



HAL
open science

Analyzing the commentator network within the French YouTube environment

Kurt Maxwell Kusterer, Sylvain Mignot, Annick Vignes

► **To cite this version:**

Kurt Maxwell Kusterer, Sylvain Mignot, Annick Vignes. Analyzing the commentator network within the French YouTube environment. 11. International Conference in Complex Networks and their Applications 2022 | CNA 2022, Università degli Studi di Palermo, Nov 2022, Palermo, Italy. hal-03799185

HAL Id: hal-03799185

<https://hal.science/hal-03799185v1>

Submitted on 5 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyzing the commentator network within the French YouTube environment

Kurt Maxwell Kusterer¹, Sylvain Mignot², and Annick Vignes³

¹ Université Gustave Eiffel, Paris, France

LISIS (Laboratoire, Interdisciplinaire, Sciences, Innovations et Sociétés)
Cité Descartes 5, boulevard Descartes Champs-sur-Marne 77454 Marne-la-Vallée Cedex 02,
France

`kurt.kusterer@univ-eiffel.fr`

² Lille Catholic University, 51 rue du port, 59000, Lille, France

`sylvain.mignot@univ-catholille.fr`

³ INRAE,

LISIS (Laboratoire, Interdisciplinaire, Sciences, Innovations et Sociétés)

`annick.vignes@enpc.fr`

1 Introduction

YouTube is the largest video hosting platform. The site has emerged in 2005 and has achieved a continuous pattern of growth since its conception (Burgess & Green 2018).

A high number of creators, viewers, subscribers and commentators act in this specific eco-system which generates a huge amount of money.

The viewers, subscribers and commentators help to increase the popularity of content creators. The most popular creators are sponsored by brands and participate in marketing campaigns. For a few of them this becomes a financially rewarding profession. This has been reinforced by the YouTube Partner Program (YPP) which organizes the sharing of advertising revenue with some of the content creators (Caplan & Gillespie 2020). With such a program, the level of competition for views/subscribers on YouTube grew immensely and we believe that the role of comments in increasing the popularity is to be emphasized.

(Rihl & Wegener 2019) underline the importance of reading and making comments in the building of fans' attachment and then the building of stars visibility. Most of the influential content creators are regularly adding comments to their own videos, or replying to followers' comments. For example, when Hassan, one of the most influential French content creators⁴ asked a very simple question as a comment on one of his videos, he received 500 replies in a very short space of time. We understand how important 'promotion' is to content creators. Comments and the intrinsic characteristics of commentators are playing an important role in the fact that a video (and then the creator who produced it) succeeds or not in the YouTube world.

In what follows, YouTube is considered as a bilateral network between the videos and the commentators.

Analyzing a detailed data set focused on French YouTubers (section 2), we consider each comment as a link between a commentator and a video. The main objective of this

⁴<https://www.youtube.com/watch?v=tGndysi405s>

paper is to understand the determinants of the creation of these links. This is to say, what can explain the choice of an agent to comment a specific video instead of another one, taking into account characteristics of commentators, videos, topics, channels as well as recommendations. This work is different from the classic NLP studies, using text mining techniques to analyze the contents of the comments and the kind of information they diffuse⁵.

2 The data set

2.1 Building the data set

The data we have has been collected via a data intermediary giving us access to a detailed data base of channels and their videos. The data is relevant for the period of 2020-2021 and focuses on the French YouTube environment.

From this set of 391 588 videos, we extract the channels which can be monetized, according to YouTube regulations (channels with at least 1000 subscribers and more than 4000 hours of viewing time during the last twelve months).

For each video, our data set provides the number of subscribers and the number of views after 180 days (after the 1st day of uploading). It also provides the name of each commentator, the number of comments he made for this video, the YouTuber's name, the topic of the video (humor, gaming...) and other specific characteristics.

Screening the monetized channel. Keeping only channels with more than 1000 subscribers is trivial, keeping those with more than 4000 hours of viewing is trickier. We only have data about the first 6 months after release of each video. Because the vast majority of views occur during the first month after the creation of a video⁶, we formulate the assumption that after six months, the number of views increase at a non significant rate and approximate the annual time of viewing by a six months time of viewing. At the end, we keep 4093 channels and 128462 videos.

The comments Our data set contains 2012966 comments. Each observation within our comments data set is a record of a comment made, indicating the name of the commentators, the video ID and the channel (or YouTuber) who has provided the video.

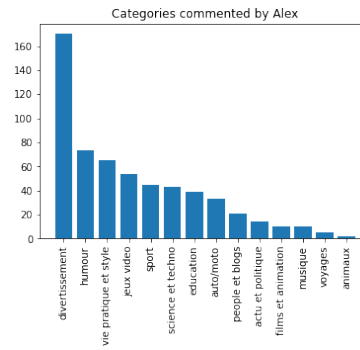
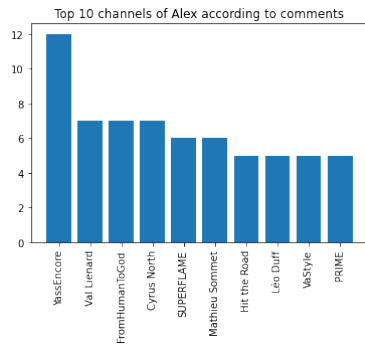
It is possible to achieve a great level of detail with the data we have, which will prove essential to understand consumer behavior within the commentator network.

We can calculate total number of comments made by each commentator. For example, we observe that 'Alex' is the commentator with the most comments and can follow his personal strategy. These statistics will help to categorize the commentators, and describe their strategies (topics, channels, number of videos...).

Figure 1 displays Top 10 channels of Alex and Figure 2 lists the categories most commented by the Alex.

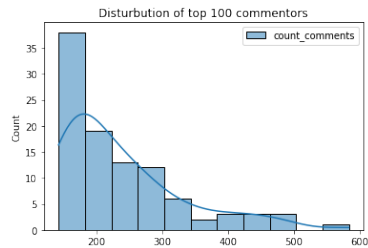
⁵For a recent review of this literature, refer to (Teng et al. 2020)

⁶We calculate that between two and six months after the release the number of views only increase on average by 11%.



3 Stylized facts

Some descriptive statistics help to have a broad overview of the commentators. We have 1032771 unique commentators in our data set, with an average number of comments 2. The sample has a standard deviation of 4, a minimum of comments equal to 1 and maximum number of comments equal to 584 (The top commentator 'Alex'). Figure 3 displays the distribution of the top 100 commentators.



4 Methodology

Quick statistics show that 96% of the commentators comment only one time a certain video. In what follows, we study a non weighted bipartite network between commentators and videos, built on the sub-sample of 96% of unique comments. There is a link between two nodes when a commentator makes a comment on a video.

We run an Exponential Random Graph Model(ERGM) approach to evaluate which characteristics influence the probability of commenting a video. The creation of a link will be explained in terms of common video features, such as duration, quality, number of likes, number of views etc.

Our method is similar to that of (Gruzd & Hodson 2021), in which they employ network analysis along with an ERGM, followed by the content analysis of metadata from YouTube channels. The originality of our work consists is based around the topic we intend to study as well as the data we analyze.

Bibliography

- Burgess, J. & Green, J. (2018), *YouTube: Online video and participatory culture*, John Wiley & Sons.
- Caplan, R. & Gillespie, T. (2020), 'Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy', *Social Media+ Society* **6**(2), 2056305120936636.
- Gruzd, A. & Hodson, J. (2021), 'Making sweet music together: The affordances of networked media for building performance capital by youtube musicians', *Social Media+ Society* **7**(2), 20563051211025511.
- Rihl, A. & Wegener, C. (2019), 'Youtube celebrities and parasocial interaction: Using feedback channels in mediatized relationships', *Convergence* **25**(3), 554–566.
- Teng, S., Khong, K. W., Sharif, S. P., Ahmed, A. et al. (2020), 'Youtube video comments on healthy eating: Descriptive and predictive analysis', *JMIR public health and surveillance* **6**(4), e19618.