Attention dynamics on YouTube : conceptual models, temporal analysisof engagement metrics, fake views

Maria Castaldo

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)
Spécialité : Automatique - Productique
Unité de recherche : Grenoble Images Parole Signal Automatique

Dynamique de l'attention sur YouTube : modèles conceptuels, analyse temporelle des métriques d'engagement, fausses vues

Attention dynamics on YouTube: conceptual models, temporal analysis of engagement metrics, fake views

Présentée par :

Maria CASTALDO

Direction de thèse :

Paolo FRASCA Chargé de Recherches, Université Grenoble Alpes	Directeur de thèse
Tommaso Venturini INRIA	Co-directeur de thèse

Rapporteurs :

Jean-Philippe COINTET
CHERCHEUR HDR, SCIENCES PO PARIS
Alessandro FLAMMINI
PROFESSEUR, Indiana University

Thèse soutenue publiquement le **17 novembre 2022**, devant le jury composé de :

Jean-Philippe COINTET CHERCHEUR HDR, SCIENCES PO PARIS	Rapporteur
Alessandro FLAMMINI PROFESSEUR, Indiana University	Rapporteur
Gilles BASTIN PROFESSEUR DES UNIVERSITES, INSTITUT D'ETUDES POLITIQUES DE GRENOBLE	Président
Kibangou ALAIN MAITRE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES	Examinateur
Béatrice ROUSSILLON MAITRE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES	Examinatrice
Claudio ALTAFINI PROFESSEUR, Linköpings Universitet	Examinateur

Invités :

Floriana Gargiulo
CHARGE DE RECHERCHE, CNRS
Tommaso Venturini
CHARGE DE RECHERCHE, CNRS

UNIVERSITÉ GRENOBLE ALPES

ÉCOLE DOCTORALE EEATS

ELECTRONIQUE ELECTROTECHNIQUE AUTOMATIQUE TRAITEMENT DU SIGNAL

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'Université Grenoble Alpes

SPÉCIALITÉ: AUTOMATIQUE-PRODUCTIQUE

Présentée et soutenue par

Maria CASTALDO

Attention dynamics on YouTube: conceptual models, temporal analysis of engagement metrics, fake views

Thèse dirigée par Paolo FRASCA et Tommaso VENTURINI

préparée au laboratoire Grenoble Images Parole Signal Automatique (GIPSA-lab)

soutenue le 17 Novembre 2022

Directeur :	Paolo FRASCA	-	GIPSA-lab, Université Grenoble Alpes, CNRS
Directeur :	Tommaso VENTURINI	-	CIS, CNRS
Encadrante :	Floriana GARGIULO	-	GEMASS, Université Paris Sorbonne, CNRS

Jury :

Rapporteurs:	Jean-Philippe COINTET	-	Medialab, Science Po, Paris
	Alessandro FLAMMINI	-	CNetS, Indiana University
Examinateur :	Gilles BASTIN	-	PACTE, Science Po Grenoble
	Alain KIBANGOU	-	GIPSA-Lab, Université Grenoble Alpes
	Béatrice ROUSSILLON	-	GAEL, Université Grenoble Alpes
	Claudio ALTAFINI	-	Linköping University

Contents

List of Figures

Introduction

Over the past 20 years, social networks have permeated our society, revolutionizing every aspect of our daily lives: social interactions, communication, and access to news. In a fairly short time, social networks have reached impressive proportions; as of today, 37% of the world's population accesses Facebook at least once a month, and 33% watches at least one video on YouTube. Names like Twitter, Tumblr, Instagram, TikTok, and LinkedIn are all familiar to the large majority of us, whether or not we are active social media users.

Social networking platforms soon abandoned their initial role of mere outlets to share experiences or communicate with friends to become key players in information dissemination and opinion formation in our societies [NR20]. Nowadays, they constitute one of the primary ways to access information for most people and consequently one of the most incisive tools in shaping public opinion. For this reason, they make up the ground on which many politicians carry out their election campaigns, where parties disseminate propaganda, and political debates are ignited.

With such an important role comes great responsibility, but over the years, social networks have not always lived up to it. The 2016 U.S. election provided the most glaring example of the repercussions that the lack of controls on online platforms can have on the democratic process. At the time, Facebook was loaded with fake news, being re-shared more than the content published by traditional media [Sil]. Thousands of fake accounts crowded Facebook and Twitter [Sha+17b], spreading fake news that influenced the political debate around the election [BF16]. This massive amount of misinformation, often favoring Donald Trump over Hilary Clinton [AG17], led many to wonder what the election outcome would have been without the influence of fake news [Out][Dew16][Rea16].

The fact that an event of such impact on society could be distorted [BF16] by the way social networks manage the dissemination of content is certainly alarming and has attracted the interest of many researchers. Understanding what governs the dynamics of content dissemination on these platforms, understanding the mechanisms for suggesting content profiled by user, understanding what makes a piece of content go viral, have become fundamental necessities to ensure the normal course of democratic processes and hence deserve the full attention of the scientific community.

Thesis Objective

The goal of this thesis is to study the dynamics of online dissemination of content, first from a conceptual point of view, and then with a data-driven approach, based on data collected from YouTube. As for the conceptual approach, we are interested in building a model through which we can study the impact that certain variables related to news consumption can have on the collective dynamics of attention. As per the data-driven approach, we aim both to develop models that explain the temporal evolution of popularity on YouTube, but also to analyze all the evidence that can emerge from the data, through tools proper to Computational Social Science [Laz+09] [Con+12].

Main Contributions

Bridging Media Studies and Computational Social Science, the first contribution of this thesis concerns the mathematical formalization of the "public arenas model" developed by Hilgartner and Bosk in 1988. Through this formalization and through a reinterpretation of the model applied to social networks, we discuss what are the risks of media arenas that overly reward trendy content with higher visibility. Such emphasis on trending matters, we claim, can have two detrimental effects on public debates: first, it shortens the amount of time available to discuss each matter; second, it increases the ephemeral concentration of collective attention.

Alongside this theoretical formalization, another significant contribution made by this thesis consists of the collection of unprecedented data on the temporal evolution of YouTube views. The collection of these temporal evolutions was only possible by querying hour-by-hour YouTube's API over the past three years, as historical datasets of this kind are in no way retrievable a posteriori through the official API. Given the difficulty of collecting temporal evolutions of this kind and the consequent lack of studies about them in the literature, these data have brought out interesting evidence from multiple points of view.

First, these data allowed us to build a model of content dissemination, when models of this kind had not been studied since 2017 (the year in which YouTube restricted access to views count timeseries). The model we have constructed allows us to distinguish the weight played by innovation, i.e., independent search for content by users, and imitation, i.e., suggestion of content by others or by YouTube, in content dissemination. Specifically, in our data we observe that the videos in which imitation plays a greater role are on average more popular and reach their audience faster then other videos.

Another interesting evidence that emerges from the data concerns a much understudied

YouTube policy: the removal of illegitimate views attributed to automated programs (bots). In our study we look at the extent of this phenomenon (which affects more than 50 percent of the videos) and its characteristics. We discuss the risks that altering engagement metrics could cause: seemingly more popular, content could be more extensively shared through human and algorithmic recommendations and thus reach a larger audience.

A final contribution linked to the data comes from having collected them in a unique period in the history of social networks: the Covid-19 pandemic. Analyzing this period allowed us to study and recognize the "natural rhythms" of access to the platform, in contrast to the "exceptional" ones linked to the effect of the pandemic on users' lives. More in general, it allowed us to analyze how themes and emotions shared online have changed because of this unprecedented shock.

Manuscript Outline

The remainder of this manuscript consists of seven chapters. Chapter 1 concerns a review of the existing literature on online attention dynamics. Very different and sometimes distant scientific communities deal with this topic: epidemiology, physics of complex systems, media studies, and marketing science. In the first chapter of this manuscript, we aim to expose the major results obtained in the different disciplines, observe the consistency or possible inconsistency of independent studies, and propose a synthesis of the knowledge recognized to date.

In Chapter 2, we present a mathematical formalization of the Hilgartner and Bosk "public arena model" and reinterpret it in the light of social networks, aiming to conceptualize an information disorder based on temporal aspects of content dissemination.

In Chapter 3, we introduce what is YouTube, its history, the controversies that involved it, and how it evolved in response to the problems that have emerged over time. We describe, to the possible extent, its recommendation system and the variables it considers when choosing which content to propose to users. We discuss the limitations imposed by the platform on data collection and discuss possible techniques to overcome these restrictions. Finally, we present the data collected in the last three years and on which the results of the following chapters are based.

In Chapters 4, 5 and 6 we discuss the evidence emerging from the data. In Chapter 4 we apply a Bass model to the evolution of views counts to identify the role that imitation and innovation play in content diffusion on the platform. In Chapter 5 we focus on YouTube policy of decreasing views counts when it deems views to be done by automated programs. In Chapter 6 we compare the period of the first French lockdown with the previous period

to study the impact of Covid-19 on the activity, themes and emotions shared online.

Finally, Chapter 7 summarizes our contributions, discusses the limitations of our work, and proposes some relevant open problems and concrete venues for future research.

List of publications

The following is an exhaustive list of publications written during these three years of Ph.D., that are either published or under review. Some of them are not directly related to the study of online attention dynamics or YouTube and are consequently not discussed in this manuscript.

Publications discussed in the manuscript

1. **Maria Castaldo**, Paolo Frasca, Tommaso Venturini. "On Online Attention Dynamics". In *Cyber-Physical-Human Systems: Fundamentals and Applications"*, John Wiley & Sons / IEEE Press. (2022, In press) This paper corresponds to Chapter 1 of this thesis.

2. **Maria Castaldo**, Tommaso Venturini, Paolo Frasca. "Junk News Bubbles: Modelling the Rise and Fall of Attention in Online Arenas". *New Media and Society* **24**, 9 (2022), pp. 2027-2045. This paper corresponds to Chapter 2 of this thesis.

3. **Maria Castaldo**, Floriana Gargiulo, Tommaso Venturini, Paolo Frasca. "The rhythms of the night: increase in online night activity and emotional resilience during the spring 2020 Covid-19 lockdown". *EPJ Data Science* **10**, 7 (2021). This paper corresponds to Chapter 6 of this thesis.

4. Krasimira Bozhanova, Yoan Dinkov, Ivan Koychev, **Maria Castaldo**, Tommaso Venturini, Preslav Nakov. "Predicting the Factuality of Reporting of News Media Using Observations about User Attention in Their YouTube Channels". In *Proceedings of the International Conference on Recent Advances in Natural Language Processing* (2021), pp. 182–189. The results of this paper, partially related to the thesis work, will be briefly discussed in Chapter 7.

Under review:

5. **Maria Castaldo**, Paolo Frasca, Tommaso Venturini, Floriana Gargiulo. "Doing data science with platforms crumbs, an investigation into fakes views and YouTube attention

cycles." *Journal of Computational Social Science*, in review. This paper corresponds to Sections 5.1 to 5.4 of this thesis.

Publications not discussed in the manuscript

During the thesis, I could also collaborate in the analysis of platforms other than YouTube. In particular, I studied the Polymath blogs, a collaborative science platform. This study has allowed me to refine my knowledge of applied network science tools, useful for the analysis of any social network.

6. Floriana Gargiulo, **Maria Castaldo**, Tommaso Venturini, Paolo Frasca. "Distribution of labor, productivity and innovation in collaborative science." *Applied Network Science* **7**, 19 (2022).

Following a research line started with my master thesis, I kept working in network dynamics and in particular on the study of a network formation game in which each node aims to maximize its Bonacich centrality. This study gave rise to two papers, a conference paper and a journal paper, the latter currently under review.

7. **Maria Castaldo**, Costanza Catalano, Giacomo Como, Fabio Fagnani. "On a Centrality Maximization Game". In IFAC-PapersOnLine 53.7 (2020), pp. 2844-2849.

8. Costanza Catalano, **Maria Castaldo**, Giacomo Como, Fabio Fagnani. "On a Network Centrality Maximization Game." *Mathematics of Operations Research*, (2022) under review.

Chapter 1

State of the Art

The goal of this chapter is to illustrate how in different disciplines the scientific community has tried to answer crucial questions about the dissemination of content online: How does collective attention concentrate and dissipate in modern communication systems? How do subjects and sources rise and fall in public debates? How are these dynamics shaped by media infrastructures? In the perspective of addressing these questions, this chapter provides a review of the literature on the dynamics of online content dissemination. Our goal is to prepare the ground for answering outstanding questions, which will be addressed in the remainder of this manuscript, through empirical investigations, mathematical modeling, and numerical simulation.

The interest in dynamics of collective attention is as old as sociology. Already in the 19th century Gabriel Tarde [Tar90]; [Tar93], argued that these fleeting dynamics (rather than the more stable structures and norms) should make up the core of social research [Lat02]. Attention dynamics rose again in sociological preoccupations in the '70s and '80s, when the major problem of the nascent media research was to describe the competition for the limited bandwidth of radio and television broadcasting. Concepts such as "attention cycles" ([Dow72]; [HB88]) and "agenda setting" ([MS72]; [McC05]) became prominent to investigate media schedules and their consequences on public debate. With the advent of digital media, the interest for collective attention shifted from the supply to the demand side. Vindicating Herbert Simon's 1971 prophecy [Sim71a], media scholars (and commercial actors) realized that in an information-rich environment, attention becomes a scarce and therefore valuable resource. This gave rise to many critical reflections on the consequences of the rise of the 'attention economy' and its way of transforming collective attention and debates into a marketable commodity [CK12]; [Cit14]).

The research on attention economy is extremely interesting for its effort to conceptualize a very large phenomenon (the way in which collective attention flows through the media system) through the continuous convergence and divergence of a myriad of individual choices [Ter12]. At the same time, and for the same reason, the literature on attention economy has remained largely theoretical. Until recently, the empirical investigation of the dynamics of collective attention has been hindered by the difficulty to procure data sets broad and

representative enough to account for an entire media population, but rich enough to distinguish each fleeting individual choice [VL10]; [Lat+12]. In the last few years, however, the massive investments by commercial and governmental actors into the surveillance of media interactions [Zub19] have generated the data necessary for the empirical and computational study of the flows of collective attention, and scholars have begun to seize this possibility.

Based on this growing literature, this chapter aims to present a synthesis of the widely recognized elements regarding the diffusion of online content. To do so, at first we will attempt to summarize the general findings concerning collective attention: we will discuss collective attention as a limited resource for which different news stories compete, discuss what outcomes emerge from this competition, and in particular how collective attention is distributed among the available content. We will present some of the major drivers of collective attention and how these have been included in models of various kinds to explain the evolution of online content popularity.

In a second phase, we will focus on the terrain on which we study collective attention: online platforms. Although we will devote a separate chapter to YouTube, it seems essential to us in this place to stress how all platforms constrain and strongly influence the way news propagates within them. Understanding through what means this influence occurs is of crucial importance when it comes to online content dissemination.

1.1 Collective Attention

This section gives an overview of what is known about collective attention, presenting the previous literature aggregated by concepts, not disciplines. Given the variety of scientific communities that have dealt with these issues, we will discuss how each feature of collective attention has been treated in various communities. We will discuss the empirical evidence to support them and, when possible, how they can be integrated into inferential models.

The characteristics of collective attention that we identified in literature and that we are going to discuss are the following:

1. it consists of a limited resource;
2. it is highly concentrated on a few objects;
3. it is attracted by novelty;
4. it is influenced by popularity.

Collective Attention is a scarce resource

When Herbert A. Simon first theorized the concept of attention economy, he built his reasoning on the statement that human attention can be treated as a scarce commodity. Despite the impressive complexity and processing power of the human brain, it is undeniable that its capacities are limited: we can barely attend to over one object at a time, and we can hardly perform two tasks at once [MI05]. Much research in the cognitive science has investigated the limitations of our brain, and we refer the interested reader to [MI05]. For our purpose, it suffices to point out that these limitations have become standard assumptions in many works, not only to analyze the patterns of online engagement [LS+19] [Wen+12] [Qiu+17] but also in modeling opinion dynamics in social groups [RF20]; [CFR21]; [Cer+21]. Crucially for this chapter, these assumptions have become an essential starting point for the research on collective attention, as the scarcity of individual cognitive resources has turned attention into the object of a increasingly competitive market, in which attention ceases to be an individual feature and becomes a collective commodity that is consumed online.

Attention is highly concentrated

As a result of the limitedness of individual and collective attention, news items have to compete with each other to gain the consideration of the public. This competition rewards few items which become over-popular, while the vast majority of them remain unnoticed. As largely discussed in the last years, online popularity is highly skewed, with a relatively small number of participants getting most of public attention. The distribution of popularity among online items has often been found to respect the "80-20 rule", also known as the Pareto rule: the 20% of the online content accounts for the 80% of the popularity. Evidence of this have been brought out on many platforms: it turned out to be true for videos on Metacafe, Yahoo!, Dailymotion, Veoh [Mit+09] and YouTube [Cha+07], and for retweets in Twitter [Bil+15] [Lu+14].

It is legitimate to wonder what causes this skewness and whether it exists also outside of user-generated content platforms like the ones mentioned above. A first answer to these questions is given by Cha et al. in 2009 [Cha+09], where they compared the consumption of videos on platforms of User Generated Content (UGC), like YouTube, and Professionally Generated Content (PGC), like Netflix or Yahoo! Movies. They outlined that in UGC platforms attention is less equally distributed among items. More precisely, at that time, on YouTube 10% of the most popular videos were accounting for nearly 80% of the total views, while on-demand videos presented a less skewed distribution of popularity. In practice, while on PGC platform it never happens that a content is left with no public, on YouTube there is a significant quantity of videos that do not receive any view at all. But that is not the only

difference between UGC and PGC: the authors also stress an enormous difference in the quantity of content uploaded on the two kinds of platforms, with UGC platform collecting a significantly higher quantity of material. The massive amount of content present on YouTube, together with the human cognitive limitations and the peer influence, might be the cause of the stressed skewness in UGC platforms: people, only disposing of a limited attention, have to choose among an excessive variety of contents and may end up relying on imitation for their choices. As a result, collective attention is more concentrated and many items cannot arouse the slightest interest.

Acknowledging that the distribution of attention is skewed is crucial not only to conduct research on attention dynamics but also to contextualize previous works in the field: working with empirical data on social platforms often means dealing with a majority of items that received no or very little attention. When aiming at modeling popularity trends, it then becomes important to remove the non-relevant observations. For instance, Crane and Sornette in [CS08] based their model of content diffusion on only the 10% of YouTube videos in their dataset, as the remaining 90% either showed a too low number of views or could be accurately described as (purely random) Poisson processes. Similarly, Kampf at al. in [Kä+12] had to perform a significant filtering of their Wikipedia dataset: the vast majority of articles they monitored were rarely accessed and almost never experienced significant bursts of activity. To filter their data, they focused on articles that exhibited a minimum rate of 256 views at least in one hour, over the period of observation. This threshold, that might not seem particularly demanding, was met by only the 0.17% of the Wikipedia articles studied by the authors. Such a low percentage is, again, evidence of the quantity of content available on the web but never or rarely accessed.

The role of novelty

Once acknowledged that few items capture most of the public attention, it comes natural to wonder which are the factors that concentrate everyone's interest on specific items. In the "Attention Economy" literature, novelty is often presented as one of the main factors [Sim71b] [Gol97]. Indeed, as Goldhaber stated, since it is hard to get new attention by repeating exactly what has already been done in the past, a key role in the attention economy is played by novelty.

Online platform managers are well aware of the importance of novelty. Its promotion is expressly sought by platforms and specifically encoded in recommendation systems and in particular in the algorithms that select and suggest content to users, trying to meet their tastes and interests. Covington et al. [CAS16], developers at YouTube in 2016, include *freshness* among the three major needs of YouTube recommendation system. In particular, they acknowledge that, as "many hours' worth of videos are uploaded each second to

YouTube", "recommending recently uploaded ("fresh") content is extremely important for YouTube as a product". In fact, they observe that users constantly prefer fresh content and, hence, to keep their engagement high and make them spend time on the platform, YouTube has to satisfy their need for novelty.

The preference for novel content has been observed also in other contexts. In 2012 on Twitter, among the total tweets published in a week, the 45% had never been published before [Wen+12]. Similarly, according to Roth et al. [RMM20] in 2020, two-thirds of the suggestions of YouTube given at a certain moment were not anymore associated with the same video after 2 days. Novelty can thus be listed as a key factor for capturing people's attention and it should be considered when modeling popularity trends and shifts of collective attention from one topic to another.

The role of popularity

Another factor that certainly draws attention to specific items is their popularity. Content already popular is much more likely to be discussed with friends, shared online, or recommended, and hence it has more chances to reach a larger audience. This link between current and previous popularity has been the subject of many studies, both predictive and inferential. Various models, stemming from different branches of science, have been adopted to describe the evolution of popularity online. By classifying these models according to research field that generated them, we could distinguish: *epidemic models*, issued from the tradition of mathematical models describing the spread of infectious diseases, *Bass-like diffusion models* stemming from the theory of innovation diffusion, and *self-exciting processes* belonging to the larger family of counting processes in statistics.

Regardless of the specific model used, the core concept is the same: people influence each other. To describe this occurrence, different terms have been adopted in different fields. When dealing with innovation diffusion models, we usually refer to the tendency of people to be influenced by others as an *imitation processes*. When adopting epidemic models instead, we usually refer to it as a *contagion* effect or a word-of-mouth effect. Despite its different names, the concept is the same: when many people are aware of a content, it becomes more likely for an individual to encounter it. We could also talk of *popularity effects*: the future spread of a piece of information is influenced by its previous success. In the following section, we are going to explain how this popularity effect has been considered by researchers in different contexts.

- **Epidemic models**. Despite originally designed to model the spread of infectious diseases, epidemic models can effectively describe the propagation of content online. Daley and Kendall [DK64], in 1964, first proposed the analogy between epidemics and

the spread of rumors, suggesting the same mathematical model might apply to both fields of study. The foundation of these models resides in partitioning a population into different classes of individuals. Among the most common classes we find Susceptible, Infected and Recovered individuals. Susceptible individuals are those who still have not been in contact with the disease. Infected individuals have caught the disease and can spread it, while recovered individuals healed from the disease and cannot transmit it anymore. Of course, when adapting these models to attention dynamics, the disease is replaced by a piece of information and healing is replaced by forgetting. We can elaborate on different models, depending on the class of individuals considered. For example, the SI model considers only susceptible or infected individuals. In an SI model, the fraction of infected individuals I evolves in time according to:

$$\dot{I} = \alpha(1 - I)I.$$

Here, the fraction of newly infected individuals is given by the probability of a susceptible individual to meet an infected one, multiplied by a transmission rate α.

Many variations of this basic model have been proposed in literature with the specific aim of explaining information and rumor diffusion. In 2006 Bettencourt et al. [Bet+06] proposed a variation of the SI model (which they called SEIZ model), to fit the spread of the use of Feynman diagrams through the theoretical physics communities in USA, Japan, and USSR in the period immediately after World War II. In their SEIZ model actors can either be susceptible (S), i.e. still unaware of an idea, exposed (E), i.e. having been in contact with the idea, infected (I), i.e. adopters of the idea, or skeptic (Z), namely aware of the idea but unconvinced by it. They prove that introducing exposed individuals consistently increases the capability of the model to fit the data: in fact, inserting a delay between the moment physicists first met Feynman diagrams and the moment they adopted them brought major improvements to the data explanation. Jin et al. later applied the same model [Jin+13] to the spread of rumors online, with similar outcomes: they confirmed the improvement due to the introduction of exposed individuals in a simple SI model. Another confirmation of the importance of introducing an exposure delay between the reception and the adoption of an idea comes from the work of Xiong et al. in 2012 [Xio+12]. The authors proposed a diffusion model with four different states: susceptible (S), contacted (C), infected (I), and refractory (R). Contacted individuals behaved exactly as the exposed individual in the SEIZ model: they acknowledged the information but have not decided yet whether to spread it or not.

Besides the above mentioned variations of the SI model, there also exists some which do not include the addition of new classes of individuals. In [Ric+14], the authors propose different biologically inspired models which proved to fit at least the 90% of videos of a conspicuous YouTube dataset. Among the considered models, we find a

variation of the SI model called *Gompertz model* and governed by:

$$\dot{I} = \alpha I \log(M/I) \tag{1.1}$$

where M represents the potentially interested public and I represents the number of users that viewed a video. They compared it with a simple *exponential model* $\dot{I} = \alpha(M - I)$ and with some variations of the Gompertz and the exponential model where the authors added a term kt to the temporal evolution of $I(t)$. While the simple Gompertz and exponential model failed at fitting the majority of the videos, the modified models brought a sensible improvement to the fittings: they explained almost 75% of videos popularity evolution. Even though adding a term kt to $I(t)$ seems rather contrived and difficult to interpret, we can explain the need of adding a further parameter k as the necessity of higher degrees of freedom when explaining complex dynamics. In this respect, adding a term kt to the evolution of infected individuals, or adding new classes of individuals to the SI model, play the same role.

- **Bass diffusion models**. The Bass diffusion model owns its name to Frank Bass, an academic pioneer of marketing research in the second half of the XX century. It was first introduced in [Bas69] to model the process of adoption of new products in a market. It is based on a simple classification of individuals into two groups: *innovators* and *imitators*. The Bass model found its main application in forecasting innovations or technology sales. The model formulation is the following:

$$\dot{F}(t) = p(1 - F(t)) + qF(t)(1 - F(t))$$

where $F(t)$ is the fraction of adopters in a population at time t, p is the coefficient of innovation and q is a coefficient of imitation. As we can see, also in this formulation of the problem, we have a term of contagion $F(t)(1 - F(t))$ that models the influence of users on each other. The analogy between infected individuals in epidemic models and innovators in models of adoption is glaring, and it has been made explicit in many works [Bas69] [CKM57] [TCG12]. That is one of the reason why, especially when these models are complemented with additional assumptions, it is hard to make a sharp distinction between adoption models and epidemic models.

  The strengths of Bass model certainly include its simplicity, which makes it a particularly interpretable model, and the fact that it possesses a closed solution that can be easily applied to data though a number of estimation methods, including ordinary least squares and maximum likelihood. At the same time, however, its simplicity has been counted by many [Kie+12] as a limitation, since it does not take into account some of the complex aspects of diffusion processes, such as the heterogeneity of consumers or their different capacity for social influence. For this reason, numerous alternatives have been proposed over the years to integrate temporal and spatial heterogeneity within the model, either from a *macroscopic* point of view or by translating it into an

agent-based model, more designed for what-if type questions. In particular, Gao et al. [Gao+21] propose a heterogeneous variant of the standard Bass model by introducing in the expression of $\dot{F}(t)$ some quantities related to users habit and content characteristics. For instance they took into account the number of followers per user, the number of favorites, the average number of tweets per day, and so on. Similarly, Hoang et al. [HL21] propose a macro version of Bass model with the addition of variables related to the users characteristics. In [Ran+15], Rand et al. apply an agent based Bass model to the diffusion of information in Twitter during four major events happened in the U.S. in 2011-2012. They obtained pretty satisfactory fitting of the increase in time of the number of people talking about each topic. In [Luu+21], Bass is used in its agent based version to study the effects of degree-changes in a network on diffusion patterns. Kim et al. [KNC13] propose an extension of the Bass model taking into account the existence of various communities and uses it to infer the diffusion process between and within different populations on the Web. All these examples give us an idea of the flexibility that can arise from the agent-based version of Bass's model and how much this can be exploited to explain real data. The examples brought so far concern only studies in which the Bass model has been applied to social networks; for a complete review of its studies in marketing science we refer the reader to [Kie+12].

- **Self-exciting processes**. The best example of how self-exciting processes can model content diffusion is given by Crane and Sornette [CS08]. The authors provide a model to fit the evolution of videos on YouTube, and they base it on three essential assumptions: (1) the relevance of human interactions in spreading a piece of information, (2) the existence of influences external to social media, and (3) the fact the humans activity follows very specific patterns inhomogeneous in time [DS05] [JS00] [Joh01]. Crane and Sornette's model consists of a self-excited Hawkes conditional Poisson process [HO74] with an instantaneous rate of views given by:

$$\lambda(t) = \lambda_0(t) + \sum_{i,t_i<t} \mu_i \varphi(t - t_i)$$

where the term $\sum_{i,t_i<t} \mu_i \varphi(t - t_i)$ models the contagion/imitation process, and $\lambda_0(t)$ represents and exogenous source of views, which captures all spontaneous engagement not triggered by epidemic effects. The parameter μ_i represents the number of potential viewers who will be influenced by person i that views a video at time t_i. The kernel $\varphi(t)$ is chosen to be equal to

$$\varphi(t) = \frac{1}{t^{1+\theta}} \tag{1.2}$$

and it represents the rate at which individuals consume information. Such formulation stems from a wider literature investigating the rhythms of individual human activity, which, in many contexts, is characterized by power law distribution of waiting times between consequent actions, for instance, between receiving an email and replying

to it [Bar05] [Vá+06] [OV09]. Here, the memory kernel $\varphi(t)$ describes the distribution of waiting times between acknowledging the existence of a video and actually watching it. We could consider it as another way to model the *latency* time between acknowledgment and adoption of an idea proposed in the SEIZ model by Bettencourt [Bet+06].

1.2 Online Platforms' Influence on Content Dissemination

Having discussed some key characteristics of collective attention, we can now discuss how it is influenced by the social networks that convey it online. Clearly, since they make up the medium through which information is propagated, their structure is in itself a constraint on its dissemination. On Twitter, for instance, we can only write brief messages and share links to external content. On Instagram, we can post almost exclusively images and video content, and on YouTube and TikTok almost exclusively videos. In addition to constraints on the type of content that can be posted, there are also constraints on how to share it. On Twitter, for example, users can retweet someone else's tweet. On Facebook, users can share others' posts, while on Instagram there is no proper way to share other people's story. Beyond these tools made available (or not) to users to circulate content, we must consider another essential actor when talking about online attention dynamics: recommendation systems. Recommendation systems are those tools that allow platforms to choose what content to present to users from the myriad of possibilities that exists. They are, for example, the algorithms behind the "Groups You Should Join" and the "Discover" feature on Facebook. They sort posts on the Twitter home page and suggest "Related Videos" next to what a user is watching on YouTube. In this section, we want to discuss the importance of recommendation system's role and how they have evolved in reaction to the criticism addressed to them. The purpose is twofold: on the one hand, we want to understand the role this actor plays in the dissemination of content, and on the other hand, we want to point out how frequent these changes have been and how quickly social network structures have evolved. This speed of change is undoubtedly relevant in dealing with the dynamics of online content dissemination as it invites us to contextualize the literature and constantly monitor the effects of the evolution of recommendation systems.

Recommendation systems as a gateways for information

Recommendation systems undoubtedly play a vital role in the business model of all social networks. Platforms sell attention to advertisements, and, to do so, they must attract it. Recommendation systems are designed for this purpose, to attract users' attention and keep the public glued to the screen. Studying what weight recommendation systems have in

deciding what people watch online can be difficult for researchers. It is often impossible to distinguish which content is carried by the recommendations and which is not. In this regard, YouTube turns out to be a rather interesting platform because it allows us to clearly distinguish what traffic comes from the recommendation system and what comes from other sources. YouTube, as we will resume in a devoted chapter, is one of the few platforms where it has been possible to quantify the impact of the recommendation system. Already in 2010, Zhou et al. [ZKG10a] confirmed that the related video recommendation on YouTube was the main source of views for most of the videos on YouTube. In 2014, Fugueiredo et al. [FBA11] confirmed that the most likely way to get to watch a video on YouTube is either "the related video" list, which displays a list of 20 videos that the platform suggests based on the previously watched video, or the "home page". These two internal sources of views account for the $32\% - 43\%$ of the total views. On the other hand, external sources like links to the video in other social media, or suggestions made by friends, account for only the 8%-16% of the total views of a video. Nowadays, YouTube itself, on its official blog, declares that "recommendations drive a significant amount of the overall viewership on YouTube, even more than channel subscriptions or search". In 2018 YouTube Chief Product Officer Neal Mohan admitted that, for over the 70 percent of the time users spend watching videos on the platform, they are lured in by one of the service's AI-driven recommendations [Sol]. The example case of YouTube, certainly very relevant, provides evidence of the importance recommendation systems can have in disseminating information. Although on other social networks it is more difficult to quantify their impact, recommendation systems make up a pervasive essence of platforms: they select and order most of the content users view. Unless typing directly in Facebook's or TikTok's search engine, unless entering Twitter through an external link, almost everything we see on the home pages of these platforms is the result of algorithmic selection, which, as we will see below, has not always worked out at its best.

1.2.1 Recommendation systems as ever-changing actors

Given the importance of recommendation algorithms in defining what people look at, we devote this section to discussing how they changed over time and why researchers should monitor their behavior constantly.

We recognize two primary drivers of the continuous updating of recommendation systems: technological improvement and response to emerging ethical needs. Regarding the first one, the goal of constant updates is always the same: enhancing the efficiency of recommendations, increasing the engagement of the public and the time users spend online. Most changes that recommendation systems undergo are not disclosed, but still some have been documented. In [CAS16] [Zha+19], for example, developers at YouTube suggest a way to remove biases towards the past in collaborative-filtering. In [Nau+19] developers at Facebook propose a way to mitigate memory constraints in deep-learning recommendation

models. In [Che+16] developers at Google propose a solution to the over-generalization problem fro users with sparse interactions, and in [Gup+13] Twitter developers succeed in reducing architectural complexity on a previously used recommendation system. All of these efforts testify to the centrality that research in recommendation systems has in companies such as Meta and Google, and they make up but a part [Hao21] of a constant process of enhancing the efficiency of recommendation systems.

In addition to technological improvement, a major player in the evolution of social networks and their recommendation systems concerns ethical issues. Over the years, platforms have been accused of promoting the diffusion of misinformation and divisive content, and of polarizing people by suggesting extremist contents. To cite some example, during the 2016 US Presidential elections false news stories outperformed real news on Facebook [Sil]. Few years after, the platform admitted to have "incite offline violence" in Myanmar in 2018 [Hao18] [Hao21] and, in 2021, it was held accountable, by a whistle-blower, for fanning ethnic violence in Ethiopia [Hao21]. To make matters worse, its own developers admitted in 2019 that they had introduced measures in the recommendation system that unintentionally favored the dissemination of divisive and violent content [MO21]. On top of the criticism about incitement to violence, both Facebook and YouTube received heavy criticism regarding the radicalization of users [Rib+20]. Already in 2016 Facebook's own researchers found that "64% of all extremist group joins [were] due to recommendation tools" [Pau21].

In response to these problems these platforms have been forced to correct, at least in part, their errors. Today, according to their official website, videos about discrimination, segregation or exclusion have been banned from YouTube [Woj]. Creators can no longer monetize videos using inappropriate language or dealing with controversial content [Ser]. Violent content and hate speech is not welcomed anymore on Facebook [Metb] and the same holds for Instagram [Meta]. Ethical concerns have thus helped shape platforms' policies, incentivized them to improve their recommendation systems, and forced them to take into account at least part of the risks they may generate.

In a nutshell, both ethical and business needs make for constant work around recommendation systems. This constant work implies constant changes in the rules underlying the dissemination of online content and makes social networks terrains of study in continuous evolution. To rephrase Heraclitus: no man ever steps in the same *platform* twice. This continuous evolution of platforms should be kept in mind when considering previous literature to better contextualize the findings and avoid improper generalizations. It also justifies the need for ongoing studies to monitor the effect of these proprietary algorithms on the dissemination of information and assess their impact in the formation of opinions in our societies.

1.3 Conclusions

In this chapter, we tried to summarize what has been studied in many disciplines regarding the dissemination of online content. We first discussed collective attention, how we can consider it a limited resource, how contents compete to earn it, and what its drivers are. In addressing these topics, we observed how the evolution of content popularity has been modeled in different disciplines, from marketing science to epidemiology. We then briefly discussed the salient features of recommendation systems, an essential vehicle of attention when discussing online dissemination. We emphasized the critical role that recommendation systems play and briefly recounted some episodes that marked and changed their implementations. We recounted the significant criticisms moved to recommendation systems along their history, criticisms that, to date, have mainly concerned the type of content that they propagate. In contrast, we should emphasize that recommendation systems are responsible not only for the type of content they diffuse, but also for shaping the dynamics of dissemination from a temporal point of view. What are the consequences of finding content that interests us more quickly? What consequences do recommendation systems have on the more or less rapid toggle of topics in the public debate? The next chapter will be devoted precisely to answering these questions, conceptually elaborating on how recommendation systems can foster over-accelerated attention dynamics and what detrimental effects can come from this acceleration.

Chapter 2

Junk News Bubbles: A Conceptual Model

Contents

As briefly mentioned at the end of the previous chapter, many of the concerns raised so far about recommender systems have been about the type of content suggested to users. Less attention has been dedicated to the risk of generating over-accelerated dynamics, whose danger relies on the ephemeral concentration of public attention. We believe that besides false contents disguised as mainstream news and explicitly directed at deceiving their receivers, media scholars should worry about the avalanche of memes, click-baits, trolling provocations and other forms of ephemeral distractions that prevent online audiences from engaging in a thoughtful public debate. This type of "information disorders" [WD17] cannot be defined on the basis of its content or its style as misinformation but relies in temporal profiles of content dissemination. If the dangers of this information disorder are neglected by current media research, it is not necessarily because they are lesser than that of misinformation [Sha+17a] and radicalization [RMC21], but because of the lack of a precise conceptualization of such phenomenon. In the following we try to make up for this lack by giving a name to this information disorder and discussing its main characteristics.

2.1 Conceptualizing Junk News Bubbles

To outline a conceptualization of this information disorder, we propose the definition of junk news bubbles: *adverse media dynamic in which a large share of public attention is captured by items that are incapable of sustaining it for a long time.* Both elements of this definition are crucial. Popular stories and even viral contents are not necessarily junk news bubbles, no matter how quickly and largely they spread in online networks. To qualify as junk news bubbles, contents must fade away as quickly as they rose, so that they distract public debate rather than nourishing it. Note that our definition is agnostic about the quality of junk contents. Even a patent piece of misinformation such as the infamous claim that "Brexit will make available 350 million pounds per week for the NHS" plastered onto a red bus during the UK-EU referendum campaign, can end up generating productive discussions if it sticks long enough in the public debate [Mar18]. Vice versa, newsworthy stories cannot contribute to democratic conversation if they are too quickly pushed out of the public agenda. In other words, the notion of junk news bubbles applies less to specific pieces of content, than to a general acceleration of online attention cycles.

Central in the '70s and '80s, the question of "attention cycles" **Downs1972** has lost steam in current media research because of the advent of digital technologies and the extension of the media system that they brought with them. Because of this extension, the question of the occupation of public debate has begun to be formulated in spatial rather than in temporal terms (i.e. where something is discussed rather than when). Temporal dynamics, however, remains crucial for, as in the words of McLuhan, "the 'message' of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs... amplif[ying] or accelerate[ing] existing processes" [McL22]. As noted by scholars working on the attention economy [Lan64] [Ter12] [CK12], digital technologies are particularly inclined to amplify "media hypes" [Vas05] and to concentrate public attention on widespread but ephemeral trends.

In this chapter, our goal is to provide a formal description of these attention dynamics in order to encourage their further empirical study. With a few remarkable exceptions (see in particular [Les10] and [LS+19]), no large-scale research has been devoted to attention cycles, despite the growing availability of traces produced by digital media [Laz+09] [Lat+12] [VJB15].

To facilitate such research, we propose a mathematical formalization of one of the most influential accounts of attention dynamics: the "public arenas model" introduced in 1988 by Stephen Hilgartner and Charles Bosk [HB88] (in the following H&B). Despite its clarity and insightfulness, H&B's framework has so far found no mathematical formalization for its complexity and lack of formal description. In this chapter, we streamline H&B's model focusing on the rise and fall of attention matters (and ignoring the linkages across different

arenas and the actors within each arena). Doing so we propose a "toy model" [RHH18], whose function is not to be applied or fitted to empirical data nor to offer an accurate description of the phenomenon that it presents, but to help in defining it and setting the conceptual bases for the further study of junk news bubbles.

2.2 Model Description in Hilgartner and Bosk Formulation

In the following we will take up the main elements of the H&B model and give them a reinterpretation from a social network perspective.

2.2.1 Population of Matters of Attention

The first ingredient of our model is a population of "matters of attention" (or "social problems" as in H&B original formulation) defined self-referentially as the entities that compete to capture public attention. The non-essential nature of this definition is crucial for H&B, who contend that "social problems are projections of collective sentiments rather than simple mirrors of objective conditions" (H&B p.54). In other words, matters of attention are defined by their visibility and not the other way around ("we define a social problem as a putative condition or situation that is labeled a problem in the arenas of public discourse and action" p.55). Three corollaries descend from this non-essentialist definition:

- First, all attention matters are equal before our model and their rise and fall depend exclusively on the competition between them and not on any substantial features ("social problems exist in relation to other social problems" p.55)
- Second, our model focuses on attention dynamics internal to media arenas, deliberately disregarding the influence of exogenous shocks. This does not mean, of course, that these shocks do not exist (clearly the breaking of a war or of an earthquake will command attention in all attention arenas). Yet, their influence is both obvious and insufficient to account for all media dynamics ("if a situation becomes defined as a social problem, it does not necessarily mean that objective conditions have worsened. Similarly, if a problem disappears from public discourse, it does not necessarily imply that the situation has improved" p.58). This is particularly true of the kind of junk news we are interested in, which may occasionally surf the drama of external events, but is more often entirely self-referential. For these reasons, exogenous shocks are deliberately excluded from our model (but empirical applications should, of course, control for them).

- Third and similarly to H&B framework, our model can be applied to different media and at different scales. Attentions matters are broadly defined as recognizable units of content in a particular forum of collective debate (the attention arena). Examples could be different videos in a given YouTube channel or different threads in a given Reddit subreddit. To be sure, we are not promising that our model will fit all media debate but inviting scholars to test it empirically on different phenomena to determine to which it can be fruitfully applied.

2.2.2 Competition Mechanisms

The second ingredient of our model are two competition mechanisms that favor some attention matters over others. The four different "principles of selection" distinguished by H&B find in our model a formalization in two main mechanisms:

- *Exogenous influences.* Three of the four "principles of selection" distinguished by H&B, "drama" (pp.61-62), "culture and politics" (H&B p.64) and "organizational characteristics" (pp. 65,66) are rendered in a deliberately coarse way in our model. The dramatic value of attention matters as well as the way in which they resonate with the general culture or with the specific organization of the medium are important, but their influence falls outside the self-induced media dynamics that constitute the focus of our model. In our formalization, the influence of these features is thus rendered as a noise which randomly increases or decreases the visibility of each item at each iteration. This solution allows for accounting for this type of influence (and to explore the effect of its variation) under the assumption that its specific nature does not affect the dynamic of junk news bubbles.

- *Endogenous trending.* The last selection principle identified by H&B, "novelty and saturation", is crucial to our model. At each iteration, the model increases or decreases the visibility of each matter, repeating its previous variation, multiplied by a parameter that accelerates or decelerates such variation. The model therefore rewards rising items and penalizes declining ones. This mechanism works as a Matthew effect [Mer68] [New01], but a dynamic one which rewards not the most visible matters, but the ones that have increased the most since the previous iteration. This boosting of trendiness is consistent with the way in which online platforms "emphasiz[e] novelty and timeliness... [by] identifying unprecedented surges of activity" and "reward[ing] popularity with visibility" (Gillespie, 2016, p.55&60). Such partiality for trendiness is characteristic of both social media and their users, in a sociotechnical loop in which the visibility granted by platform algorithms both depends on and is influenced by the number of views generated by different contents.

2.2.3 Attention Boundaries

The third ingredient of our model are the attention boundaries. At each iteration, after adding (or subtracting) to each attention matter its random variation and its trending acceleration, the model corrects the potential visibility of each item to make sure that it remains within two inflexible boundaries:

- *Lower boundary*: exclusion of negative visibility. Because it is impossible to conceptualize such a thing as negative attention, when noise or acceleration push the visibility of a matter of attention below zero, the item is removed from the arena and replaced with a new one with null initial visibility. Because a new attention matter can enter the arena only when an old one leaves it, the number of items in the model remains fixed (but some items can have visibility equal to zero).

- *Upper boundary*: saturation of the attention capacity. After having applied noise and acceleration and corrected for negative attention, the model divides the potential visibility of each item by the sum of the potential visibilities of all items. This normalization makes sure that the sum of all computed visibilities remains equal to one. This boundary implements a key ingredient of H&B framework, the idea that each debate arena has a fixed attention capacity (or "carrying capacity", in H&B terms). The fixity of the global "carrying capacity" is crucial to ensure that our model does not converge to a trivial winner-takes-all equilibrium. While raising attention matters are pushed to an increasing visibility by their trendiness, they all end up reaching a point where they exhausted their potential for growth, begin to slow down and are penalized by competition mechanisms. The inelasticity of attention capacity also ensures that the visibility gained by one matter of attention is always lost by some other so that "the ascendance of one social problem will... be accompanied by the decline of one or more others" (H&B p.61). While we are, of course, aware that public attention fluctuates with circadian and professional rhythms, we believe that these cyclical fluctuations can be discounted for the sake of simplicity. Following H&B, we think that good reasons for a fixed attention capacity can be found in the limited staging capacity of media ("the prime space and prime time for presenting problems publicly are quite limited" p.59) and, more importantly, in the limited capacity of the public to attend to public ("members of the public are limited not only by the amount of time and money they can devote to social issues, but also by the amount of 'surplus compassion' they can muster for causes beyond the usual immediate concerns" p.59). This second element is crucial to understand why the assumption of a limited carrying capacity remains relevant for online media even if digital technologies removed most of the barriers of conventional news gatekeeping [SV09a].

2.3 Model Formalization

In the following we give a mathematical formalization of the model described so far, giving each of the variables a name and translating the competing mechanisms into relationships among them.

2.3.1 Variables Definition and Dynamics

1. We call x_i each item of our population of matters of attention, with $i = 1, \ldots, n$, where n is the maximum number of items in the population. We call "visibility" or π_i^t the share of attention captured by x_i at time t. By a mechanism explained below, at each timestep, the sum of π_i^t for all i is fixed and equal to one. This allows to interpret each π_i^t as the percentage of the total attention captured by each item i at time t.

2. We model the two competition mechanisms as follows:

   - *Endogenous trending.* At every timestep $t + 1$, the visibility π_i^t of each item i is modified by adding to its current visibility π_i^t a term which repeats its previous variation (i.e. π_i^{t-1} - π_i^t) multiplied by a positive factor α, which could be interpreted as a boost of trendiness.
   - *Exogenous influences.* In our formalization, we render all external influences on media dynamics as a noise ε_i^t which increases or decreases the visibility of item i randomly at timestep t. The noise ε_i^t is a realization of a normal distribution with mean equal to zeros and standard deviation $\sigma = \sqrt{c}n$, where c is a positive parameter. We can therefore write the potential visibility of each item after the iteration p_i^{t+1} as the output of the two above mechanisms as follows:

   $$p_i^{t+1} = \pi_i^t + \alpha(\pi_i^t - \pi_i^{t-1}) + \varepsilon_i^t$$

3. At each iteration t, the potential visibility p_i^{t+1} is replaced with its corrected version $\hat{p}_i^{t+1}$ to abide by the model's attention boundaries:

   - *Exclusion of negative visibility.* $\hat{p}_i^{t+1}$ equals p_i^{t+1} if p_i^{t+1} is positive. Otherwise it is set to zero. Hence,

   $$\hat{p}_i^{t+1} = \max(0, \hat{p}_i^{t+1})$$

   - *Saturation of the attention capacity.* The limited capacity of an arena is represented by the constraint of having a fixed sum of popularities at each timestep. Therefore, each visibility is obtained from the non-negative $\hat{p}_i^{t+1}$ by normalization.

   $$\pi_i^{t+1} = \frac{\hat{p}_i^{t+1}}{\sum_j \hat{p}_j^{t+1}}$$

2.3.2 Initialization

At the first step of the model, the visibility of every i (i.e. π_i^1) is initialized with a random numbers drawn from a uniform distribution between 0 and 1 and normalized to satisfy the constraint $\sum_i \pi_i^1 = 1$. At the second step, the visibility every i (i.e. π_i^2) is obtained by adding to π_i^1 a noise ε_i^t drawn from the normal distribution $N(0, \frac{1}{cn^2})$ and normalizing. After the first two steps, the dynamics is self-sustained by evaluating equations (1), (2) and (3) at each iteration.

2.3.3 Parameters Interpretation

Inspecting the equations above, it is easy to observe that our model has only three parameters:

- α, the *trendiness boost*, which decides whether the visibility variation at the previous iteration is amplified at the next one and by how much, is the key parameter of our model. Conceptually, α can be interpreted as the keenness of media algorithms and media users to identify and promote trendy matters of attention. The bigger is α, the more important is the role played by trendiness in the sociotechnical choices that influence the visibility of media items. High values of trendiness boost thus simulate the attention dynamics occurring in debate arenas prone to junk news bubbles.
- n, which represent the maximum number of attention matters simultaneously present in the simulation,
- c, which represent the size of noise, that is to say the importance of exogenous influences.

Both n and c are used in the realization of noise and, because they appear in the denominator of the distribution that generates noise, the higher they are, the smaller are the variations due to noise.

2.4 Model Results and Discussion

Despite its simplicity, our model is able to generate patterns comparable with the empirical observations of media systems [Les10] [LS+19]. In particular, our formalization supports the H&B intuition that the "shifting waves of social problems" (H&B p.67) typical of media attention cycle can be explained by the interaction between the push of trendiness and

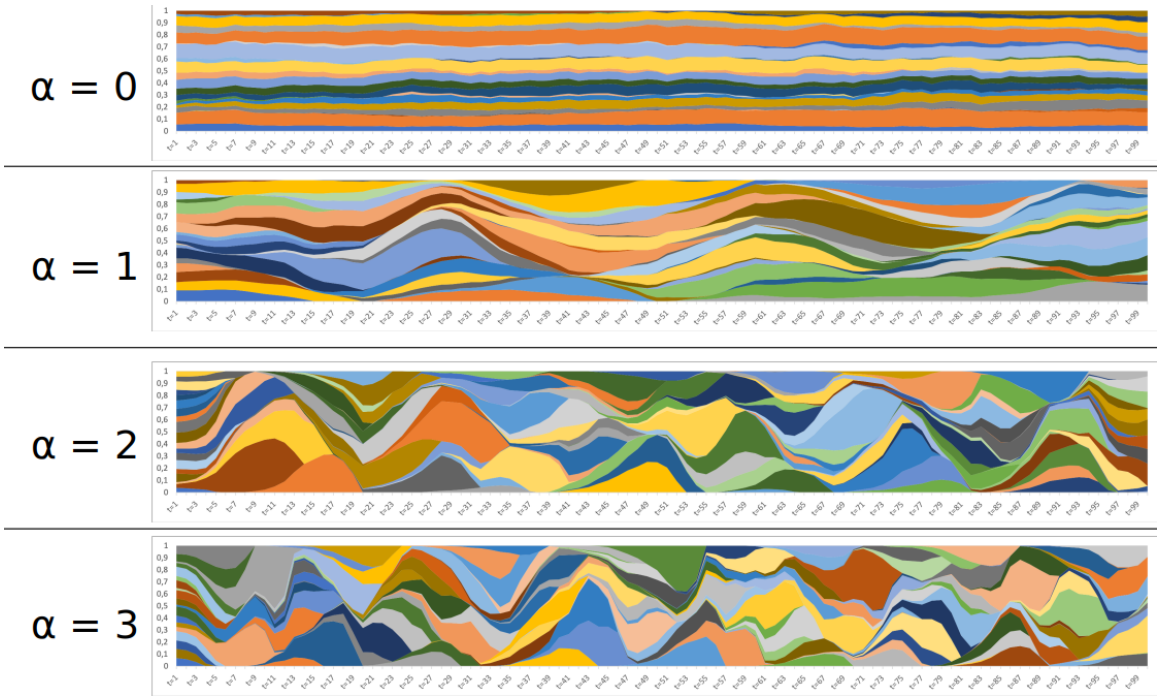


Figure 2.1: Evolution of the Junk News Model for trendiness boosts = 0, 1, 2, and 3 (with $N = 20$ and $c = 12$). Each color area corresponds to the attention received by an item. The first 100 iterations are shown.

the saturation of the carrying capacity: "if we explore these complex linkages, we find a huge number of positive feedback loops, 'engines', that drive the growth of particular problems. Growth is constrained, however, by the negative feedback produced by the finite carrying capacities of the public arenas, by competition among problems for attention, and by the need for continuous novel drama to sustain growth" (H&B p.67). Previous studies [Wen+12] [Gon+10] [CLP07] considered the role of users' limited attention in media competition assigning users a maximal number of possible interactions (a sort of Dunbar number for individual attention). In most of these models, the fall of popularity is obtained forcing an aging process of media items through an explicit time decay term. This aging process, however, is difficult to justify theoretically and empirically. One of the most original aspects of our model is that it dispenses with this aging process: items' popularity decays naturally through the interplay between the trendiness and the saturation of the overall attention capacity.

The comparison between the graphs in Figure 2.1 suggests that, as the boost of trendiness grows, the rise and fall of attention matters becomes steeper. This relation can be tested by computing the mean steepness of attention curves (the absolute increase or decrease by

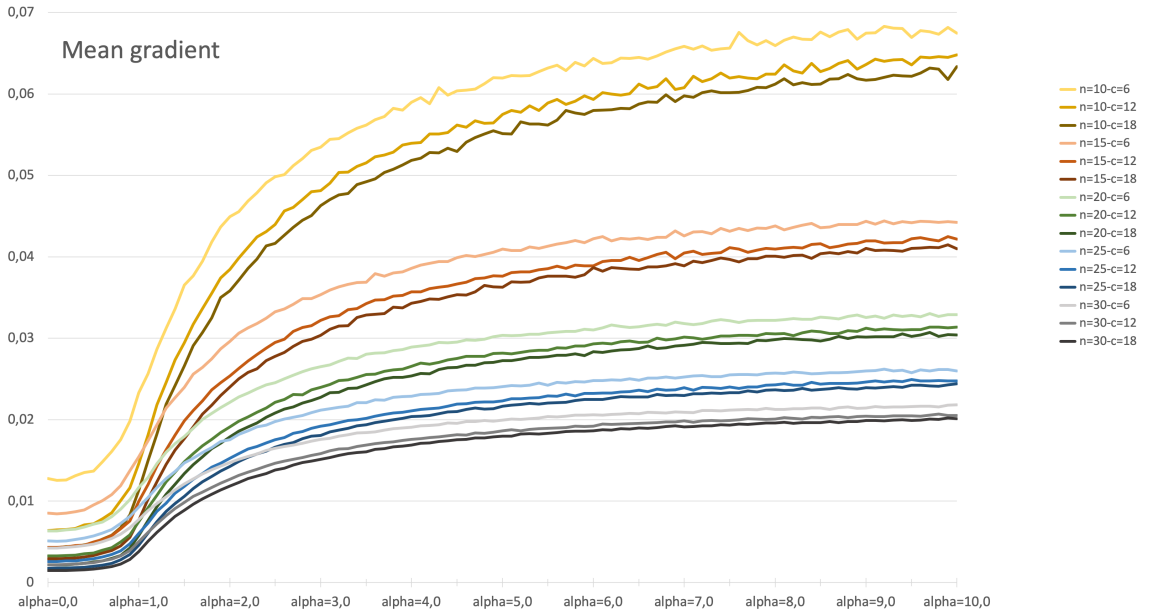
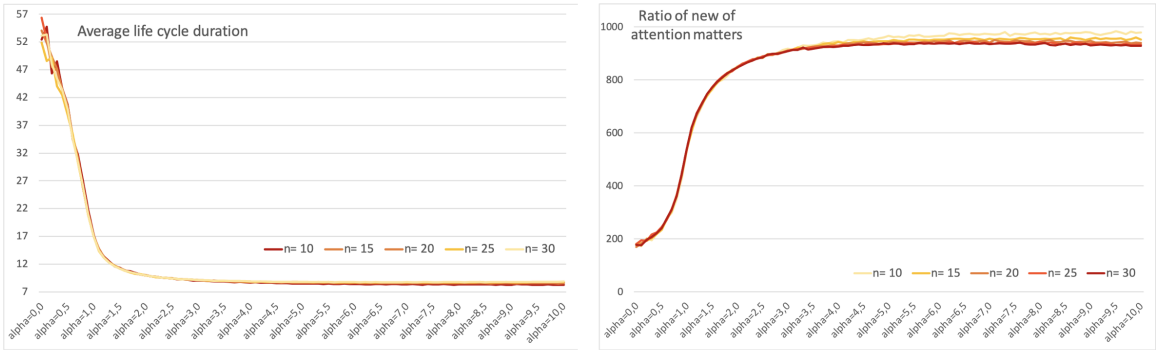


Figure 2.2: Mean increment of attention curves as function of the trendiness boost (for different values of n and c)

unit of time) and observing that it increases monotonically with the increase of alpha before reaching a plateau (probably due to the upper and lower constraints on the state and to the impossibility of compressing the width of curve beyond a certain point).

Figure 2.2 confirms that the relation between the steepness of attention curve and trendiness boost is not substantially transformed by the other parameters of our model. The number of attention matters and the importance of exogenous influences shift the position of the curve, but do not change its shape. Also, because both n and c affect the curve in the same way, only n will be explored in the next figures. Considering together Figure 2.1 and 2.2, it is also interesting to notice that trendiness boost increases rise- and-fall steepness by affecting both dimensions of the media cycle: the height of attention curves and their width. This suggests that junk news bubbles can combine features that may appear contradictory.

- Regardless of the number of items or the level of noise, the stronger is trendiness boost the shorter is the lifecycle of individual attention matters (Figure 2.3(a)). Remarkably, this is true for all attention matters: even items that reach very high levels of visibility end up falling as quickly as they rose. As a consequence of the shortening of attention weaves, a higher number of matters enter and exit the arena (Figure 2.3(b)). This may contribute to making platforms more attractive by increasing the dynamism of



a. b.

Figure 2.3: (a) Mean length of attention matters' life cycle and (b) ratio of new attention matters entering the model in its first 10.000 iterations, at the variation of trendiness boost (for different values of n and with c set to 12)

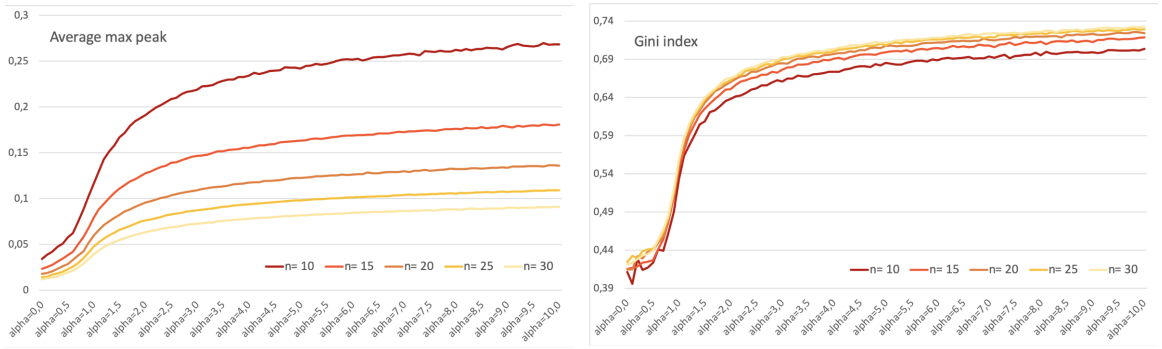
their offer of information and entertainment.

- On the other hand, higher trendiness boost increases the maximum visibility reached by attention matters (Figure 2.4(a)) and, most importantly, amplifies the difference between successful and unsuccessful attention matters, creating a situation in which, at each iteration, most of the available attention is captured by a minority of over-visible items (Figure 2.4(b)). "There is a huge 'population' of potential problems-putative situations and conditions that could be conceived of as problems. This population, however, is highly stratified. An extremely small fraction grows into social problems with 'celebrity' status... [while] the vast majority of these putative conditions remain outside or on the extreme edge of public consciousness" (H&B p. 57).

2.5 Conclusions

Taken alone, none of the consequences of junk news bubbles highlighted by our model is particularly surprising: being an acceleration, trendiness predictably shortens the lifespan of attention matters and, being a positive feedback, it increases their maximum visibility. Their combination, however, is remarkable as it creates a shoaling of attention waves which reduces the width and increases the height of attention curves. Debate arenas characterized by stronger trendiness may therefore end up displaying a syncopated rhythm of attention that is at the same time increasingly fast and increasingly concentrated. [1] Junk news bubbles

[1] Empirical evidence on the coupling of speed of consumption and concentration of collective attention can be found in the paper this chapter is based on [Cas+22]. Here we leave the empirical discussion aside



a. b.

Figure 2.4: (a) Mean height of the attention curve peaks and (b) Gini index of attention concentration at each iteration of the model, at the variation of trendiness boost (for different values of n and with c set to 12)

are characterized by the same attention skewness of Boydstun's "media storms" [BHW14], but not by the same persistence in time. As such, they are particularly worrying because they take attention away from other discussions (because of their skewness), without producing the heightened public awareness created by media storm (because of their ephemerality). While evidence exists that the syncopation described by our model can be found in numerous online platforms [WH07] [CS08] [YL11] [Cas+14] [BSH15], little empirical research has been carried out on the consequences of such attention regime. The risks of distraction related to screens and online media have been decried at the individual and cultural level (cf. for instance, [Gol97], [Has11], [Cit14]), but hardly investigated through the records increasingly made available by digital platforms themselves. This Chapter hopes to facilitate such line research by providing a formal description of a distracted attention regime: a situation in which attention waves becomes both higher and narrower and in which public debate is trapped in a continual succession of hot button issues.

Such a situation, arguably, is not particularly propitious to quality. While our model defines junk news bubbles independently from their content value, we suspect this attention regime to be associated with misinformation and poor quality. A regime in which visibility is granted and withdrawn with great rapidity unsurprisingly favours click-baity content designed to catch the attention more than to retain it. This observation may explain why, in political discourse, traditional propaganda is increasingly replaced by political trolling aimed at drowning opponents' discourse in noise [FSCS18] or simply to monetize political outrage [BE19]. Being a simplified formalization of a relatively abstract framework, our mathematical model does not allow substantial claims about actual attention dynamics. It allows, however, to advance a precise hypothesis about the junk news bubbles and their

to focus on the conceptualization junk news bubbles.

detrimental effects on public debate: the fascination with trendiness of digital platforms and their users may create an over-accelerated public debate in which a disproportionate share of media attention is captured by matters which are incapable to sustain it. As the shoaling of sea waves is associated with the entering in shallower waters, so junk news bubbles may be associated with a shallower public debate, a risk that raises concerns and normative implications different from those associated with misinformation. While the latter can be (and has been) addressed by tweaking the recommendation algorithms to favour mainstream sources of information, this solution does not necessarily solve the problem highlighted in this chapter.

Describing an attention regime that is increasingly pervasive in online media, junk news bubbles cannot be fought by censoring specific content or specific sources, but demands a deep restructuring of the system of incentives that characterize digital communication. Until social media will obtain the largest share of their profits from selling metrics of shallow and ephemeral engagement (e.g. impressions, views, clicks and shares) and until their users will be rewarded according to the same metrics, little are the chances to avoid dynamics of hyper-acceleration. This does not mean that all content producers will play the game of click bait and junk news – think of the many amazing works produced on YouTube by both mainstream and native creators [SV09b] [BG10] – nor that online platforms can only promote superficial forms of engagement – think of how Twitter has been invested by all sorts of political activists [Ger12]. It does mean, however, that in the old opposition between a distracted public opinion [Lip22] [Lip27] and engaged public inquiry [Dew27], accelerated attention regimes stack the odds against the latter.

Chapter 3

YouTube and Data Collection

Contents

In the previous chapter, we provided a conceptualization of over-accelerated collective attention regimes in which attention waves become higher and narrower and in which public debate is trapped in a continuous succession of hot-button issues. We have discussed how this ephemeral concentration of media attention can prevent the public from digesting successive public issues, reducing the quality of public debate. The conceptualization of junk news bubbles leads to a reconsideration of the importance of the temporal aspects of content dissemination. It gives rise to the need to collect large-scale data to investigate what form collective attention regimes take in the real world.

To meet this need, over the past three years we have collected a large temporal dataset on YouTube. There are many reasons behind the choice of this platform. First of all, YouTube constitutes the second largest social network in the world after Facebook. It has 2.1 billion monthly active users worldwide i.e., 1 in 4 people in the world accesses YouTube at least once a month. For many of these users, over the years YouTube has become an important source to get news. A 2020 survey [Sto+20] reports that about a quarter (26%) of U.S. adults get their news on the platform, either through established news organization channels (like CNN or Fox News) or through independent channels native to the platform.

These numbers give us an idea of the impact YouTube can have in the formation of opinions in a society and how influential it can be in the democratic process. Despite YouTube's reach in terms of users and its importance in disseminating news, the scientific community has paid little attention to it so far compared to other social networks. The reasons behind this lack of studies in the literature can be sought in the difficulty of analyzing video content (compared, for example, to text content, such as posts on Facebook) and also in the difficulty of obtaining data, made hardly accessible by the platform.

In the following we will first give some context information about YouTube, its history and the controversies that affected it. In a first section we recount the main stages of the platform's evolution and the challenges it has faced over the years. Then we analyze the heart of how YouTube works: its recommendation system, or the set of algorithms that suggests personalized content to users based on their interests. We then discuss, in Section 3.2, the constraints that have prevented and restrained scientific studies on YouTube: we present the official tools, i.e., the application programming interface (API), through which data can be collected, and we address its limitations and how these have increased over the past decade or so. We then present other data collection possibilities that attempt to overcome the limitations imposed by the official API and that provide a real alternative for giving YouTube the attention it deserves in scientific research: in particular, we present several methods of scraping from the source code by which YouTube web pages are programmed and discuss the limitations of this approach. In a final section we introduce the data we have collected thanks to these very methods. On this data, unprecedented in the history of YouTube, we will base the empirical results presented in the rest of the manuscript.

3.1 YouTube as a Platform

3.1.1 History and Controversies

YouTube was born on February 14, 2005, the brainchild of Chad Hurley, Jawed Karim, and Steve Chen, three former Paypal employees. The service was designed with a completely different purpose than it has today, that of a dating platform where users could introduce themselves through video content. Soon, however, partly because of the lack of users willing to make such use of it, the developers realized that the strength of their creation laid precisely in providing a space for uploading video content, no matter what kind [Dre16]. Abandoning the idea of a dating platform, YouTube became the largest platform on which to upload amateur videos, thanks in part to the absence of major competitors other than Vimeo, which remained, however, a project poorly followed by its developers. It grew at an unprecedented rate, so much so that only a year after its creation Google decided to buy it for a value of 1.65 billion [SP06]. YouTube soon became the second most used social network

in the world, after Facebook. It grew from 6 hours of new videos uploaded per minute in 2007 to 300 hours of videos uploaded per minute in 2014. The number of monthly active users grew from 20 million in 2006 to 1.3 billion in 2014 and now stands at 2.1 billion.

YouTube, as well as other incredibly successful social networks that emerged in the same years, was born out of the work of developers who perceived it as a technology provided to the general public to use in the way they preferred. The creators of YouTube, as well as those of Facebook in the same years, saw themselves as providers of an online technology; their responsibility was to make operations such as video uploading, sharing, and dissemination work, certainly not to filter the content on the platform. YouTube was in charge of the container, not the content. It was a tech company, certainly not a media company. But things did not take long to change. YouTube was, and still is, a host for all kinds of content: from entertainment videos to do-it-yourself tutorials, from music videos to opinion discussions about news and politics. Especially the presence of the latter content and the fact that anyone with internet access had the ability to upload videos online did not hesitate to raise issues: soon both developers and the public opinion were faced with the danger coming from fake news, hate speech, and any other kind of information manipulation that could be spread through YouTube.

Central to this transition toward a greater focus on the type of content YouTube hosted was certainly the 2016 U.S. presidential election. Although the major scandals regarding the amount of misinformation disseminated online on that occasion mainly concerned Facebook, this period profoundly marked the history of all social networks because it gave birth to a collective awareness of the role that social media plays in shaping individual opinions and the consequent responsibilities that come with this role. Beginning in 2016, for the first time in its history, Facebook began to intervene in the content it hosted, accepting to identify itself no longer with the definition of *tech company* but rather with that of *nontraditional media* [NR20].

As for YouTube, the line between freedom and regulation of content has always been kept rather opaque. A rich, though not exhaustive, history of the controversies faced by YouTube regarding disinformation was compiled by Bloomberg in 2019 [Ber19], thanks to multiple interviews done with YouTube employees. According to the report, even YouTube would not have been spared from the disinformation campaigns that characterized the 2016 election. In that year, some of the platform's employees studied which news channels were the most followed during the election campaign: channels such as Breitbart News and Infowars, known for their provocative and outrageous nature, dominated information on the platform. The problem, although raised by these employees (who remained anonymous) was not considered a priority. It had to wait until the following year for the platform to begin to address the prevention of the spread of dangerous content (both from the standpoint of incitement to hate and the spread of disinformation). A first attempt was made in 2017, through the

introduction of a metric called "responsibility"[BS19], which was aiming to assess the quality of time spent on the platform through rating questionnaires proposed to users at the end of videos. Unfortunately, as with many other policies, YouTube not only does not explain how it works, but also does not provide the data to be able to analyze the effectiveness and impact of these changes.

Along with the introduction of this new metric, in November 2017 YouTube carried out what many youtubers have called "The Purge"[Mon17]: all of a sudden millions of videos were removed from the platform or demonetized (i.e., deprived of the ability to place advertisements and thus make money from their content). This swift and decisive action was in response to a media scandal about disturbing videos aimed at or about children. The worst episode concerned a channel called "Toy Freaks" in which a father posted videos of his daughters in situations of extreme pain. The main problem arose as this channel was by no means a niche channel, but YouTube itself suggested it and reported it among the 100 most watched on the platform. In response to this and other scandalous children's content, companies like Adidas, Mars, and Hewlett-Packard in March 2017 had stopped advertising on YouTube, boycotting the platform, to prevent their advertisements from being associated with such videos. Despite the attempt to purge the platform made by CEO Susan Wojcicki to regain the trust of advertisers, such incidents, which proved the platform's inability to handle violent/controversial/misinformation content, did not hesitate to reoccur. Another rather scandalous example was the dissemination on the "Trending Videos" pages of an anti-gun control conspiracist video that claimed that the mass shooting at a school in Parkland, Florida, never happened and was instead staged.

This was the time when the CEO proposed the addition of a small box still present below the videos (see Figure 3.2) in which the source was discussed and funding was made explicit. Today, for example, if we watch a video posted by Al Jazeera, YouTube suggests the link to the Wikipedia page about the Quatar media outlet and highlights the fact that the channel receives funding from the Qatari government [BS19].

Shortly thereafter, in 2019, YouTube again claimed to have made further strides in combating disinformation [Dwo19] [Ben19]. According to them on their own blog [The19]: "In January, we piloted an update of our systems in the U.S. to limit recommendations of borderline content and harmful misinformation, such as videos promoting a phony miracle cure for a serious illness, or claiming the earth is flat. We're looking to bring this updated system to more countries by the end of 2019. Thanks to this change, watch time that this type of content gets from recommendations has dropped by over 50 percent in the U.S.".

In recent years, YouTube has shown much greater responsiveness to misinformation problems related to specific events and has acted much more quickly than in the past. For example, after the appearance of Covid, they have introduced rules that "prohibit content

that spreads misinformation in the medical field by placing itself at odds with information provided on COVID-19 by local health authorities or the World Health Organization (WHO)." Likewise, at the outbreak of the war in Ukraine YouTube blocked the access around the world to channels associated with Russian state-funded media, like Russia Today and Sputnik.

In general, what platform's managers are nowadays committed to doing, according to the official web-site [The], can be summarized as follows:

1. YouTube implements measures aimed at countering political interference during elections;
2. YouTube establishes rules against identity theft or the use of deep fakes to manipulate the diffusion of information;
3. YouTube enacts hate speech norms that ban content claiming that well-documented, major violent events never happened;

Despite the platform's commitment on its official pages [The], it is difficult to quantify the effectiveness or extent of these measures. For example: what is considered political interference? What, in the case of item 3, is considered a well-documented event? To these questions, as to many others, YouTube does not answer openly. These policies, likely implemented through a complex mix of artificial intelligence and human moderators, are not fully disclosed and so it is difficult for researchers to verify their effectiveness or fairness. Complicating the situation are significant limitations on researchers' access to data, which make it very hard to study the impact of the platform's choices on misinformation, as we will discuss in the next sections.

3.1.2 Recommendation System: Functioning and Controversies

On YouTube, to date, 500 hours of videos are uploaded every minute [Cec20]. To keep users from getting lost in this jungle of content and risking soon desisting from looking for what interests them, YouTube has a solution: it suggests personalized content. Personalization is everywhere on the platform: videos on the home page are carefully selected to capture our attention, suggested videos next to a video we just watched are designed to meet our tastes. Bearing in mind that the platform's main business is to sell advertising, and to do that it needs to capture the attention of Internet users, it therefore becomes evident that the mechanisms used to recommend content are of central importance to YouTube's business model. On YouTube, recommendations are made through various *recommendation systems*, artificial intelligences programmed to predict what each user might be interested

in at a given time. Understanding how recommendation systems work is of paramount importance: in fact, it has been reported repeatedly [Zho+16][ZKG10b] by researchers how recommendation systems are the primary audience aggregator for the platform, i.e., they are the primary cause why a user ends up viewing a certain piece of content. That is why, in the following, we look in more detail at how they are implemented and discuss some of the controversies surrounding them.

Functioning

As the reader might expect, the recommendation systems used by the platform are proprietary and we don't have much information about them. What we do know comes mainly from two articles published by the platform's developers: the first "Deep Neural Networks for YouTube Recommendations"[CAS16] was published in 2016 while the second "Recommending what video to watch next: a multitask ranking system"[Zha+19] dates back to 2019. These articles, although fundamental to understanding how personalization works, remain rather vague on some key aspects of the implementation. In addition, given the speed with which recommendation systems are updated, the two papers are rather dated and thus may differ from current recommendation systems in some aspects. Despite this, in the following we try to extract the main components underlying these recommendation systems that should be retained to understand the dynamics of attention on the platform.

YouTube recommendation systems are deep learning algorithms, built on Google Brain [Dea+12], a library know by developers as its open source version, Tensor Flow [AIM17]. In general, recommendation consists of two phases: a first phase of candidate selection and a second phase in which candidates are sorted and then proposed to the end user. In the candidate generation phase, the huge YouTube corpus is narrowed down to hundreds of videos. To do this, YouTube builds a classifier to predict which video w_t will be watched by the user at time t. The video then becomes a class, a label to be assigned to the user.

Regarding the architecture of the classifier, a schematic can be found in Figure 3.1. Obviously the structure is very complex, and the interested reader can find more information in [CAS16]. To get a general idea: several features are concatenated into a large first layer, followed by several layers of fully connected Rectified Linear Units (ReLU) [GBB11]. We do not want to dwell on the complexity of the structure and implementation of the recommender system. In fact, these are deep learning algorithms, i.e., black boxes of which it is very difficult if not impossible to understand the reasons that lead to certain outputs. What we are most interested in is not understanding *how* the algorithm chooses but *from* what data: so we want to better understand what the input features are. According to what is discussed in this article, the features on which the recommendation system makes its choices are as follows:

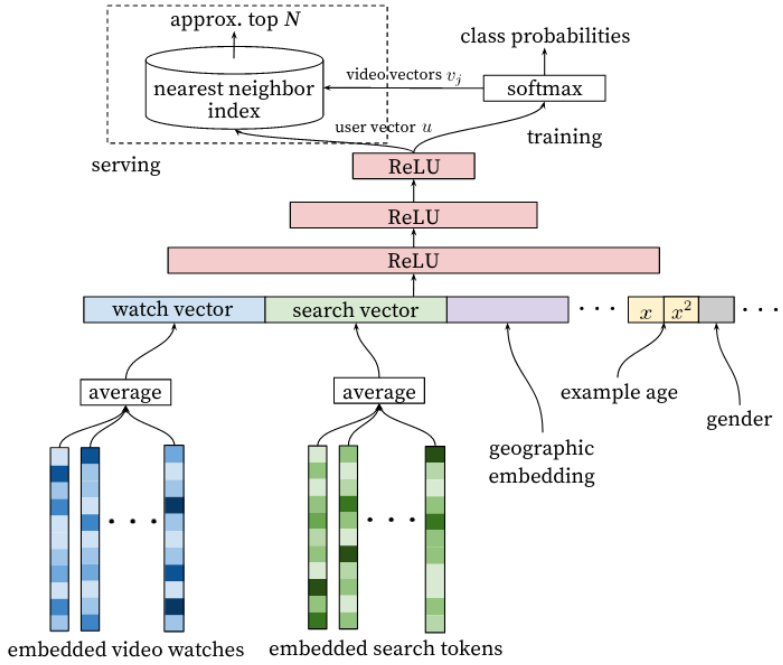


Figure 3.1: Deep candidate generation model architecture for YouTube recommendations, as pictured in [CAS16]

- the *novelty* of the content;
- the similarity between users;
- previous popularity;

As for *novelty*, in the article [CAS16], Covington et al. devote an entire paragraph to a feature they call "Example Age." As they explain, "recommending recently uploaded ("fresh") content is extremely important for YouTube as a product. We consistently observe that users prefer fresh content, though not at the expense of relevance." To add this variable to the recommendation system, they add the age of the training set during the training. At serving time, they set the eta variable to zero so that the rater knows to make predictions at the end of the ti time period covered by the training.

Another very important feature (rather a set of features) is user *similarity*. In fact, before being passed as input to the classifier, users and content are mapped together in a dense dimensional N-space where the distance rappresents the similarity and affinity between two objects (be they videos or users). The user then becomes a point in this space and these

coordinates are passed to the classifier as input. This embedding is learned together with all the other parameters of the classifier. The mapping is done from the history of videos watched and searches made on the YouTube search engine. Somewhat like it is done for word embedding, starting from a user's views history (where each video ID is considered a word), users and videos are mapped to a space where they are more or less close based on how similar they are. Taking into account the history of other users to suggest content to a specific user falls under the concept of "collaborative filtering." This concept is very important and will be discussed in the next chapter, when we will try to give weight in content dissemination to the different features taken as input by the recommendation system

Last but not least, the previous popularity appears. In fact, the developers affirm the "In addition to the first-order effect of simply recommending new videos that users want to watch, there is a critical secondary phenomenon of boot-strapping and propagating viral content." In general, this aspect is quite familiar to researchers: rich gets richer type dynamics have been repeatedly observed in social networks that featured similar recommender systems.

Now that we have seen what the features of YouTube's recommendation system are, we can account for some of the criticisms that have been raised against it over the years, some of which have yet to be clearly answered.

Controversies

To understand how delicate the role of the recommendation is system we devote this section to one of the most heated and still debated controversies about YouTube, a controversy that concerned its recommendation system directly and has seen it cited as "one of the most powerful radicalizing tools of the 21st century." In a 2018 article published in the New York Times, Zeynep Tufekci claims that, starting watching videos of Donald Tramp rallies during the 2016 election campaign, she was guided by the recommendation algorithm (through autoplay) toward supremacist, holocaust-denying, and conspiracy content.

After this episode, the journalist began to create other channels to see if this was a phenomenon related only to the far right. On a new channel, after watching videos of Berni Sanders and Hillary Clinton, the algorithm steered her toward leftish conspiratorial videos, which, for example, claimed that the U.S. government was behind the attack on the Twin Towers. She then tried to create new channels and create a history of videos watched that were not necessarily political: videos about vegetarianism led to videos about vegans. Videos about jogging led to videos about ultra-marathons.

As Tufekci rightly argues, this is not to say that there is a willingness on YouTube's part to extreme the audience. Rather, it means that the recommendation algorithm, which as

we have just seen is trained to get people to stay online as much as possible, has learned that increasingly extreme content intrigues us and is the most likely to keep us glued to the screen a little longer. In her words, "YouTube leads viewers down a rabbit hole of extremism, while Google racks up the ad sales." To support her argument, the reporter refers to an investigation conducted by the Wall Street Journal with the help of a developer of YouTube's recommendation system [LM18]. In the study it was stated that the recommendation algorithm often "fed far-right or far-left videos to users who watched relatively mainstream news sources" and that this was true for a great many topics.

Since the publication of this NY Times article, a number of studies have been pursued by the scientific community to test whether there are indeed patterns of radicalization incentivized by the suggestions. However, mainly due to a lack of representative data, the soundness of these studies is questionable and, in any case, the results are mixed and do not allow for a clear conclusion on the matter.

In particular, two studies were published in 2020 that sought to investigate the potentially radicalizing role of the recommendation system: "Auditing Radicalization Pathways on YouTube" by Ribeiro et al. [Rib+20] and "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization" by Ledwich and Zaitsev [LZ16] the first study focuses particularly on right-wing content, from the more moderate Alt-lite and Intellectual Dark Web (I.D.W.) to the extreme Alt-right. It claims to observe significant migrations of users to increasingly extreme content by pointing out that a large percentage of users who consume Alt-right content today once consumed Alt-lite or I.D.W. content. On the opposite side we have the study carried out by Ledwich and Zaitsev, where instead the authors refute the popular radicalization claims. According to their analysis, YouTube's recommendation algorithm actively discourages users from visiting radicalizing or extreme content. On the contrary, according to them, the algorithm favors mainstream media over independent YouTube channels in its suggestions.

The differences in the results of these two studies can be explained both (1) in terms of the channels studied and (2) in terms of the type of data returned by the YouTube API. For example, in terms of the differences among channels (1), the first study considers about 300 channels, some of which are very small and niche. The second study, on the other hand, considers 800 channels that have at least 10 thousand subscribers. To this difference in terms of type and quantity of channels, it should be added the fact that the data used in the second study are probably unrealistic (2) : they are officially provided by YouTube (through its API, which we will discuss later) but do not concern real suggestions made to real users. The suggestions studied in [LZ16] are suggestions made to a user with no history, about whom we have no information. These are "wide-ranging" suggestions, much more generic than those actually made to users on the platform. This absence of real personalization in the data collected weakens, in our opinion, the results obtained by Ledwich and Zaitsev.

In contrast, the results obtained by Robeiro et al. are based in part on paths of real users who moved from certain types of channels to others. Hence the absence of personalization in the data provided by YouTube could therefore be another cause of the discrepancy in the results of these contemporary studies.

Thus, it becomes evident how the provision of data is crucial to be able to do analyses and audits on the role of the recommendation system in the dissemination of online content. For this reason, we believe that it is important to devote specific space to discussing the possibilities that YouTube provides for data collection and the stringent limitations it imposes.

3.2 Collecting YouTube Data: Challenges and Tools

3.2.1 YouTube's API

An API (Application Programming Interface) is a set of procedures that enables communication between different software or different components of a software. If, for example, a programmer, in developing his new app, wants to integrate YouTube content and give his users the ability to comment on or like videos within his app, then he would have to use the YouTube API. In general, the API is the official tool for exchanging information with the platform, allowing developers to download and upload information to YouTube. For this reason, APIs are widely used by researchers to download data from social networks. There are, however, two important limitations that are introduced by APIs, and in particular by YouTube API:

1. only certain types of requests can be made;
2. the amount of requests each user can make is limited.

In the following we illustrate what types of requests can be made and what the limits are on their daily amounts on YouTube.

What can we ask?

The type of information that can be requested from the API changes considerably whether we want to collect information on third-party channels or whether we want to request information on a channel we own. In general, as you might expect, the interest of researchers is to collect data on many third-party owned channels, so in the following we will focus on

these types of requests. Regarding third-party channels, a good rule of thumb is that the platform only returns publicly visible information. For example, it returns the number of views, comments, and commenters on a video, or it returns, for instance, a channel description and the (rounded) number of subscribers. What it does not return is information that is not displayed on the channel's or video's web page: for example, the names of people who subscribed to a channel or watched a video, the time users spend watching a video or how they ended up on that content. Some examples of data that can be queried to the API are shown in Figure 3.2.

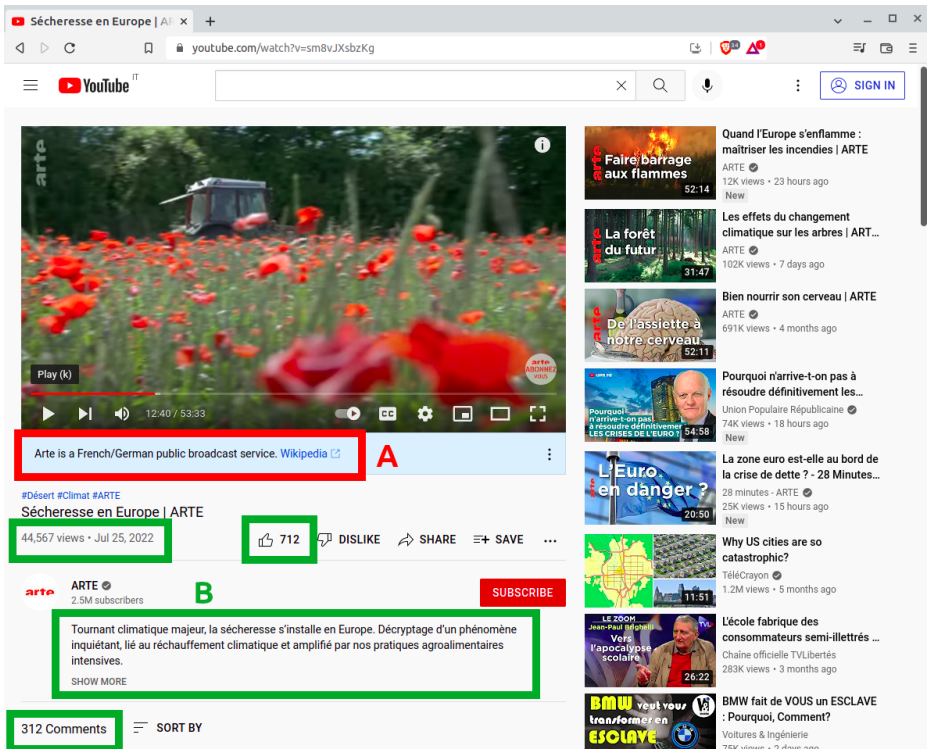
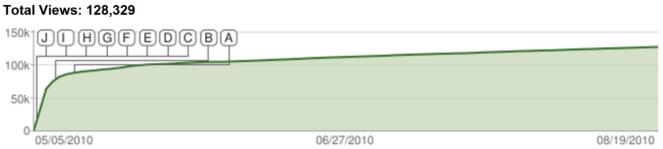


Figure 3.2: Sample video watch webpage. **A**: the box with information on the source. **B**: Examples of information retrievable though the API.

Another rather interesting piece of information that can be collected through the API is the list of "related videos," or videos suggested alongside a video that the user is watching (the videos column on the right of Figure 3.2). However, as discussed earlier, the suggested videos are different depending on the logged user, because they are designed specifically to meet his tastes and preferences. Thus, the videos returned by the API are the videos that would be suggested to a general user about whom we have no information. This makes it very difficult to study the personalization implemented by YouTube and its possible biases because there is no way to collect the suggestions that are proposed to real users with different profiling. Hence the difficulty of verifying whether YouTube actually acts or has

ever acted as a "Great Radicalizer", as discussed in the previous section.



Links

	Date	Link	Views
A	May 13, 2010	First referral from related video - MILEY CYRUS CAN'T BE TAMED SPOOF SHOOT!	4,600
B	May 10, 2010	First referral from related video - Did it all by myself	1,995
C	May 07, 2010	First referral from related video - Ke$ha BTS music video "Your love is my drug"	12,408
D	May 06, 2010	First referral from a subscriber module	34,601
E	May 06, 2010	First view from a mobile device	13,044
F	May 06, 2010	First referral from YouTube - Homepage	12,429
G	May 06, 2010	First embedded on - ijustine.com	3,010
H	May 06, 2010	First referral from YouTube search - zhu zhu pets	2,685
I	May 06, 2010	First referral from YouTube - /user/ijustine	2,430
J	May 06, 2010	First referral from YouTube - /videos	1,492

Figure 3.3: Example of statistics not anymore disclosed by YouTube, source [ZKG10a]

Now that the reader has an idea of the data that can be collected, it is appropriate to discuss those that cannot be retrieved (anymore). For example, as documented until at least 2017 [Zho+16] [HAK17], YouTube presented below each video a "Statistics & Data" section, which included the video's views count evolution over time, along with the top ten sources those views where coming from, as shown in Figure 3.3.

With the removal of this data, many scientific questions can no longer be answered. By removing the main sources of user flow, we can no longer assess what impact the recommendation system has on the platform and what share it generates. The latest studies on this subject date back to 2017 [ZKG10b][Zho+16]. More relevant in the context of this manuscript, the temporal evolution of viewcount has also been removed and can no longer be collected by researchers. This prevents any study of diffusion dynamics or prediction of future popularity based on the previous time series. In fact, the last studies dealing with temporal dynamics of content dissemination on YouTube stop at 2017 [PAG13][HAK17].

How much can we ask?

Leaving aside the limitations on the *type* of queries allowed by the API, YouTube also imposes a limit on the *amount* of information requested. For example, for each project (and thus for each API access key) the owner is entitled to requests with a total value of 10,000 *quota units* per day. Quota units are the "cost" associated with each request: for example, there are requests costing 1 quota unit, such as the number of current views collected by a video, and requests costing 100 quota units, such as requests on videos related to a specific video. As the reader can imagine, these limits are often quite restrictive and cause significant

slowdowns in data collection on the platform. Certainly these restrictions on data access are one of the reasons behind the paucity of scholarly literature on YouTube, despite the social media's reach. Other platforms, among which Twitter in particular, are much more generous in giving access to data. To give the reader an idea on Twitter the API 'request limit' is 900 requests every 15 minutes, or 86,400 per day, 8 times more than YouTube. In addition, from 2021 Twitter proposes an API specifically for academics, through which it provides real time and historical data with higher limits for number of queries (10 million Tweets / month, instead of 2.5 million with normal access). This explains why Twitter, while ranking 17th in the world in number of active users each month, is the most studied platform by academics.

3.2.2 Overcoming API Limitations

This section is intended to summarize all the possibilities for researchers to collect data on YouTube without going through the API. The huge advantage of not using the API lies in being able to collect much more data, often much faster. On the other hand, the techniques we are about to discuss reside mainly on the HTML/Javascript structure of the YouTube.com domains. This makes them very ephemeral, since even small changes in the structure of web pages can render these data collection methods ineffective. They require a lot of maintenance and are rarely reusable over time.

The non-API based methods that we adopted for the analysis we will present in the next chapters are mainly (1) web scraping and (2) RSS flows. Web scraping refers to the practice by which data is collected by reading the source code of a web page. When it goes well, the code is simple html that can be read by Python libraries such as BeautifoulSoup. Other times unfortunately, if we need information that is contained in elements of CSS or Javascript embedded in the html, then a simple read of the source code is not enough but we have to render the entire website through a real driver. This operation is much more expensive in terms of time, and therefore such requests are often difficult to handle in short time or in massive amounts. For example, to read information in JavaScript code we can use Selenium, a python (but not only) library that functions as a general-purpose web page rendering tool designed for automated testing. We can think of it as an essential web browser that executes JavaScript and returns HTML to your script. Selenium waits for client-side technologies, such as JavaScript, to load first, essentially waiting for the entire page to load. The general rule is that the elements you can interact with are in Javascript, while the parts that are just displayed are in HTML. If, for example, we are on the main page of a channel, we normally see the latest published videos and various playlists of videos. If we move with the cursor over the video, it starts playing for a few moments. This dynamic content is obtained through elements in Javascript. As a result, collecting for example the latest videos posted by a channel can be time-consuming because it requires the simulation

of a real browser. One possible solution is to use RSS feeds. To the best of our knowledge, data from RSS feeds have not been used in any of the YouTube studies, and in general there seems to be little knowledge about it. RSS feeds are files that contain a summary of the latest updates to a certain web page. RSS feeds exist on YouTube for both channels and playlists. These files, in the specific case of YouTube, collect the last 15 videos published by the channel or introduced in the playlist[1]. RSS feeds are by their nature much faster to use than Selenium. Therefore, in order to collect real-time data on the latest publications from thousands of videos, RSS feeds are the only viable option for researchers.

3.3 Collected Datasets

The datasets collected over these three years are particularly interesting especially because they are difficult to obtain through the YouTube API. They mainly concern the temporal evolution of views and commenters on a list of channels representing the French media sphere on YouTube. Below we elaborate on how these channels were chosen, what they represent, and then go into the specifics of the statistics collected.

The channel list

The goal of this collection was to have a fairly truthful representation of the French media and political space on YouTube. The choice of channels was based on recent research by the LISIS lab (Laboratoire Interdisciplinaire Sciences Innovations Sociétés) in Paris, whose researchers have been working on the French media sphere on the platform for years. The corpus is not limited to traditional journalism channels, but also opens up to content produced by creators native to the platform. It also includes all those channels that in a general way produce opinion and current affairs analysis and all those channels that are explicitly linked to economic, political, or associative entities. We also find all public service organizations that run their public relations on YouTube. The collection of these channels was made from a list of channels of professional media, well-known youtubers addressed to politics, activist associations, parliamentarians, channels of candidates for the 2019 European elections, channels of political parties, Gile Jaunes channels, channels of associations addressed to public causes and channels of large public or private institutions. Starting from an initial list compiled by hand by snooping around YouTube, the list was then expanded following the effect of the platform's recommendations, until it reached 800 channels. To these were then added smaller channels, which were searched in a database called Wizdéo, until reach-

[1]For example, the last 15 videos published by Le Monde can be found here: `https://www.youtube.com/feeds/videos.xml?channel_id=UCYpRDnhk5H8h16jpS84uqsA`

ing 1400 channels. More information about the channel selection and the discussion of its possible biases can be found in [Ben+20].

Starting with this list of channels from December 2019, through the techniques explained in Section 3.2, we began to collect (1) the temporal evolution of the views of the newly published videos and (2) all their comments and commenters. In the next two sections we go into the details of collecting these two datasets.

3.3.1 Temporal Evolution of Engagement Metrics

As discussed earlier, as of 2017 YouTube removed the ability to observe the evolution of views of a video. This time evolution is very interesting for us since it constitutes the footprint of the diffusion in the audience of certain content. Consequently, given the lack of recent studies on this issue, we were interested in collecting these time series despite the limitations introduced by YouTube. As discussed in Section 2.2 the platform to date only returns the number of current views, that is, the number of views a video has at the time the API is queried. Therefore, in order to collect time series of the evolution of views, it becomes necessary to create an *online* script that constantly queries the API or scrapes the HTML at a predetermined frequency. Over the past three years, we have collected the time evolutions of views at two different frequencies: every hour and every 5 minutes. The two datasets were obtained with different techniques and have very different scope in terms of size, as we explain in more detail below. In the rest of the manuscript we will refer to them as "the hourly dataset" and the "5-minutes dataset."

The hourly dataset

The hourly dataset covers all the channels in our list. For each video published by these channels as of December 2019, thanks to a collaboration with the Qatar Computing Research Institute, we collected the number of views every hour. The collection stops once one week has passed since publication. The choice to collect only one week's evolution of views is justified by the fact that news channels often collect most views in a few days after publication, presenting a strong initial explosion followed by a power-law decay [YXS15]. Indeed, our data confirm this rapid decline in user engagement, as only 3 percent of views are obtained in the last 24 hours. This collection was obtained through the official YouTube API, by means of multiple different access keys. The limits imposed by the API on 10,000 requests per day had a strong impact in determining the frequency of collection. Wanting to follow 1400 channels and all their new videos for a week forced us to be unable to collect at a higher frequency because we would have exceeded the daily limit of allowed requests.

This dataset now has, after three years of collection, an impressive scope. We have collected the time series of more than 900.000 videos for three years. Three years constitute a rather impressive span of time when considering that YouTube is but 17 years old. During these three years we have been able to observe events unprecedented in human history, such as the Covid-19 pandemic and the return of war to Europe with the Russian invasion of Ukraine. Therefore, we believe that our dataset has considerable value for scientific research and for investigating both YouTube and the changes these major events have had on their users.

Given the scope of our dataset, throughout various analyses we focused on specific periods of the collection. Throughout the manuscript therefore, whenever a specific segment of the hourly dataset will be used to provide analyses, the time frame of collection and the number of channels active in that time frame will be specified. The latter information is not trivial because we ha to keep in mind that not all channels publish with high frequency and some can result inactive for months. More importantly, some channels might not be on the platform during time windows: they might be shut down or temporarily removed from the platform (as happened, for example, to Russia Today and Sputnik France a few days after the invasion of Ukraine).

The 5-minutes dataset

For reasons that will be detailed in Chapter 5, to assess the amount of information lost by collecting data every hour, we collected a smaller dataset at a frequency of 5 minutes. This frequency was chosen to minimize the loss of information and is in practice the most useful frequency of data collection, since we empirically observed that view counts are updated no faster than every 5 minutes. The dataset is significantly smaller than the previous one: it contains 1012 videos posted between February 2, 2022 and February 16, 2022. It is available at [Cas+21a], where video identifiers have been anonymized. This dataset, given the much higher frequency of collection, and consequently the higher number of daily requests was not collected through the official API but through reading RSS feeds and web scraping with BeautifoulSoup. RSS feeds were used to collect new videos published by the channels in our list while web scraping was used to read every 5 minutes the view count appearing on videos web pages. The code used for this collection is available on Github.

3.3.2 Comments and Similarity among users

As we saw in Section 3.1.2, YouTube's recommendation system is one of the major drivers of attention toward specific content. One of the main features considered by the recommendation algorithm when making suggestions is the similarity among users. In other words, a

video is more likely to be suggested to a user if many similar users have already watched it. For this reason, putting effort into overcoming the restrictions of YouTube's API, we decided to build a network of similarity between users. Since we could not know users' view history, we relied on comment history: for each video published from our channel list from December 1, 2019 to December 30, 2021, we collected comments and commenters through web scraping. We then constructed the similarity network between users by creating a graph in which each node represented a user. The weight associated with a link was equal to the number of videos the linked users both commented on. This rather large network (with more than a million users) is interesting as it can be considered a proxy of the paths followed by recommendation systems when making suggestions. Studying diffusion dynamics over this network could be a way to simulate the diffusion of content through algorithmic recommendations.

3.4 Conclusions

In this Chapter we wanted to emphasize the importance of YouTube in the international media sphere. We recounted its history and the negative effects it can cause without content regulation. We discussed how poorly, compared to other platforms, YouTube has been studied through data-driven approaches, and we identified the API's restrictions as one of the major responsible for this lack. We have presented alternative methods of data collection that can overcome the limitations imposed by the API. Through these methods, we were able to collect data on temporal aspects that had not been studied for years. Precisely because of the rarity of previous studies, these data made interesting results emerge, as we will see in the next chapters.

Chapter 4

Bass Diffusion Model

Contents

In Chapter 2, we conceptualized a potentially harmful regime of public attention, emphasizing the importance of studying the spread of online content from a temporal perspective. This conceptualization prompted us to collect data on the temporal evolution of online popularity on YouTube, as we explained in Chapter 3. In this Chapter, we are about to examine these time series, keeping in mind the characteristics of the platform's recommendation system discussed in Section 3.1.2. As far as possible, we would like to distinguish what role the input features of the recommendation algorithm play in content diffusion and how they shape different attention regimes. To do so, we choose to explain the views count evolution through a Bass model. In fact, we believe its components can easily be reinterpreted in light of our knowledge of the recommendation system, providing the key to its investigation.

4.1 A Bass Model for Attention Dynamics

Although Section 1.1 already introduced the Bass model, let us briefly recall the scope it was conceived for and its formulation. It originated in management science to describe the dynamics of adopting new products in a market [Bas69]. It identifies two type of new costumers for a product: *imitators*, who discover the new product through someone who already owns it, and *innovators*, who discover it independently from others. If we call M

the final amount of adopters, and S the amount of people who have adopted the product, then the amount of new adopters $\dot{S}$ is described by the following equation:

$$\dot{S} = \alpha(M - S) + \beta\frac{S}{M}(M - S) \tag{4.1}$$

where α represents a *coefficient of innovation* and β a *coefficient of imitation.*

4.1.1 Interpretation for YouTube

In the context of YouTube, these quantities can take on a new interpretation:

1. Instead of adopters of a new product, S represents users who have watched a new video;

2. The quantity $\beta\frac{S}{M}(M - S)$ will thus represent users who became aware of this new video through those who have already watched it. This phenomenon, which effectively mimics contagion in an epidemic model, can occur either through (1) *human recommendation* or through (2) *algorithmic recommendation.* Human recommendation (1) can be though as a word-of-mouth phenomenon in which those who have seen the video suggest it to others. This exchange can occur in many ways: it can occur verbally, it can occur by sharing the link in a private chat, or by sharing it on a public profile. It does not matter the medium through which the sharing takes place, we talk about human recommendation every time the actor promoting a content is a real person, voluntarily choosing to promulgate it. Algorithmic recommendation (2) on the other hand is defined as such when the actor disseminating a content is not a person but the recommendation system. As we saw in Chapter 3, recommendation systems based on collaborative filtering (such as YouTube's) incentivize outreach among like-minded people. They somewhat mimic the "word-of-mouth", that in reality occurs between people who know each other, between people who don't. For this reason, we think that the term $\beta\frac{S}{M}(M - S)$, being a term of contagion, can well represent the collaborative filtering part of the recommendation system.

3. The $\alpha(M - S)$ quantity indicates all those users who did not learn about a content from recommendations (human or algorithmic). In the case of YouTube, they can be users who search for a content directly on YouTube search engine, or, for instance, users that subscribed to a channel and receive notifications once a new piece of content is posted. In fact, in this case it is not YouTube's recommendation that chooses the content but the user himself who decides to be notified for something.

4.1.2 Model Strengths

The choice of this model has several strengths:

1. it does not require many parameters and therefore remains simple to interpret;
2. it clearly distinguishes the role of innovation and imitation in determining the dynamics of dissemination;
3. its closed solution is known, enabling parameter estimations through data fitting.

4.2 Data Fitting

Data. For data fitting, we used 25,000 videos randomly drawn from our time dataset and published between December 2019 and May 2021. We define the time series of video views as the series $\{v_t\}_{t\in\{1,\dots,170\}}$ where t represents the number of hours elapsed since the video was published.

Parameters estimate methods. To estimate the parameters of model (4.1) we use its closed-form solution

$$S^{\alpha,\beta,\gamma,M}(t) = M\frac{1-\gamma e^{-(\alpha+\beta)t}}{1+\frac{\beta}{\alpha}e^{-(\alpha+\beta)t}}. \tag{4.2}$$

where γ is introduced to define the general solution of equation 4.1. Imposing $S^{\alpha,\beta,\gamma,M}(0) = s_0$, we can then express γ in terms of the initial condition $\gamma = 1+\frac{s_0}{\hat{N}}(1-\beta)$. To fit parameters, we use the least squares criterion, hence, we look for the set $(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{M})$ that minimizes:

$$\min_{\alpha,\beta,\gamma,M}\sum_t (S^{\alpha,\beta,\gamma,M}(t)-v_t)^2$$

To do this we use the Levenberg-Marquardt method [Mor78], implemented Python's library SciPy [Vir+20]. For each video we use different initial poits for the optimization, to avoid local minima.

Goodness of fit: Before analyzing the values our parameters are set to, we analyze the goodness of fit of our model. To do this we use two measures proposed in previous literature [Gao+21] [Ric+14]:

- MdAPE. It is the Median Absolute Percentage Error, which is the median value of APE(t), defined, at each time, as

$$\text{APE}(t) = \frac{|S^{\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{M}}(t) - v_t|}{v_t + 1}.$$

  where $S^{\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{M}}$ is the estimation of our model and v_t are the actual views at time t. APE(t) represents the percentage of error at every time t. The choice of dividing by $v_t + 1$, instead of the more intuitive v_t prevents problems of definition when $v_t = 0$.

- MAPE. It is the Mean Absolute Percentage Error. As out timeseries have 170, observation then the MAPE of is defined as:

$$\text{MAPE} = \frac{1}{170} \sum_{t=1}^{170} \frac{|S^{\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{M}}(t) - v_t|}{v_t + 1}$$

The distributions of these two errors are shown in Figure 4.1.

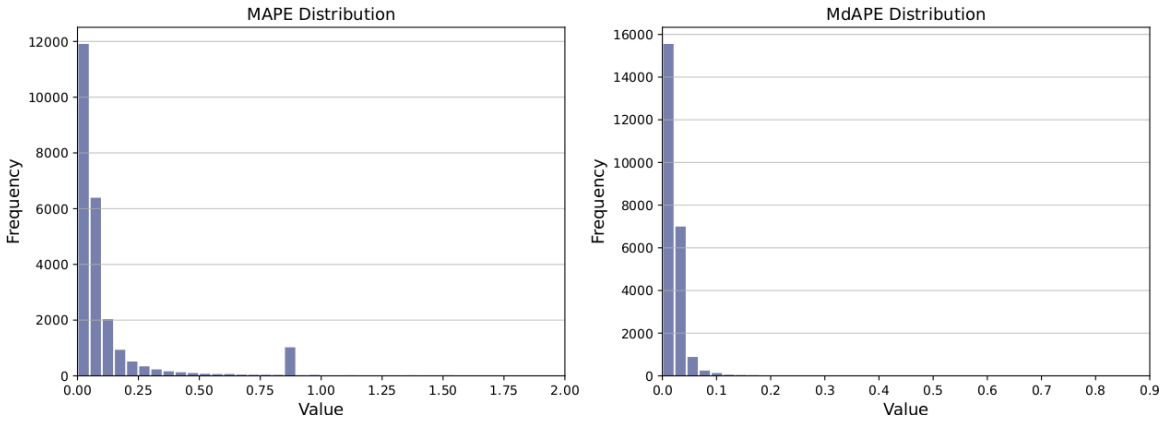


Figure 4.1: MAPE and MdAPE distribution.

In previous works applying a Bass model on Twitter, values of MdAPE around 0.1 and MAPE around 0.12 have been considered satisfying [Gao+21]. Other works on YouTube [Ric+14], applying six different bio-inpired models, succeeded in fitting 95% of views with a MAPE smaller than 0.05. We believe that considering well explained only the curves with a MAPE under 0.05 would be too demanding in our case: the MAPE is sensitive to the APE distibution tail of each video, and grows easily even if only few hours are not well fitted. Given the comparison with the literature and the sensitivity of MAPE to high but concentrated errors, we decide to set a dual threshold of goodness-of-fit, higher for MAPE (0.12) and lower for MdAPE (0.04). This choice seems reasonable when looking at videos with errors close to these thresholds (Figure 4.3): the model is still capable of explaining

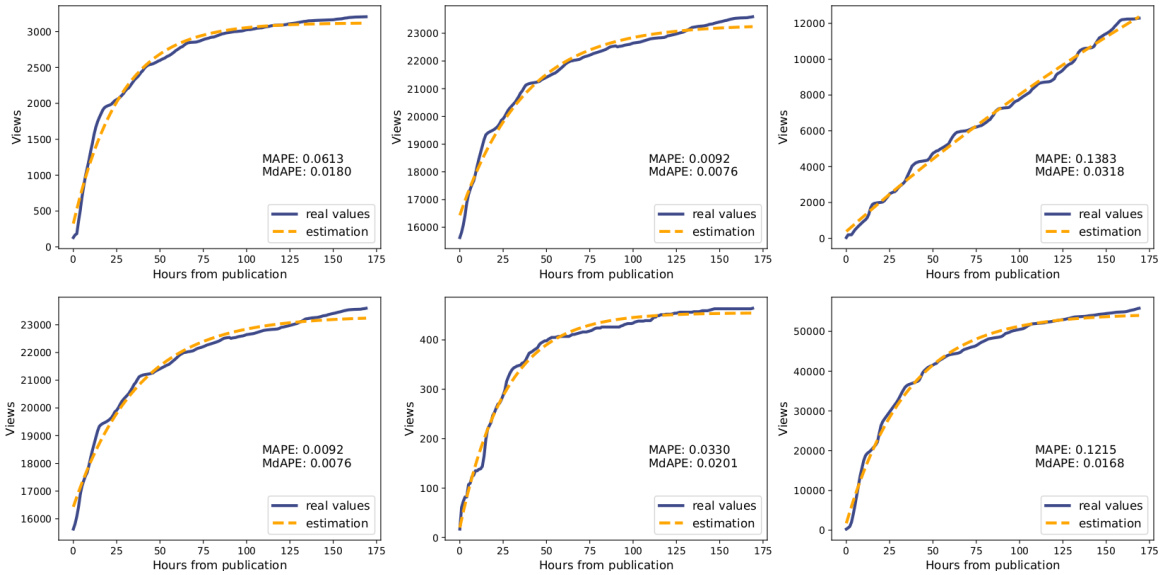


Figure 4.2: Examples of good video fittings with low MAPE.

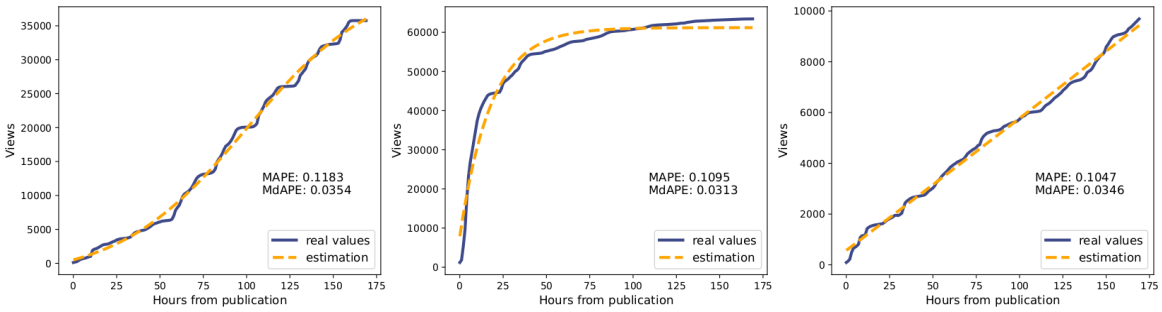


Figure 4.3: Examples of video with MAPE and MdAPE around the thresholds

their trends. Setting the requirement MAPE < 0.12 and MdAPE < 0.04 leaves us with the 78% of our videos being well explained by Bass model.

Concerning the remaining 22%, some observations could be made in order to understand the limitations Bass model has on our data. We identify three typical profiles that the model is unable to explain. The first is views evolutions which presents some drops in the cumulative count, as show in the two left plots of Figure 4.4. We will discuss this phenomenon extensively in Chapter 5, for the moment it is enough to know that these drops in cumulative counts are made voluntarily by the platform: they represent an artificial intervention, interfering with the natural evolution of engagement that we are modeling. Hence we should not be too concerned about the low quality fitting in this case. The second type of videos not explained by the Bass model is the one shown in the central

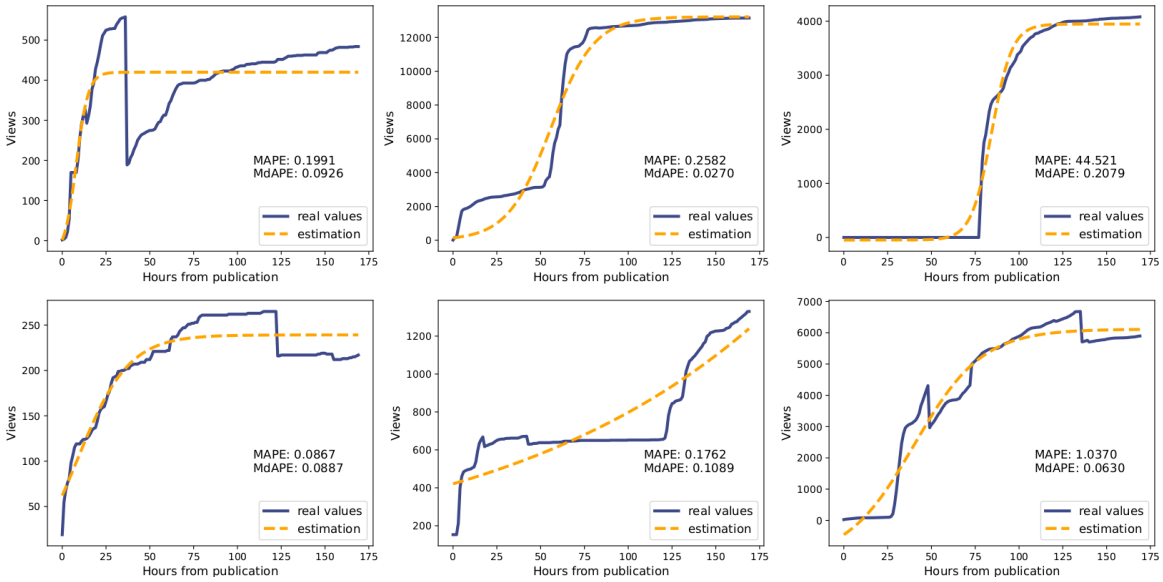


Figure 4.4: Examples of bad video fittings with high MAPE. (left) Views count with drops. (center) Views count with distinct growing phases. (right) Views count stable at zero in a first phase.

plots of Figure 4.4: the views, after an apparent stabilization, start to rise abruptly again, defining two different stages of growth. The third type of videos not well modeled are those whose growth begins several hours, if not days, after publication (right plots of Figure 4.4). Regarding these types of videos, the Bass model, given its simplicity, fails to model subsequent phases of stabilization and growth. To overcome this problem, one might consider superimposing multiple Bass models, each of which could indicate, for example, the spread within an online community. Although this may be an interesting line of research for the future, here we will limit ourselves to analyzing the curves well fitted by the natural Bass model 4.2.

4.3 Stronger Recommendation Means Higher Popularity and Shorter Life

Reinterpreted from a social network perspective, Bass allows us to distinguish human and algorithmic recommendation from personal initiative to view content. To distinguish these two parts, we can look at the model's total number of imitators and innovators, evaluated by integrating in time $\alpha(M - S)$ and $\beta S(M - S)/M$. Figure 4.5(a) shows the distribution of the proportion of imitators per video, with logarithmic bins. As we can see, the curve

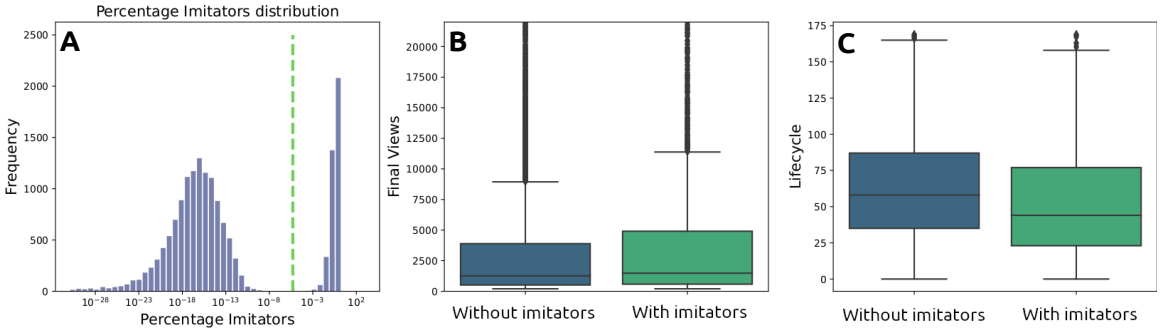


Figure 4.5: **A** Distribution of fraction of imitators. **B** Final views distribution for videos with and without an imitation component. **C** Lifecycle distribution for videos with and without an imitation component

is bimodal: most of our videos (76%) fall in the first bell-shaped part of the distribution with a percentage of less than 1% of imitators. Hence, recommendations, either human or algorithmic, do seem pretty marginal for the majority of the videos. This result should come as no surprise: even though it is hard to quantify the percentage of total videos the recommendation system promotes, we know from different sources that recommendations are skewed and concentrated on a few videos [STK18] [San+21]. It therefore makes perfect sense that the existence of recommendation is observable mainly in a minority of total videos.

Setting a threshold at 1%, we can then define two classes of videos, those with and without a significant presence of imitators, and study their characteristics. Figure 4.5(b) shows that videos with recommendations are typically more popular than those without, with a statistically significant difference in their distributions verified through a Kolmogorov-Smirnov test of p-value 0.0007. This is pretty reasonable as, once a contagion-like dynamic is activated, it can contribute more to the diffusion of a content and to its popularity.

What is perhaps lees expected is that, along with this increase in popularity, videos reduce the time they take up in collective attention. To prove that, let us define the *lifecycle* of a video as the time it takes to get 90% of its total views. This measure expresses the concentration in time of view acquisition. When it is low the video gathers most of its audience in a short time, when it is high the content circulates for longer. As shown in Figure 2.3, videos with an active imitation dynamics generally have shorter lifecycles. This behavior is quite interesting and, from some points of view, counter-intuitive. One would expect that, to collect more views, a piece of content would have to circulate longer. What seems to be happening here instead is that recommendations increase the popularity of a video while decreasing its lifecycle.

This behavior is preserved if, instead of dividing the videos into two classes, we consider a variable threshold on the percentage of imitators. In Figure 4.6, on the x-axis we consider

different percentages p of imitators and on the y-axis we display the mean and median final views and lifecycles for videos showing more than p imitators. As we can see once again, the percentage of imitators causes an increase in the popularity of the content, accompanied by a reduction in its lifecycle.

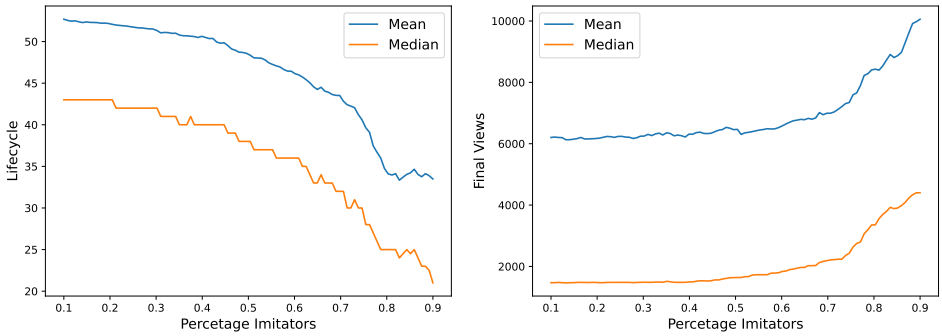


Figure 4.6: (left) Mean and median lifecycles (in hours) of videos with more than an x percentage of imitators. (right) Mean and median of final views of videos with more than an x percentage of imitators.

4.4 Discussion

In this Chapter, we proposed a simple model distinguishing between user initiative and human and algorithmic recommendation. We observed that in only 28% of cases recommendations play a role in the dynamics of content dissemination, consistently with some studies on the skewness of recommendations on YouTube [STK18] [San+21]. In these cases, the stronger the recommendation, the more popular the content is, and the faster it reaches its audience.

These results align with a study from Richier et al.[Ric+14]: in their paper, the authors explain the evolution of view counts on YouTube through 6 different bio-inspired models. Two of them, in particular, disentangle our imitation and innovation components and represent them separately: the exponential and the Gompertz model. The exponential model represents spontaneous adoption, and the Gompertz model represents the adoption due to contagion/imitation. The authors observe that non-popular videos are best explained by the exponential model (in a slightly modified version). At the same time, popular videos are mostly explained by the Gompertz model (also in a slightly modified version). Our results are thus consistent with these findings. Videos well explained by models with a contagion/recommendation component are more popular than others.

In addition to this finding, our work highlights the less obvious observation that recommendations result in faster exhaustion of interest. This outcome is not trivial. Higher popularities are, in general, associated with longer lifecycles. If we consider, for instance, the 50% of most popular videos in our corpus, their expected lifecycle is 80.8 hours, versus 71.0 for 50% of the less popular videos. Let us go back to our formulation of junk news bubbles as attention regimes in which a large share of public attention is captured by objects incapable of sustaining it. Using the Bass model applied to real data, we could conclude that platform recommendations can be responsible for generating this kind of attention.

Undoubtedly, what we have presented in this chapter is a preliminary work that has limitations. First of all, the goodness of fit obtained can be improved, for example, by generalizing the Bass model to cases with an elastic interested audience $M(t)$ [JGK08]. This elasticity could model the adaptative aspect of the recommendation system that updates its targets based on the previous history. Another limitation associated with the model is given by the impossibility of distinguishing between human and algorithmic recommendations. As a solution, introducing a veritable network of user similarity, like the one collected in Section 3.3.2, would allow to mimic algorithmic recommendations better and better pinpoint their role. This introduction would be possible only by adopting an Feature-Driven Heterogeneous Bass model [Gao+21], which will be further discussed in Conclusions and Perspectives Chapter.

A final limitation of the Bass model consists of its possible alternative interpretations. As noted in Section 3.1.2, recommendation systems, besides prior similarity and popularity, strongly hold for novelty. In the Bass model, it could be argued that the term $\alpha(M - S)$ partially accounts for this novelty factor since it plays its role mainly at the beginning of the dynamics, and its importance decreases with time. According to this interpretation, the recommendation system would appear in both the imitative and the innovative components, the latter being reinterpreted as a novelty term. This interpretation does not significantly change our results: the imitative term of our model represents profiled recommendations occurring between similar users through collaborative filtering. This specific feature of the recommendation system is what we account responsible for fostering popularity and activating faster dynamics. The contagion-like functioning of the recommendation system is what we believe could lie behind the rise of junk news bubble dynamics.

Chapter 5

Fake Views, Real Trends

Contents

5.1 Introduction: An Emerging Evidence from the Data

As anticipated in the previous sections, the data collected over the past three years constitute an unprecedented source of information in terms of scope and, more importantly, in terms of novelty. In fact, given the scarcity of data accessible through the API and the consequent general scarcity of analysis about it, our data become a good opportunity to analyze phenomena that are still little studied on the platform.

The first surprising evidence that emerged from the time series is that the number of views can decrease over time (Figure 5.3AA). A count that, in theory, could only increase when new users watch a video instead often shows drops. The reasons behind such decreases can be found on some YouTube official web pages: "We want to make sure that videos are viewed by actual humans and not computer programs"[Noac]. Several times in the past

there have been reported occasions where the platform has removed significant amounts of views. For example in December 2012, the platform deleted 2 billion views from the channels of record companies such as Universal and Sony [Gay12] [Hof12] [Noab] [Dre14]. Over the years, countless youtubers have suffered sudden and drastic cuts to their views (and many have complained about it, often through YouTube videos). According to YouTube's policies [Noaa] [Pfe14] [Noac], these interventions aim at preserving a "meaningful human interaction on the platform" and to oppose "anything that artificially increases the number of views, likes, comments or other metric either through the use of automatic systems or by serving up videos to unsuspecting viewers" [Noaa]. To put it another way, the platform wants to make sure that views are not artificially recorded, and when it detects some it removes them from the count.

Before 2017, when views evolutions were public, some media interest was raised about these corrections [Kam15] [Qui15] but not much scientific research was concerned with verifying the fairness of this policy. Up to our knowledge, the only previous work concerning views count corrections is that of Marciel et al.[Mar+16] in 2016. In the paper, the authors study the phenomenon of views corrections in relation to video monetization and the identification of possible frauds, drawing on research carried out on ads frauds in other social media [CZC14] [NS19]. To test how YouTube removes artificially created views, Marciel et al. created a bot able to record views automatically on videos of their creation, some monetized (i.e., earning money from ads) and some not. The authors showed YouTube recognizes illicit views much more accurately on non-monetized videos. On monetized videos, YouTube removed a lower percentage of the views made by the researchers. One possible explanation for that, discussed by Marciel et al. in their work, is that YouTube earns a percentage on monetized videos and is therefore interested in maintaining their view counts high (so that advertisers would pay more).

Although we believe that this investigation of the correlation between monetization and view correction is a useful first step toward understanding YouTube's policy, we do believe that other pressing questions should be answered. For instance, can fake views, even when removed, have an impact on the success of a video and be used to influence YouTube's content dissemination? As already discussed in Chapter 1 and 2, social media tends to favor rich-gets-richer dynamics, and future visibility is highly dependent on past popularity, given that trending contents tend to be favored by human influencers [Rog18] and recommendation algorithms [Gil16].

This is where fake views come into play. Indeed, if the correction of illegitimate views happens too late, these views have the potential to weight in the cycle of trendiness [Cas+21c] and unfairly propel their targets. If YouTube fake views correction is significantly slower than its recommendation dynamics, then artificially promoted videos risk to be favored by human and algorithmic recommendations, and thus reach larger audiences and collect

extra real views. If, before being deleted, fake views are able to trigger a cascade effect that increases the visibility of some content, then they may be used to manipulate online debate. Not unlike social bots [Fer+16] and paid commentators [KPR17], fake views could give the false impression that some content is highly popular and endorsed by many, thus distorting public debate and ultimately endangering democratic processes: these risks of social bots have been highlighted in multiple papers [Zha+13] [Sha+17a] [Rat+11] [MM10] [BF16] [Lle+19] [BM19] [RM21] [LSS16].

5.2 Overcoming Information Loss

To study the extent of the phenomenon of fake views and investigate their possible role in propelling content, we used the hourly dataset presented in Chapter 3, narrowed down (to make it more manageable) to a time interval from January 1, 2021 to May 10, 2022. Narrowing the time window to this interval, our observations cover 1064 channels out of the 1400 in our full list. On this period, we collected the time series of 270,133 videos. The dataset is available at [Cas+21b], where video identifiers have been anonymized. Through this data we would therefore like to study the phenomenon of removing fake views. A problem arises, however: what we can observe in our data is nothing more than the difference in views between two successive hours, which does not correspond exactly to the amount of fake views removed. In hours when, for example, the number of views accumulated by a video is greater than the number of views removed, we are unable to identify the intervention implemented by the platform. Obviously, one of the ways to reduce this loss of information is by increasing the frequency at which data are collected. However, as discussed in Chapter 3, the limitations imposed by the API are quite stringent and prevent us from collecting data with finer granularity. However, having at our disposal a smaller dataset, the 5-minutes dataset presented in Chapter 3, allows us to estimate the loss of information associated with hourly aggregation and to devised a method to reconstruct the true amount of fake views removed at each hour, as we will explain next.

Method to estimate corrections

In the hourly dataset, the amount of views $\tilde{v}_h^i$ we observe in one hour on video i is nothing more than the difference between the non-observable views v_h^i actually recorded on the video and the corrections c_h^i made by the platform (i.e. $\tilde{v}_h^i = v_h^i - c_h^i$). Wanting to study the phenomenon of corrections, a first way to estimate c_h^i might be to consider it equal to the changes in observable views when these are negative (i.e. $-\tilde{v}_h^i 1_{\{v_h^i > 0\}}$). Obviously, this choice constitutes an underestimation of the number of corrections since (1) their magnitude might be greater in the case $v_h^i > 0$ and especially (2) many corrections are likely to escape

our attention when they are not greater than the number of views recorded in the same hour.

Therefore, it becomes important, before analyzing the phenomenon of corrections, to understand how large the information loss associated with hourly collection is. To estimate this amount, we can use the 5-minute dataset. We can use the corrections visible in the 5-minutes frequency timeseries as a proxy for the real ones c_h^i. We then can aggregate the 5-minutes timeseries to see how they would look like if we sampled every hour and not every minute, and use these aggregated series to estimate the information loss.

To information loss can be defined though two errors associated with the estimated corrections $\hat{c}_h^i$:

- the fraction of *lost corrections*, consisting of the fraction of real corrections that are lost in the reconstructed series;

$$\frac{\sum_{h,i:c_h^i>\hat{c}_h^i}(c_h^i - \hat{c}_h^i)}{\sum_{h,i}(c_h^i)}$$

- the fraction of *lost interventions*, consisting of the fraction of interventions, namely the hours in which the platform intervened, no longer visible in reconstructed series;

$$\frac{\sum_{h,i:c_h^i=0} 1_{\hat{c}_h^i>0}}{\sum_{h,i} 1_{c_h^i>0}}$$

When approximating the corrections with the negative views visible in the hourly collection, i.e. by taking $\hat{c}_h^i = -\tilde{v}_h^i 1_{\{v_h^i>0\}}$, we obtain the errors shown in Table 5.1. As the table shows, the loss of information is far from being negligible. We loose 66.31% of the corrections and 60% of the interventions.

In the following we present two methods to reduce this information loss. The first, which we will refer to as the *benchmark method*, is simpler and uses a heuristic. The second method, which we will refer to as the *reconstruction method*, uses an XGBoost classifier [CG16] to improve the benchmark method and reduce information loss in the reconstructed series.

Since these two methods can introduce errors of overestimating both the corrections and the number of interventions, we should monitor two other errors to validate them. The errors are as follows:

- the fraction the *added corrections*, consisting of the corrections mis-added by the re-

construction methods, divided by the total real corrections;

$$\frac{\sum_{h,i:\hat{c}_h^i > c_h^i} (\hat{c}_h^i - c_h^i)}{\sum_{h,i}(c_h^i)}$$

- the fraction of *added interventions*, consisting of the number of mis-added interventions by reconstruction methods, divided by the total number of real interventions.

$$\frac{\sum_{h,i:\hat{c}_h^i = 0} 1_{c_h^i > 0}}{\sum_{h,i} 1_{c_h^i > 0}}$$

	Hourly Aggregation	Benchmark Method	Reconstruction Method
Lost Corrections	66.31%	50.27 %	35.64%
Added Corrections	0%	1.63%	4.93%
Lost Interventions	60.00%	60.00 %	45.27%
Added Interventions	0%	0 %	1.31%

Table 5.1: **Validation of Reconstruction Method on the 5-minute-frequency dataset.** The table shows the loss of information with hourly aggregation and with the estimates done with the proposed Reconstruction Method and the Benchmark method.

5.2.1 Benchmark Method

As a benchmark method we propose the following heuristic: in hours with negative views, corrections are approximated by the number of negative views increased by the expected views at that hour. To estimate the expected views in a certain hour, we considered various quantities, such as the average and minimum number of views over a time window around the given hour. The time window considered are shown in Figure 5.1 and they should be intended symmetrical, e.g. a window size of two hour means that both the two hours preceding and following the hour of interest have been considered. As shown in Figure 5.1 the errors considered in Table 5.1 are minimal when the minimum over a one-hour time window is chosen as the approximation. Hence we approximate:

$$\hat{c}_h^i = (-v_h^i + \min(v_{h+1}^i, v_{h_1}^i)) 1_{\{v_h^i > 0\}} \tag{5.1}$$

This benchmark method reduces the lost corrections from 66.31% to 50.27% by introducing only the 1.63% of added corrections.

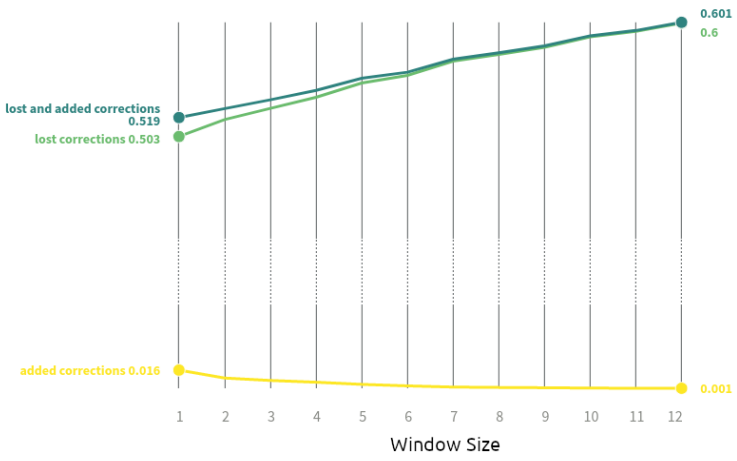


Figure 5.1: Error introduced by the benchmark method, varying the time window.

5.2.2 Reconstruction Method

The benchmark method adjusts negative views to estimate the platform corrections, but it fails to detect correction events that have occurred when the observed views $\hat{v}_h^i$ are non-negative. Hence we developed a method meant to detect anomalies in the views evolution and attribute them to concealed corrections. The method consists of an XGBoost classifier that can detect hours with unobserved corrections. Below we present how this classifier was trained and we analyze its performance.

We constructed the train dataset as follows:

- each row represents one hour of our time-series;
- for each hour we extracted the evolution of views in the 12 hours before and after that hour and added them as features;
- we added as features the time of day (since, are we will see shortly, corrections are roughly periodic and occur mostly between 4 and 6 p.m.) and the number of hours elapsed since publication.

In order to do parameter tuning, we chose to optimize the F1 score, a metric defined as:

$$\text{F1} = \frac{2}{\frac{1}{\text{precision}} \times \frac{1}{\text{recall}}}$$

where *precision* stands for the rate of true positives over all the samples classified as positive, while *recall* stands for the rate of true positives over all the really positive samples in the data. In this way we can limit the number of false positives and false negatives introduced by our classifier. We divided our dataset into a train set of 90075 observations and a test set of 24674 observations. We used the train set to perform a 5-fold cross validation to choose the optimal values of *maximum depth* of the decision trees, the *learning rate* and *alpha* parameters of the XGBoost classifier. The performances in terms of F1 score associated with different combinations of values are shown in Figure 5.2. The best parameters, able to grant an F1 score equal to 0.649, are a maximum depth equal to 25, a learning rate equal to 0.2 and alpha equal to 1.

The classifier allows us to reconstruct which hours have had YouTube interventions without these being visible in the hourly aggregation. Once we identify the hours with interventions, we estimate the magnitude of the corrections by formula (5.1). In this way, as shown in Table 5.1, we are able to reduce the lost corrections from 66.31% to only 35.64%.

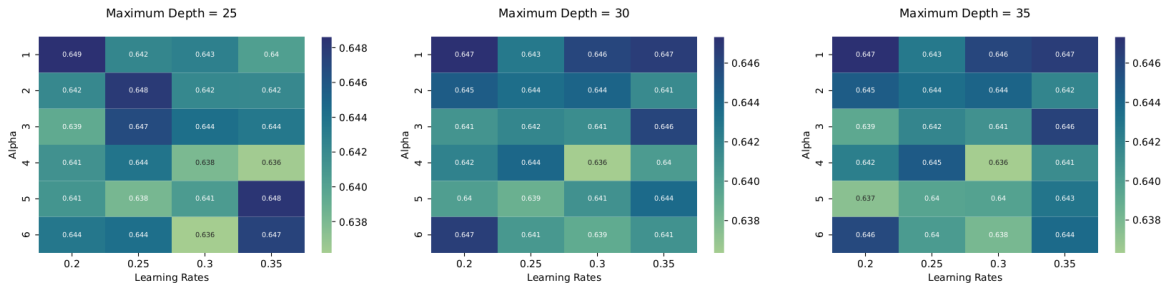


Figure 5.2: **XGBoost parameter tuning.** Performances in terms of F1 score associated with different combinations of parameters' values.

5.3 Fake Views Corrections

Once we have devised a method for reconstructing the time series of corrections, we are ready to analyze them. In particular, for the reasons anticipated at the beginning of the chapter we are interested in understanding what the extent of this phenomenon is and whether illegitimate views may have an effect on the propulsion of content through human and algorithmic recommendation.

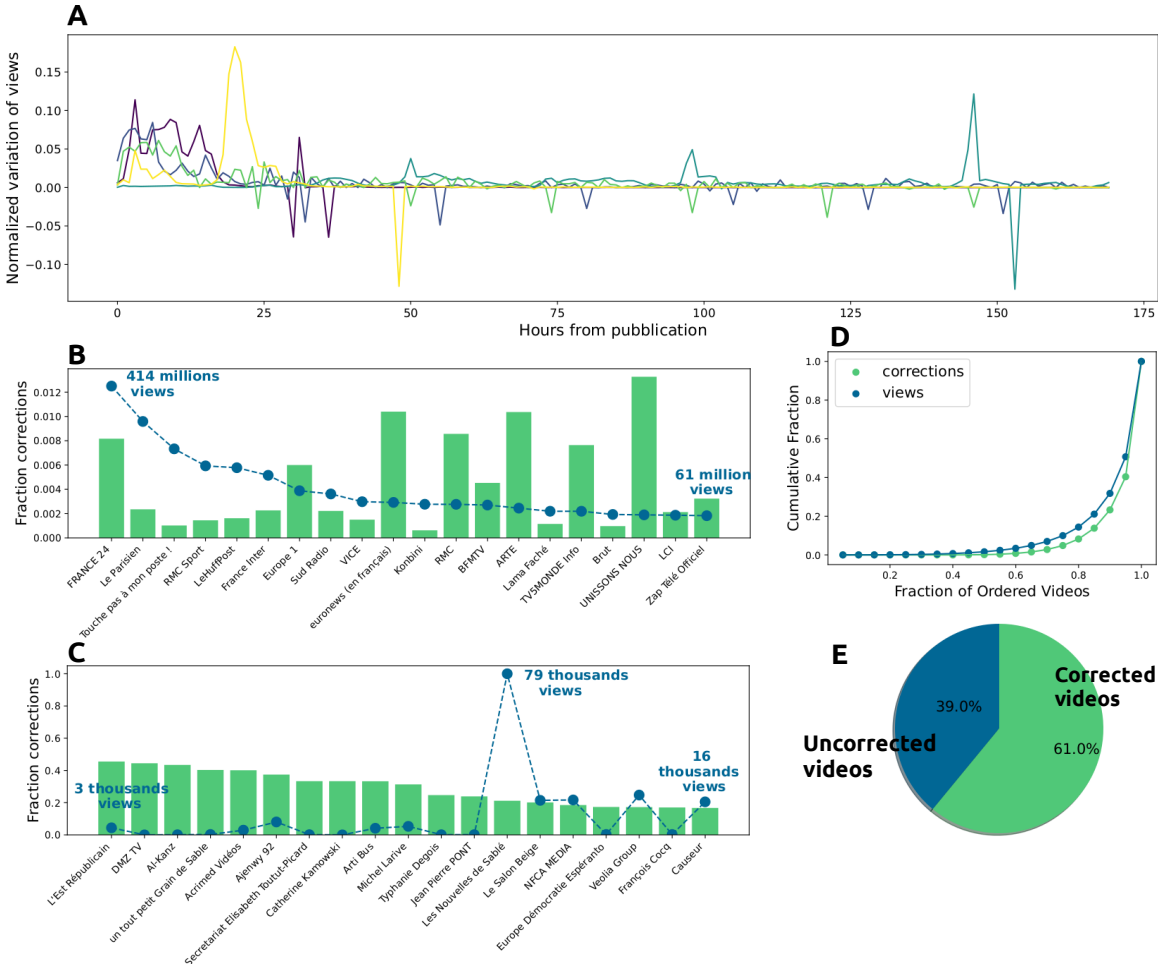


Figure 5.3: **A:** 5 sample videos and their hourly evolution of views. **B:** The 20 most viewed channels in our dataset and their fraction of corrections over real views. **C:** The most corrected channels in terms of fraction of corrections over real views. **D:** Lorenz curve of the distribution of views and corrections among different videos **E:** Percentage of videos concerned by the policy.

5.3.1 Scale of the phenomenon

A rather impressive first aspect about the policy of removing views is its extent: we detected corrections for almost all monitored channels (90% of them) and for 61% of the videos in our corpus. Such a large scope underscores the importance of better understanding how these corrections are made. In fact, corrections in our corpus amount to about 22.5 millions. Although they represent, on average, a seemingly modest 0.5 percent of the total views, their number remains impressive and, more importantly, their distribution is very uneven. If we look at the Lorenz curve (Figure 5.3D) of the distribution of corrections among different videos, we can see that most of the corrections (more than 80%) are concentrated on only 20% of the videos. In comparison, the concentration of corrections appears to be stronger than that of legitimate views.

The heterogeneity of corrections is confirmed when we examine the most popular and the most corrected videos. Figure 5.3B shows the 20 most popular channels in our dataset and the percentage of actual views that corrections account for. These very popular channels, which are mainly traditional media channels such as TV stations, newspapers, and radio stations, still show marked differences in their corrections (between 0.1 percent and 1.3 percent). In contrast, if we look at the 20 channels with the highest fraction of corrections to actual views (Figure 5.3C), we find channels with 40 percent corrections. These channels, which are mostly platform-native youtubers, collect substantially fewer views than the top channels (averaging about half a million views over the collection period).

5.3.2 Correction rhythms

The view correction activities by YouTube have some interesting recurrences. If we look at Figure 5.4B, we can see how corrections are distributed according to the time of day. We can see, for example, that while the median number of removed views hovers around a few dozen at most hours, it rises to more than 10,000 at 5 p.m. and hovers around 5,000 at 4 and 6 p.m. This observation is confirmed if we study the number of videos corrected at each hour of the day (Figure 5.4C). Again, while normally the median number of corrected videos is close to zero, between 4 p.m. and 6 p.m. it is around 150 to 250. These rhythms of correction are peculiar and completely different from the rhythms at which views are made on the platform. In fact, if we look at Figure 5.4C, we can see how the views are distributed at different times of the day: they present a completely different behavior, with a minimum of hourly views around 7 a.m. and peaks of views in the evening or night hours (9 p.m. to 2 a.m.).

In summary, views are distributed quite evenly during the day, according to circadian rhythms [Cas+21d], while corrections are concentrated in specific time slots. This fact sug-

gests that most correction activities take place once a day, every 24 hours. This rhythm not only is unrelated to the rhythms of views production, but also seems rather slow given the fast pace at which content is propagated on the platform and the speed at which suggestions from recommendation systems are updated. For comparison, most of the videos in YouTube's "trending" section[1] are less than 24 hours old.[2] Therefore it is legitimate to ask whether such a frequency of corrections is too low to prevent possible interference of non-legit views with human and algorithmic recommendation.

If instead of looking at the distribution by time of day we study how corrections and views are distributed starting from publication, another interesting result emerges. In Figure 5.4D, on the x-axis, we considered a time starting from midnight prior to the publication of a video. For each hour starting from midnight prior to publication, we summed the number of views, corrections, and corrected videos (with the convention of considering these quantities zero prior to publication). The expedient of measuring time from the midnight before publication allows us to maintain the periodicity of the corrections phenomenon, which would be lost if we summed the time series from their time of publication. What can be seen in this graph (normalized to make corrections and views comparable) is that views are much more concentrated in the first few hours after a video is published, while corrections are more spread out over the life of the videos. In fact, most views occur before 5 p.m. on the second day, the time when substantial corrections begin to occur. This apparent delay in the corrections, together with their low frequency, prompt us to investigate more deeply the relationship between corrections and popularity.

5.3.3 Late Corrections and Popularity

To gain a deeper insight into correction mechanisms and investigate possible interference with recommendations, we should consider the timing of corrections not only in absolute terms, as done in the previous section, but also relative to when videos collect the most views. In particular, we should ask whether corrections are made before or after a video reaches its peak popularity. If corrections occur after this peak, one might suspect that by inflating the number of views, illegitimate views could make a video appear more popular and unfairly boost its human and algorithmic recommendation.

Figure 5.4(left) presents the percentage of corrections made before videos reach a certain percentage of real views. For each video, we calculated the percentage of views collected at each hour. We then counted the number of fake views that occur before a certain percentage of real views. As the graph shows, most of the illegitimate views are corrected after videos

[1] https://www.youtube.com/feed/trending

[2] After collecting the top 20 videos in the trendy section in 7 different days, we found out that only the 25% have been published since more than a day.

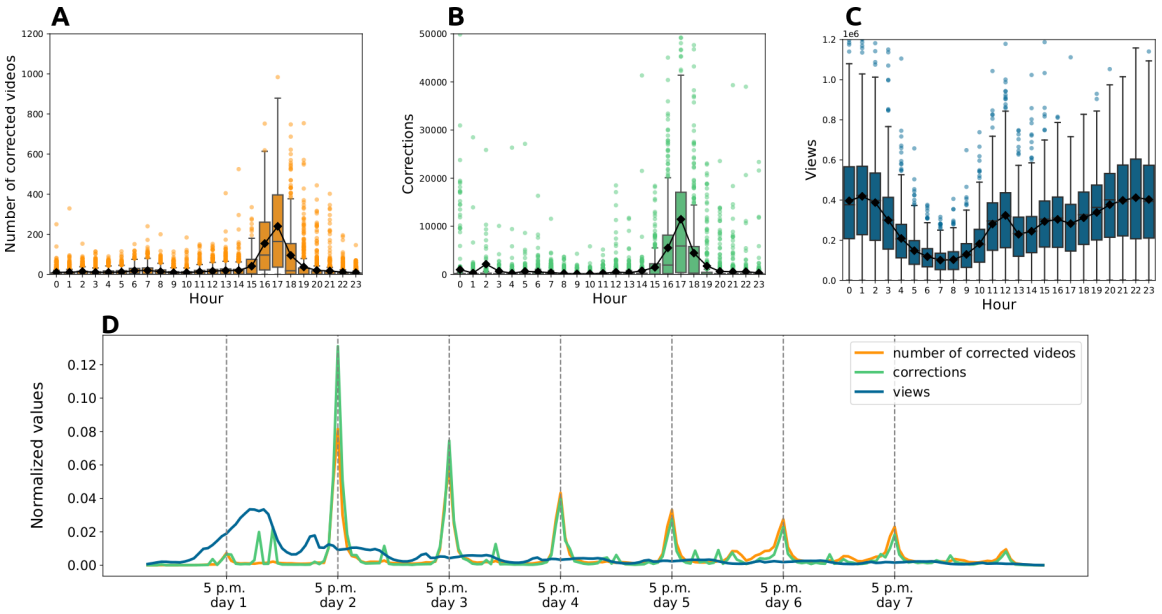


Figure 5.4: **A:** Distribution of the number of corrected videos per hour of the day. **B:** Distribution of corrections per hour of the day. **C:** Distribution of views per hour of the day. **D:** Normalized number of corrected videos, corrections and views since midnight before publication.

have collected most of their real views. On average, only about 10 percent of the corrections are made before the videos have achieved 80 percent of the views. Even more striking is the fact that as many as 54 percent of corrections are made after the videos have stopped collecting real views, at the end of their lives. It is therefore quite clear that most illegitimate views are removed very late in the lives of most videos, well after their popularity has peaked and begun to decline.

In order to understand whether these delays have an effect on video popularity, we need to better investigate the relationship between fake views and popularity. In Figure 5.5(right) we examine the relationship between the number of fake views and the total number of legitimate views per channel. In principle, these two quantities should be independent: in fact, if YouTube's correction policy were sufficiently fast and efficient, illegitimate views should have no impact on real views and, at the same time, there is no apparent reason why more popular videos should attract more fake views than others. Yet, Figure 5.5(right) shows a strong linear correlation between the logarithms of the two quantities. With a R-square equal to 0.819, the logarithms of fake and real views are related by a linear regression with intercept equal to 1.7704 and slope equal to 1.0574. The p-values associated with these quantities are significant, specifically less than 0.001. This linear correlation between the

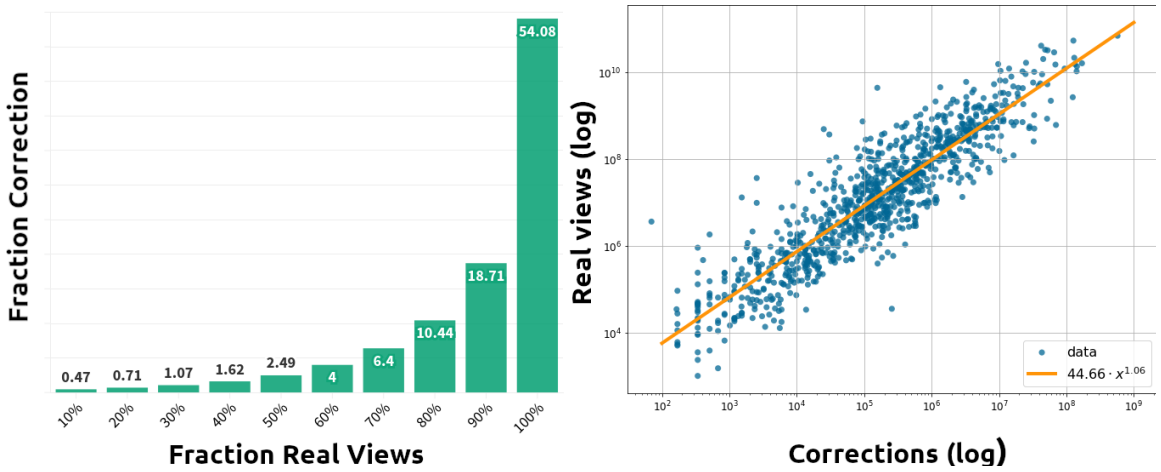


Figure 5.5: **Left:** Fraction of corrections occurring after different percentages of real views. **Right:** Correlation between real and fake views per channel.

logarithms of the two quantities results in the following relationship between corrections (c) and real views (v):

$$c = 58.94 \cdot v^{1.06}$$

The relationship between these two quantities is hence slightly more than linear, with a 95% confidence interval on the exponent being within $[1.026,\ 1.089]$.

In view of this clear correlation, it becomes natural to wonder about a possible causal relationship between fake and real views. In particular, we would like to know whether fake views have an impact on popularity or not. Unfortunately, with the information we have, we are unable to answer this question. Although we have been able to reconstruct the time series of the *removal* of fake views, we are unable to gather any information on the timing of *generation* of these fake views. Without knowing when fake views are made, we cannot in any way know whether they are recorded before or after the videos become popular, and thus we cannot investigate the *causal order* of the two phenomena. This information is not provided by the platform in any form and we hence cannot investigate the causal relationship between fake views and popularity.

5.4 Discussion

The analysis of our data reveals that fake views are widespread and that most fake views are corrected relatively late in the life of YouTube videos. This observation rises concerns because of the key role of early popularity in determining overall visibility. Indeed, rich-

get-richer dynamics have been repeatedly observed in the evolution of views counts on the platform [Bor+12], and the total number of previous views has been credited as the most important predictor of future popularity [Bor+12] [SH10] [PAG13]. Furthermore, rewarding trendy contents with more visibility is a central feature of the recommendation system, which –according to YouTube developers– aims at "bootstrapping and propagating viral content" [CAS16]. While encouraging diffusion of trending videos, the recommendation system also constitutes the main source of views for most YouTube videos [ZKG10b] [Zho+16] and hence regulates the attention economy of the platform. Its suggestions are updated relatively quickly: according to Roth et al., for instance, two thirds of the suggestions are associated with a given video for less than two days [RMM20]. Such a fast refreshing of suggested videos, along with the massive impact of recommendation system, warrants concerns about delays in fake views correction. With a delay in correction, videos has handsomely enough time to be recognized as viral and thus be pushed to a wider audience than the one they would have achieved without illegitimate boosting.

Our work has brought to light the vast amount of channels and videos concerned by this phenomenon and its features in term of rhythms and frequency. We have also highlighted the existence of a positive correlation between illegitimate and real views. In the absence of first-hand data on fake views corrections, we cannot rule out the possibility that it is fake views that boost the popularity of some videos. Given the importance of the subject and the potential harm from the malfunctioning of the correction policy, our findings should –at the very least– encourage YouTube to include in its API the number of corrected views for each videos, as well as their history. As we have shown, this information is crucial to investigate the alarming possibility that techniques of views inflation could be used to manipulate video visibility by triggering viral dynamics sustained by human and algorithmic recommendation.

5.5 A Bot Experiment

Goal. To understand what level of control YouTube applies to genuine views and when it corrects illegitimate ones, we wrote a bot able to record views automatically. To check its effectiveness, we created a dedicated channel where we uploaded short new videos. In this way, we tested whether our views were recorded or not, excluding the presence of some outside audience as much as possible. By writing this bot, we could distinguish two types of checks made by the platform. The first ones are *real-time* checks: some automated views are spotted immediately and not even added to the count. Then there are some *deferred checks*, that result in the view count drops discussed in the previous sections. In the following we discuss how the real-time controls performed by YouTube work in practice. We think it is important to discuss them in order to understand that bypassing them is not trivial: given the amount of precautions that must be taken to register illegitimate views, it is very

unlikely that they can be recorded by accident, hence their amount in our hourly dataset becomes even more relevant.

YouTube real time checks on fake views

By implementing an algorithm that records views automatically, we could empirically identify some of the real-time checks the platform makes on new views. In practice, certain browser configurations did not allow us to record views one after the other. YouTube tries to make sure that it is not always the same user viewing a piece of content in a loop. To do that, it does not count:

1. consecutive views made by the **same logged-in user**;
2. consecutive views made from the **same IP address**;
3. consecutive views made by the **same browser** (recognizable by cookies and cache).

Overcoming some of these checks is easy, others require more effort. The IP address (2) control is easily overcome through a VPN with enough IP addresses to change them each time we make a new view. Check (3) can be overcome through a clean browser, opened in incognito mode, without cookies, and with empty cache memory. In our case, we used Brave Browser. The hardest control to overcome is the one related to not logging in (1). Most YouTube users access the platform through an automatic log-in, sometimes without being aware of it, so our readers will probably not be familiar with what happens to not logged-in users. In this case, YouTube proposes pop-ups that ecourage users to log in and ask them to accept the platform's Terms and Conditions (see Figure 5.6). In order to access the page and view the videos, it is necessary to accept or close these pop-ups. This need for interaction with the browser makes this the trickiest control to overcome: pop-ups are different and do not always appear in the same order, which makes it difficult to automate their closing. A way to overcome this check is to integrate some CSS code on the browser (though an extension) to automatically close pop-ups on pages belonging to the YouTube.com domain. However, we notice empirically that, in this way, many views do not get counted. Therefore, we opted for using the openCV (open source computer vision) library for Python, able to recognize screen areas, like the buttons to close the pop-ups.

In addition to these controls on the users, the platform also applies controls on how long a video is watched. This choice is in the wake of the platform's need to keep interactions meaningful [Noaa]. Asking to watch a video for a certain amount of time is a tool the platform uses to test if there is genuine interest from the user, to avoid the situation in

which he/she ends up on a content by accident (e.g., through Autoplay). Empirically, we observed that spending a minute on a video suffices to get the views counted.

It is important to point out that despite our efforts to pass checks (1) (2) and (3), there is one last control that is difficult to circumvent. Suppose much traffic to the same YouTube page comes from the same IP address, which is often the case using VPN services. In that case, YouTube places higher barriers to access, such as Captcha tests [Goo] asking to identify some images to proceed. In this case, we cannot overcome the barrier imposed by YouTube, and consequently our views are not recorded. However, we observe empirically that YouTube proposes Captcha tests less than once in 10 times and thus, they do not dramatically affect our algorithm.

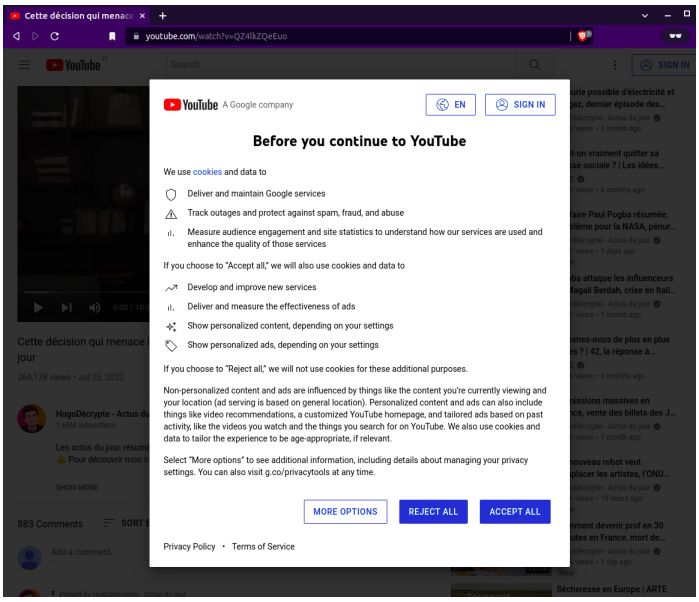


Figure 5.6: Example pop-ups proposed to non-logged users.

Bot implementation

Once we have discussed the checks performed by YouTube and how to overcome them, we can move on to describe how our algorithm worked in practice. To record more than one view at a time, we used two computers connected in parallel through an SSH port. The "main" computer chose the video on which to perform the views and scheduled them. For each of the scheduled views, the algorithm:

1. connects both computers to a different server of a VPN;

2. opens the URL of the selected video with a clean browser in incognito mode. The browser was previously set to allow auto-play on videos and not show ads;
3. waits a minute and a half to be sure to have enough time to upload the web page and watch the video for at least one minute;
4. closes the browser;
5. disconnects the VPN.

To ensure that step (3) allows us to watch the targeted video for at least a minute, we need additional foresight at the browser level: we need to install an extension that can block advertisements at the beginning of videos. Since every time we open the browser a different advertisement (with a different length) may show up, without removing advertisements we would not know how long to wait on the page to view the video for at least one minute.

5.5.1 An Online Experiment

Once the bot was set, we tested it on real videos. This test served to verify that our views were recorded and passed YouTube's real-time checks. It also constitutes a first attempt to study the time it takes for YouTube to recognize illegitimate views. The videos on which we tested our bot and their hourly time evolution are presented in Figure 5.7.

Video selection. We chose videos published by channels in our list (see Section 3.3) so that, at the same time, we could collect their views evolution. We chose videos published no more than 24 hours earlier in order to always intervene in the first hours of their lives.

Scheduling of views. We decided to schedule on each video a number of views proportional to the average views of its channel. In particular, we decided to schedule views for 0.01 times the average views per channel.

Results. The resulting view counts are shown in Figure 5.7. The thick blue line indicates the cumulative evolution of views on our target videos. The vertical light-blue area shows the bot's activity period, while the red line indicates the number of views scheduled during the activity period. Videos are sorted based on the quantile they represent in the distribution of views of their channels, to give an idea of their relative popularity. The first ones, framed in green, are relatively unpopular, and their number of views is lower than that of 50% of videos on their channels. The others are relatively popular videos, representing quantiles above the median of their channels. From these plots, we can draw the following conclusions:

1. **Not all the views we programmed are recorded**. For example, we see that in Figure 5.7 (a), (b), (c), (d) we planned to record more views than those that we see

at the end of the bot's activity window. Harder interposed checks, such as Captcha tests, could be the cause of this mismatch between scheduled and recorded views and explain why some were not counted.

2. **Delays in corrections**. Corrections are particularly visible in some cases, for instance in Figure 5.7(a), (b), (c), (d), (e), (f). In those cases, we see that they occur about 24 hours after the start of the bot action. To study the phenomenon of corrections in the cases in which it is less visible, we reconstructed the corrections' signal through our "Reconstruction Method" presented in Section 5.2.2. Figure 5.8 presents the time series obtained. Also on the reconstructed signals, we observe that, in most cases, corrections occur about 24 hours after the bot starts recording its views.

3. **With popularity comes fewer corrections**. Unpopular videos get corrected for the majority of the recorded fake views (see Figure 5.7(a), (b), (c), (d)). Instead, videos with more than a thousand views, show fewer visible (and reconstructed) corrections. For instance, for video (w), only 10 views out of the 1032 we scheduled are removed. For video (p), only 151 views out of the 1314 planned are removed. In addition, video (z) does not show any sign of corrections. We can explain this evidence in different ways: the first, more probable explanation is that, by collecting views every hour, we lose information about the corrections. In that case, we should remark that also the reconstruction method presented in Section 5.2.2, does not succeed in recovering all the information lost. Another explanation, less probable but still interesting, could be that unpopular videos are actually more corrected because they make fake views are easier to detect. Let us look, for example, at Figure 5.7(b) and let's put ourselves in YouTube's shoes. If, out of 800 views, 655 are made by non logged-in users and they are all done exactly 1 and a half minute apart, it seems rather probable that they are fake. On the other hand, in the case of Figure 5.7(z), where we recorded 1.340 views starting when the video already had more than 6.000, it might be more difficult for YouTube to detect the fake ones, hidden among the real ones. It seems plausible, and not to be ruled out, that fake views are easier to detect in the absence (or near absence) of real views.

5.5.2 Discussion

We discussed these examples to show how the bot works and how it could be used to study YouTube's correction policy in the future. Compared to only studying corrections in the hourly dataset, the advantage it brings is the knowledge of when fake views are recorded. This knowledge could in the future help investigate the existence of a causality behind the correlation shown in Figure 5.5. The 28 examples in Figure 5.7 are still too few to draw conclusions about how YouTube corrects fake views, especially because some of the outcomes

lend themselves to different, and hence unclear, interpretations. In this section we want to summarize what emerged from this experience what the limitations of our bot are when applied to real videos.

To begin with, we have experienced firsthand that recording views is not trivial. This makes the 22.5 million illegitimate views in our dataset even more interesting since it seems very difficult for them to have been recorded by chance or by unaware users. It seems more probable that behind most of them lies a specific intention, probably motivated by the profits that come along with views. Although recording fake views is not trivial, at the same time, it is also not impossible. A person with some average programming knowledge, a VPN access, and a little goodwill may be able to pass YouTube's real-time checks. Fortunately, real-time checks are not the only ones the platform performs, and deferred checks could stop programmers from boosting targeted videos.

Applying our bot to real videos, we observed that, on average, YouTube removes our fake views within 24 hours. The first question we could ask is whether this is a sufficiently short time. The second question is whether this result concerning our bot implies that all fake views are corrected within 24 hours. Regarding the first question, as discussed in previous sections, we know that 24 hours is often sufficient time for a video to reach most of its audience. We can look at videos (o) (p) and (s) in Figure 5.7 to realize that most of the growth, even for popular videos, can occur in the first 24 hours. Consequently, we cannot rule out the possibility that in 24 hours fake views could already affect human and algorithmic recommendations. Regarding the second question, clearly, we cannot exclude the possibility that other types of bots could be corrected after 24 hours. Our bot is a rather rudimentary version of a bot that can record views on YouTube. One flaw among all, it does not log in. This limitation certainly makes it easily detectable since most real users do log in. To overcome this problem, we would have to create fake accounts, which is surely more complex. Without going that far, options for making our bot less identifiable could include making the frequency of access to videos less deterministic, avoiding a constant cadence of accesses, and making the time spent watching the video random. Testing different versions of our bot could help better understand the time needed for YouTube to identify fake views.

Finally, in future applications of this bot, it would be necessary to collect the time series of views with a frequency every 5 minutes. Indeed, as shown by the previous examples, some results are difficult to interpret without the exact data on the number of views removed. Increasing the frequency of collection could give clearer answers and help the interpretation of the results.

Figure 5.7: Bot at work. Blue line indicates the cumulative evolution of views on our target videos. The vertical light-blue area indicates the period of activity of the bot, while the red line indicates the amount of views scheduled during the activity period. Videos in the green box are less popular with respect to half of the videos in their channels.

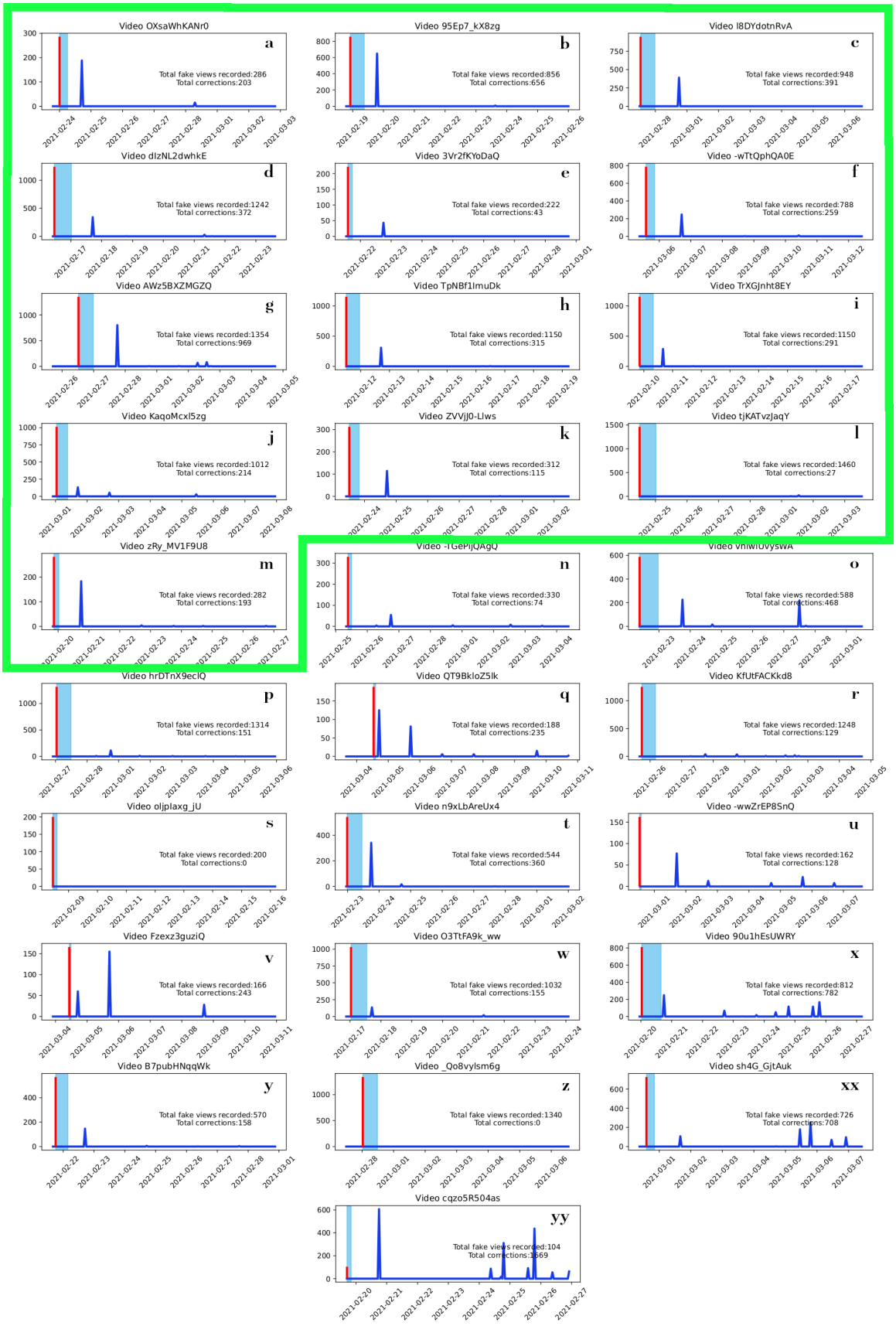


Figure 5.8: Bot at work. Blue line indicates the corrections reconstructed with our "Reconstruction Method". The vertical light-blue area indicates the period of activity of the bot, while the red line indicates the amount of views scheduled during the activity period. Videos in the green box are less popular with respect to half of the videos in their channels.

Chapter 6

How Covid Disrupted Online Rhythms

Contents

6.1 Introduction

Collecting data on social networks is always affected by the historical moment in which it is done. Indeed, social networks constitute a snapshot of society and people's habits: they are a mirror of fashions, trends, debates, and issues dear to the communities that produce content, whether these are videos on YouTube or posts on Facebook or any other platform. When major events, capable of distorting trends and diverting topics, hit the lives of users obviously the content shared online bears the scars of such upheavals.

The Covid-19 pandemic undoubtedly constituted one, if not the largest, of such events throughout social media history. The pandemic, accompanied by the first lock-downs established in various countries around the world, pervaded every aspect of people's daily lives, constituted a huge upheaval in human habits, and its effects are certainly observable through

what users shared and disseminated in the months following the outbreak of the pandemic. No event in the (fairly recent) history of social media has ever been of this magnitude, no event has ever so closely affected the lives of every single person across the entire planet.

As the reader might imagine, the data we have collected over the past 3 years have evidently been affected by the Covid-19 pandemic in multiple ways. For many months the pandemic partly prevented us from observing YouTube in a "normal" context. People found themselves overnight unable to leave their homes, unable to have well-defined work schedules, unable to practice their sports or hobbies, unable to meet in bars among friends. This has incredibly increased the use of online platforms to communicate and share content, to keep in touch or to get informed. Online platforms actually experienced a boom in users and content, which led them to experience a period that we might call extra-ordinary and not very representative of the normal use made by people.

Hence, when looking at our YouTube data, we cannot ignore what has happened in the world since March 2020. What we can do, however, is take advantage of the incredible richness given by having collected data during this very period to study and understand what happened in people's lives during the early lockdowns. That is the reason why a few months after the end of the first lockdown, we decided to analyze the visible disruptions in our data in terms of platform access rates and topics covered by the content creators. Specifically, we leveraged this opportunity to study what the ordinary rhythms of platform access are, compared with the extraordinary rhythms induced by the disruption of people's daily lives. Furthermore we analyzed the changes in terms of themes and emotions shared online. We sought to understand how Covid-19 affected YouTube users in emotional terms and took the opportunity to observe what features of the content covered remained resilient to a shock of such magnitude.

To give the study a broader scope and verify that the results were not specifically related to YouTube, we decided to compare and integrate the results with a dataset from Twitter. Introducing this dataset has been very useful: it allowed us to recognize some peculiar characteristics of YouTube in terms of circadian rhythms and topics covered on the platform. In light of the data, YouTube appears marked by higher nighttime activity compared to Twitter, and its share peaks occur later in time. At the level of emotions and topics, YouTube also shows differences from Twitter, coming across as a social network marked by a strong presence of emotions, positive ones in particular.

6.2 A Comparative Study with Twitter

Since the activity and the topics discussed online are strongly dependent on users' demography and on platforms' scope, we decided to analyze them not only on our YouTube hourly

dataset but also on Twitter. As already discussed when talking about the limitations of APIs, Twitter is actually more permissive in terms of data collection and it allowed us to retrieve a rather large dataset.

6.2.1 The Twitter Dataset

The Twitter dataset comprises about 8 millions tweets, retrieved through the official Twitter API, posted by 5161 active but non-professional users from February 17 to April 14. These non-professional users have been identified within a wider dataset of about 33 millions tweets containing Covid-19-related content. This corpus of tweets was collected with the help of Science–Po MediaLab in Paris, by using the python based scraper Gazouilloire [Med], a tool developed by Dime Web for systematic and configurable Twitter data collection through Twitter's official API. The data were collected based on a query of Covid-19-related words in French. All twitting and re-twitting times were collected in European Central Timezone (UTC +1). Being Covid-19 the most trending communication topic in France (as we can observe for example from Google Trends), since its first diffusion in Europe and even more around the lockdown decision, we can assume that a relevant part of the French Twittosphere is potentially present in this database.

Differently from the YouTube database, which does not allow extracting information on the users, for the Twitter database we can define precise profiles of the users we are interested to study, in order to have a more homogeneous, even if reduced, population. Since we are interested in the activity of common Twitter users, we decided to exclude newspapers, bloggers, radios, associations, etc. and only consider non-professional users. Moreover, recalling that we needed the Twitter dataset to be comparable to the YouTube one, we had to restrict the content production on Twitter to the sole production taking place in France. The requirements listed above, together with the filtering they set off, can be clarified as follows:

- facing the need to exclude professional users, we only considered those (1) whose profile did not contain keywords associated to professional use of the platform (e.g. "media", "blog", "official", etc.); (2) with a number of followers lower than the median of the whole dataset; (3) with an activity lower than the median activity ($\sim$400 tweets by week);
- targeting a significant statistical analysis, we discarded users who published less than 100 tweets in the whole period;
- focusing on France, we filtered the sole users who explicitly declared their location to be in France, searching in the location descriptions multiple translations of "France",

the name of all the cities in France with a population larger than 10000 inhabitants (with eventual translations), the French regions and departments.

This filtering left us with 5161 users: their Twitter profiles' descriptions have been manually checked to be sure that the non-professional selection mechanism was effective. Finally, in order to construct the dataset of tweets, the entire timelines of the selected users have been collected through the Twitter API, including tweets not related to Covid-19. Even though this sample is not meant to be representative of the full French Twittosphere, it provides a complete perspective of a sizable and homogeneous set of active non-professional users of the platform over the relevant time frame.

6.2.2 Overcoming Fake Views Corrections on the YouTube Dataset

For the analyses provided in this Chapter, we used the hourly dataset presented in Chapter 3. The hourly dataset was restricted to a time window that runs from February 17 to April 14, 2020. Narrowing our dataset to this time window allows us to study as many lockdown weeks as previous weeks. With this choice we find in our dataset 99,992 videos published by 1031 channels.

As we want to investigate online viewing rhythms and the types of content watched, it is essential to understand how many people viewed a video and when. The time series collected in our dataset, combined with the title and description of videos, are excellent for this analysis. Nevertheless, as discussed in Chapter 5, they sometimes present negative deltas in the number of views, representing fake views recorded in previous hours. Hence, when looking at an hour with a negative difference in views, we know that some previous hours did not have as many views as they seemed, as some were fake. Since these illegitimate views risk altering our results, we must preprocess our data before proceeding with the analysis. In the absence of information about the time illegitimate views are recorded, we decided to redistribute them uniformly in the hours before their corrections. More precisely, if we call v_h the views collected by a generic video at hour h after publication and T_h the total number of views at hour h, if $T_{h+1} < T_h$, we correct the time series as follows:

$$\hat{T}_j = (1-p)\, T_j \qquad j = 1, \ldots, h$$

where $\hat{T}_j$ is the corrected time series until hour h and $p = \frac{T_h - T_{h+1}}{T_h}$ is a percentage of correction.

6.3 Habit Changes During the Spring 2020 Covid-19 lockdown

6.3.1 An Increased Online Activity

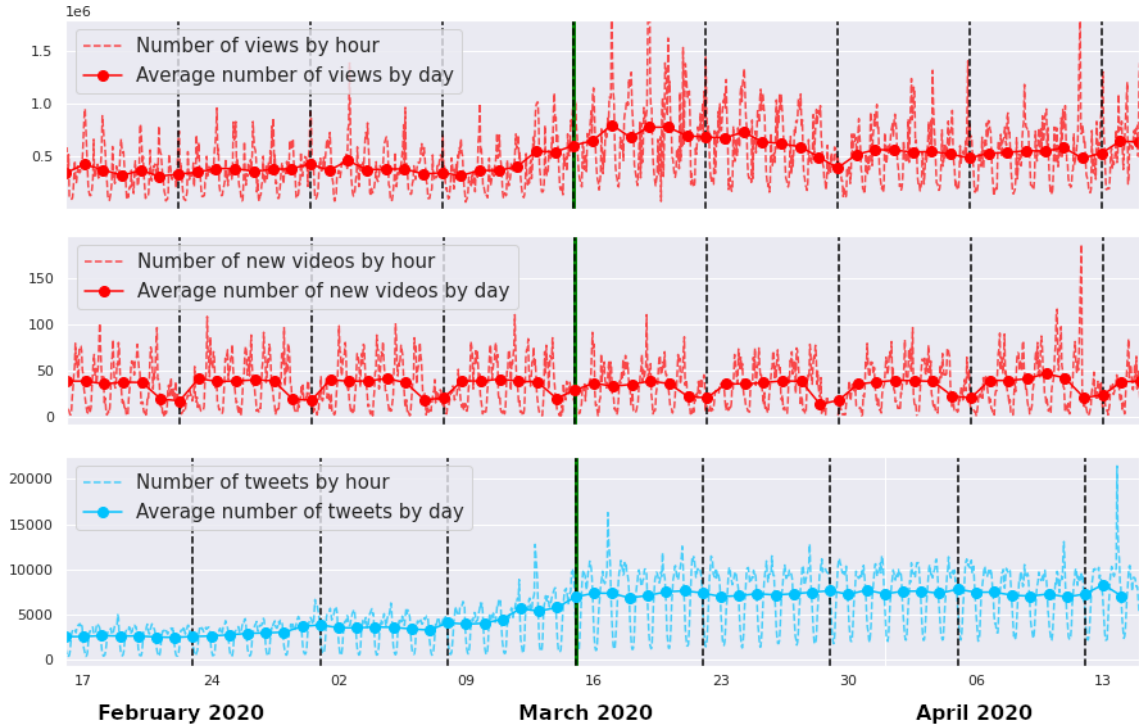


Figure 6.1: Increase of platforms activity after the lockdown. (top) Evolution of number of visualizations by day on YouTube. (middle) Evolution of number of new published videos on YouTube. (bottom) Evolution of number of Covid-19 related tweets or re-tweets

The first major habit change generated by the 2020 spring lockdown in France is a considerable increase in the online activity. Recalling that the French first lockdown was announced on March 15 and enforced on the 17th, we can better understand the time series for Twitter posting, YouTube posting and YouTube views displayed in Fig. 6.1. All three time series show weekly and daily fluctuations. The average daily signals, smoothed by a moving average over a 7 day rolling window, reveal an increase of activity for Twitter posting and YouTube watching around the beginning of the lockdown. Notice that on both platforms the increase of activity started from the very moment the lockdown was announced (which itself sparked much debate).

As for the posting of videos on YouTube, such activity is less casual and more stable than tweeting and therefore conserved the same weekly and daily rhythms during the lockdown. Since the video production is hardly affected by the lockdown, we will not consider this

dimension in the rest of analysis. In the following, we will refer to the period *before* the lockdown as to the three weeks from February 17 to March 9. At the same time we will refer to the period *after* the lockdown enforcement as to the three weeks from March 23 to April 14. To exclude the transient effects of the transition phase, we discard the data about the two weeks around the lockdown onset (from the 9th to the 22nd of March).

6.3.2 Lack of Sleep and Night Online Activity

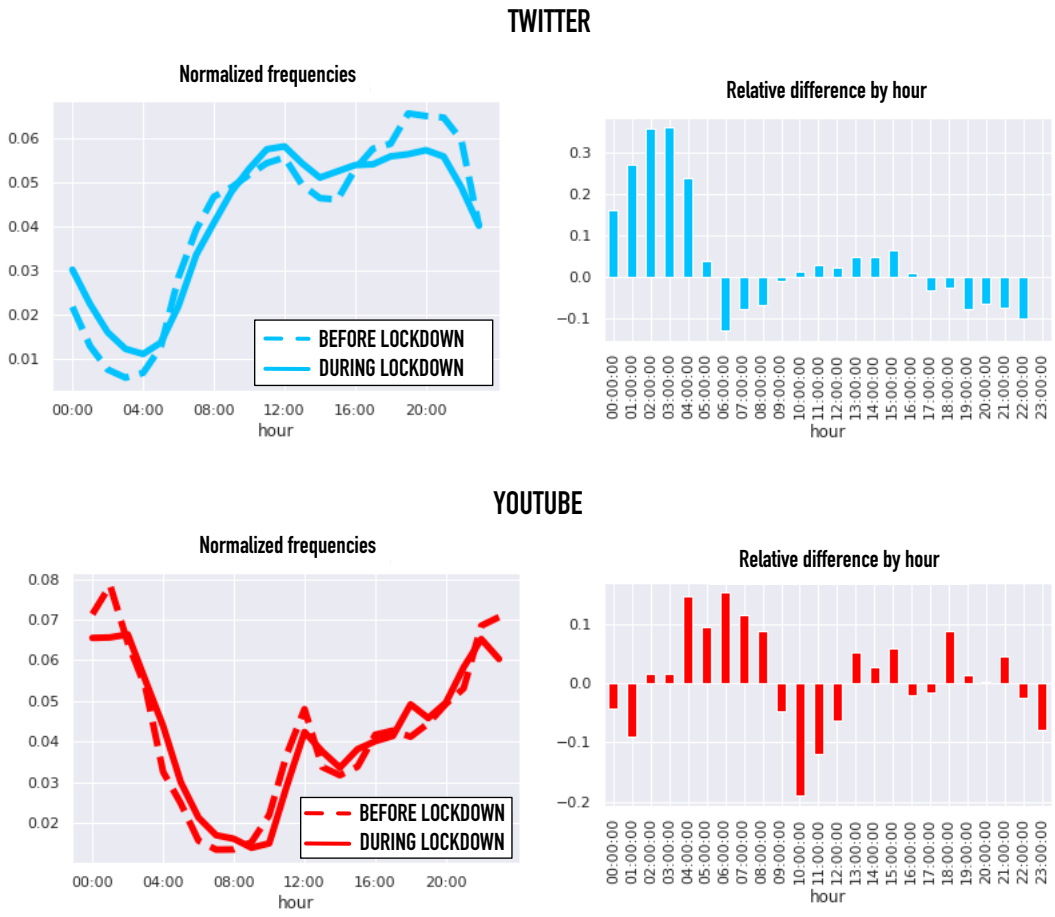


Figure 6.2: Circadian Rhythm Changes. On the left, Twitter and YouTube circadian rhythms before and after the lockdown. On the right, we explicitly evaluate the relative differences between rhythms before and after the lockdown

To understand whether the increase in the activity was uniform during the day or it affected some moments specifically, we calculated the normalized daily activity profiles before

and after lockdown for each hour of the day by

$$f^{\text{Twitter}}(h) = \frac{\sum_{d\in\text{days}} N_{\text{tweets}}(d,h)}{\sum_{d\in\text{days}}\sum_{h\in\{0,\dots,23\}} N_{\text{tweets}}(d,h)} \tag{6.1}$$

and

$$f^{\text{YouTube}}(h) = \frac{\sum_{d\in\text{days}} N_{\text{views}}(h)}{\sum_{d\in\text{days}}\sum_{h\in\{0,\dots,23\}} N_{\text{views}}(d,h),} \tag{6.2}$$

where $h \in \{0,\dots,23\}$ are the hours of day, d are the days considered, and $N_{\text{tweets}}(d,h)$ and $N_{\text{views}}(d,t)$ are respectively the number of tweets and of YouTube views at hour h of day d. The results are reported in the left plots of Fig. 6.2. We first observe that the profiles for Twitter and YouTube are quite different: while Twitter is mostly used during the day, with a strong activity decrease after midnight, YouTube is characterized by a higher night activity. While Twitter is an active media, characterized by a debating and prosuming culture [RDJ12] that encourages participation at the time of the day when engagement is maximum, videos watching on YouTube is, for many users, a more passive activity [Kha17] which can easily fit more relaxed late hours.

To quantify the differences between profiles before and after the lockdown, we calculated the relative differences of the normalized profiles:

$$\delta(h) = \frac{f^{\text{after}}(h) - f^{\text{before}}(h)}{f^{\text{after}}(h) + f^{\text{before}}(h)} \tag{6.3}$$

This quantity is reported in the right plots of Fig. 6.2. Both YouTube and Twitter experienced an activity increase during the night and a smaller decrease of the activity in the early morning (6am-9am for Twitter and 9am-12am for YouTube). We observe that the morning decrease in Twitter is much smaller than the night increase. This suggests that, with the lockdown, people stayed longer awake during the night but without oversleeping in the morning.

To appraise the variations in the activity during lockdown nights, we compare this variation with another factor known to impact online circadian patterns: the weekly cycle of weekends and working days. We aggregate the data at the level of day and night: based on the shapes of the curves in Fig. 6.2, we consider night on Twitter the hours between 11PM and 6AM of the following day and on YouTube the hours between 1AM and 8AM. We decompose the aggregated activity into weekends and working days. To obtain comparable measures we divide the activity counts by the number of hours of the corresponding time period (night=7, day=17) and by the number of days of week parts (weekend=2, working days=5). The average number of tweets/views in the different categories are represented

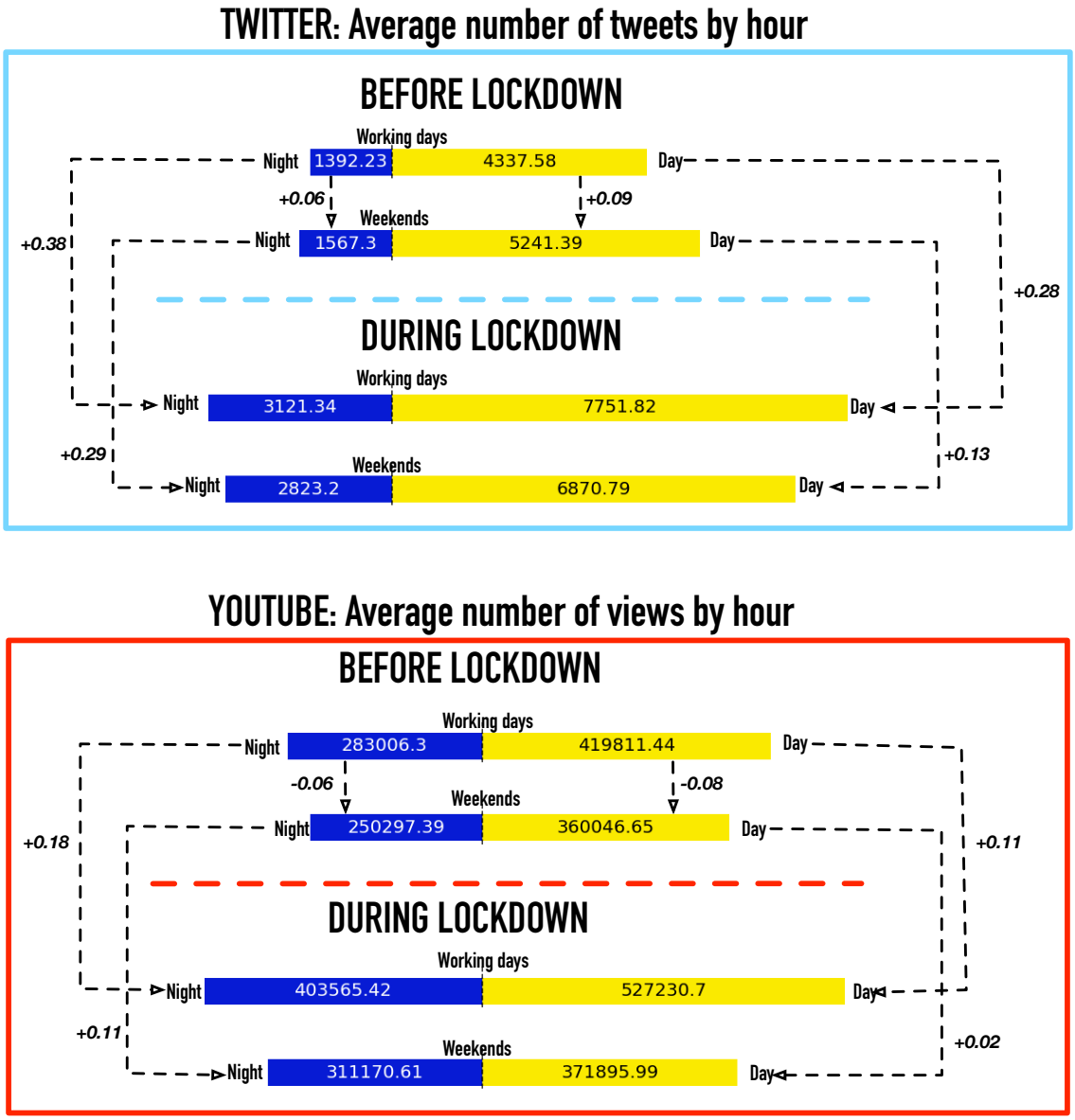


Figure 6.3: Night-vs-day and working day-vs-weekend patterns. Average number of Tweets (upper plot) or YouTube views (lower plot) by hour, during day and night, working days and weekends. The numbers next to the dotted lines represent the relative increment between the related quantities.

in Fig. 6.3, together with the relative changes among the classes. As we already observed, the night variations are the largest changes, both in weekends and working days. The ac-

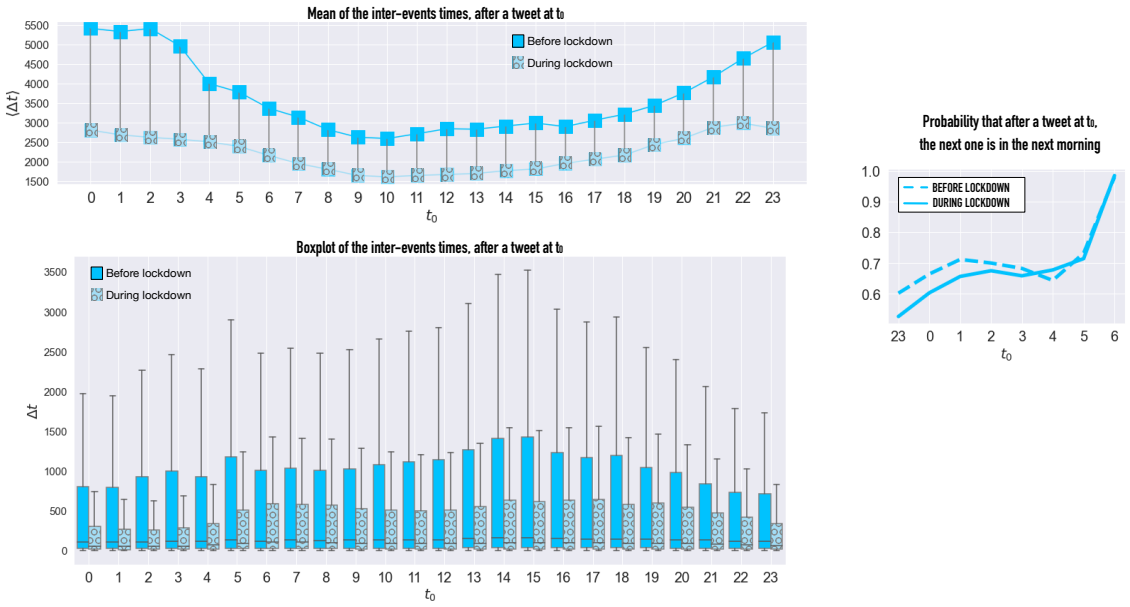


Figure 6.4: Interevents. Left plots: Average inter-event distance for an event starting at time t_0 and boxplot of the interevents starting at t_0. Right plot: Probability that a tweet following a tweet at t_0 is posted on the following morning (dashed=before lockdown, solid=during lockdown).

tivity increments due to lockdown are significantly larger than the variations associated with the normal week cycle. In YouTube the variations also have opposite signs: while the activity normally decreases during weekend, it increases for lockdown. For both the platforms, and both for day and night, the most significant difference concerns the working days: the augmented social media usage seems to be replacing the time previously dedicated to daily routines (like commuting to work, going to sleep early, etc.) more than the time of recreational activities.

To confirm the hypothesis of reduced sleep during the lockdown, we analyze the situation at the individual level, when possible. As we do not have information on the users that watched a video on YouTube, we limited this analysis to Twitter. For each Twitter user, we calculate the average time lag between two consecutive Tweets. Fig. 6.4 shows the hourly average of this measure (and the boxplots to appreciate its variability) before and after the lockdown. While, in normal times, the average inter-event times are much higher during the night (because of sleeping breaks), the lockdown flattened the curve, thereby suggesting a shortening of sleep intervals [Bec+20]. In the right plot of Fig. 6.4 we display, for each hour of the night, the probability that the subsequent tweet is posted the following morning (between 7 AM and 12AM) rather than during the same night. This quantity represents the probability to "go to sleep" after an event at t_0. For this analysis we only considered the

inter-events that are longer than 1 hour, in order to get the last action only in a potential activity burst and reduce the noise. We also excluded all the inter-events finishing after the next morning. We see that, until 4AM, the probability to get asleep is lower during lockdown. These findings are based on an aggregated statistical study of the inter-events. Unfortunately, the short time span of our data collection (2 months) does not allow a more sophisticated analysis, for example, of the users chronotypes, which, as shown by [ALS18]; [Ale+15] can be extremely important for understanding the individual sleep behavior.

6.3.3 Themes and Emotional Change during the Lockdown

Once analyzed how the lockdown has changed the rhythms of access to the platform, we can now turn to how the content on the platform has changed. In order to do this, we needed to label each video and each tweet according to the themes covered and the emotions present. To do this we used a well-known and tested tool: the LIWC [TP10] dictionary. The LIWC dictionary classifies words according to more than 70 emotional, stylistic and thematic dimensions and has been used in several similar studies such as [GM11]; [DLC17]. More precisely, the LIWC dictionary provides a list of words associated with different dimension (e.g., the category "positive emotions" contains words such as: happy, beautiful, and good). Since the texts analyzed are written in French, we used the French version of the LIWC dictionary [Pio+11]. Using the categories of the LIWC dictionary we will consider three separated analytic dimensions:

- General Affects: Positive Affect, Negative Affect;
- Specific Emotions: Sadness, Anger, Anxiety and Accomplishment:
- Thematic contents: Work, Social life, Religion, Death, Fun, Exclusion, Biology, Money.

We first assign the tweets and YouTube videos (based on the words contained in their titles and descriptions) to one category for each dimension. To do so, we count how many terms from each category are contained in each tweet/video, and we label it with the prevalent category. For example, for the dimension "General Affect" each content is categorized as either Positive or Negative Affect or not classified if the items contains no categorical words or similar proportions of positive and negative terms. We also consider the global emotional level ("Affect") of the items, by counting together the positive and negative words. For each dimension, we compute the fraction of tweets and retweets in each category during the lockdown and the difference compared to the previous period. In the same way, for YouTube, we evaluate the fraction of views of videos in each category over the total number of views. The results are reported in the left plot of Fig. 6.5.

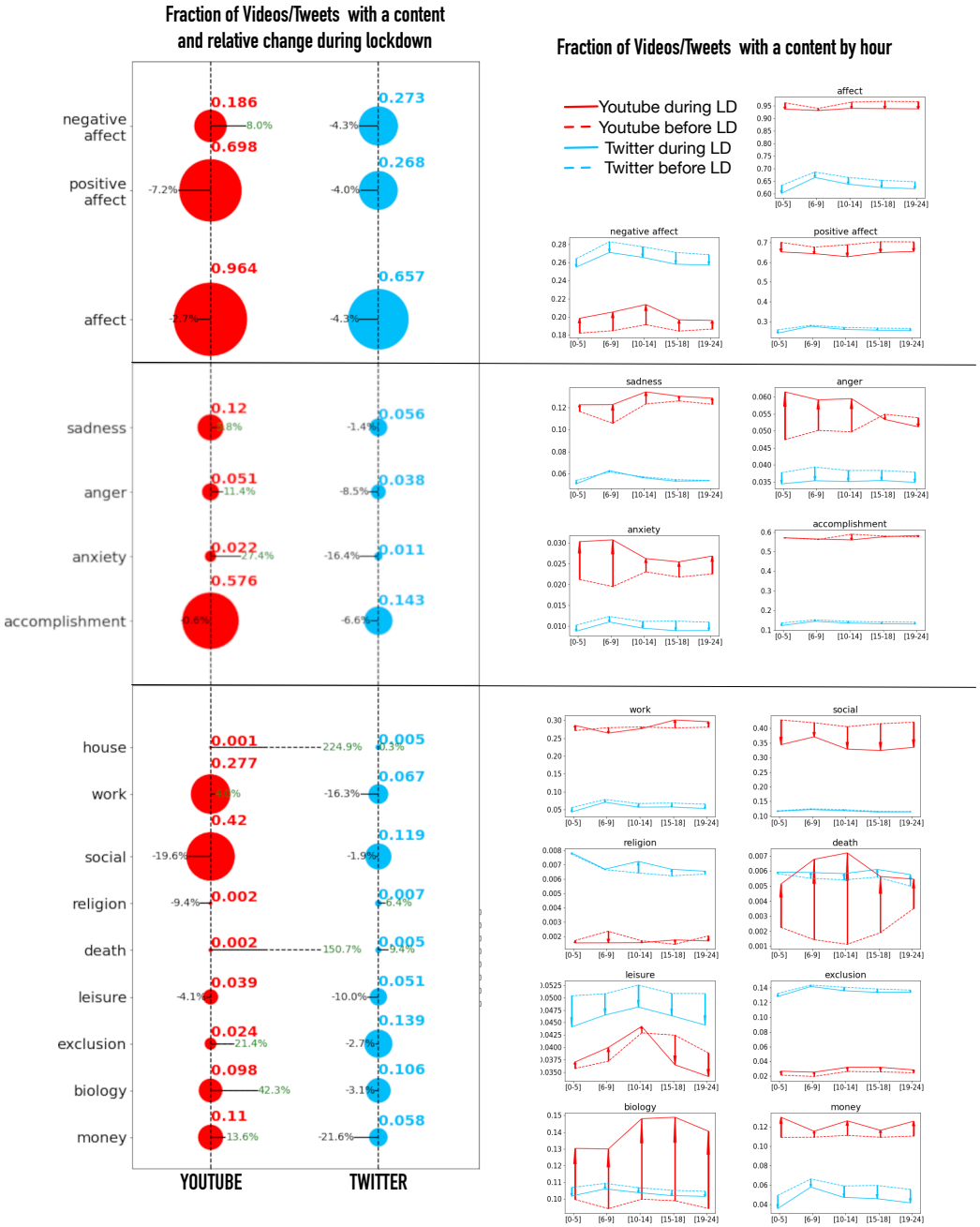


Figure 6.5: Themes and Emotions before and after lock-downs Left plot: Fraction of videos/Tweets with a content and relative change with lockdown. The size of the points is proportional to the fractions (quantified by the upper numbers). The orientation of the line indicates if there was an increase (orientation toward right) or decrease (toward left) with the lockdown. The length of the line is proportional to the percentage increase/decrease with the lockdown. Right plot: Fraction of videos/Tweets with a content by hour. The continuous line indicates the fractions after the lockdown, the dotted lines before. An arrow starts from the before to the after line for each hour period: if the arrow is oriented towards the top, the lockdown increased the content fraction in the selected hours and viceversa. In both plots YouTube is in red, Twitter is in blue.

Table 6.1: p-values of Kolmogorov-Smirnov tests for YouTube content differences displayed by Fig. 6.5, i.e. content differences before and during the lockdown

Category	Avg Before Lockdown	Avg During Lockdown	KS p-value
Negative Affect	0.187	0.202	$2.29 \cdot 10^{-9}$
Positive Affect	0.693	0.647	~ 0
Affect	0.961	0.938	~ 0
Sadness	0.118	0.127	$7.56 \cdot 10^{-8}$
Anger	$5.17 \cdot 10^{-2}$	$5.78 \cdot 10^{-2}$	$3.16 \cdot 10^{-12}$
Anxiety	$2.15 \cdot 10^{-2}$	$2.88 \cdot 10^{-2}$	$2.31 \cdot 10^{-7}$
Accomplishment	0.575	0.571	0.075
Work	0.278	0.285	0.0025
Social Life	0.419	0.341	~ 0
Religion	$1.91 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.29 \cdot 10^{-4}$
Death	$1.96 \cdot 10^{-3}$	$5.64 \cdot 10^{-3}$	~ 0
Leisure	$3.95 \cdot 10^{-2}$	$3.77 \cdot 10^{-2}$	$1.52 \cdot 10^{-10}$
Exclusion	$2.35 \cdot 10^{-2}$	$2.85 \cdot 10^{-2}$	$1.82 \cdot 10^{-14}$
Biology	$9.70 \cdot 10^{-2}$	0.14	~ 0
Money	0.109	0.123	$2.22 \cdot 10^{-16}$

Comparing the two platforms, we first observe that YouTube is more "emotional" than Twitter and is generally populated by more positive content. Both platforms experience a decrease of the emotional sphere during the lockdown, but on YouTube in particular we observe an increase of emotionally negative contents. Regarding specific emotions, we notice that expressions of accomplishment decline in both platforms. Instead, while Twitter experienced a decrease of all specific emotions, YouTube, which was already characterized by a higher level of anger, sadness and anxiety before the lockdown, goes through an important increase of these sentiments. From a thematic point of view we observe, unsurprisingly, a decrease of the contents related to social life and leisure and an increase of contents related to death and house. On Twitter we also have a significant increase of religion-related contents. All differences in distribution of contents before and after the lockdown have been tested statistically with Kolmogorov-Smirnov (KS) tests. For both platforms before and after the lockdown an hourly aggregation has been performed in order to get two different samples of emotion distribution. The results of the KS test over those distributions are reported in Table 1 and Table 2 and are most of the time statistically significant.

Table 6.2: p-values of Kolmogorov-Smirnov tests for Twitter content differences displayed by Fig. 6.5, i.e. content differences before and during the lockdown

Category	Avg Before Lockdown	Avg During Lockdown	KS p-value
Negative Affect	0.275	0.261	$3.11 \cdot 10^{-15}$
Positive Affect	0.268	0.256	$3.11 \cdot 10^{-15}$
Affect	0.659	0.628	$3.11 \cdot 10^{-15}$
Sadness	$5.55 \cdot 10^{-2}$	$5.54 \cdot 10^{-2}$	0.984
Anger	$3.92 \cdot 10^{-2}$	$3.50 \cdot 10^{-2}$	$3.11 \cdot 10^{-12}$
Anxiety	$1.15 \cdot 10^{-2}$	$9.39 \cdot 10^{-3}$	$3.11 \cdot 10^{-12}$
Accomplishment	$1.46 \cdot 10^{-1}$	$1.32 \cdot 10^{-1}$	$3.11 \cdot 10^{-12}$
Work	$6.96 \cdot 10-2$	$5.58 \cdot 10-2$	0.0025
Social Life	0.119	0.117	$1.46 \cdot 10^{-6}$
Religion	$6.87 \cdot 10^{-3}$	$7.14 \cdot 10^{-3}$	$3.85 \cdot 10^{-3}$
Death	$5.51 \cdot 10^{-3}$	$5.90 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$
Leisure	$5.34 \cdot 10^{-2}$	$4.57 \cdot 10^{-2}$	$3.11 \cdot 10^{-15}$
Exclusion	$1.37 \cdot 10^{-1}$	$1.34 \cdot 10^{-1}$	$1.73 \cdot 10^{-8}$
Biology	$1.06 \cdot 10^{-1}$	$1.03 \cdot 10^{-1}$	$4.20 \cdot 10^{-8}$
Money	$5.91 \cdot 10^{-2}$	$4.51 \cdot 10^{-2}$	$3.11 \cdot 10^{-15}$

6.3.4 Emotional Resilience

An interesting final analysis can be conducted by comparing the results on circadian rhythm with those on themes and emotions. We could ask, for instance, whether the changes in terms of emotions characterized only certain times of the day or were uniform across different hours. Previous work has shown interest in the association between the type of content and the time users share it, to identify different types of social media users [GM11]; [DLC17]; [Lam+13]: for instance, we know that content shared at night has lower emotional engagement. To perform this analysis, we first cluster the hours of the day into larger intervals to reduce possible noise.

Hours clustering

In this paragraph we cluster the hours of the day by means of the similarity of content shared. For Twitter, we build the sets $K(d, h)$ containing all the hashtags posted in day d at hour h. For YouTube, we build the sets $K(d, h)$ containing all the videos visualized in day d at hour h. We define the time similarity matrix, Θ, between two hours h_1, h_2, based

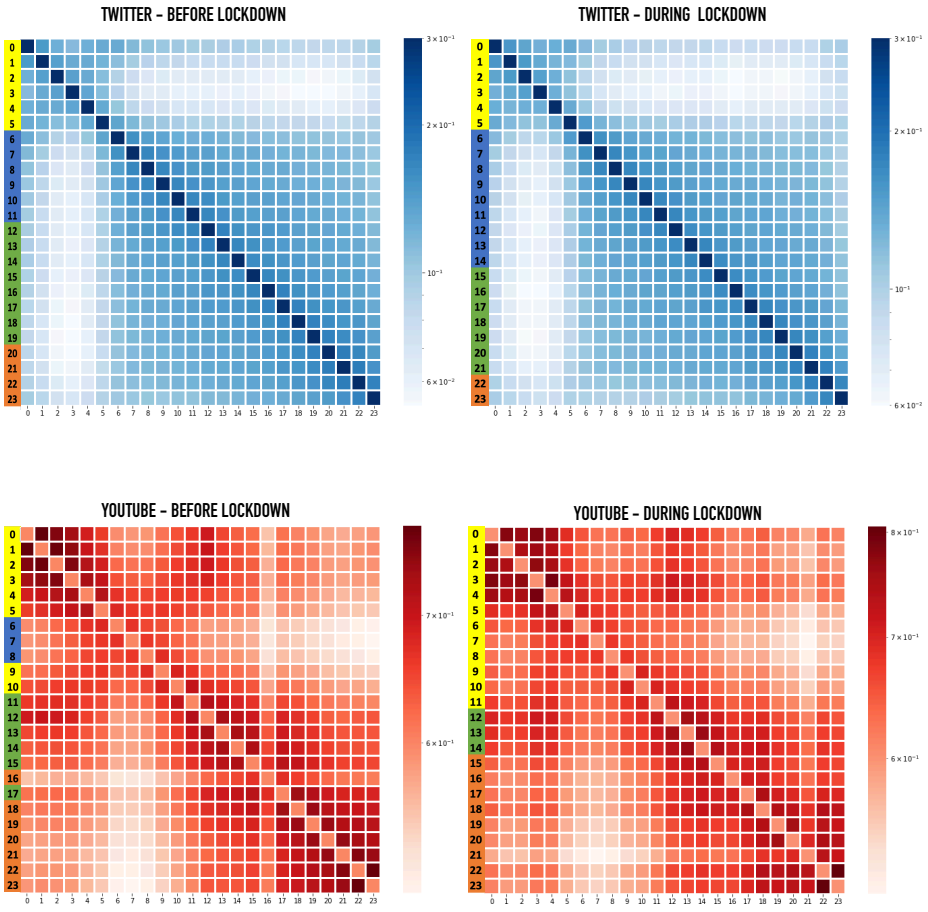


Figure 6.6: Hours Correlation. Heat-maps of the content hour similarity before and during the lockdown for Twitter and YouTube. The colors of the hours on the left of the graph represent the partition of the k-means clustering.

on the Jaccard similarity between the sets $K(d, h_1)$ and $K(d, h_2)$ as:

$$\Theta(h_1, h_2) = \frac{1}{N_{\text{days}}} \sum_{d \in \text{days}} J(K(d, h_1), K(d, h_2)) \tag{6.4}$$

where $J(K(d,h_1),K(d,h_2)) = \frac{|K(d,h_1)\cap K(d,h_2)|}{|K(d,h_1)\cup K(d,h_2)|}$ is the Jaccard similarity. Matrix Θ, represented in Fig. 6.6, indicates how the content shared or viewed at a certain hour is similar to the content in all the other hours. We perform a k-mean clustering procedure on this matrix to better identify the relationships between the hours of the day. The results of the clustering are represented by the colors of the hours on the left of the plots of Fig. 6.6. These clusters identify different periods of the day, and show how contents evolve along the hours. Both for Twitter and YouTube, and both before and after lockdown, night hours (0am-5am) are characterized by contents distinctively different from the rest of the day. Before the lockdown, morning hours (6am-10am for Twitter and 6am-8am for YouTube) were the most different from night-time and constitute a well definite cluster. For Youtube before the lockdown, the late morning hours 9am-10am showed a return of the night contents. For both platforms before the lockdown, we observe a lunch-afternoon cluster and an evening cluster. Lockdown affected the morning cluster, though in opposite ways for the two platforms. In Twitter, the morning cluster extended to first hours of the afternoon cluster (lunch hours), while in YouTube the night vibe extended into the morning (until 11am). In Twitter, the lockdown shifted the afternoon cluster after lunch time (3pm-9pm) and consequently reduced the evening hours (10pm-11pm). In YouTube, we observe the emergence of contents dedicated to the lunch hours (12am-2pm) and a second uniform block until covering the afternoon and the evening (3pm-11pm).

Emotions and themes variation per time of the day

Drawing on our previous hour clustering, we divide the day according to five time periods: [0am-5am],[6am-9am],[10am-2pm],[3pm-6pm],[7pm-12pm] to identify a circadian profile for each of our categories (right plot of Fig. 6.5). For several categories related to emotions, we can first notice an interesting Twitter/YouTube difference: what peaks in the early morning on Twitter [6am-9am] tends to peak in the following interval in YouTube [10am-2pm], thereby suggesting that YouTube content is consumed later in the day. Going into more detail, in agreement with the findings of [GM11]; [DLC17]; [Lam+13] we observe that nights are characterized by low emotional levels, especially positive ones, while both positivity and negativity tend to peak at the moment of the awakening. This pattern is more evident on Twitter also at the level of specific emotions, while on YouTube a significant portion of negative contents is consumed during the night.

An interesting observation, again in phase with the precedent findings of [GM11]; [DLC17]; [Lam+13], is that the daily emotional patterns seem to be resilient to the Covid-19 disruption: even if the volumes of some emotions changed during the lockdown, their daily distribution generally maintained the same shape, as demonstrated by a rough parallelism of the lines before and during the lockdown (with some exceptions regarding anger and anxiety on YouTube). This fact confirms the observation made in

[GM11] that external factors, even as important as the Covid-19 lockdown, influence the emotional patterns less than the sleep-wake cycles. Interestingly, this is not the case for the distribution of topics which has been more significantly influenced by the lockdown.

6.4 Discussion: YouTube, an Emotional and Nightly Platform

Collecting data over the past three years was an unexpected opportunity to observe an unprecedented phenomenon in human history: Covid-19 pandemic. It allowed us to investigate how human online behavior changed, which habits remained constant, and which were influenced by this external shock. In addition to that, given its comparative nature, this study helped shed more light on which features are peculiar to YouTube with respect to Twitter.

Regarding the changes in human habits, we observed that the distinction between night and day in online activity becomes less sharp in the absence of external constraints (like school or work). Circadian rhythms seem, therefore, strongly related to lifestyles and working schedules. Taking about the change of themes and emotions shared online, we find the decrease in the emotionally charged content rather unexpected. Although negative feelings such as anger and anxiety increased on YouTube, revealing a stressful and maybe even painful time for the users, the overall level of online emotivity decreased. One plausible explanation for this could lie in the disorientation following the pandemic outbreak. People, astonished by the news and afraid of the impact Covid might have had on their lives, awaited answers, living in a state of suspension in which it was not clear what was happening and how it would affect them. While, more expectedly, topics shifted to themes like biology and death, we observed users refraining from expressing emotions, probably because they did not know how to react to a situation they had never experienced before. Analyzing human behavior also allowed us to capture characteristics resilient to external shock. While the general emotional charge of online content decreased during the lockdown, it maintained its normal daily distribution. After March 17th, nightlife continued to be characterized by less emotional content despite the stress caused by the Covid-19 crisis. Even if the circadian rhythms changed and people stayed awake longer, at night they continued to share and consume the same content.

Putting the effects of Covid-19 aside for a moment, we can remark some of the differences between YouTube and Twitter brought to light by our comparative study. First, the typical access rhythms to the two platforms seem pretty different. Twitter is more used at lunchtime and in the early evening, up to 9 p.m., while YouTube is more accessed between 8 p.m. and 1 a.m. There are many possible explanations behind this finding. (1) We can argue that posting is a rather fast activity compared to watching a video. Watching a video requires

some time to dedicate to it, which might happen more likely at the end of the day. (2) Twitter's user base is generally older than YouTube's and likely more tied to steady sleep-wake rhythms dictated by work schedules. (3) Watching a video could be considered a more passive action than writing a post and, like other forms of passive entertainment, it is likely to be sought in the late evening hours to relax.

In addition to circadian rhythm differences, YouTube and Twitter are also different in terms of content type. YouTube comes across as a platform where emotions, especially positive ones, are much more represented. More than 9 out of 10 views go to videos with some emotional connotation. The reasons behind this result could be different and would be worth some future deeper investigation. They could range from differences in the audiences of the two platforms, to differences in creators' needs when drawing users' attention. What is interesting to notice here is that, even during a shock like Covid-19, although decreased, the quantity of emotional content remained a distinctive trait of the platform when compared with Twitter.

Chapter 7

Conclusions and Perspectives

7.1 Main Contributions

In the last 20 years, social networks have completely revolutionized how we communicate and share experiences and information. In a relatively short time, they have become one of the primary sources of news circulation in our societies [Sto+20]. Throughout their history, we have witnessed how their malfunction and deregulation can have detrimental effects on our societies: misinformation can influence the outcome of election campaigns [Sil] [AG17] [BF16], and incitement to hate can arouse offline violence [Hao18][Hao21] [Hao21]. Aware of these consequences, researchers have shown increasing interest in understanding online content diffusion and the role platforms play in shaping it. A large part of this interest has so far focused on the type of content suggested: false or biased news [AG17], radicalizing content [Pau21], among others. Sitting in the framework of information disorder studies, this thesis proposed to shift the focus from the type of content diffused to the temporal aspects of its diffusion. We devoted our work to study the dynamics of online content diffusion, investigating the role recommendations play in driving collective attention, and the way artificial inflations of engagement might interfere with information spread.

To highlight the need of a greater focus on the temporal aspects of content diffusion, in Chapter 2 we provided a conceptualization of an information disorder related to the speed of news consumption. We defined junk news bubbles as situations in which a large share of public attention is captured by a few objects incapable of sustaining it. We discussed the harmful effects of such attention regime in public debates: focusing on a few constantly changing issues, the audience has no time to digest or discuss new information. Through formalizing the public arena model of Hilgartner and Bosk [HB88], we discussed how, in principle, recommendation systems may enhance the emergence of such regimes. We hence emphasized the need to investigate their role through empirical data analysis.

To begin a line of research in this regard, we collected a large dataset on the temporal evolution of view counts on YouTube. In particular, we collected all videos published in the last three years by a list of 1400 French channels dealing with politics and information. Temporal data of this kind had not been collected since 2017, when YouTube significantly

restricted the accessibility to platform's statistics. Besides allowing a preliminary study of the role of human and algorithmic recommendation in the dissemination of online content, given their novelty, the data brought interesting and almost unexplored evidence to light.

Our first empirical finding is that the views count time series are well approximated by the Bass model. Through the Bass model, we distinguished two types of users who watch a video: users who watch it on their initiative (innovators) and users who discover it through human or algorithmic recommendations (imitators). In agreement with the observed skewness of suggestions on the platform [STK18] [San+21], in our analysis we find out that the role of recommendations is significant only in a minority of videos. These videos are, on average, more popular than the others, and they reach their interested audience faster. This evidence supports the intuition that recommendation systems might foster over-accelerated dynamics, reducing the time collective attention is committed to particular issues.

In addition to allowing for this modeling, our data brought to light some interesting evidence: YouTube intervenes in adjusting view counts, reducing the total number of views accumulated by a video when it deems them to be done by automated programs (bots). In Chapter 5 we revealed the extent of this phenomenon: one in two videos is affected. We discussed some features and some periodicities of the interventions made by the platform, but more importantly, we raised a broader question: can these illegitimate views foster content diffusion? Like social bots [Fer20] [Fer+16], these types of automated programs could give the false impression that a piece of information, regardless of its accuracy, is highly popular and endorsed by many. They could, in other words, encourage further human and algorithmic recommendations and propel the diffusion of targeted videos. Proving the existence (or not) of this fostering effect is hard, especially without data allowing the analysis of a causal relationship between fake views and popularity. To overcome this limitation, we implemented a bot able to record automatic views on YouTube and studied the platform's reaction to its activity. The application of this bot in different contexts in the future could help to further investigate YouTube's policy on views counting.

Last but not least, we seized the opportunity of having collect data during an unprecedented historical period: the Covid-19 pandemic. Given the moment's uniqueness, we dedicated some attention to the major changes lockdowns established worldwide had on online activity. This analysis allowed us to distinguish "natural" behaviors on the platform from "extraordinary" ones, helping to shed more light on a platform less studied from a data-driven approach than others.

Overall, with our study, we hope to have ignited interest in the role of recommendations in the temporal diffusion of content and the effects that possible alterations of engagement metrics may cause in the diffusion of online content. What we have presented here are but the first steps in this investigation, and in what follows, we will discuss some potential paths

for its advancement.

7.2 Limitations and Open Questions

We would like to conclude this work by discussing its limitations and the questions it leaves open. We will devote this section to a broad reflection on the limitations that originate from the data used and the choices made along the work. The next section will be devoted to the discussion of the extensions specifically related to diffusion models.

As interesting as our dataset is, it yet covers only one platform, though crucial in the diffusion of news and the formation of public opinion, and certain types of channels, which we have identified as the most relevant to the public debate (in France). Even though our findings about online content dissemination are indicatory of the dynamics in the whole YouTube, broadening the research to different contexts should help distinguish the general facts from those specific to our case study. For example, considering other channels could highlight which dissemination behaviors are typical of news and information videos and which cut across other content categories. In fact, it is reasonable to expect that news channels are more influenced by novelty than other types of content, such as music or how-to videos. Taking these types of videos into account could help to understand the general behavior of content circulation on YouTube, regardless of the topic.

In addition, we could extend the research to other platforms. YouTube is indeed a rather unique platform, because, unlike other social media such as Instagram, Facebook, and Twitter, it lacks a veritable network of friendships. Consequently, on YouTube, it is rather difficult to share videos of third-party creators with friends. People can, for example, create playlists and share the link with friends, but this is certainly not the primary purpose YouTube was built for and remains a pretty marginal tool. On other platforms, instead, friendship plays a more central role. On Facebook, for instance, users can re-post a piece of content on their profile so that it appears on their friends' walls, and friendship networks have been widely studied [Bas+17] [WL10] [TMP12]. YouTube, instead, using the terminology introduces in Chapter 4, does not encourage human recommendations on the platform. Using a suitable diffusion model, we could investigate how the presence or absence of this friendship network affects the diffusion of online content, shedding more light on the effects of under-represented human recommendations.

Although in this thesis we chose to focus primarily on the temporal aspects of content dissemination, the link between these temporal aspect and the quality of disseminated content should certainly be investigated in the future. Even though we defined junk news bubbles from a purely temporal point of view, it is likely that low-quality content such as click-bait or trolling provocations follow a more sped-up diffusion dynamic, given their inflaming nature.

We could wonder if other low-quality contents behave similarly. For instance, we could be interested in investigating if misinformation spreads according to particular temporal patterns and whether it follows more sped-up dynamics. A first step in this direction comes from work on misinformation detection. Previous studies in this field [Ngu+22] [RSL17] have shown that, among a combination of other features, temporal aspects can be good predictors of the reliability of contents. Following these results, in [Boz+21] we proposed a classifier aiming at labeling the factuality of YouTube channels taking into account a combination of temporal and content-based features. Also in this case, many temporal features proved to be good predictors, thus suggesting the existence of a link between content reliability and the temporal aspects of its diffusion. These results encourage further study of the relationship between content quality and attention dynamics.

Last but not least, the empirical analyses presented in this thesis focused primarily on the basic YouTube unit: its videos. When it comes to collective attention, however, a large part of the interest lies instead in understanding how different themes or topics spread. Extensive literature has been devoted to the dynamics of topic diffusion [HHN00][Les10][LBK09][Zha+15], while less has studied the dynamics of single objects of attention [FCL+11][Ric+14]. Observations frequently made about the spread of topics, or "memes," concern how much they are influenced by events outside the platforms where they spread. This literature has often highlighted that temporal trends are strongly linked to external factors. For example, different types of events generate different discussion dynamics: unexpected events present shares' peaks after they occur, while planned events show a slowly growing attention before they happen. An interesting extension of our work might be to study how video dynamics contribute to topic dynamics. Such a study would make it possible to establish a link between the role of platforms in the dissemination of individual content and the speed with which different issues follow each other into public debate.

7.3 Extended Diffusion Models

Although it has been shown in the literature to be a valid method for fitting time evolutions [HL21] [Ran+15], the Bass model remains a relatively simple tool and its performance in terms of fitting can certainly be improved. In this section, besides discussing possible better-performing extensions of the Bass model, we discuss how another class of models, Hawkes processes, can be used to weight innovation versus imitation in content diffusion. Hawkes processes allow integrating more assumption in their formulation and prove to be more flexible that the Bass model. Hence we suggest some formulations of Hawkes processes that could be used in the future to further investigate the role of human and algorithmic recommendations in content diffusion.

7.3.1 Extensions of Bass Model

The Bass model has found multiple extensions in the literature [JGK08]. However, most of them are concerned with relaxing the imitator/innovator [JGK08] distinction, which is instead fundamental in our analysis. Below we will present some ideas to extend the Bass model while maintaining this distinction clear.

Following the work of Richier et al. [Ric+14], a first attempt to improve fitting performances could be to consider adding a linear growth factor to the solution of the Bass model (4.2).

$$S^{\alpha,\beta,\gamma,M}(t) = M\frac{1-\gamma e^{-(\alpha+\beta)t}}{1+\frac{\beta}{\alpha}e^{-(\alpha+\beta)t}} + kt.$$

In their experience, in fact, Richier et al. observe that introducing this linear term significantly improves the explanatory capability of some exponential and epidemic models on real data. Despite it might be useful for improving performances, the addition of kt is difficult to justify. In fact, it makes the number of final adopters diverge in time, a fact that is hardly supported by empirical evidence (as we discussed, the majority of videos collect their views in the very first few days after publication). Moreover, adding a term kt complicates the interpretation of M, which does not represent anymore the final population reached by a video.

Another option to increase the flexibility of the Bass model while maintaining its interpretability could be to introduce some variations right into definition of the final public M, for instance considering it as a function of time. The introduction of this elasticity on the final public could represent, for instance, the change in time of the audience targeted by the recommendation system. This change in time could be justified by the learning and adjusting mechanisms of recommendation systems. Among the possible choices for $M(t)$, a reasonable one, for example, could be $M(t) = (1+\gamma e^{-\delta t})M$, so that, taking the limit for t than goes to infinity, M would maintain the interpretation of final public reached by a content. Considering $M(t)$ as a function of time, however, might result in complicating, if not preventing, the identification of a closed-form solution, needed to perform least-squares minimization. To overcome the need of a closed-form solution for fitting, a good alternative is presented in [Mat+12], and roughly consist of discretizing the differential model. Beyond these gimmicks to add flexibility to the model, it would certainly be interesting to incorporate insights about user similarity into the diffusion mechanism. Bass's model in its differential formulation assumes homogeneous contagion, but this assumption appears oversimplifying when considering how recommendation system are built. A better assumption in the model would be to make the contagion probability depend on users similarity. This similarity between users could be empirically quantified by counting the number of video two users both commented on, as discussed in Section 3.3.2. This *measures* of similarity would

naturally define a *network* of similarity (different from the real network of friendship) over which algorithmic recommendations (and not human recommendations) drive the contagion.

To include some high-level features of this similarity network, we could use the Feature-Driven Heterogeneous (FD-HBass) Bass Model presented in [Gao+21]. In this model, the authors replace the imitation and innovation coefficients with linear combinations of variables representing the content and the community viewing it. In our case, content-related features could be the popularity of the topics covered by a video, the number of subscribers to its channel, or the number of videos posted daily. The features related to the community could include, for instance, the average similarity among commenters or their average centrality in the similarity network. We would in this way introduce aggregated features of the similarity network into the Bass model while keeping the distinction between innovators and imitators.

7.3.2 Hawkes Processes

Hawkes processes are another class of models taking into account innovation and imitation dynamics. Recalling Section 1.1, they consist of counting processes in which the intensity function depends explicitly an all previously occurred events according to the rate

$$\lambda(t) = \lambda_0(t) + \sum_{i,t_i<t} \varphi(t - t_i)$$

where $t_i < t$ are all the events occurred before time t. In this formulation we recognize a term of innovation $\lambda_0(t)$, modeling the spontaneous engagement not triggered by epidemic effects, and a term of contagion/imitation $\sum_{i,t_i<t} \varphi(t - t_i)$ that depends on the time all previous event occurred.

Hawkes processes have been widely used to explain the diffusion of content online, under the assumption that the human activity of re-sharing and forwarding (i.e. human recommendations) is the main driver of content diffusion [Du+12] [GRLS13] [YL11][YZ13] [ZZS13]. This assumption is manifested through the choices of $\lambda_0(t)$ and $\varphi(t)$. In fact, when Hawkes processes are applied to social networks, $\lambda_0(t)$ is commonly considered either equal to zero or to a constant [CS08]. This latter choice models a background Poisson process that does not take into account dynamics of innovation that might depend on time: for instance, video discussing current affairs loose their interest when aging. Concerning the kernel $\varphi(t)$, the most common assumption is to take it equal to $\varphi(t) = 1/(t)^{(1+\theta)}$, a formulation known in literature to mimic the human behavior of reading and re-posting [Kar+11] [Bar05].

In other words, in the largely used formulation of Hawkes processes [Zha+15] [KL16] innovation and algorithmic recommendation are not represented. An interesting continuation

of our work would then be to include these two factors in Hawkes processes. To include innovation, we could choose a function $\lambda_0(t) = \alpha e^{-kt}$ decreasing with time, to model the decreasing effect of novelty in attracting fresh audience. For what concerns the kernel $\varphi(t-t_i)$, it is not straightforward to say what functions could better model how the effect of recommendations change over time. Some empirical studies in the wake of [RMM20], could guide the choice of the kernel by investigating how recommendations change in time.

Although in theory the addition of these elements to Hawkes' processes sounds appealing, some practical difficulties in the application to YouTube need to be discussed. In particular, we should keep in mind that, to set the parameters of Hawkes processes we usually maximize the log of the likelihood function $L(\theta)$:

$$L(\theta) = \prod_{i=0}^{n} \lambda(T_i) e^{-\int_0^T \lambda(t)dt} \tag{7.1}$$

given the realizations $\{T_1, T_2, \ldots, T_n\}$. When it comes to YouTube the exact time of views is not known and hence the vector of realizations $\{T_1, T_2, \ldots, T_n\}$ is not available. To overcome this problem, one first solution would consists to fit the evolution of the number of comments, instead of the number of views, considering them a proxy for popularity. Comments have the advantage of being associated with a date and a time of publication. Moreover they associate each event i with a user and its characteristics. Among its characteristics, as explained in the case of Bass model, its similarity to other users could be introduced in the kernel $\varphi(t - t_i)$ to better simulate the effect of recommendation systems. For example, we could consider a kernel $\varphi(t) = s_i\psi(t)$ depending on the out-degree s_i of user i in the similarity network. The final Hawkes process would result in:

$$\lambda(t) = \alpha e^{-kt} + \sum_{i,t_i<t} s_i \psi(t - t_i).$$

Results obtained through this approach on comment evolution could be compared with the ones obtained through the Bass model to check the consistency of the two approaches.

An interesting final extension of Hawkes processes could allow to distinguish the role of friendships and similarity. In platforms where we can identify a network of friendship and a network of similarity, we could introduce characteristics of both in the rate's definition. For instance, using a spatio-temporal formulation of Hawkes processes, we could define the rate of contagion as:

$$\lambda(t, j) = \lambda_0(t) + \sum_{i:t_i<t} f_{ij}\varphi_f(t - t_i) + \sum_{i:t_i<t} s_{ij}\varphi_s(t - t_i)$$

where f_{ij} is a binary variable indicating whether i and j are friends, s_{ij} is the similarity between user i and j and $\varphi_f(t-t_i)$ and $\varphi_s(t-t_i)$ represents the diffusion kernels associated

with human and algorithmic recommendations. This final formulation accounts for all three factors that have been discussed so far contributing to online content diffusion: innovation, human recommendation and algorithmic recommendation. Innovation and human recommendation represent two aspects of users' choices: users choose who to be friend with and what content to expressly look up. This final formulation could therefore distinguish the weight of users' choices as opposed to platform's choices in determining what users see. It could be a way to finally investigate the role of algorithmic recommendations, decoupling them from human ones.

Bibliography

[AG17] Hunt Allcott and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election". In: *Journal of Economic Perspectives* 31.2 (2017), pp. 211–36 (cit. on pp. 1, 97).

[AIM17] Martín Abadi, Michael Isard, and Derek G. Murray. "A computational model for TensorFlow: An introduction". In: *Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages*. MAPL 2017. Barcelona, Spain: Association for Computing Machinery, 2017, 1–7 (cit. on p. 36).

[Ale+15] Talayeh Aledavood et al. "Daily rhythms in mobile telephone communication". In: *PloS one* 10.9 (2015), e0138098 (cit. on p. 88).

[ALS18] Talayeh Aledavood, Sune Lehmann, and Jari Saramäki. "Social network differences of chronotypes identified from mobile phone data". In: *EPJ Data Science* 7.1 (2018), pp. 1–13 (cit. on p. 88).

[Bar05] Albert-László Barabási. "The origin of bursts and heavy tails in human dynamics". en. In: *Nature* 435.7039 (May 2005), pp. 207–211 (cit. on pp. 15, 102).

[Bas+17] Irène Bastard, Dominique Cardon, Raphaël Charbey, Jean-Philippe Cointet, and Christophe Prieur. "Facebook, pour quoi faire?" In: *Sociologie* 8.1 (2017), pp. 57–82 (cit. on p. 99).

[Bas69] Frank M. Bass. "A new product growth for model consumer durables". In: *Management Science* 15.5 (1969), pp. 215–227 (cit. on pp. 13, 49).

[BE19] Joshua A Braun and Jessica L Eklund. "Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism". In: *Digital Journalism* 7.1 (2019), pp. 1–21 (cit. on p. 29).

[Bec+20] Francois Beck, Damien Léger, Lisa Fressard, Patrick Peretti-Watel, Pierre Verger, and The Coconel Group. "Covid-19 health crisis and lockdown associated with high level of sleep complaints and hypnotic uptake at the population level". In: *Journal of Sleep Research* n/a.n/a (June 2020), e13119 (cit. on p. 87).

[Ben19] Greg Bensinger. *YouTube says viewers are spending less time watching conspiracy theory videos. But many still do.* 2019. URL: `https://www.washingtonpost.com/technology/2019/12/03/youtube-says-viewers-are-spending-less-time-watching-conspiracy-videos-many-still-do/` (cit. on p. 34).

[Ben+20] Bilel Benbouzid, Emma Gauthier, Pedro Ramaciotti, Bertrand Roudier, and Tommaso Venturini. "Le cadran de la visibilité de la sphere médiatique et politique sur YouTube". working paper or preprint. Dec. 2020 (cit. on p. 45).

[Ber19] Mark Bergen. *YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant*. 2019. URL: `https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant` (cit. on p. 33).

[Bet+06] Luís M A Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chavez. "The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models". English (US). In: *Physica A: Statistical Mechanics and its Applications* 364 (May 2006), pp. 513–536 (cit. on pp. 12, 15).

[BF16] Alessandro Bessi and Emilio Ferrara. "Social bots distort the 2016 U.S. Presidential election online discussion". In: *First Monday* 21.11 (2016) (cit. on pp. 1, 61, 97).

[BG10] Jean Burgess and Joshua Green. "YouTube: Online Video and Participatory Culture, by Jean Burgess and Joshua Green". In: *Popular Communication* 8.1 (2010), pp. 96–98 (cit. on p. 30).

[BHW14] Amber E. Boydstun, Anne Hardy, and Stefaan Walgrave. "Two Faces of Media Attention: Media Storm Versus Non-Storm Coverage." In: *Political Communication* 31.4 (2014), pp. 509–531 (cit. on p. 29).

[Bil+15] David R. Bild, Yue Liu, Robert P. Dick, Z. Morley Mao, and Dan S. Wallach. "Aggregate characterization of user behavior in Twitter and analysis of the retweet graph". In: 15.1 (2015) (cit. on p. 9).

[BM19] Marco T. Bastos and Dan Mercea. "The Brexit botnet and user-generated hyperpartisan news". In: *Social Science Computer Review* 37.1 (2019), pp. 38–54 (cit. on p. 61).

[Bor+12] Youmna Borghol, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. "The Untold Story of the Clones: Content-Agnostic Factors That Impact YouTube Video Popularity". In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '12. Beijing, China: Association for Computing Machinery, 2012, 1186–1194 (cit. on p. 71).

[Boz+21] Krasimira Bozhanova, Yoan Dinkov, Ivan Koychev, Maria Castaldo, Tommaso Venturini, and Preslav Nakov. "Predicting the factuality of reporting of news media using observations about user attention in their YouTube channels". In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*. 2021, pp. 182–189 (cit. on p. 100).

[BS19] Mark Bergen and Lucas Shaw. *To Answer Critics, YouTube Tries a New Metric: Responsibility.* 2019. URL: https://www.bloomberg.com/news/articles/2019-04-11/to-answer-critics-youtube-tries-a-new-metric-responsibility#xj4y7vzkg (cit. on p. 34).

[BSH15] Roja Bandari, Asur Sitaram, and Bernardo A Huberman. "The Pulse of News in Social Media: Forecasting Popularity". In: *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media.* 2015 (cit. on p. 29).

[Cas+14] Carlos Castillo, El-Haddad Mohammed, Pfeffer Jürgen, and Stempeck Matt. "Characterizing the Life Cycle of Online News Stories Using Social Media Reactions." In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work.* CSCW. 2014, pp. 211–23 (cit. on p. 29).

[CAS16] Paul Covington, Jay Adams, and Emre Sargin. "Deep neural networks for YouTube recommendations". In: *Proceedings of the 10th ACM Conference on Recommender Systems.* New York, NY, USA: Association for Computing Machinery, 2016, 191–198 (cit. on pp. 10, 16, 36, 37, 71).

[Cas+21a] Maria Castaldo, Paolo Frasca, Tommaso Venturini, and Floriana Gargiulo. *Views evolution 5 minutes frequency.* 2021. URL: https://doi.org/10.6084/m9.figshare.20080019.v1 (visited on 06/16/2022) (cit. on p. 46).

[Cas+21b] Maria Castaldo, Paolo Frasca, Tommaso Venturini, and Floriana Gargiulo. *Views evolution anonymized.* 2021. URL: 10.6084/m9.figshare.20079857 (visited on 06/16/2022) (cit. on p. 61).

[Cas+21c] Maria Castaldo, Tommaso Venturini, Paolo Frasca, and Floriana Gargiulo. "Junk news bubbles modelling the rise and fall of attention in online arenas". In: *New Media & Society* (2021) (cit. on p. 60).

[Cas+21d] Maria Castaldo, Tommaso Venturini, Paolo Frasca, and Floriana Gargiulo. "The Rhythms of the Night: increase in online night activity and emotional resilience during the Spring 2020 Covid-19 lockdown". In: *EPJ Data Science* 10.1 (Dec. 2021), pp. 1–15 (cit. on p. 67).

[Cas+22] Maria Castaldo, Tommaso Venturini, Paolo Frasca, and Floriana Gargiulo. "Junk news bubbles modelling the rise and fall of attention in online arenas". In: *New Media & Society* 24.9 (2022), pp. 2027–2045. eprint: https://doi.org/10.1177/1461444820978640. URL: https://doi.org/10.1177/1461444820978640 (cit. on p. 28).

[Cec20] Laura Ceci. *Hours of video uploaded to YouTube every minute as of February 2020.* 2020. URL: https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/ (visited on 09/15/2021) (cit. on p. 35).

[Cer+21] Francesca Ceragioli, Paolo Frasca, Benedetto Piccoli, and Francesco Rossi. "Generalized solutions to opinion dynamics models with discontinuities". In: *Crowd Dynamics*. Ed. by Bellomo N. and Gibelli L. Vol. 3. Modeling and Simulation in Science, Engineering and Technology. Birkhauser, 2021 (cit. on p. 9).

[CFR21] Francesca Ceragioli, Paolo Frasca, and Wilbert Samuel Rossi. "Modeling limited attention in opinion dynamics by topological interactions". In: *Network Games, Control and Optimization*. Ed. by Samson Lasaulce, Panayotis Mertikopoulos, and Ariel Orda. Cham: Springer International Publishing, 2021, pp. 272–281 (cit. on p. 9).

[CG16] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794 (cit. on p. 62).

[Cha+07] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. "I Tube, You Tube, Everybody Tubes: Analyzing the world's largest user generated content video system". In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. IMC '07. San Diego, California, USA: Association for Computing Machinery, 2007, 1–14 (cit. on p. 9).

[Cha+09] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. "Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems". In: *IEEE/ACM Transactions on Networking* 17.5 (Oct. 2009), pp. 1357–1370 (cit. on p. 9).

[Che+16] Heng-Tze Cheng et al. "Wide & Deep Learning for recommender systems". In: *CoRR* abs/1606.07792 (2016). arXiv: `1606.07792` (cit. on p. 17).

[Cit14] Yves Citton. *Pour une écologie de l'attention*. Paris: Seuil, 2014 (cit. on pp. 7, 29).

[CK12] Patrick Crogan and Samuel Kinsley. "Paying attention: Toward a critique of the attention economy". In: *Culture Machine* 13 (2012), pp. 1–29 (cit. on pp. 7, 20).

[CKM57] James Smoot Coleman, Elihu Katz, and Herbert Menzel. "The diffusion of an innovation among physicians". In: vol. 20. 4. 1957, pp. 253–270 (cit. on p. 13).

[CLP07] Ciro Cattuto, Vittorio Loreto, and Luciano Pietronero. "Semiotic dynamics and collaborative tagging". In: *Proceedings of the National Academy of Sciences* 104.5 (2007), pp. 1461–1464 (cit. on p. 26).

[Con+12] Rosaria Conte et al. "Manifesto of Computational Social Science". In: *European Physical Journal Special Topics EPJST* 214 (2012) (cit. on p. 2).

[CS08] Riley Crane and Didier Sornette. "Robust dynamic classes revealed by measuring the response function of a social system". In: *Proceedings of the National Academy of Sciences* 105.41 (2008), pp. 15649–15653 (cit. on pp. 10, 14, 29, 102).

[CZC14] Liang Chen, Yipeng Zhou, and Dah Ming Chiu. "Fake view analytics in online video services". In: *Proceedings of Network and Operating System Support on Digital Audio and Video Workshop*. New York, NY, USA: Association for Computing Machinery, 2014, 1–6 (cit. on p. 60).

[Dea+12] Jeffrey Dean et al. "Large Scale Distributed Deep Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012 (cit. on p. 36).

[Dew16] Caitlin Dewey. *Facebook fake-news writer: 'I think Donald Trump is in the White House because of me'*. 2016. URL: `https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/` (visited on 01/20/2022) (cit. on p. 1).

[Dew27] John. Dewey. *The Public and Its Problems*. Athens: Ohio University Press, 1927 (cit. on p. 30).

[DK64] D J Daley and D G Kendall. "Epidemics and Rumours". In: *Nature* 204.1118 (1964) (cit. on p. 11).

[DLC17] Fabon Dzogang, Stafford Lightman, and Nello Cristianini. "Circadian mood variations in Twitter content". In: *Brain and neuroscience advances* 1 (2017), p. 2398212817744501 (cit. on pp. 88, 91, 93).

[Dow72] Anthony Downs. "Up and down with Ecology: The 'Issue-Attention Cycle'." In: *The Public Interest* 28 (1972), 38–50 (cit. on p. 7).

[Dre14] Stuart Dredge. *Google goes to war on 'fraudulent' YouTube video views*. 2014. URL: `http://www.theguardian.com/technology/2014/feb/05/YouTube-fake-views-counts-google` (visited on 09/15/2021) (cit. on p. 60).

[Dre16] Stuart Dredge. "YouTube was meant to be a video-dating website". In: *The Guardian* (2016) (cit. on p. 32).

[DS05] Fabrice Deschâtres and Didier Sornette. "Dynamics of book sales: Endogenous versus exogenous shocks in complex networks". In: *Phys. Rev. E* 72 (1 2005), p. 016112 (cit. on p. 14).

[Du+12] Nan Du, Le Song, Ming Yuan, and Alex Smola. "Learning networks of heterogeneous influence". In: *Advances in neural information Processing Systems*. Ed. by F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012 (cit. on p. 102).

[Dwo19] Elizabeth Dwoskin. *YouTube is changing its algorithms to stop recommending conspiracies.* 2019. URL: https://www.washingtonpost.com/technology/2019/01/25/youtube-is-changing-its-algorithms-stop-recommending-conspiracies/?itid=lk_interstitial_manual_7 (cit. on p. 34).

[FBA11] Flavio Figueiredo, Fabrício Benevenuto, and Jussara M. Almeida. "The tube over time: characterizing popularity growth of YouTube videos". In: *Proceedings of the fourth ACM international conference on Web search and data mining.* WSDM '11. New York, NY, USA, Feb. 2011, pp. 745–754 (cit. on p. 16).

[FCL+11] Adrien Friggeri, Jean-Philippe Cointet, Matthieu Latapy, et al. "A real-world spreading experiment in the blogosphere". In: *Complex Systems* 19.3 (2011), p. 235 (cit. on p. 100).

[Fer+16] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. "The rise of social bots". In: *Communications of the ACM* 59.7 (2016), pp. 96–104 (cit. on pp. 61, 98).

[Fer20] Emilio Ferrara. "Bots, Elections, and Social Media: A Brief Overview". In: *Disinformation, misinformation, and fake news in social media: emerging research challenges and opportunities.* Ed. by Kai Shu, Suhang Wang, Dongwon Lee, and Huan Liu. Springer International Publishing, 2020, pp. 95–114 (cit. on p. 98).

[FSCS18] Claudia Flores-Saviaga, Keegan Brian C., and Saiph Savage. "Mobilizing the Trump Train: Understanding Collective Action in a Political Trolling Community." In: *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018).* 2018, pp. 82–91 (cit. on p. 29).

[Gao+21] Xiaofeng Gao, Zuowu Zheng, Quanquan Chu, Shaojie Tang, Guihai Chen, and Qianni Deng. "Popularity Prediction for Single Tweet Based on Heterogeneous Bass Model". In: *IEEE Transactions on Knowledge and Data Engineering* 33.5 (2021), pp. 2165–2178 (cit. on pp. 14, 51, 52, 57, 102).

[Gay12] Damien Gayle. *YouTube cancels billions of music industry video views after finding they were fake or 'dead'.* 2012. URL: https://www.dailymail.co.uk/sciencetech/article-2254181/YouTube-wipes-billions-video-views-finding-faked-music-industry.html (visited on 09/15/2021) (cit. on p. 60).

[GBB11] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics.* Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, 2011, pp. 315–323 (cit. on p. 36).

[Ger12] Paolo. Gerbaudo. *Tweets and the Streets: Social Media and Contemporary Activism.* London: Pluto Books, 2012 (cit. on p. 30).

[Gil16] Tarleton Gillespie. *#trendingistrending. When algorithms become culture*. Ed. by Robert Seyfert and Jonathan Roberge. London and New York, 2016 (cit. on p. 60).

[GM11] Scott A Golder and Michael W Macy. "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures". In: *Science* 333.6051 (2011), pp. 1878–1881 (cit. on pp. 88, 91, 93, 94).

[Gol97] Michael H. Goldhaber. "The attention economy and the Net". In: *First Monday* 2.4 (1997). URL: `https://firstmonday.org/ojs/index.php/fm/article/view/519` (cit. on pp. 10, 29).

[Gon+10] Marcos André Gonçalves, Jussara M. Almeida, Luiz G. P. dos Santos, Alberto H. F. Laender, and Virgílio Almeida. "On Popularity in the Blogosphere". In: *IEEE Internet Computing* 14.3 (May 2010), pp. 42–49 (cit. on p. 26).

[Goo] Google. "reCaptcha". In: (). URL: `https://www.google.com/recaptcha/about/` (cit. on p. 73).

[GRLS13] Manuel Gomez-Rodriguez, Jure Leskovec, and Bernhard Schölkopf. "Modeling Information Propagation with Survival Theory". In: *Proceedings of the 30th International Conference on Machine Learning*. Ed. by Sanjoy Dasgupta and David McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, 2013, pp. 666–674 (cit. on p. 102).

[Gup+13] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. "WTF: The who to follow service at Twitter". In: *Proceedings of the 22nd International Conference on World Wide Web*. WWW '13. Rio de Janeiro, Brazil: Association for Computing Machinery, 2013, 505–514 (cit. on p. 17).

[HAK17] William Hoiles, Anup Aprem, and Vikram Krishnamurthy. "Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data". In: *IEEE Trans. on Knowl. and Data Eng.* 29.7 (2017), 1426–1437 (cit. on p. 42).

[Hao18] Karen Hao. *Facebook admits it was used to 'incite offline violence' in Myanmar*. 2018. URL: `https://www.bbc.com/news/world-asia-46105934` (cit. on pp. 17, 97).

[Hao21] Karen Hao. *The Facebook whistleblower says its algorithms are dangerous. Here's why.* 2021. URL: `https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/` (cit. on pp. 17, 97).

[Has11] Robert. Hassan. *The Age of Distraction: Reading, Writing, and Politics in a High-Speed Networked Economy*. bingdon: Taylor Francis, 2011 (cit. on p. 29).

[HB88] Stephen Hilgartner and Charles L. Bosk. "The Rise and Fall of Social Problems: A Public Arenas Model". In: *American Journal of Sociology* 94.1 (1988), pp. 53–78 (cit. on pp. 7, 20, 97).

[HHN00] Susan Havre, Beth Hetzler, and Lucy Nowell. "ThemeRiver: Visualizing theme changes over time". In: *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings.* IEEE. 2000, pp. 115–123 (cit. on p. 100).

[HL21] Tuan-Anh Hoang and Ee-Peng Lim. "Virality and Susceptibility in Information Diffusions". In: *Proceedings of the International AAAI Conference on Web and Social Media* 6.1 (2021), pp. 146–153 (cit. on pp. 14, 100).

[HO74] Alan G. Hawkes and David Oakes. "A Cluster Process Representation of a Self-Exciting Process". In: *Journal of Applied Probability* 11.3 (1974), pp. 493–503 (cit. on p. 14).

[Hof12] Chase Hoffberger. *YouTube strips Universal and Sony of 2 billion fake views.* 2012. URL: `https://www.dailydot.com/unclick/YouTube-universal-sony-fake-views-black-hat/` (visited on 09/15/2021) (cit. on p. 60).

[JGK08] P. C. Jha, Anshu Gupta, and P. K. Kapur. "Bass model revisited". In: *Journal of Statistics and Management Systems* 11.3 (2008), pp. 413–437. URL: `https://doi.org/10.1080/09720510.2008.10701320` (cit. on pp. 57, 101).

[Jin+13] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. "Epidemiological modeling of news and rumors on Twitter". In: *Proceedings of the 7th Workshop on Social Network Mining and Analysis.* SNAKDD '13. Chicago, Illinois: Association for Computing Machinery, 2013 (cit. on p. 12).

[Joh01] Anders Johansen. "Response time of internauts". In: *Physica A: statistical mechanics and its applications* 296.3 (2001), pp. 539–546 (cit. on p. 14).

[JS00] Anders Johansen and Didier Sornette. "Download relaxation dynamics on the WWW following newspaper publication of URL". In: *Physica A: Statistical Mechanics and its Applications* 276.1 (2000), pp. 338–345 (cit. on p. 14).

[Kam15] Izabella Kaminska. *The real-world cost of YouTube's fake viewers.* 2015. URL: `https://www.ft.com/content/7a5d4b84-62af-11e5-9846-de406ccb37f2` (cit. on p. 60).

[Kar+11] M. Karsai et al. "Small but slow world: How network topology and burstiness slow down spreading". In: *Phys. Rev. E* 83 (2 2011), p. 025102 (cit. on p. 102).

[Kha17] M. Laeeq Khan. "Social Media Engagement: What motivates User Participation and Consumption on YouTube?" In: *Computers in Human Behavior* 66 (Jan. 2017), 236–247 (cit. on p. 85).

[Kie+12] Elmar Kiesling, Markus Günther, Christian Stummer, and Lea M Wakolbinger. "Agent-based simulation of innovation diffusion: a review". In: *Central European Journal of Operations Research* 20.2 (2012), pp. 183–230 (cit. on pp. 13, 14).

[KL16] Ryota Kobayashi and Renaud Lambiotte. "Tideh: Time-dependent Hawkes process for predicting retweet dynamics". In: *Tenth International AAAI Conference on Web and Social Media*. 2016 (cit. on p. 102).

[KNC13] Minkyoung Kim, David Newth, and Peter Christen. "Modeling Dynamics of Diffusion Across Heterogeneous Social Networks: News Diffusion in Social Media". In: *Entropy* 15.10 (2013), pp. 4215–4242 (cit. on p. 14).

[KPR17] Gary King, Jennifer Pan, and Margaret E. Roberts. "How the Chinese government fabricates social media posts for strategic distraction, not engaged argument". In: *American Political Science Review* 111.03 (2017), pp. 484–501 (cit. on p. 61).

[Kä+12] Mirko Kämpf, Sebastian Tismer, Jan W. Kantelhardt, and Lev Muchnik. "Fluctuations in Wikipedia access-rate and edit-event data". en. In: *Physica A: Statistical Mechanics and its Applications* 391.23 (2012), pp. 6101–6111 (cit. on p. 10).

[Lam+13] Vasileios Lampos, Thomas Lansdall-Welfare, Ricardo Araya, and Nello Cristianini. "Analysing mood patterns in the United Kingdom through Twitter content". In: *arXiv preprint arXiv:1304.5507* (2013) (cit. on pp. 91, 93).

[Lan64] Richard A. Lanham. "The Economics of Attention: Style and Substance in the Age of Information." In: *The Public Interest* 28 (1964), 38–50 (cit. on p. 20).

[Lat02] Bruno Latour. "Gabriel Tarde and the End of the Social". In: *The Social in Question. New Bearings in the History and the Social Sciences*. Ed. by Patrick Joyce. London: Routledge, 2002 (cit. on p. 7).

[Lat+12] Bruno Latour, Pablo Jensen, Tommaso Venturini, Sébastian Grauwin, and Dominique Boullier. "'The whole is always smaller than its parts': a digital test of Gabriel Tardes' monads". In: *The British Journal of Sociology* 63.4 (2012), pp. 590–615 (cit. on pp. 8, 20).

[Laz+09] David Lazer et al. "Computational social science". In: *Science* 323.5915 (2009), pp. 721–723 (cit. on pp. 2, 20).

[LBK09] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. "Meme-Tracking and the Dynamics of the News Cycle". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '09. Paris, France: Association for Computing Machinery, 2009, 497–506 (cit. on p. 100).

[Les10] an Kleinberg Jon. Leskovec Jure andLars Backstrom. "Meme-Tracking and the Dynamics of the News Cycle." In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris, France, Oct. 2010, pp. 497–506 (cit. on pp. 20, 25, 100).

[Lip22] Walter. Lippmann. *Public Opinion*. New York: Simon Schuster, 1922 (cit. on p. 30).

[Lip27] Walter. Lippmann. *The Phantom Public*. New York: The Macmillan Company, 1927 (cit. on p. 30).

[Lle+19] Clare Llewellyn, Laura Cram, Adrian Favero, and Robin L. Hill. "For whom the bell trolls: troll behaviour in the Twitter Brexit debate". In: *JCMS: Journal of Common Market Studies* 57.5 (2019), pp. 1148–1164 (cit. on p. 61).

[LM18] Paul Lewis and Erin McCormick. *How an ex-YouTube insider investigated its secret algorithm*. 2018. URL: `https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot` (cit. on p. 39).

[LS+19] Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. "Accelerating dynamics of collective attention". en. In: *Nature Communications* 10.1 (Apr. 2019), p. 1759 (cit. on pp. 9, 20, 25).

[LSS16] Eric Lipton, David E. Sanger, and Scott Shane. *The perfect weapon: how Russian cyberpower invaded the U.S.* 2016. URL: `https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html` (visited on 09/15/2021) (cit. on p. 61).

[Lu+14] Yao Lu, Peng Zhang, Yanan Cao, Yue Hu, and Li Guo. "On the frequency distribution of retweets". In: *Procedia Computer Science* 31 (2014). 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014, pp. 747–753 (cit. on p. 9).

[Luu+21] Duc Luu, Ee-Peng Lim, Tuan-Anh Hoang, and Freddy Chua. "Modeling Diffusion in Social Networks Using Network Properties". In: *Proceedings of the International AAAI Conference on Web and Social Media* 6.1 (2021), pp. 218–225 (cit. on p. 14).

[LZ16] Mark Ledwich and Anna Zaitsev. "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization". In: *First Monday* (2016) (cit. on p. 39).

[Mar+16] Miriam Marciel et al. "Understanding the detection of view fraud in video content portals". In: *Proceedings of the 25th International Conference on World Wide Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, 357–368 (cit. on p. 60).

[Mar18] Noortje Marres. "Why We Can't Have Our Facts Back." In: *Engaging Science, Technology, and Society* 4.423 (2018) (cit. on p. 20).

[Mat+12] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, and Christos Faloutsos. "Rise and Fall Patterns of Information Diffusion: Model and Implications". In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '12. Beijing, China: Association for Computing Machinery, 2012, 6–14 (cit. on p. 101).

[McC05] Maxwell McCombs. "A Look at Agenda-setting: Past, present and future". In: *Journalism Studies* 6.4 (2005), pp. 543–557 (cit. on p. 7).

[McL22] Marshall McLuhan. "Understanding Media: The Extensions of Man." In: New York: McGraw-Hill, 2022 (cit. on p. 20).

[Med] MediaLab. *Gazouilloire*. URL: `https://github.com/medialab/gazouilloire` (cit. on p. 81).

[Mer68] Robert K. Merton. "The Matthew Effect in Science." In: *Science* 159.3810 (1968), pp. 56–63 (cit. on p. 22).

[Meta] Meta. *Community Standards - Objectionable Content*. URL: `https://help.instagram.com/477434105621119` (visited on 01/20/2022) (cit. on p. 17).

[Metb] Meta. *Instagram Community Guidelines*. URL: `https://m.facebook.com/communitystandards/objectionable_content` (cit. on p. 17).

[MI05] R. Marois and J. Ivanoff. "Capacity limits of information processing in the brain". In: *Trends in cognitive sciences* 9.6 (2005), pp. 296–305 (cit. on p. 9).

[Mit+09] Siddharth Mitra, Mayank Agrawal, Amit Yadav, Niklas Carlsson, Derek Eager, and Anirban Mahanti. "Characterizing web-based video sharing workloads". In: *Proceedings of the 18th International Conference on World Wide Web*. WWW '09. Madrid, Spain: Association for Computing Machinery, 2009, 1191–1192 (cit. on p. 9).

[MM10] Panagiotis Metaxas and Eni Mustafaraj. "From obscurity to prominence in minutes: political speech and real-time search." In: *Proceedings of the 2nd international Web Science Conferences*. Raleigh, North Carolina, USA, 2010 (cit. on p. 61).

[MO21] Jeremy B. Merrill and Will Oremus. *Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation*. 2021. URL: `https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/` (cit. on p. 17).

[Mon17] Blake Montgomery. *YouTube Has Deleted Hundreds Of Thousands Of Disturbing Kids' Videos.* 2017. URL: https://www.buzzfeednews.com/article/blakemontgomery/youtube-has-deleted-hundreds-of-thousands-of-disturbing (cit. on p. 34).

[Mor78] Jorge J. Moré. "The Levenberg-Marquardt algorithm: Implementation and theory". In: *Lecture Notes in Mathematics, Berlin Springer Verlag.* Vol. 630. 1978, pp. 105–116 (cit. on p. 51).

[MS72] Maxwell E. McCombs and Donald L. Shaw. "The Agenda-Setting Function of Mass Media". In: *Public Opinion Quarterly* 36.2 (1972), p. 176. eprint: arXiv:1011.1669v3 (cit. on p. 7).

[Nau+19] Maxim Naumov et al. "Deep learning recommendation model for personalization and recommendation systems". In: *CoRR* abs/1906.00091 (2019). arXiv: 1906.00091 (cit. on p. 16).

[New01] M E. Newman. "Clustering and Preferential Attachment in Growing Networks." In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 64.025102 (2001) (cit. on p. 22).

[Ngu+22] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. "FANG: Leveraging social context for fake news detection using graph representation". In: *Commun. ACM* 65.4 (2022), 124–132 (cit. on p. 100).

[Noaa] *Fake Engagement Policy - YouTube Help.* 2021. URL: https://support.google.com/youtube/answer/3399767?hl=en (visited on 09/15/2021) (cit. on pp. 60, 72).

[Noab] *Fake YouTube views cut by 2 billion as Google audits record companies' video channels.* 2012. URL: https://www.huffpost.com/entry/fake-YouTube-views-cut-google-audit_n_2380848 (cit. on p. 60).

[Noac] *How engagement metrics are counted.* URL: https://support.google.com/YouTube/answer/2991785?hl=en\%E2\%80\%8B (visited on 07/12/2021) (cit. on pp. 59, 60).

[NR20] Philip M Napoli and Caplan Robyn. "Why media companies insist they're not media companies, why they're wrong, and why it matters". In: *First Monday* (2020) (cit. on pp. 1, 33).

[NS19] Shishir Nagaraja and Ryan Shah. "Clicktok: click fraud detection using traffic analysis". In: *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks.* WiSec '19. Miami, Florida: Association for Computing Machinery, 2019, 105–116 (cit. on p. 60).

[Out] *Click and elect: how fake news helped Donald Trump win a real election.* URL: https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies (cit. on p. 1).

[OV09] J. G. Oliveira and A. Vazquez. "Impact of interactions on human dynamics". In: *Physica A: Statistical Mechanics and its Applications* 388.2-3 (Jan. 2009), pp. 187–192 (cit. on p. 15).

[PAG13] Henrique Pinto, Jussara M. Almeida, and Marcos A. Gonçalves. "Using early view patterns to predict the popularity of YouTube videos". In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining.* WSDM '13. Rome, Italy: Association for Computing Machinery, 2013, 365–374 (cit. on pp. 42, 71).

[Pau21] Kari Paul. *It let white supremacists organize': the toxic legacy of Facebook's Groups.* 2021. URL: https://www.theguardian.com/technology/2021/feb/04/facebook-groups-misinformation (cit. on pp. 17, 97).

[Pfe14] Philipp Pfeiffenberger. *Keeping YouTube views authentic.* 2014. URL: https://security.googleblog.com/2014/02/keeping-YouTube-views-authentic.html (cit. on p. 60).

[Pio+11] Annie Piolat, Roger J Booth, Cindy K Chung, Morgana Davids, and James W Pennebaker. "La version française du dictionnaire pour le LIWC: modalités de construction et exemples d'utilisation". In: *Psychologie française* 56.3 (2011), pp. 145–159 (cit. on p. 88).

[Qiu+17] Xiaoyan Qiu, Diego F. M. Oliveira, Alireza Sahami Shirazi, Alessandro Flammini, and Filippo Menczer. "Limited individual attention and online virality of low-quality information". In: *Nature Human Behaviour* 1.0132 (2017) (cit. on p. 9).

[Qui15] Ben Quinn. *Google charges advertisers for fake YouTube video views, say researchers.* 2015. URL: https://www.theguardian.com/technology/2015/sep/23/google-advertisers-fake-YouTube-video-views-adwords-bot (visited on 09/15/2021) (cit. on p. 60).

[Ran+15] William Rand, Jeffrey Herrmann, Brandon Schein, and Neža Vodopivec. "An agent-based model of urgent diffusion in social media". In: *Journal of Artificial Societies and Social Simulation* 18.2 (2015), p. 1 (cit. on pp. 14, 100).

[Rat+11] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Alessandro Flammini, and Filippo Menczer. "Detecting and tracking political abuse in social media". In: *Proceedings of the 5th AAAI International Conference on Weblogs and Social Media (ICWSM'11).* 2011 (cit. on p. 61).

[RDJ12] George Ritzer, Paul Dean, and Nathan Jurgenson. "The Coming of Age of the Prosumer". In: *American Behavioral Scientist* 56 (Mar. 2012), pp. 379–398 (cit. on p. 85).

[Rea16] Max Read. *Donald Trump Won Because of Facebook*. 2016. URL: `https://nymag.com/intelligencer/2016/11/donald-trump-won-because-of-facebook.html` (visited on 01/20/2022) (cit. on p. 1).

[RF20] Wilbert Samuel Rossi and Paolo Frasca. "Opinion dynamics with topological gossiping: Asynchronous updates under limited attention". In: *IEEE Control Systems Letters* 4.3 (2020), pp. 566–571 (cit. on p. 9).

[RHH18] Alexander Reutlinger, Dominik Hangleiter, and Stephan Hartmann. "Understanding (with) Toy Models". In: *The British Journal for the Philosophy of Science* 69.4 (2018), pp. 1069–1099 (cit. on p. 21).

[Rib+20] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. "Auditing Radicalization Pathways on YouTube". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT* '20. Barcelona, Spain: Association for Computing Machinery, 2020, 131–141 (cit. on pp. 17, 39).

[Ric+14] Cédric Richier, Eitan Altman, Rachid Elazouzi, Tania Jimenez, Georges Linares, and Yonathan Portilla. "Bio-inspired models for characterizing YouTube viewcout". In: *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. 2014, pp. 297–305 (cit. on pp. 12, 51, 52, 56, 100, 101).

[RM21] Ludovic Rheault and Andreea Musulan. "Efficient detection of online communities and social bot activity during electoral campaigns". In: *Journal of Information Technology & Politics* 18.3 (2021), pp. 324–337 (cit. on p. 61).

[RMC21] Pedro Ramaciotti Morales and Jean-Philippe Cointet. "Auditing the effect of social network recommendations on polarization in geometrical ideological spaces". In: *Fifteenth ACM Conference on Recommender Systems*. 2021, pp. 627–632 (cit. on p. 19).

[RMM20] Camille Roth, Antoine Mazieres, and Telmo Menezes. "Tubes & Bubbles. Topological confinement of YouTube recommendations". In: *PLoS ONE* 15.4 (2020) (cit. on pp. 11, 71, 103).

[Rog18] Richard Rogers. "Otherwise engaged: social media from vanity metrics to critical analytics". In: *International Journal of Communication* 12.732942 (2018), pp. 450–472 (cit. on p. 60).

[RSL17] Natali Ruchansky, Sungyong Seo, and Yan Liu. "CSI: A hybrid deep model for fake news detection". In: CIKM '17. Singapore, Singapore: Association for Computing Machinery, 2017, 797–806 (cit. on p. 100).

[San+21] Leonardo Sanna, Salvatore Romano, Giulia Corona, and Claudio Agosti. "YT-TREX: Crowdsourced analysis of YouTube's recommender system during COVID-19 pandemic". In: *Information Management and Big Data*. Ed. by Juan Antonio Lossio-Ventura, Jorge Carlos Valverde-Rebaza, Eduardo Díaz, and Hugo Alatrista-Salas. Cham: Springer International Publishing, 2021, pp. 107–121 (cit. on pp. 55, 56, 98).

[Ser] Google YouTube Terms of Service. *Advertiser-friendly content guidelines*. URL: `https://support.google.com/youtube/answer/6162278?hl=en` (cit. on p. 17).

[SH10] Gabor Szabo and Bernardo A. Huberman. "Predicting the Popularity of Online Content". In: *Commun. ACM* 53.8 (Aug. 2010), 80–88 (cit. on p. 71).

[Sha+17a] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. "The spread of fake news by social bots". In: *arXiv preprint arXiv:1707.07592* 96 (2017), p. 104 (cit. on pp. 19, 61).

[Sha+17b] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. "The spread of low-credibility content by social bots". In: *IEEE Transactions on Automatic Control* 63.9 (2017), pp. 2898–2912 (cit. on p. 1).

[Sil] Craig Silverman. *This analysis shows how viral fake election news stories outperformed real news on Facebook*. URL: `https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook` (cit. on pp. 1, 17, 97).

[Sim71a] Herbert A. Simon. "Designing organizations for an information rich world". In: *Computers, communications, and the public interest*. Ed. by M Greenberger. Baltimore: Johns Hopkins Press, 1971, pp. 37–72 (cit. on p. 7).

[Sim71b] Herbert A. Simon. "Designing organizations for an information rich world". In: *Computers, communications, and the public interest*. Ed. by Martin Greenberger. Baltimore, 1971, pp. 37–72 (cit. on p. 10).

[Sol] Joan E. Solsman. *YouTube's AI is the puppet master over most of what you watch*. URL: `https://www.cnet.com/news/youtube-ces-2018-neal-mohan/` (cit. on p. 16).

[SP06] Andrew Ross Sorkin and Jeremy W. Peters. "Google to Acquire YouTube for $1.65 Billion". In: *The New York Times* (2006) (cit. on p. 32).

[STK18] Aaron Smith, Skye Toor, and Patrick van Kessel. *Many turn to YouTube for children's content, news, how-to lessons*. 2018. URL: `https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/` (visited on 09/15/2021) (cit. on pp. 55, 56, 98).

[Sto+20] Galen Stocking, Patrick van Kessel, Michael Barthel, Katerina Eva Matsa, and Maya Khuzam. "Many Americans Get News on YouTube, Where News Organizations and Independent Producers Thrive Side by Side". In: *Pew Research Center* (2020) (cit. on pp. 31, 97).

[SV09a] Pamela J. Shoemaker and Timothy Vos. *Gatekeeping Theory*. New York: Routledge, 2009 (cit. on p. 23).

[SV09b] Pelle Snickars and Patrick Vonderau. *The YouTube Reader*. New York: Columbia University Press, 2009 (cit. on p. 30).

[Tar90] Gabriel Tarde. *Les lois de l'imitation*. Paris: Félix Alcan, 1890 (cit. on p. 7).

[Tar93] Gabriel Tarde. *Monadologie et sociologie*. Paris: Les empêcheurs de penser en rond, 1893 (cit. on p. 7).

[TCG12] Jameson L Toole, Meeyoung Cha, and Marta C González. "Modeling the adoption of innovations in the presence of geographic and media influences". In: *PloS one* 7.1 (2012), e29528 (cit. on p. 13).

[Ter12] Tiziana Terranova. "Attention, Economy and the Brain". In: *Culture Machine* 13 (2012), pp. 1–19 (cit. on pp. 7, 20).

[The] The YouTube Team. *How does YouTube address misinformation?* URL: `https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/` (cit. on p. 35).

[The19] The YouTube Team. *YouTube is changing its algorithms to stop recommending conspiracies.* 2019. URL: `https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate/` (cit. on p. 34).

[TMP12] Amanda L Traud, Peter J Mucha, and Mason A Porter. "Social structure of facebook networks". In: *Physica A: Statistical Mechanics and its Applications* 391.16 (2012), pp. 4165–4180 (cit. on p. 99).

[TP10] Yla R Tausczik and James W Pennebaker. "The psychological meaning of words: LIWC and computerized text analysis methods". In: *Journal of language and social psychology* 29.1 (2010), pp. 24–54 (cit. on p. 88).

[Vas05] Peter L.M. Vasterman. "Media-Hype: Self-Reinforcing News Waves, Journalistic Standards and the Construction of Social Problems." In: *European Journal of Communication* 20.4 (2005), 508–30 (cit. on p. 20).

[Vir+20] Pauli Virtanen et al. "SciPy 1.0: Fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272 (cit. on p. 51).

[VJB15] Tommaso Venturini, Pablo Jensen, and Latour Bruno. "Fill in the Gap: A New Alliance for Social and Natural Sciences". In: *Journal of Artificial Societies and Social Simulation* 18.2 (2015), p. 11 (cit. on p. 20).

[VL10] Tommaso Venturini and Bruno Latour. "The Social Fabric: Digital Traces and Quali-quantitative Methods". In: *Proceedings of Future En Seine 2009*. Paris: Editions Future en Seine, 2010 (cit. on p. 8).

[Vá+06] Alexei Vázquez, João Gama Oliveira, Zoltán Dezsö, Kwang-Il Goh, Imre Kondor, and Albert-László Barabási. "Modeling bursts and heavy tails in human dynamics". In: *Physical Review E* 73.3 (2006) (cit. on p. 15).

[WD17] Claire Wardle and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking". In: *Report to the Council of Europe* (2017) (cit. on p. 19).

[Wen+12] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. "Competition among memes in a world with limited attention". en. In: *Scientific Reports* 2.1 (Mar. 2012), p. 335 (cit. on pp. 9, 11, 26).

[WH07] Fang Wu and Bernardo A. Huberman. "Novelty and collective attention". en. In: *Proceedings of the National Academy of Sciences* 104.45 (Nov. 2007), pp. 17599–17601 (cit. on p. 29).

[WL10] Andreas Wimmer and Kevin Lewis. "Beyond and below racial homophily: ERG models of a friendship network documented on Facebook". In: *American journal of sociology* 116.2 (2010), pp. 583–642 (cit. on p. 99).

[Woj] Susan Wojcicki. *My mid-year update to the YouTube community*. URL: `https://www.theverge.com/2021/4/9/22375702/google-updates-youtube-ad-targeting-hate-speech` (visited on 01/20/2022) (cit. on p. 17).

[Xio+12] Fei Xiong, Yun Liu, Zhen-jiang Zhang, Jiang Zhu, and Ying Zhang. "An information diffusion model based on retweeting mechanism for online social media". In: *Physics Letters A* 376.30 (2012), pp. 2103–2108 (cit. on p. 12).

[YL11] Jaewon Yang and Jure Leskovec. "Patterns of Temporal Variation in Online Media." In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. WSDM 2011. 2011, pp. 177–86 (cit. on pp. 29, 102).

[YXS15] Honglin Yu, Lexing Xie, and Scott Sanner. "The lifecyle of a YouTube video: phases, content and popularity." In: *Ninth International AAAI Conference on Web and Social Media*. 2015 (cit. on p. 45).

[YZ13] Shuang-Hong Yang and Hongyuan Zha. "Mixture of mutually exciting processes for viral diffusion". In: *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. ICML'13. Atlanta, GA, USA: JMLR.org, 2013, II–1–II–9 (cit. on p. 102).

[Zha+13] Jinxue Zhang, Rui Zhang, Yanchao Zhang, and Guanhua Yan. "On the impact of social botnets for spam distribution and digital-influence manipulation". In: *Proceedings of the 2013 IEEE Conference on Communications and Network Security, CNS 2013*. IEEE Computer Society, 2013, pp. 46–54 (cit. on p. 61).

[Zha+15] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. "Seismic: A self-exciting point process model for predicting tweet popularity". In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015, pp. 1513–1522 (cit. on pp. 100, 102).

[Zha+19] Zhe Zhao et al. "Recommending what video to watch next: A multitask ranking system". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19. Copenhagen, Denmark: Association for Computing Machinery, 2019, 43–51 (cit. on pp. 16, 36).

[Zho+16] Renjie Zhou, Samamon Khemmarat, Lixin Gao, Jian Wan, and Jilin Zhang. "How YouTube videos are discovered and its impact on video views". In: *Multimedia Tools Appl.* 75.10 (May 2016), 6035–6058 (cit. on pp. 36, 42, 71).

[ZKG10a] Renjie Zhou, Samamon Khemmarat, and Lixin Gao. "The Impact of YouTube Recommendation System on Video Views". In: *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*. IMC '10. Melbourne, Australia: Association for Computing Machinery, 2010, 404–410 (cit. on pp. 16, 42).

[ZKG10b] Renjie Zhou, Samamon Khemmarat, and Lixin Gao. "The impact of YouTube recommendation system on video views". In: *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*. Association for Computing Machinery, 2010, 404–410 (cit. on pp. 36, 42, 71).

[Zub19] Shoshana Zuboff. *The Age of Surveillance Capitalism*. New York: Hachette, 2019 (cit. on p. 8).

[ZZS13] Ke Zhou, Hongyuan Zha, and Le Song. "Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes." In: *AISTATS*. Vol. 31. JMLR Workshop and Conference Proceedings. JMLR.org, 2013, pp. 641–649 (cit. on p. 102).

Résumé — Cette thèse étudie la diffusion de contenus en ligne, et plus spécifiquement leurs aspects temporels, avec une forte approche interdisciplinaire et une attention particulière à YouTube. Tout d'abord, dans un contexte d'études des médias, nous discutons de l'importance d'étudier les rythmes de consommation de contenu et soutenons que, de la même manière que d'autres troubles de l'information, certains régimes d'attention peuvent être défavorable à au développement d'un débat public sain et florissant. Nous fournissons un nouveau concept de régime d'attention sur-accéléré, dans lesquels la majorité de l'attention est obtenue par une minorité d'objets, sans pour autant être capable de la pérenniser, donnant lieu à des débats publics éphémères et fragmentés. Cette conceptualisation justifie l'intérêt d'étudier la dynamique temporelle en ligne avec une approche empirique. Pour ce faire, nous collectons et étudions une large base de données YouTube, contenant l'évolution du nombre de vues de 1400 chaînes représentatives de la sphère médiatique française. Nous proposons un modèle de Bass pour expliquer l'évolution du nombre de vues et nous étudions le rôle que les mécanismes d'imitation et d'innovation jouent dans la diffusion sur YouTube. L'imitation joue ainsi un rôle significatif seulement dans une minorité de vidéo, caractérisées par leur popularité et la rapidité de leur diffusion. Nous observons la présence d'une composante d'imitation significative uniquement dans une minorité de vidéos, qui sont en moyenne plus populaires et se diffusent plus rapidement. Par ailleurs, en plus de rendre possible cette modélisation, les données YouTube ont mis en évidence des point intéressants : la plateforme réduit souvent le nombre de vues, en supprimant celles attribuables à des programmes automatisés. Ce phénomène touche une vidéo sur deux et la plupart des chaînes étudiées. Nous fournissons une analyse des taux de correction et discutons de la possibilité que, si elles sont corrigées trop tard, les fausses vues puissent interférer avec les recommandations humaines et algorithmiques, favorisant le contenu ciblé. Comme contribution finale, nous saisissons l'opportunité d'avoir collecté des données pendant une période historique sans précédent : la pandémie de Covid-19. Étant donné le caractère unique de ce moment, nous consacrons une attention particulière aux changements majeurs que les confinements mis en place dans le monde ont eu sur l'activité en ligne. Cette analyse nous permet de distinguer les comportements "naturels" des comportements "extraordinaires" sur la plateforme, contribuant ainsi à éclairer davantage le fonctionnement de YouTube en tant que média social.

Mots clés : Dynamique de l'attention, YouTube, dynamique de diffusion, fausses vues, bots, réseaux sociaux, sciences sociales computationnelles, études des médias

Abstract — This thesis investigates the temporal aspects of online diffusion dynamics with a strongly interdisciplinary approach and a specific focus on YouTube. First, in a Media Studies context, we discuss the importance of investigating content consumption rhythms, arguing that, in the same way as other information disorders, certain attention regimes can restrain the development of a healthy public debate. We provide a conceptualization for over accelerated attention regimes in which few contents get most of the collective attention but are incapable of sustaining it for long, giving rise to ephemeral and fragmented public debates. This conceptualization justifies the interest in studying online temporal dynamics with an empirical approach. To do so, we collect and study a large YouTube dataset containing the evolution of views-count for 1400 channels representative of the French media sphere. We propose a Bass model to explain views-count evolutions and investigate the role that imitation and innovation mechanisms play in shaping the diffusion on YouTube. We observe the presence of a significant imitation component only in a minority of videos, which are on average more popular and spread faster. Besides allowing for this modeling, the YouTube dataset brings forth some interesting evidence: the platform often reduces the number of views, removing those attributable to automated programs. This phenomenon touches almost all the channels studied and one of every two videos. We provide an analysis of correction rates and discuss the possibility that, if corrected too late, fake views could interfere with human and algorithmic recommendations, promoting targeted content. As a final contribution, we seize the opportunity of having collect data during an unprecedented historical period: the Covid-19 pandemic. Given the moment's uniqueness, we dedicate some attention to the major changes lockdowns established worldwide had on online activity. This analysis allows us to distinguish "natural" from "extraordinary" behaviors on the platform, helping shed more light on YouTube as a social media outlet.

Keywords: Attention dynamics, YouTube, diffusion dynamics, fake views, bots, social networks, computational social science, media studies.

Gipsa-lab, 11 rue des Mathématiques
Grenoble, France