupère certaines données en fonction de différents paramètres.

Enfin, étant donné que lorsque l'humain ou un robot se connecte à Netflix le processus est un enchaînement d'action, une trace est enregistrée par le système. Ces logs informatiques peuvent ensuite être exploités et analysés en utilisant la fouille de processus, un domaine au croisement du data mining et de l'approche processus.

Afin d'analyser le contenu que propose Netflix et sa conformité par rapport à la directive Européenne, je fus amené à utiliser deux processus d'analyse différents : la data analyse et le process mining.

2. Data Analyse

Comme nous avons pu le voir dans la section précédente qui présente les différentes données disponibles grâce au projet Streaming for Good, de nombreuses données sont accessibles avec des outils visuels. Cependant, ceux-ci ne permettent pas une analyse quantitative et qualitative sur un ensemble de données précises. C'est pourquoi, une partie de cette recherche consiste à utiliser l'exploration des données / fouille de données : la data analyse. Tout comme le Data Mining qui consiste à identifier des patterns, l'analyse des données permet de traiter des données et les transformer en informations utiles et exploitables.

En utilisant l'API VoD-Prime, les fonctions disponibles avec la bibliothèque client Python S4Gpy, qui permet de récupérer certaines métadonnées à l'aide de l'identifiant d'un contenu (depuis la base de données du projet Stream4Good), l'outil d'automatisation open source Selenium qui donne la possibilité d'écrire des scripts de récupération des données depuis un site (comme IMDB), j'ai pu extraire différentes informations. Par la suite, il est possible d'obtenir des représentations graphiques en utilisant la bibliothèque Python Matplotlib ou bien de prendre les données et les traiter avec Excel.

3. Process Mining

Définition du Process Mining

A la croisée du data mining et la modélisation et analyse de processus, le process mining est une discipline de recherche encore jeune. Les recherches dans ce domaine ont débuté en 1999. Au début la discipline n'avait que peu de donnée à analyser et les techniques d'exploration des processus étaient assez naïves et inutilisables. Cependant lors de la dernière décennie les sources des données se sont multipliées et les techniques d'analyses se sont perfectionnées. Aujourd'hui des algorithmes d'exploration de processus sont utilisés dans différents secteurs et systèmes (commerce, marketing, publicité...). L'informaticien et professeur titulaire à l'Ecole supérieure polytechnique de Rhénanie-Westphalie, Willibrordus Martinus Pancratius van der Aalst présente les nombreux travaux dans le domaine du process mining du groupe (qu'il dirige) Process and Data Science (PADS) dans un ouvrage simplement intitulé « Process Mining, Data Science in Action ».

Au fur et à mesure des années, de nombreuses solutions indépendantes pour l'exploration des processus furent développées. Afin d'éviter la construction d'outils dédiés uniquement à un

nouvel algorithme, l'initiative ProM fut initiée. Cet environnement qui offre une base commune à tous les chercheurs de la discipline et facilement adaptable. Toutes les nouvelles techniques peuvent être ajoutées comme une simple extension (plugin). La première version (ProM 1.1 – 2004) accueillait 29 plugins différents, alors que la dernière version publiée en 2010 (ProM 6) possède pas moins de 300 extensions.

En plus de cet solution open source, de nombreuses entreprises proposent leurs propres solutions à l'aide de logiciels le plus souvent payants. On a par exemple Lexmark qui propose Perceptive Process Mining, ou encore Fluxicon et sa solution Disco, qui offre la possibilité de charger des document CSV, mapper des colonnes de manière automatique. Selon les dires de Wil van der Aalst, l'algorithme qui est utilisé par Disco est une version amélioré du Fuzzy Miner de ProM.

Dans le cas du projet Streaming for Good, la récolte des données suit un processus bien défini. Lorsqu'un utilisateur ou un robot arrive sur une des pages de Netflix (Accueil, Programmation Direct, visionnage d'un contenu) tout un processus d'affichage est déclenché. La génération de la page qui est alors constituée d'un ensemble d'éléments et le résultat d'un processus. Bien que celui-ci ne soit pas humain mais dirigé par des algorithmes (Ted Sarandos - 2014 : « *Les algorithmes pilotent l'ensemble de notre site web - il n'y a pas de pouce d'espace éditorial non calculé* ») il est possible d'en extraire une trace, un log informatique. Par la suite, afin de comprendre l'organisation des différents éléments affichés, la séquentialité et l'organisation du contenu on peut utiliser des algorithmes dédiés à la fouille de processus.

Différents types d'algorithmes

Aujourd'hui, lorsqu'on parle d'analyser des traces de nombreuses solutions sont disponibles. Entre les 300 extensions de la bibliothèque open source ProM, et les solutions développées par certaines entreprises on peut vite se perdre. Toutes basées sur des algorithmes, il est important de voir les différentes versions de ces méthodes afin de trouver celle qui correspond le mieux à l'objet de recherche.

Pour illustrer le fonctionnement des algorithmes qui seront présentés, nous utiliserons une trace simple (au format XES) afin de mettre en pratique les méthodes. Celle-ci est composée de trois cas et quatre activités (A, B, C, D).

- 1x Cas 1 A B C D
- 1x Cas 2 A C B D
- 1x Cas 3 A E D

Algorithme alpha

Lors de l'initiation à la fouille de processus nous sommes généralement amenés à utiliser l'algorithme alpha (α -algorithm) afin de comprendre le fonctionnement de cette discipline de recherche. Composé de seulement 8 lignes (Process Mining, Data Science in Action - Page 186) cet algorithme qui génère un réseau de Petri fut l'une des premières méthodes développées pour le process mining. Mais afin d'analyser et d'identifier des comportements, des boucles et autres actions celui-ci a vite montré ses limites. Il est en effet possible d'obtenir un même modèle avec deux traces (logs informatiques) différentes. La Figure 8 montre un réseau de Petri qui est le résultat de l'analyse d'une trace.

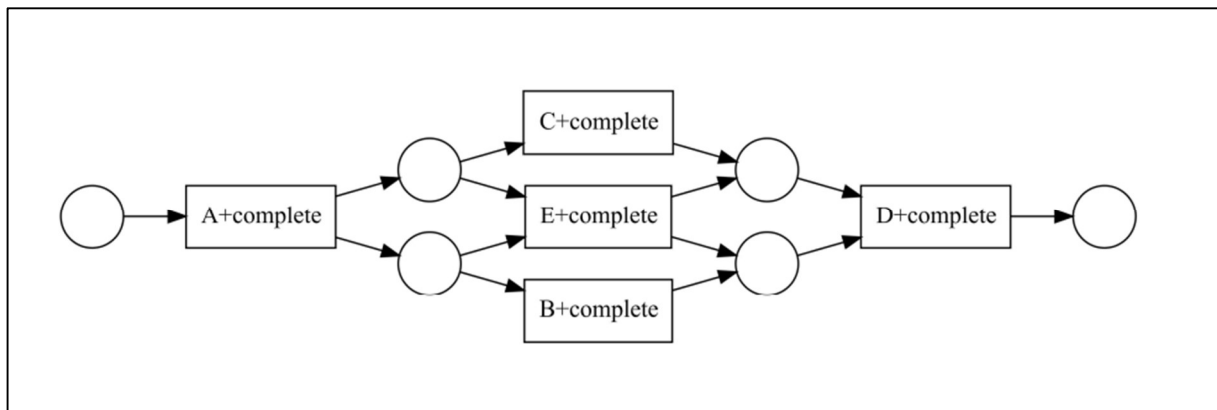


Figure 8 Réseau de Petri - Algorithme Alpha - ProM

Heuristic Miner

Après des évolutions de l'algorithme alpha (α^+ et α^{++}), Ton Weijters propose l'Heuristic Miner. Avec une approche heuristique, il répond avec sa solution à différents problèmes / limites de l'algorithme alpha, ce qui rend alors sa méthode plus pertinente. En effet, contrairement à son prédécesseur qui se basait sur les successions immédiates, l'Heuristic Miner identifie dans un premier temps les relations au sein du log et utilise ensuite les relations identifiées pour construire un modèle de processus. Toujours en utilisant la même trace la Figure 9 montre le réseau de Petri que l'on obtient en utilisant cette fois-ci l'Heuristic Miner. On voit clairement une évolution dans la représentation des deux résultats. Dans le cas de la Figure 8, si on déroule manuellement le réseau il est possible d'obtenir par exemple une activité C et E dans un même cas, or cela est impossible à obtenir si on fait le même exercice avec la Figure 9.

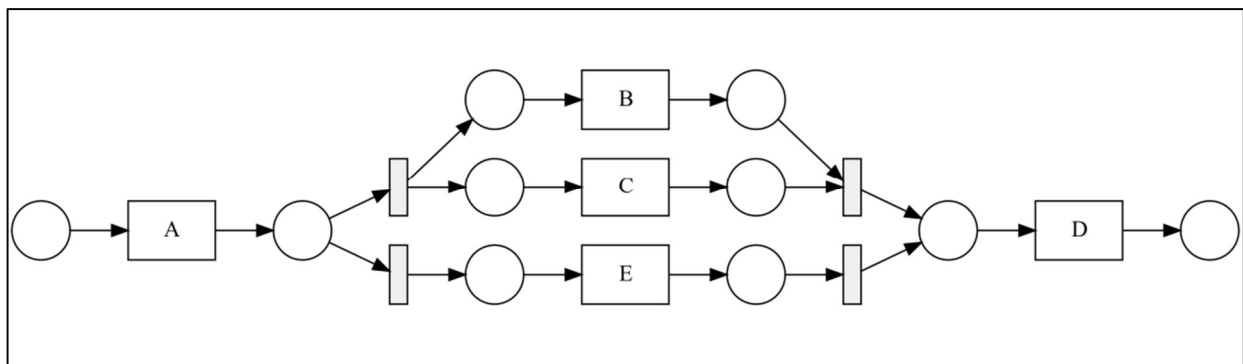


Figure 9 Réseau de Petri - Algorithme Heuristic Miner - ProM

Fuzzy Miner

Développé par Christian W. Günther, le Fuzzy Miner est considéré comme l'un des algorithmes les plus performants. Ce fut d'ailleurs le premier à pouvoir exploiter de manière efficace les traces à grand nombre d'activités et peu structurées. Avec la possibilité de faire des abstractions ou des agrégations, il est possible de regrouper certaines activités et leurs relations afin de ne garder (lors de la représentation finale) que celles qui sont importantes dans tout le log. Ainsi, cet algorithme se base sur la signification (l'importance d'une activité) et la corrélation (la proximité de deux activités successives)³⁰ des activités.

Lorsque l'on passe la trace à travers l'algorithme Fuzzy Miner, on obtient cette fois-ci une autre représentation. Il ne s'agit plus d'un réseau de Petri mais d'un Fuzzy model (Figure 10). Lors la sélection du traitement par Fuzzy Miner, ProM nous propose alors la possibilité de sélectionner des métriques (indicateurs). Mesurer l'importance des événements par leur fréquence relative, mesurer l'importance de deux événements par la fréquence de leur observation consécutive ou encore mesurer la corrélation de deux événements par leur proximité temporelle.

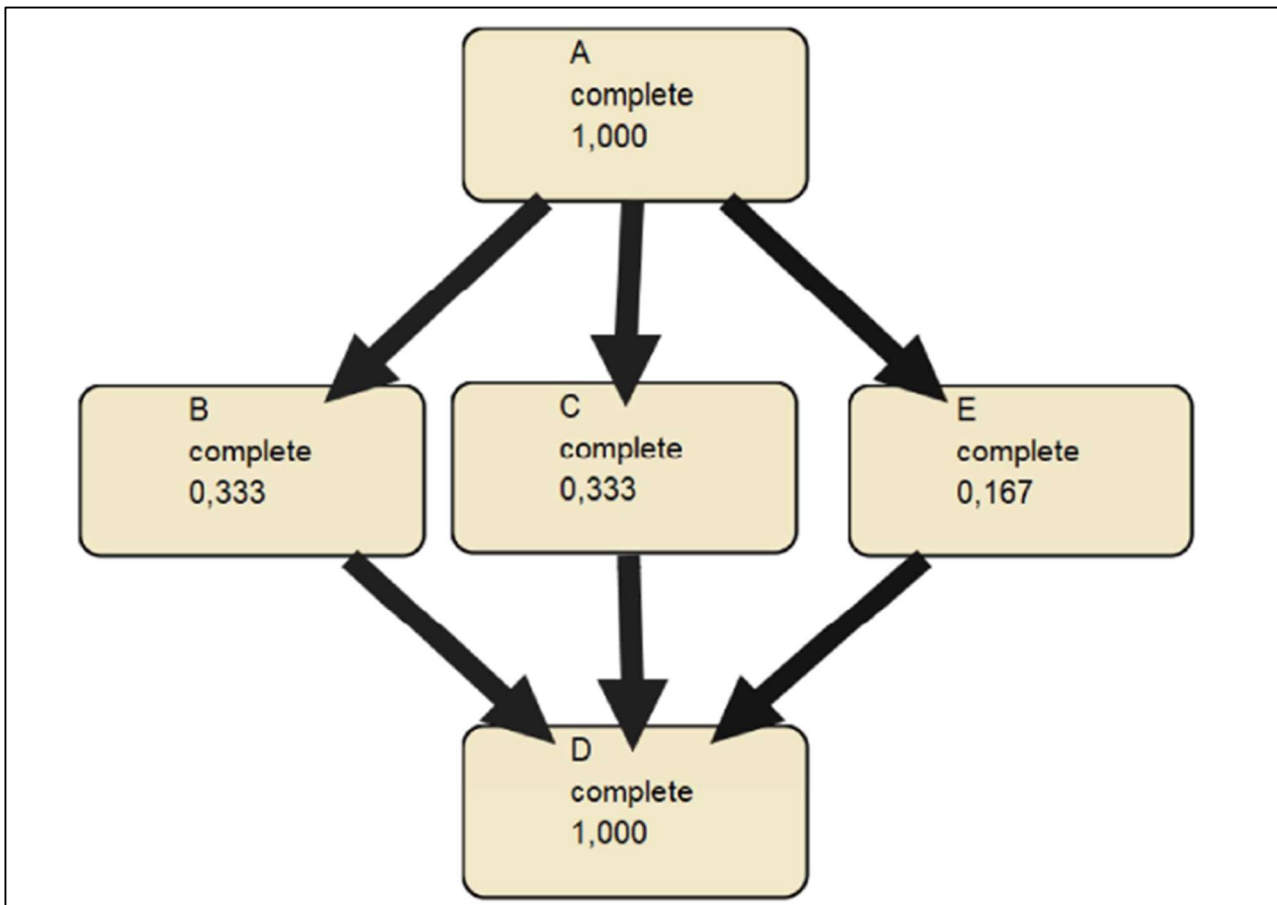


Figure 10 Fuzzy Model - Algorithme Fuzzy Miner - ProM

Disco Miner

Basé sur les travaux de Christian W. Günther et son algorithme Fuzzy Miner, le Disco Miner a été développé par Fluxicon. L'entreprise a repris le Fuzzy Miner et ses fonctions de simplification et de mise en évidence des activités importantes afin d'y ajouter des nouveaux outils de modélisation et des indicateurs de performance (process metrics). Le logiciel Disco donne accès à de nombreux outils comme la découverte de processus automatisé, des statistiques sur les processus, l'analyse des variantes et des cas individuels, le filtrage, ou encore la mise en évidence des performances des différents processus. Dans un document³¹ qui s'approche d'un « libre blanc », Fluxicon souligne la facilité d'utilisation, la fidélité ainsi que la performance de leur solution.

Après le traitement de la trace à trois cas, le Disco Miner nous donne en résultat la Figure 11. On y retrouve nos différentes activités, et un indicateur de la fréquence absolue de chacun. Mais ce résultat n'est pas la chose que nous donne Disco après le traitement de la trace. Il est en effet possible d'identifier des variants, mettre en valeur des statistiques etcetera. Toutefois ces outils ne nous sont pas d'une grande utilité dans l'analyse d'une petite trace.

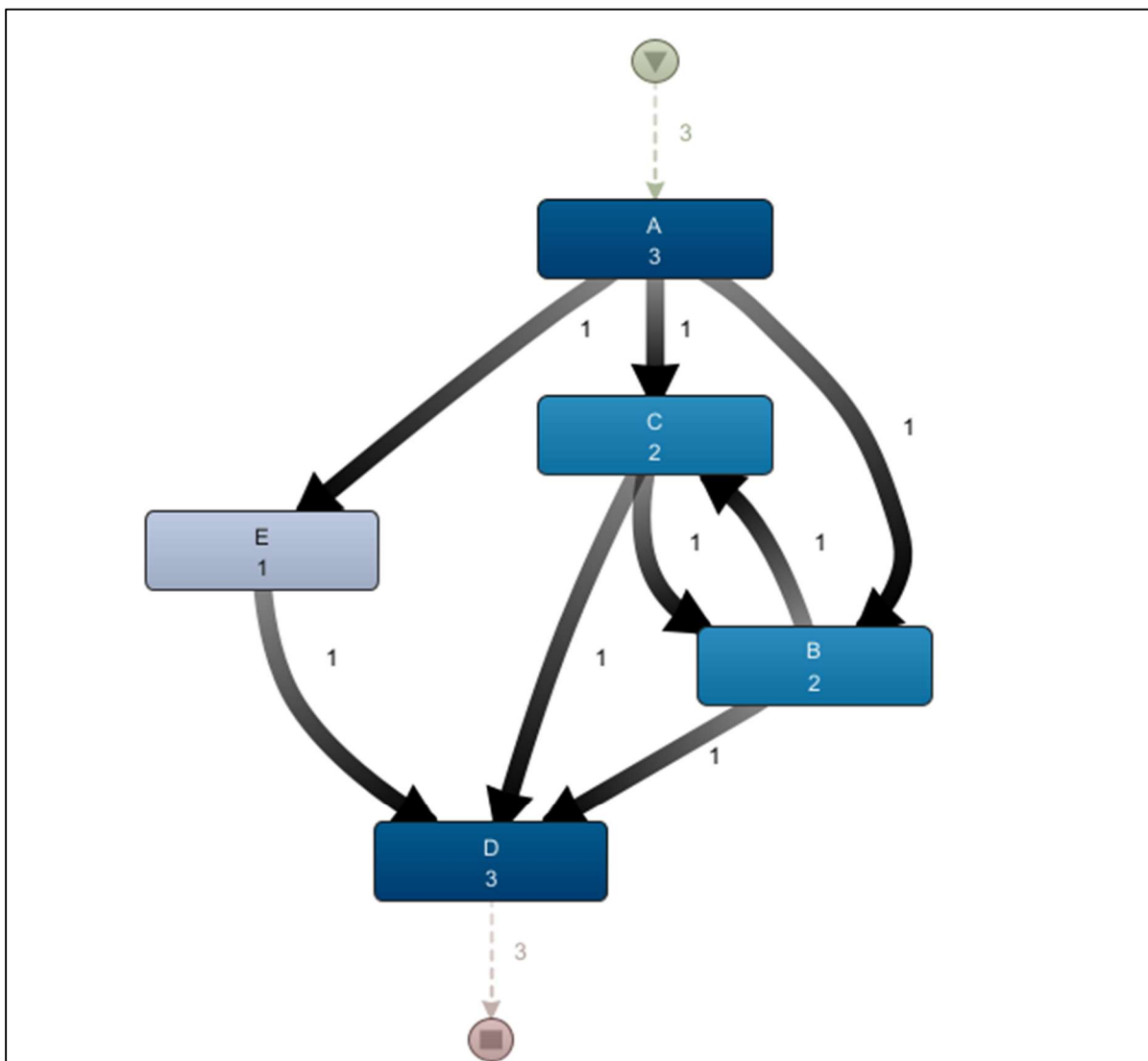


Figure 11 Disco Miner Model - Algorithme Disco Miner - Disco

III. Résultats et analyses

Afin de répondre à la question de recherche, nous allons dans un premier temps analyser les différentes données récupérées sur l'onglet Direct de Netflix, et ensuite nous passerons à l'exploitation des traces générées par les utilisateurs et les robots lors de leur navigation et utilisation du service.

1. Analyse du Direct

Instaurée progressivement en France depuis la fin de l'année 2020, la fonctionnalité « Direct » (Figure 12) est ce qui se rapproche le plus de la télévision traditionnelle. Avec une programmation connue à l'avance, Netflix offre la possibilité aux spectateurs de se laisser porter par une programmation linéaire où il ne faut plus faire de choix, car celui-ci est fait au préalable par le géant Américain.

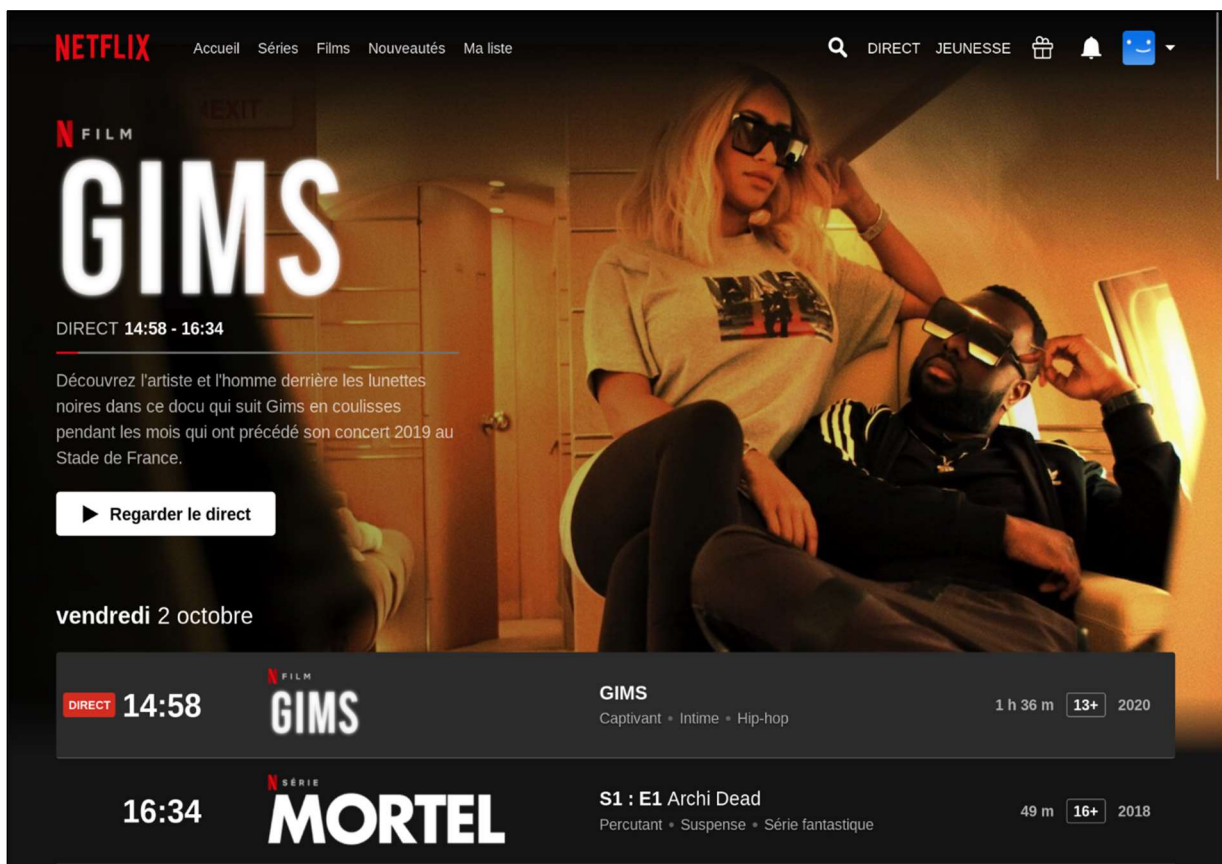


Figure 12 En Direct - Netflix - Source : about.netflix.com

Afin d'analyser la mise en avant du contenu européen au sein de cette nouvelle fonctionnalité, il est dans un premier temps important de présenter la notion des heures de grande écoute. Cette traduction des termes anglais prime time désigne une partie de la journée (le soir) où l'audience est la plus importante. Après une longue journée de travail, d'éducation... ce moment de la journée est l'occasion pour les différents membres d'une famille (par exemple) de se rassembler afin de se divertir. Dans le domaine télévisuel ces heures représentent une période fortement appréciée par les annonceurs publicitaires³².

Comme la Directive Européenne impose de mettre en avant du contenu européen, il est intéressant de comparer la mise en avant durant les heures de grande écoute et durant le reste de la journée. Généralement il s'agit de la période qui suit les journaux télévisés de la soirée. Cependant, il n'y a pas de définition précise qui indique un début et une fin à cette période. D'autant plus que la télévision peut modifier la durée des émissions et des publicités afin de se s'adapter à un horaire de diffusion précis. En fonction de la chaîne et d'événements d'actualité, ces horaires ne sont pas fixe. En revanche, on parle souvent de soirée en deux voire trois parties avec la première qui débute tout de suite après le journal (première partie de la soirée) qui est diffusé vers 20h.

Pour cela, dans un premier temps nous proposons un premier tableau récapitulatif des contenus dont la diffusion commence dans l'intervalle 19h/1h puis un autre tableau qui se concentre cette fois-ci sur la première partie de soirée, avec les contenus dont la diffusion débute entre 21h et 22h.

Comme cela fut expliqué dans la partie dédiée aux données du projet Streaming for Good (p.17), un Robot Direct (programme sur un système embarqué) se charge de récupérer la programmation de l'onglet Direct de Netflix. Par la suite, afin d'analyser cette dernière, on extrait les données par le biais de la bibliothèque python S4Gpy. Après un traitement est une extraction sous format JSON il est possible d'extraire des statistiques relatives à la mise en avant du contenu de la fonctionnalité Direct (Voir Annexe 2).

Diffusion dans les deux parties du Prime Time

			NA	EU	Autre
Semaine 1	Prime Time	Nombre de Contenu	5	23	5
		Fréquence	15,15 %	69,70 %	15,15 %
	Non Prime Time	Nombre de Contenu	26	70	17
		Fréquence	23,01 %	61,95 %	15,04 %
Semaine 2	Prime Time	Nombre de Contenu	4	17	3
		Fréquence	16,17 %	70,83 %	12,50 %
	Non Prime Time	Nombre de Contenu	26	31	18
		Fréquence	34,67 %	41,33 %	24,00 %
Semaine 3	Prime Time	Nombre de Contenu	6	13	12
		Fréquence	19,35 %	41,94 %	38,71 %
	Non Prime Time	Nombre de Contenu	22	66	20
		Fréquence	20,37 %	61,11 %	18,52 %
Semaine 4	Prime Time	Nombre de Contenu	3	18	7
		Fréquence	10,71 %	64,29 %	25,00 %
	Non Prime Time	Nombre de Contenu	22	40	28
		Fréquence	24,44 %	44,44 %	31,11 %
Semaine 5	Prime Time	Nombre de Contenu	2	12	3
		Fréquence	11,76 %	70,59 %	17,65 %
	Non Prime Time	Nombre de Contenu	11	35	26
		Fréquence	15,28 %	48,61 %	36,11 %
Semaine 6	Prime Time	Nombre de Contenu	8	0	9
		Fréquence	47,06 %	00,00 %	52,94 %
	Non Prime Time	Nombre de Contenu	21	13	15
		Fréquence	42,86 %	26,53 %	30,61 %
Semaine 7	Prime Time	Nombre de Contenu	4	1	5
		Fréquence	40,00 %	10,00 %	50,00 %
	Non Prime Time	Nombre de Contenu	9	11	5
		Fréquence	36,00 %	44,00 %	20,00 %

Diffusion dans la première partie du Prime Time

			NA	EU	Autre
Semaine 1	Prime Time	Nombre de Contenu	1	2	5
		Fréquence	12,50 %	62,50 %	25,00 %
	Non Prime Time	Nombre de Contenu	30	88	20
		Fréquence	21,74 %	63,77 %	14,49 %
Semaine 2	Prime Time	Nombre de Contenu	2	6	0
		Fréquence	25,00 %	75,00 %	00,00 %
	Non Prime Time	Nombre de Contenu	28	42	21
		Fréquence	30,77 %	46,15 %	23,08 %
Semaine 3	Prime Time	Nombre de Contenu	0	4	5
		Fréquence	00,00 %	44,44 %	55,56 %
	Non Prime Time	Nombre de Contenu	28	75	27
		Fréquence	21,54 %	57,69 %	20,77 %
Semaine 4	Prime Time	Nombre de Contenu	3	4	3
		Fréquence	30,00 %	40,00 %	30,00 %
	Non Prime Time	Nombre de Contenu	22	55	31
		Fréquence	20,37 %	50,93 %	28,70 %
Semaine 5	Prime Time	Nombre de Contenu	1	3	0
		Fréquence	25,00 %	75,00 %	00,00 %
	Non Prime Time	Nombre de Contenu	12	44	29
		Fréquence	14,12 %	51,76 %	34,12 %
Semaine 6	Prime Time	Nombre de Contenu	0	0	5
		Fréquence	00,00 %	00,00 %	100,00 %
	Non Prime Time	Nombre de Contenu	29	13	19
		Fréquence	47,54 %	21,31 %	31,15 %
Semaine 7	Prime Time	Nombre de Contenu	0	0	4
		Fréquence	00,00 %	00,00 %	100,00 %
	Non Prime Time	Nombre de Contenu	12	13	6
		Fréquence	41,94 %	38,71 %	19,35 %

Comme nous pouvons le voir avec les deux tableaux, le contenu européen occupe une place majeure lors des heures de prime time et non prime time durant de nombreuses semaines.

Dans le cas de l'observation sur les deux périodes de Prime Time on peut voir que les semaines 1 à 5 sont des semaines où le contenu européen occupe avec une moyenne de 63,47 % l'espace prime time de 19H00 à 1H00 du matin.

Résumé des fréquences de présence des contenus (Deux parties du Prime Time)			
	EU	NA	Autre
Période Prime Time	46,76%	22,96%	30,28%
Période non Prime Time	46,85%	28,09%	25,06%

Par la suite, lorsqu'on se concentre sur la première partie du Prime Time (début de diffusion entre 21H00 et 22H00) nous pouvons observer une légère diminution du contenu européen sur cette période dans le tableau récapitulatif ci-dessous. En regardant le tableau semaine par semaine, nous pouvons voir que les chiffres sont similaires dans les deux fourchettes horaires que nous avons décidé d'analyser. Cette diminution peut s'expliquer par la semaine 6 et 7, où le contenu européen ne fut présent dans l'intervalle de temps étudié.

Résumé des fréquences de présence des contenus (Première partie du Prime Time)			
	EU	NA	Autre
Période Prime Time	42,42%	12,21%	44,37%
Période non Prime Time	47,19%	28,29%	24,52%

Pour conclure, l'analyse de plusieurs semaines de diffusion a permis de comparer la mise en avant des contenus diffusés en fonction de leur origine. Avec les données disponibles, nous avons pu voir que peu importe la fourchette horaire étudiée, le contenu européen occupe une majeure partie de la programmation linéaire telle qu'elle a été programmée par Netflix.

Toutefois, il est important de noter le fait que le robot de scraping du Direct est encore un robot jeune en cours de développement. L'extraction du programme de diffusion peut-être par moment incomplet et ne représente donc pas la programmation définitive telle qu'elle a été programmée par Netflixⁱ.

ⁱ Les données de la semaine dite « 6 » ne furent récupérées uniquement pour 3 jours de la semaine et non 7 jours.

2. Analyse de la page d'accueil

Après avoir analysé l'une des dernières fonctionnalités ajoutées à Netflix : Direct, concentrons-nous sur la page d'accueil du service de streaming américain. Tout au long du fonctionnement l'extension Prime Space sur le navigateur Google Chrome des différents utilisateurs, des logs informatiques sont enregistrés.

Lolomo, Thumbnail, Watch

Avec un premier log dit « non traité » (tel qu'il a été enregistré) on retrouve 26 activités et 2 083 cas différents. Avant d'observer les Figure 14 et 14 obtenues avec le Disco Miner, voyons dans un premier temps les différentes activités que l'on va retrouver. Tout d'abord, il y a l'activité « start_session » qui marque l'arrivée de l'utilisateur sur Netflix. Ensuite, on retrouve un ensemble d'activités qui commence par « lolomo ». Il s'agit là d'un acronyme qui signifie liste de liste de film (List of List of Movies), qui est tout simplement une catégorie de film ou série sur une ligne de contenu (voir Figure 13).

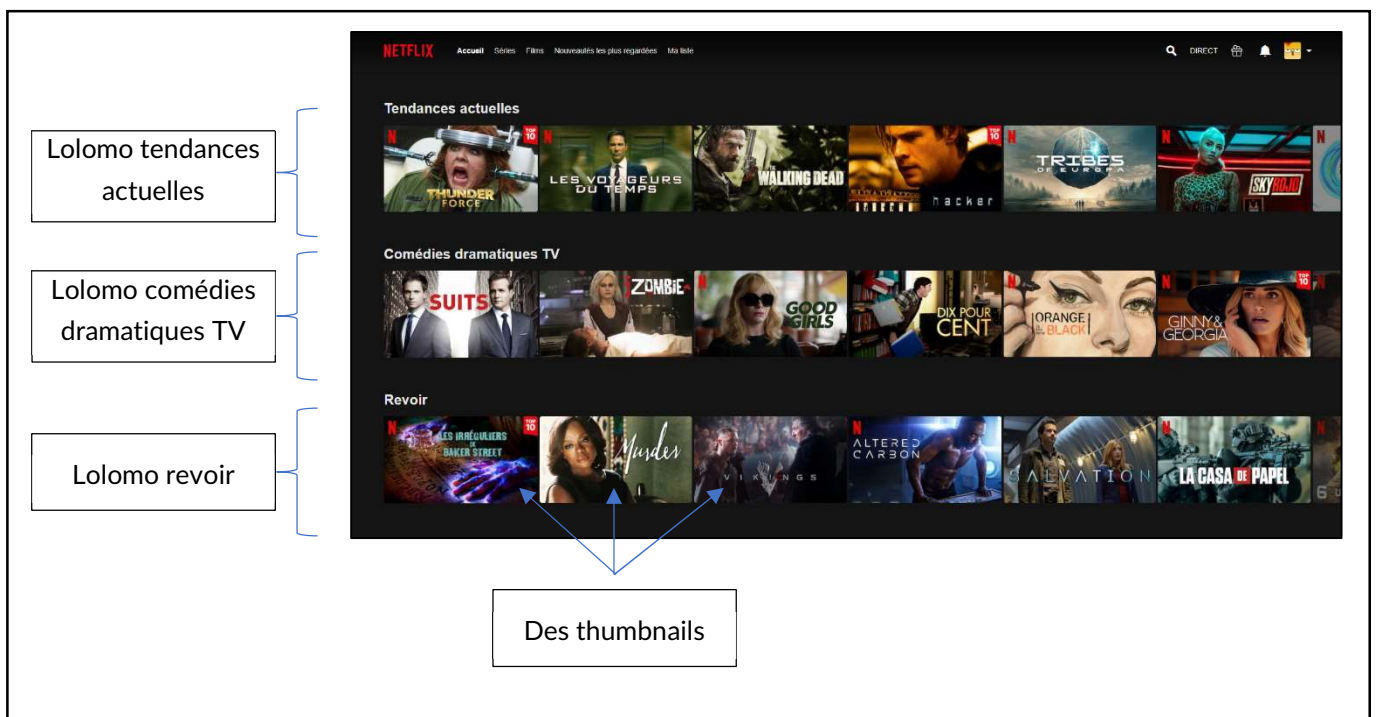


Figure 13 Netflix, les Lolomos et thumbnails

Après les activités lolomo, on retrouve les thumbnails qui sont les vignettes qui présentent le contenu. Enfin pour finir, avant la fin du processus, on observe la présence des activités « watch ». Au nombre de trois, elles représentent l'action de l'utilisateur qui va regarder un contenu Européen, Américain ou autres. L'étude de ces activités permet de se faire une idée sur le taux de transformation des contenus en fonction de leur provenance.

Etant donné que la trace est importante, il existe un grand nombre de chemin (path) entre les différentes activités. Afin donc de rendre le résultat lisible on a fixé le nombre d'activité à 100% et le nombre de chemin à 0%. Le traitement avec le Disco Miner procède alors à des abstractions et/ou agrégations pour afficher uniquement les chemins les plus présents dans l'ensemble de la trace. Ensuite, si on réduit le nombre d'activité à 10% afin de se concentrer sur les éléments les plus fréquents on obtient la « map » (représentation) de la Figure 155.

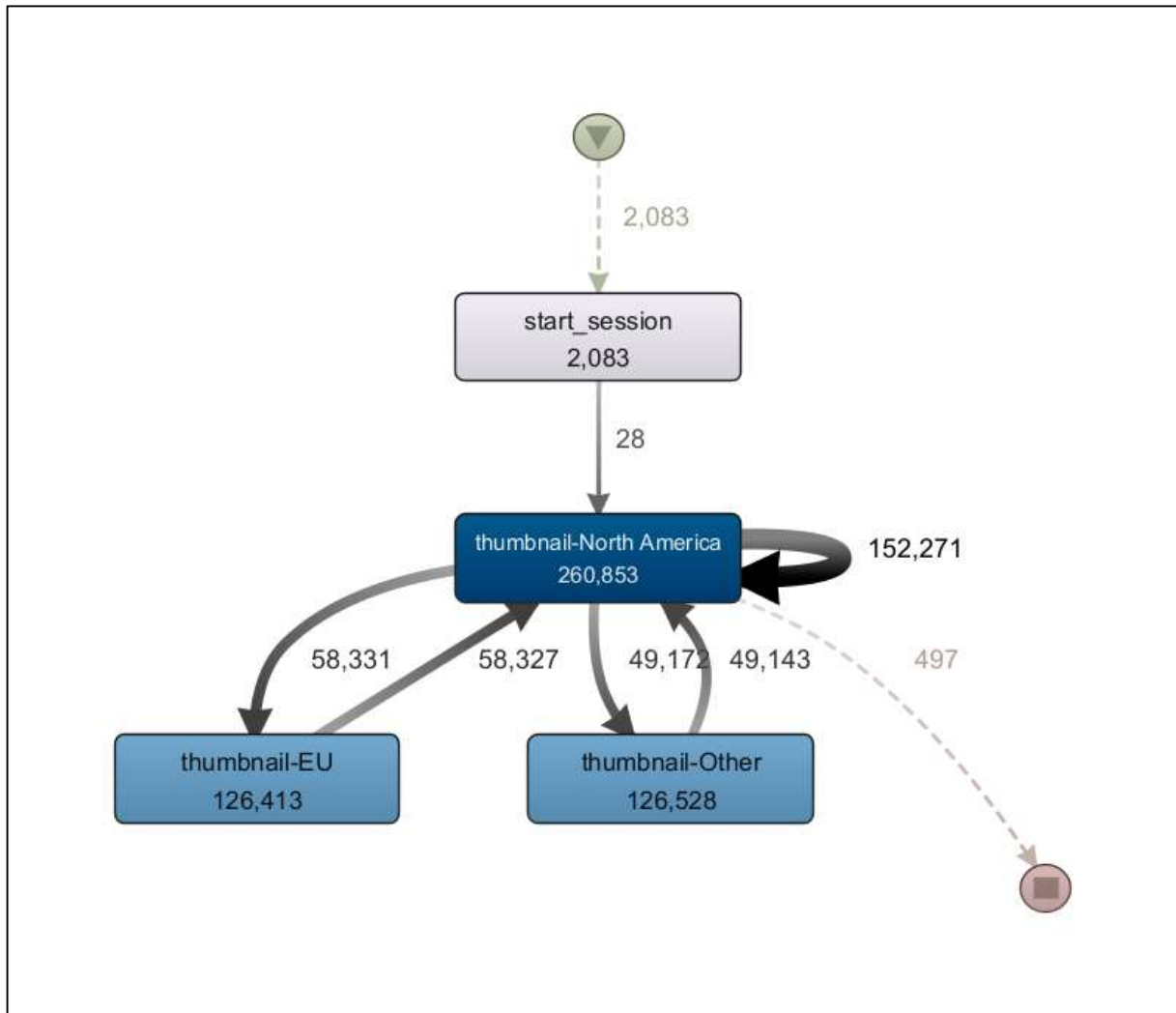


Figure 15 Trace Netflix : Lolomo, Thumbnail, Watch - Fréq. Absolue -10% - Disco Miner

Avec cette première version qui met en valeur la fréquence absolue des évènements on remarque que sur les 2 083 cas, les vignettes des contenus Américains ont une fréquence absolue de 260 853 (50,77%), les vignettes « autres » ont une fréquence absolue de 126 528 (24,63%) et les Vignettes EU, une fréquence absolue de 126 413 (24,60%).

Focus par lolomo

Après cette vision globale obtenue avec une trace « brute » qui nous montre l'ensemble des lolomos, les thumbnails de l'ensemble des sessions et la conversion (le fait de regarder un contenu) penchons-nous sur une analyse niveau par niveau (lolomo par lolomo). Comme l'a souligné l'étude présentée dans la partie Découvrabilité du contenu (1.4) les deux tiers de l'espace sur un écran de la page Netflix et réservé à promouvoir son propre contenu avant de passer à des éléments issus des algorithmes de recommandation. Voyons donc la composition en thumbnails des premiers lolomo afin d'observer la mise en avant ou non du contenu européen.

Afin de faire une telle analyse, nous avons pris une trace (50 activités et 2 357 cas différents) qui contient des activités : lolomo(s), avec un numéro (ce qui permet d'identifier la position sur la page), ainsi que des activités thumbnails (NA/EU/Other) avec des attributs (row/col - ligne/colonne) permettant d'identifier l'appartenance ou non de la vignette au lolomo. Cet attribut donne aussi la possibilité d'appliquer un filtre précis sur la position des éléments que l'on veut analyser.

Lolomo 0

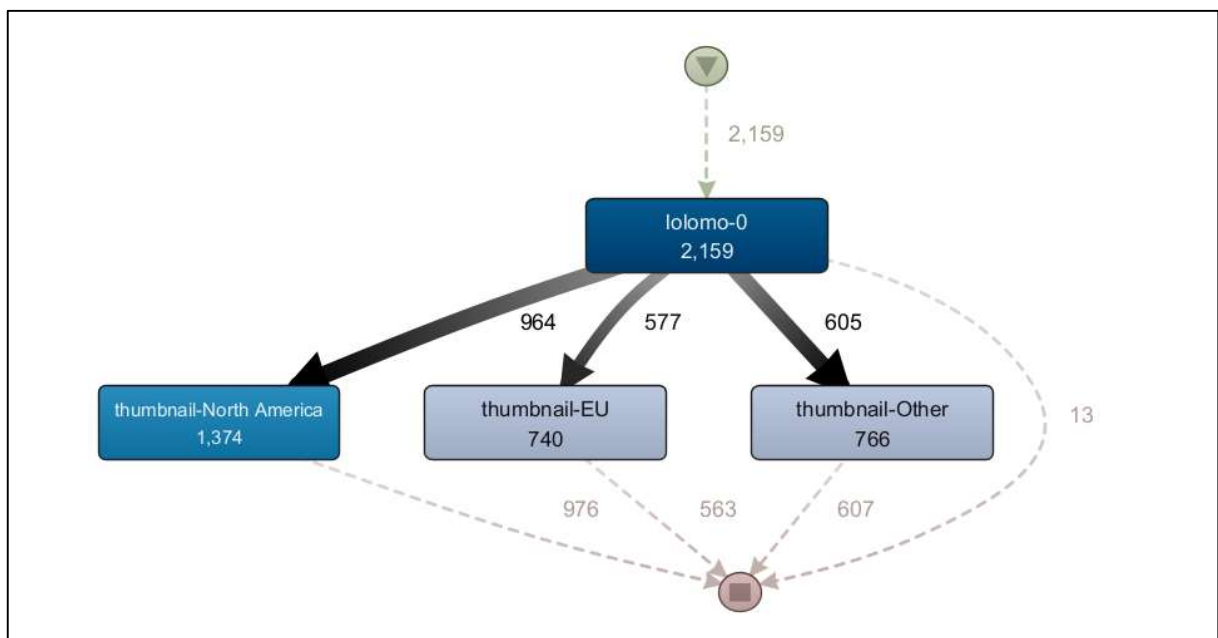


Figure 16 Lolomo 0 Thumbnails NA/EU/Other

Lolomo 1

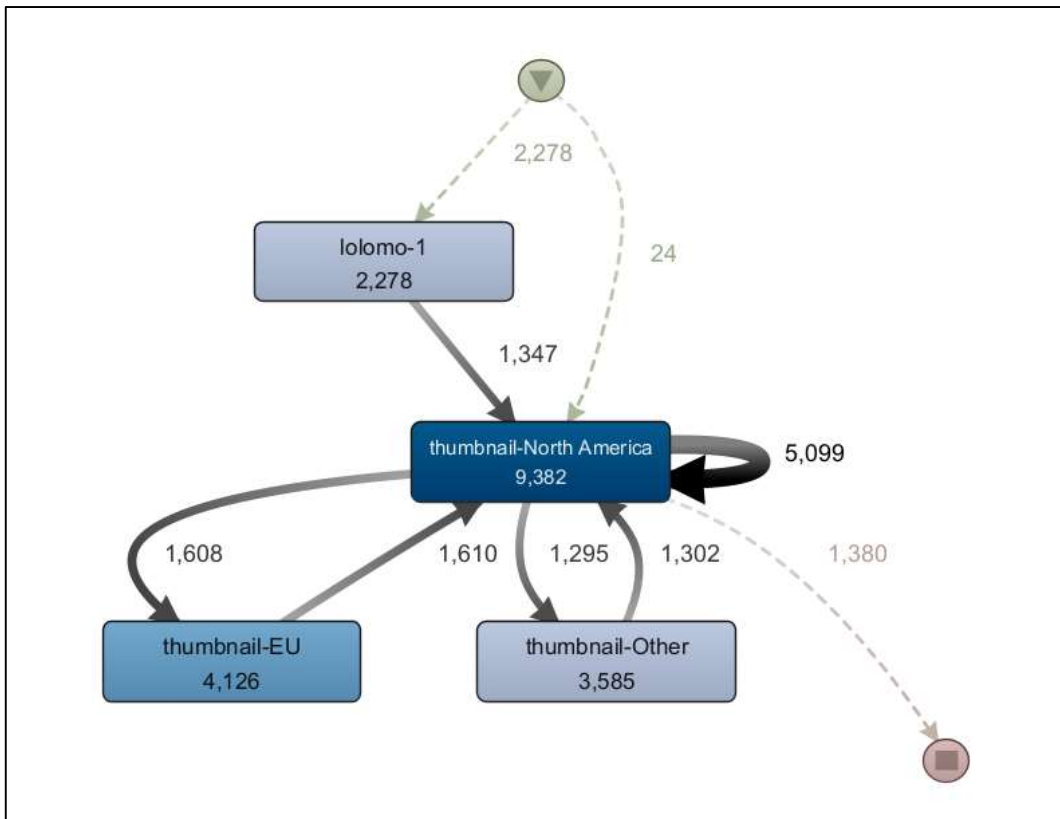


Figure 17 Lolomo 1 Thumbnails NA/EU/Other

Lolomo 2

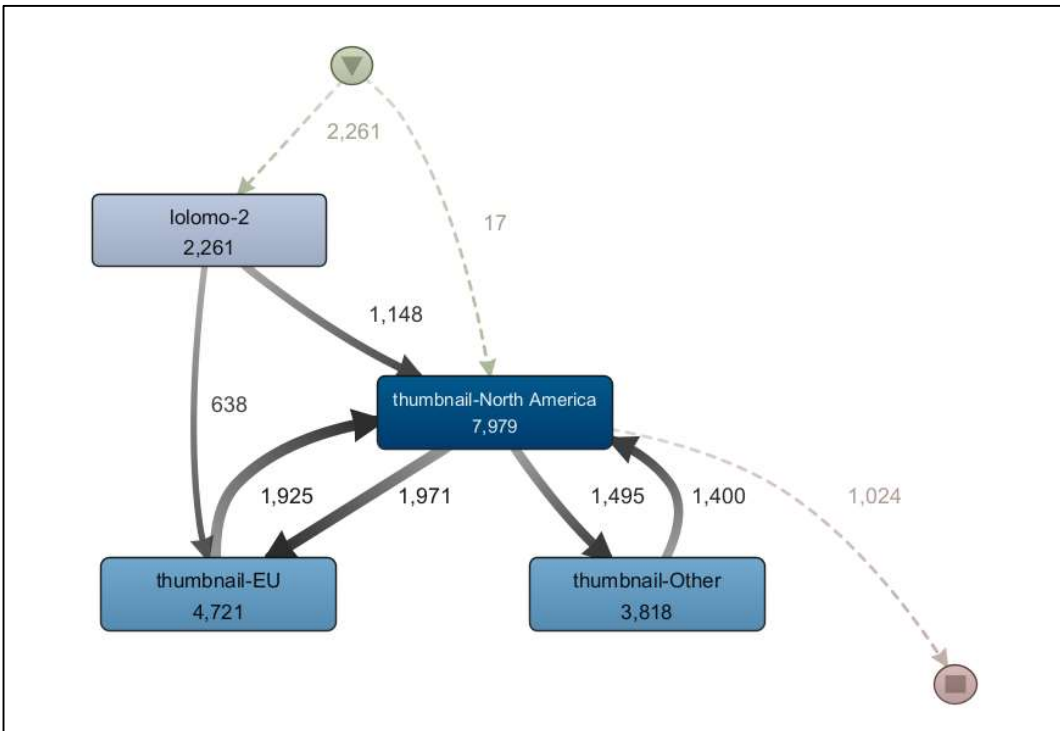


Figure 18 Lolomo 2 Thumbnails NA/EU/Other

Lolomo 3

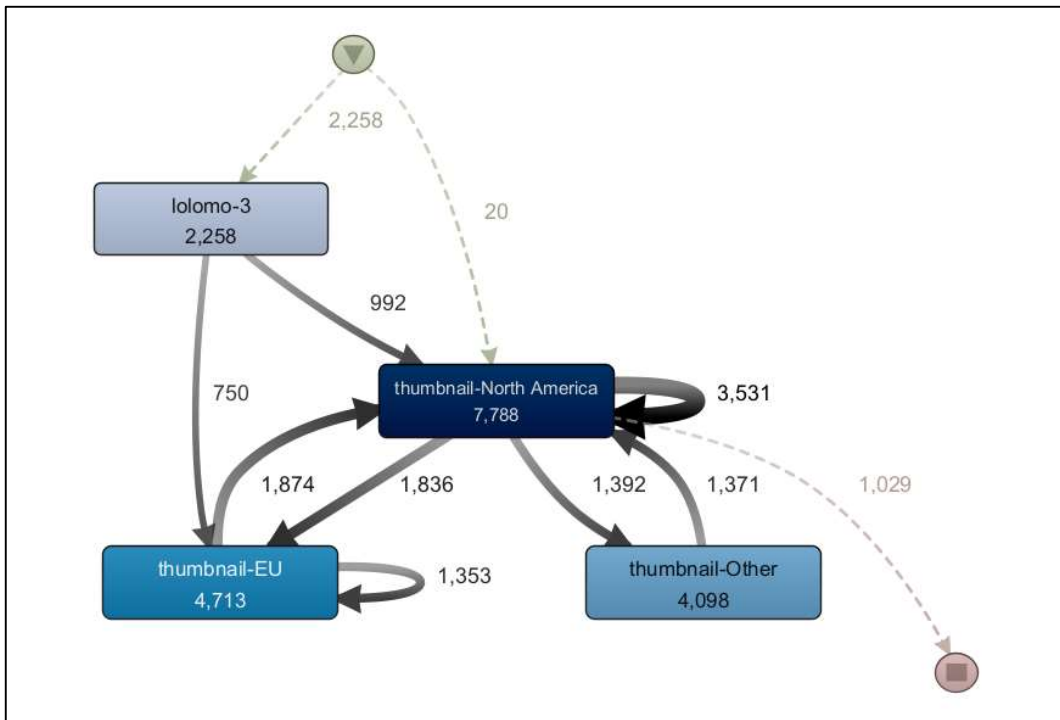


Figure 19 Lolomo 3 Thumbnails NA/EU/Other

Lolomo 4

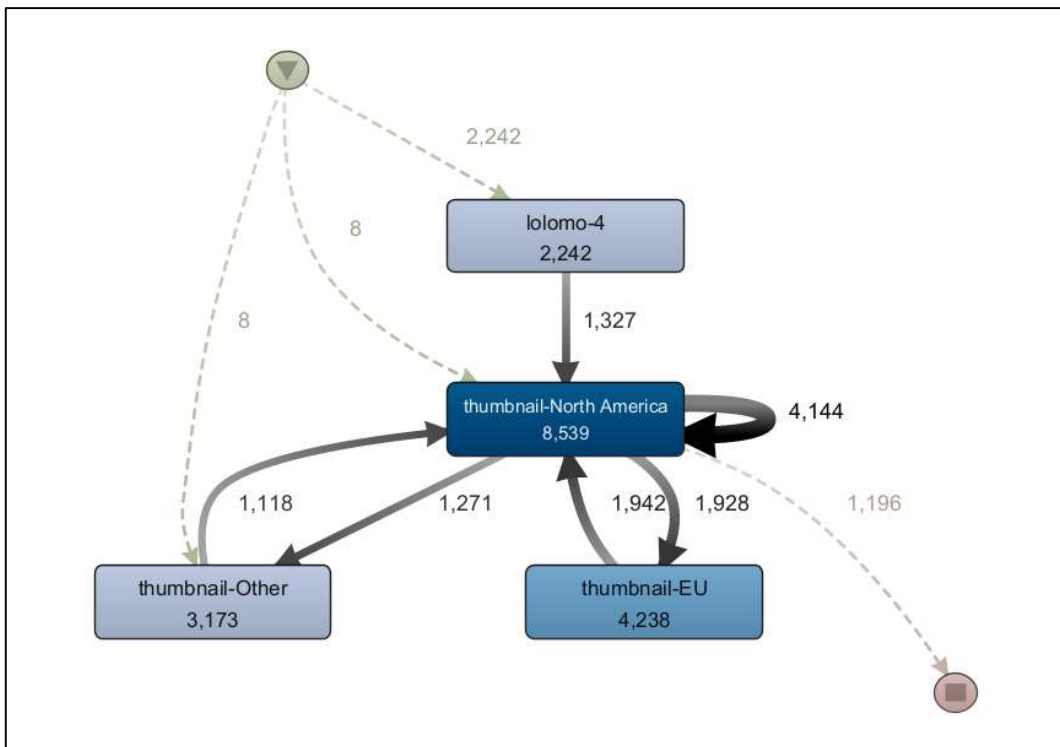


Figure 20 Lolomo 4 Thumbnails NA/EU/Other

Lolomo 5

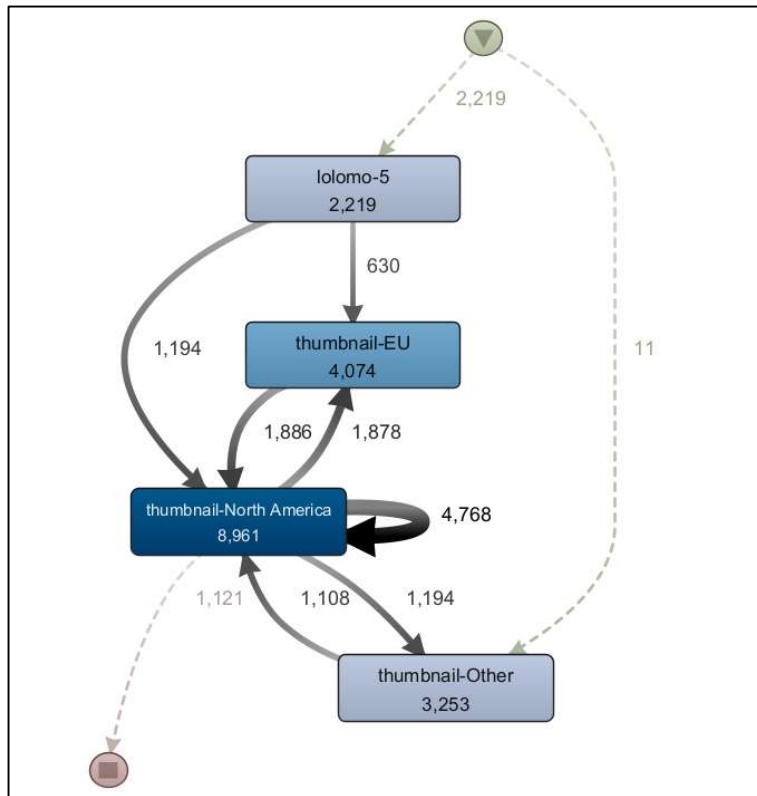


Figure 21 Lolomo 5 Thumbnails NA/EU/Other

Lolomo 6

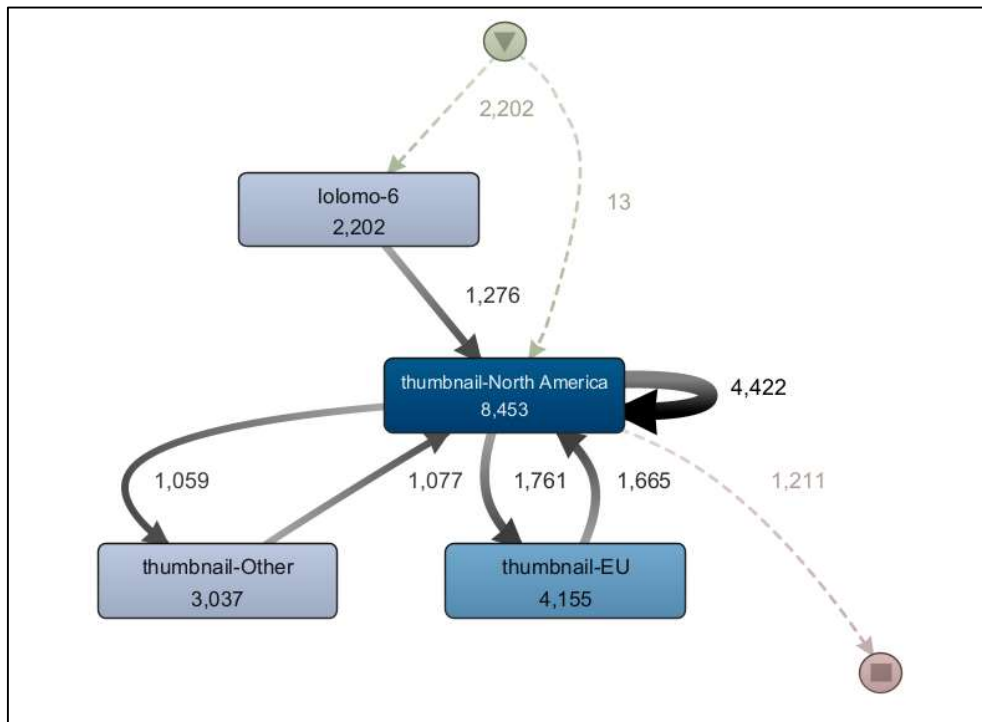


Figure 22 Lolomo 6 Thumbnails NA/EU/Other

Figure et nom	Informations	Fréquence relative
Figure 16 : Lolomo 0	1 374 : Vignettes NA 766 : Vignettes autres 740 : Vignettes EU	47,71% : Vignettes NA 26,60% : Vignettes autres 25,69% : Vignette EU
Figure 17 : Lolomo 1	9 382 : Vignettes NA 4 126 : Vignettes EU 3 585 : Vignettes autres	54,89% : Vignettes NA 24,14% : Vignettes EU 20,97% : Vignettes autres
Figure 18 : Lolomo 2	7 979 : Vignettes NA 4 721 : Vignettes EU 3 818 : Vignettes autres	48,3% : Vignettes NA 28,58% : Vignette EU 23,11% : Vignettes autres
Figure 19 : Lolomo 3	7 788 : Vignettes NA 4 713 : Vignettes EU 4 098 : Vignettes autres	46,92% : Vignettes NA 28,39% : Vignette EU 24,69% : Vignettes autres
Figure 20 : Lolomo 4	8 539 : Vignettes NA 4 238 : Vignettes EU 3 173 : Vignettes autres	53,54% : Vignettes NA 26,57% : Vignette EU 19,89% : Vignettes autres
Figure 21 : Lolomo 5	8 961 : Vignettes NA 4 074 : Vignettes EU 3 253 : Vignettes autres	55,02% : Vignettes NA 25,01% : Vignette EU 19,97% : Vignettes autres
Figure 22 : Lolomo 6	8 453 : Vignettes NA 4 155 : Vignettes EU 3 037 : Vignettes autres	54,03% : Vignettes NA 26,56% : Vignette EU 19,41% : Vignettes autres

	Fréquence des vignettes par origine du contenu sur les 7 premiers lolomos		
	NA	EU	Autres
Moyenne	51,49 %	26,55 %	21,96 %
Médiane	53,54 %	26,57 %	20,97 %

Avec cette trace passée à travers le Disco Miner, on peut se faire idée sur la composition en contenu niveau par niveau de chaque lolomo grâce à la fréquence moyenne et la médiane des fréquences relatives observées. Nous pouvons voir une présence importante du contenu d'origine NA (51,49%) au détriment du contenu EU (26,55%) et autre (21,96%).

Figure et nom	Fréquence relative
Figure 14 et Figure 15 : Lolomo, Thumbnail, Watch – Fréq. Absolue	50,77% : Vignettes NA 24,63% : Vignette autre 24,60% : Vignette EU
Figure 16 à Figure 22 : Moyenne 7 Lolomo	51,49% : Vignettes NA 26,55% : Vignettes EU 21,96% : Vignettes autres

Avec 2 357 cas analysés par le Disco Miner afin de d'obtenir la composition des lolomos, la mise en avant du contenu européen n'a pas au même niveau que dans le cas de la fonction Direct, fraîchement instaurée par Netflix. Ce même constat, on peut le faire avec le résultat général obtenu auparavant avec la trace dite « non traité » et ses 2 083 cas.

On constate qu'au fur et à mesure des sessions, le lolomo 0 qui est la première chose que l'on voit lorsqu'on se connecte sur Netflix propose différentes vignettes où le contenu européen à une fréquence de 25,69 % sur l'ensemble des sessions enregistrés. Ensuite on retrouve le lolomo 1 et 2 ou le contenu qui nous intéresse, fut observé avec une fréquence relative de 24,14% et 28,58% respectivement.

Concentrons-nous sur le lolomo 1, qui présente la fréquence relative du contenu EU la plus faible (sur les 7 premiers lolomo que voit l'utilisateur). Avec Disco, il est possible d'appliquer un certain nombre de filtres afin d'affiner la recherche et extraire des éléments. Si l'on rend « mandatory » la présence d'un thumbnail EU dans le lolomo 1, Disco nous indique que sur les 2 357 cas, 1 566 cas ont obligatoirement au moins une vignette EU au sein du lolomo 1, ce qui représente 66 % des cas (Annexe 1). De plus, lorsque celle-ci est mandatory, cette vignette obtient alors une fréquence relative de 31,74 % (NA = 46,00%, Autre = 22,26%). En revanche, cela veut dire que 791 cas et donc sessions utilisateurs (par la même occasion) n'ont pas eu de thumbnails EU dans le lolomo 1.

Fréquence des vignettes par origine du contenu sur les 7 premiers lolomos - Vignette EU « Mandatory »				Pourcentage des cas
	NA	EU	Autres	
Lolomo 0	28,49 %	59,49 %	12,3 %	25%
Lolomo 1	46,00 %	31,74 %	22,26 %	66%
Lolomo 2	45,78 %	32,82 %	21,39 %	80%
Lolomo 3	45,35 %	33,38 %	21,27 %	79%
Lolomo 4	50,33 %	32,46 %	17,22 %	74%
Lolomo 5	53,33 %	30,98 %	15,68 %	72%
Lolomo 6	49,11 %	35,66 %	15,23 %	66%

Le fait de rendre obligatoire la présence d'une vignette EU change de manière significative la présence du contenu EU au sein des différents lolomos. L'application du filtre exclus en effet toutes les traces qui n'avaient pas de thumbnails EU au sein des 7 premiers lolomos.

Avec le tableau ci-dessus on observe alors une mise en avant du contenu EU supérieure à 30% est donc en accord avec la Directive Européenne. Toutefois il faut nuancer le fait que ce cas de figure ne représente que 66 % (73% si on exclut le lolomo 0 qui est un lolomo d'accueil) des sessions enregistrées par le projet Streaming for Good.

IV. Conclusion

Ainsi comme nous avons pu le voir grâce aux différentes statistiques obtenus avec les traces du projet Streaming for Good et les données enregistrées lors des différentes sessions de visionnage d'utilisateur ou de robot, la fouille de processus et la data analyse permet d'extraire différentes données à des fins d'analyse. Bien que les données du projet ne sont pas représentatives de l'entièreté du catalogue, il est question ici d'étudier la mise en avant du contenu européen au sein d'un environnement dynamique, évolutif, régi par des algorithmes. Les différentes informations que le projet S4G collecte sont des éléments que l'utilisateur ou un robot contrôlé va pouvoir observer durant sa visite de la plateforme de streaming. Alors le contenu enregistré par le projet reprend ce qui est mis en avant par le service à différents endroits et différents niveaux.

L'étude de la programmation du Direct proposée par Netflix (qui retire le choix de l'utilisateur) montre que le service répond à la Directive Européenne en mettant en avant bien plus que 30 % de contenu européen. Outre l'étude de cette fonctionnalité nouvelle, nous nous sommes intéressés à la partie la plus importante et la plus visible de Netflix, la page d'accueil. Afin de savoir si la règle des 30 % est respectée nous avons eu recours à la fouille de processus. Bien que les méthodes du domaine ont été conçues pour analyser les comportements d'un humain qui génère un log informatique, les outils que propose la discipline peuvent être utilisés afin d'étudier les traces issues de l'affichage de la page d'accueil. Avec un nombre considérable d'information à traiter, le logiciel Disco a permis de mettre en évidence des statistiques en fonctions de différents critères. Comme évoqué précédemment, la mise en avant du contenu consiste à mettre en vue, faire connaître quelque chose. C'est pourquoi, après l'étude de l'ensemble du contenu enregistré par le projet S4G (P.31 - Lolomo, Thumbnail, Watch), nous nous sommes penchées sur l'étude des 7 premières lignes de la page d'accueil et les contenus qui les composent. Dans un premier temps, nous avons pu voir que la règle des 30 % n'est pas respectée. Mais lorsque nous avons exclu (avec un filtre) les traces qui ne comportaient pas un contenu EU par lolomo (on garde les traces où un lolomo doit obligatoirement avoir au minimum une vignette EU), nous nous sommes retrouvés avec une page d'accueil où les 7 premiers lolomos respectaient la mise en avant à hauteur minimum de 30% dans 66% des cas observés.

Enfin, la trace (Lolomo, Thumbnail, Watch) étudiée qui regroupe les différents lolomos présents lors des nombreuses sessions des utilisateurs nous montre qu'il n'y a pas de lolomo dédiée au contenu européen. L'ajout d'une telle ligne dans une position haute de l'écran permettrait de mettre en avant cette catégorie et son contenu aux yeux de l'utilisateur.

V. Limites

Bien que l'étude ait pu mettre en avant des résultats, ils sont, comme indiqué précédemment, non représentatifs de l'ensemble de contenu que propose Netflix. En plus de cela, les différents robots qui alimentent la base de données du projet en informations supplémentaires sont encore en cours de développement, perfectionnement et automatisation. Lorsque ceux-ci seront pleinement opérationnels il sera possible de préciser et affiner les différentes études du projet S4G.

VI. Recherches futures

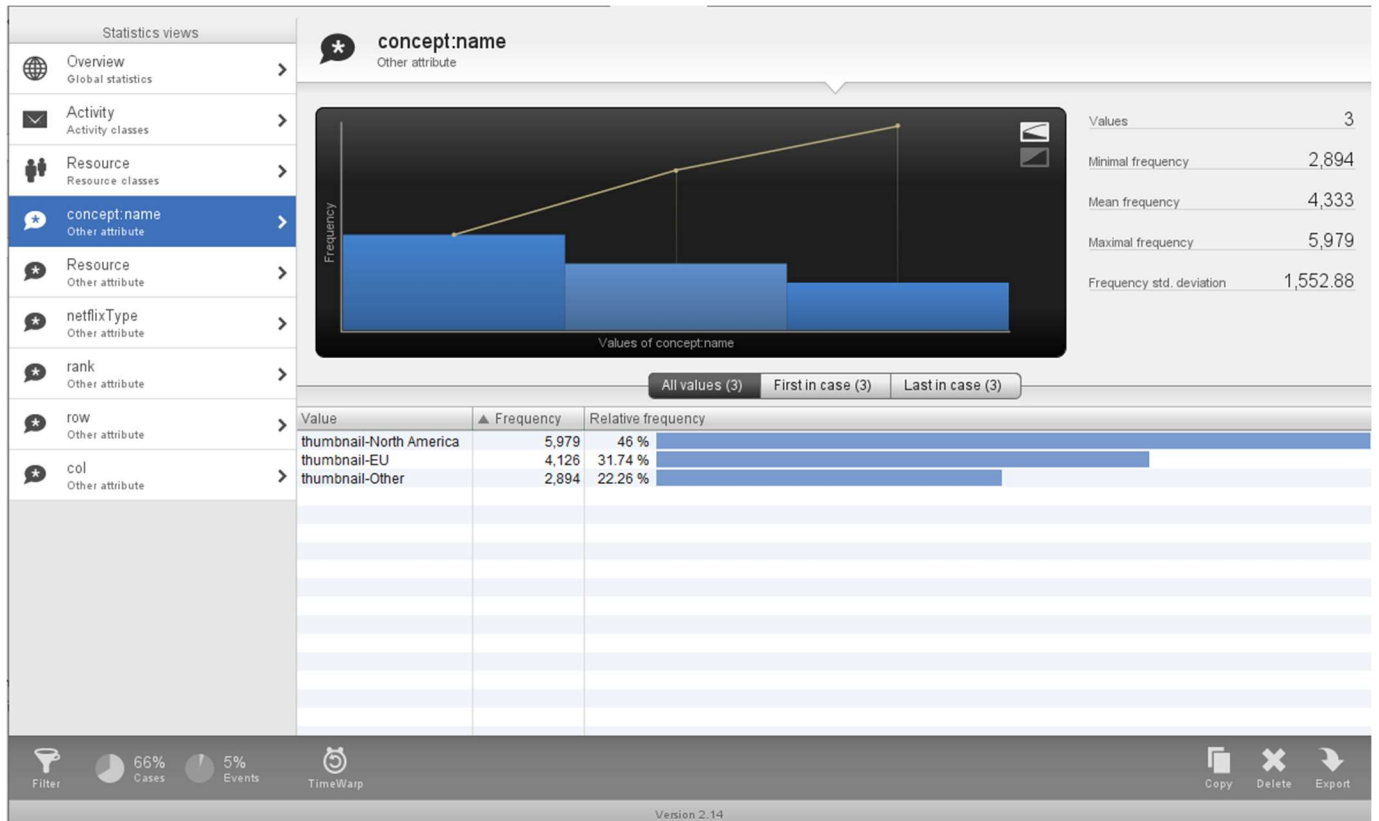
Le projet Streaming for Good n'en est qu'à ses débuts. Avec l'implication d'étudiants de différents niveaux qui apportent chacun leur expertise, leur point de vue et leurs idées, de nombreuses questions de recherches sont possibles. Alimentée de jours en jours avec de nouveaux éléments et de nouveaux outils, le projet permettra dans le futur de décortiquer de manière plus précise et plus poussée différents aspects de plateformes numériques. Comme nous avons pu le voir, le contenu qui compose la base de données est en constante évolution. L'équipe de développement est à l'affût des nouveautés ainsi que des demandes formulées par les membres du projet afin de répondre à de nouvelles questions d'études et de recherches. L'étude de la conformité à la Directive Européenne est elle aussi qu'à ses débuts. Avec ce mémoire qui tente de proposer une étude statistique à l'aide d'outils pour le process mining et de la data analyse, il reste d'autres aspects à analyser et mettre en valeur. Il est en effet possible de s'interroger sur la dimension juridique, le taux de transformation d'un contenu mis en avant ou encore l'attractivité des vignettes en fonction de leur origine.

VII. Table des illustrations

Figure 1 « Les cercles de la réussite » de la création Original Netflix House of Cards – Source : The New York Times	8
Figure 2 Pays où Netflix est disponible – Source Netflix.com	10
Figure 3 Streaming 4 Good schéma.....	15
Figure 4 - Schéma architecture technique Stream4Good avec Humains	17
Figure 5 - Schéma architecture technique Stream4Good avec Robots	18
Figure 6 Dataviz Dashboard VOD Prime Space.....	19
Figure 7 API Direct - Stream for Good	20
Figure 8 Réseau de Petri - Algorithme Alpha - ProM	23
Figure 9 Réseau de Petri - Algorithme Heuristic Miner - ProM.....	23
Figure 10 Fuzzy Model - Algorithme Fuzzy Miner - ProM	24
Figure 11 Disco Miner Model - Algorithme Disco Miner - Disco.....	25
Figure 12 En Direct - Netflix - Source : about.netflix.com.....	26
Figure 13 Netflix, les Lolomos et thumbnails	31
Figure 14 Trace Netflix : Lolomo, Thumbnail, Watch – Fréq. Absolue - Disco Miner	32
Figure 15 Trace Netflix : Lolomo, Thumbnail, Watch – Fréq. Absolue -10% - Disco Miner....	33
Figure 16 Lolomo 0 Thumbnails NA/EU/Other.....	34
Figure 17 Lolomo 1 Thumbnails NA/EU/Other.....	35
Figure 18 Lolomo 2 Thumbnails NA/EU/Other.....	35
Figure 19 Lolomo 3 Thumbnails NA/EU/Other.....	36
Figure 20 Lolomo 4 Thumbnails NA/EU/Other.....	36
Figure 21 Lolomo 5 Thumbnails NA/EU/Other.....	37
Figure 22 Lolomo 6 Thumbnails NA/EU/Other.....	37

VIII. Annexes

Annexe 1 : Application d'un filtre vignette EU « mandatory » au lolomo 1



IX. Références

¹ ACREP : Accueil > L'Arcep : <https://www.arcep.fr/larcep.html>

² Qu'est-ce-que le CSA : <https://www.csa.fr/Informer/Qu-est-ce-que-le-CSA>

³ Confinement : Netflix limite ses débits en Europe pour désengorger internet - Raphaël Grably – BFMTV - 19/03/2020 - https://www.bfmtv.com/tech/vie-numerique/confinement-netflix-limite-ses-debits-en-europe-pour-desengorger-internet_AN-202003190102.html

⁴ Netflix, champion du trafic internet français devant google - Paul Louis avec AFP - 27/06/2019 - https://www.bfmtv.com/tech/netflix-champion-du-traffic-internet-francais-devant-google_AV-201906270104.html

⁵ Création audiovisuelle : la "concurrence déloyale" des géants du Net - Source AFP - 30/09/2017 - https://www.lepoint.fr/high-tech-internet/creation-audiovisuelle-la-concurrence-deloyale-des-geants-du-net-30-09-2017-2160988_47.php

⁶ Directive (UE) 2018/1808 du Parlement Européen et du conseil - Document 32018L1808 – 14/11/2018

⁷ Bilan aux actionnaires Netflix du 19/01/21 - https://s22.q4cdn.com/959853165/files/doc_financials/2020/q4/FINAL-Q420-Shareholder-Letter.pdf

⁸ Avec Netflix, la « télévision du futur » arrive en France - Stéphane Lauer – 13/09/2021 – Le Monde - https://www.lemonde.fr/economie/article/2014/09/13/netflix-que-le-carnage-commence_4487006_3234.html

⁹ Srivatsa Maddodi, & Krishna Prasad, K. (2019). Netflix Bigdata Analytics- The Emergence of Data Driven Recommendation. International Journal of Case Studies in Business, IT, and Education (IJCSBE), 3(2), 41-51. DOI: [org/10.5281/zenodo.3510316](https://doi.org/10.5281/zenodo.3510316)

¹⁰ Fonctionnement du système de recommandations de Netflix – Netflix.com - Consulté le 24/06/2021 - <https://help.netflix.com/fr/node/100639>

¹¹ Leçon de big data par Netflix – Pubosphere.fr - CULTURALLY- 27/08/2021

https://pubosphere.fr/lecon_de_big_data_par_netflix/

¹² Les recettes de Netflix - Thibault Henneton - Le Monde diplomatique - Février 2019 -

<https://www.monde-diplomatique.fr/2019/02/HENNETON/59557>

¹³ Direct : une nouvelle fonctionnalité testée en France – Netflix.com - Emily Grewal (Product

Manager) – 05/11/2020 - <https://about.netflix.com/fr/news/direct-new-feature-tested-in-france>

¹⁴ Intel reiterates chip supply shortages could last several years- Ben Blanchard – 31/05/2021

- <https://www.reuters.com/technology/intel-reiterates-chip-supply-shortages-could-last-several-years-2021-05-31/>

¹⁵ Netflix anticipe une croissance limitée de ses abonnés – LesEchos.com - Marina Alcaraz –

21.07.2021 - <https://www.lesechos.fr/tech-medias/medias/netflix-parie-sur-les-jeux-video-pour-conserver-lattention-des-consommateurs-1333606#xtor=CS1-26>

¹⁶ Quatre ans de Netflix en France : Quelles conséquences ? - Pierre Dezeraud - 18/09/2018

- <https://www.ozap.com/actu/quatre-ans-de-netflix-en-france-quelles-consequences/566808>

¹⁷ Bruno delecour (filmotv) à #media : netflix fait de la concurrence déloyale – 09/03/2016 -

<https://www.cbnews.fr/medias/image-bruno-delecour-filmotv-media-netflix-fait-concurrence-deloyale-16367>

¹⁸ Lange-Médart André, « Vers une révision a minima de la directive SMA », Les Enjeux de

l'information et de la communication, 2016/2 (N° 17/2), p. 91-112. DOI : 10.3917/enic.021.0091. URL : <https://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2016-2-page-91.htm>

¹⁹ Exception culturelle européenne en vue pour Netflix et la VoD- Nicolas Madelaine –

17/05/2016 - <https://www.lesechos.fr/2016/05/exception-culturelle-europeenne-en-vue-pour-netflix-et-la-vod-229122>

²⁰ Audience Internet Global en France en juin 2021 – Médiamétrie – 29.07.2021 -

<https://www.mediametrie.fr/sites/default/files/2021-07/2021%2007%2029%20CP%20Audience%20Internet%20Global%20Juin%202021.pdf>

²¹ 'The Queen's Gambit' Chess Boom Moves Online – 16.12.2020 – Bloomberg.com - Rachael Dottle - <https://www.bloomberg.com/graphics/2020-chess-boom/>

²² Drive to thrive: Netflix's docuseries a boost for Formula 1 - Jason Abbruzzese – NBC News – 22/06/2021 - <https://www.nbcnews.com/tech/tech-news/netflix-f1-espn-boost-tv-ratings-espn-rcna1237>

²³ "Lupin" sur Netflix entraîne aussi un succès littéraire XXL - Louise Wessbecher – 15/01/2021 - https://www.huffingtonpost.fr/entry/lupin-sur-netflix-entraîne-aussi-un-succès-littéraire-xxl_fr_600170acc5b697df1a04c3e7?ncid=tweetlnkfrhpmg00000001

²⁴ Hallinan B, Striphas T. Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society*. 2016;18(1):117-137.doi:10.1177/1461444814538646

²⁵ Solsman, J. (2018, January 10). CES 2018: YouTube's AI recommendations drive 70 percent of viewing. CNET - <https://www.cnet.com/news/youtube-ces-2018-neal-mohan/>

²⁶ Television had prime time. How about prime space for on-demand services? – Grégoire Bideau – 30/09/2020 - <https://medium.com/@gbideau/television-had-prime-time-how-about-prime-space-for-on-demand-services-ee534c7cdd5d>

²⁷ Site Streaming for Good – Nicolas Herbaut - <https://stream-for-good.miage.dev/>

²⁸ Project prime space distributed scraping, Oualid GUENNIF, Grasaan GRATION, Nasser HARTI – Etudiant de L3 MIAGE Université Panthéon Sorbonne Paris 1 – 2021 - <https://stream-for-good.miage.dev/sites/stream-for-good.miage.dev/files/Distributed%20Scraping%20-%20Rendu%20Projet%20Commun%20-%202020.2021-.pdf>

²⁹ Documentation S4Gpy version 0.0.10 – Nicolas Herbaut – 16/05/2021 - <https://pypi.org/project/s4gpy/>

³⁰ Fouille de processus auto-définis : cas d'étude d'un moteur de recherche d'une bibliothèque numérique – Aout 2019 - DOI:10.3166/INFORSID..1-16 - https://www.researchgate.net/publication/335401657_Fouille_de_processus_auto-definis_cas_d%27étude_d%27un_moteur_de_recherche_d%27une_bibliotheque_numerique

³¹ Disco Tour – Fluxicon - Anne Rozinat - <https://fluxicon.com/disco/files/Disco-Tour.pdf>

³² Prime time, prime times – Définitions Larousse - https://www.larousse.fr/dictionnaires/francais/prime_time/63946