La Vulgarisation Scientifique sur YouTube

Vincent Barbosa Vaz [1]; Cécile Pov [1]; Guillaume Triqueneaux-Fleury [2]; Aïchatou Sow [2]
[1]ESIEE Paris, [2]UPEM

Introduction

Aujourd'hui, l'essor du numérique offre un nouvel accès à la culture. En effet, non seulement il popularise les modes de diffusion de l'information (vidéo, audio), mais il participe également à la vulgarisation de disciplines peu connues ou pointues : c'est particulièrement le cas des sciences, souvent considérées comme peu abordables pour un public non-spécialisé.

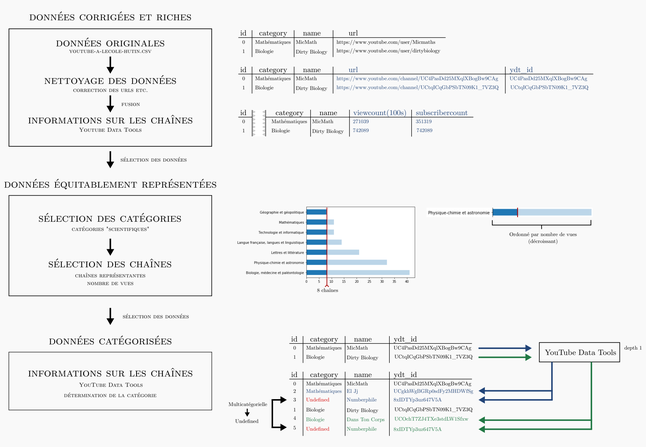
Dans cette étude, nous nous intéresserons à la vulgarisation scientifique sur Youtube, qui est à la fois un média et un réseau social. L'objectif est d'étudier, via l'analyse de réseaux, les interactions entre les différentes communautés scientifiques de vulgarisateurs, et de discerner des disciplines émergentes. Nous avons choisi comme document de référence un rapport de 2018 du ministère de la culture, qui recense 350 chaînes scientifiques et culturelles francophones classées par catégorie.

Méthode et caractérisation d'ensemble

Nous avons choisi de limiter l'étude à une analyse fixée dans le temps. Une étude temporelle nécessiterait une période d'analyse (et de récolte de données) sur une année pour observer des variations au sein des communautés. Nous sommes partis du document réalisé par Mathilde Hutin et disponible au format CSV. Nous avons procédé à un traitement des données en Python pour identifier et corriger les urls incorrectes, ajouter de l'information (à partir de YouTube Data Tools) et déterminer la catégorie des nouvelles chaînes du réseau.

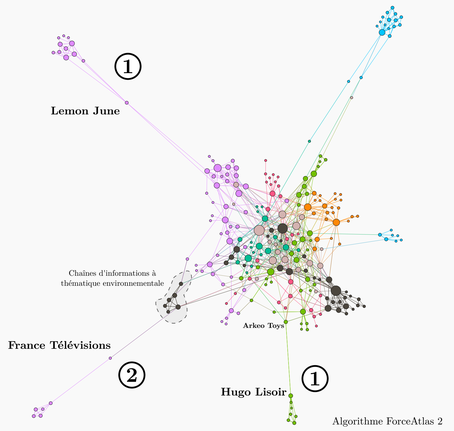
Le réseau comporte 280 sommets (des chaînes) et 893 liens. Un lien A -> B signifie “A est abonné à B”. Il y a 15 composantes connexes : une giant component (253 chaînes), 5 autres composantes (une avec 9 chaînes, et les autres avec 2 ou 3 chaînes), et 9 noeuds isolés. Le degré moyen des noeuds mesuré est 3.189, le diamètre est de 12 et la moyenne des plus courts chemins vaut 4.672, suggérant une forte connexité globale du réseau et de possibles interactions entre les catégories.

Collecte et nettoyage des données



Visualisations

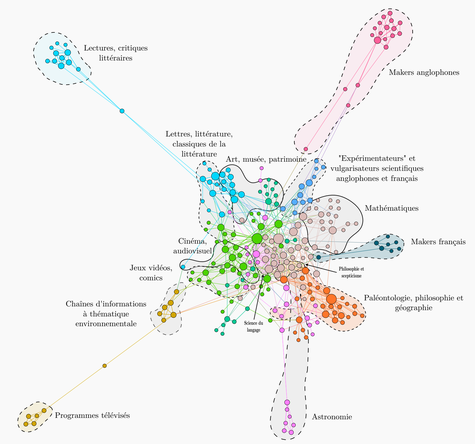
Réseau des chaînes



Lettres et littérature
Technologie et informatique
Physique-chimie et astronomie
Langue française, langues et linguistiques
Biologie, médecine et paléontologie
Géographie et géopolitique
Mathématiques
Undefined (+2 catégories)

Figure (1)

Les communautés de chaînes



Lettres, littératures et histoire des arts
“Expérimentateurs” et vulgarisateurs
Patrimoine architectural, découvertes et écriture
Makers Français
Makers anglophones
Chaînes d'actualités
Astronomie et géographie
Archéologie
Audiovisuel
Sciences naturelles et cognitives

Figure (2)

Partition du réseau : interactions entre groupes de chaînes

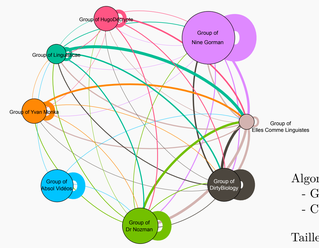


Figure (3)

Algorithmes :
- Groups by Partition
- Circular Layout

Taille des noeuds : degree

Interactions entre chaînes

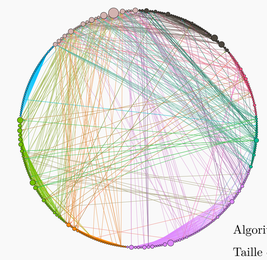


Figure (4)

Algorithmes Circular Layout
Taille des noeuds : betweenness

Analyse et Mesures

Figure (1) - Zones 1 : Communautés isolées, reliées à la Giant component par un point d'articulation (une chaîne qui fait "pont"). Zone 2 : La chaîne "France Télévisions" lie deux communautés que sont "Lettres et littérature" et "Biologie, médecine et paléontologie”.

Figure (2) - Le score de modularité (0.61) montre que le réseau a une structure de communauté significative. 26 communautés ont été détectées. Une première analyse manuelle des thématiques au sein des communautés montre la pertinence de la plupart de ces regroupements, mais également la limite de “résolution” de l'algorithme.

Figure (3) - Les arcs réflexifs indiquent que les communautés interagissent entre elles. Les fortes interactions passent par les chaînes appartenant à plusieurs catégories, ce sont des points d'articulations, des points de passage.

Figure (4) - L'algorithme Force Atlas 2 nous a permis d'identifier plusieurs communautés, mais il est moins adapté pour l'étude des interactions entre les différentes catégories. Nous optons pour l'algorithme Circular Layout : on projette sur un cercle les nœuds de notre réseau, en les ordonnant selon un critère prédéfini (ici, la catégorie). En couplant les fonctions de filtrage Intersection, SelfLoop et Mask sur des données spécifiques, nous constatons que les liens inter-catégoriels représentent 37 % des liens du réseau, mais seuls 20 % de ces liens sont mutuels, soit 8 % sur l'ensemble des liens du réseau. La majorité de ces liens est entretenue par des chaînes ayant soit un certain prestige (degré élevé d'arcs entrants), soit une betweenness élevée.

Les chaînes de “Technologie et informatique” ne possèdent quasiment aucun lien inter-catégoriel. A l'inverse, la catégorie “Langue” possède de nombreux liens avec les autres communautés, mais avec des liens intra-catégoriels plus faibles. Par ailleurs, un ordonnancement des noeuds par pays montre que les chaînes anglophones interagissent uniquement entre elles (barrière linguistique).

Conclusion

L'analyse de notre réseau montre que les interactions entre les catégories se font majoritairement via des chaînes multi-thématiques et populaires. Ces chaînes constituent donc des canaux de diversification de contenu pour leur reste de leur communauté. Nous avons pu détecter des communautés, mais des pistes d'améliorations existent pour confirmer la pertinence de notre démarche : travailler sur d'autres réseaux avec des chaînes-seed différentes, augmenter la profondeur (2) sur Youtube DataTools - il faut alors faire attention au “bruit”, automatiser la catégorisation des chaînes (extraction d'informations depuis l'API de Youtube et Machine Learning). Une plus-value de notre travail est notre programme d'extraction et nettoyage des données sur Python, qui rend notre travail facilement exploitable et réutilisable pour d'autres analystes.

Références

• YouTube Data Tools, https://tools.digitalmethods.net • Les chaînes YouTube culturelles et scientifiques francophones, Mathilde Hutin, https://www.datastro.eu/explore/dataset/youtube-a-lecole-hutin/information

Pour en savoir plus : https://github.com/v-barbosavaz/youtube_network_analysis