SciencesPo
CHAIR DIGITAL, GOVERNANCE AND SOVEREIGNTY

Project
Liberty
Institute
//

SOCIAL MEDIA RESET

Redesigning the infrastructure
of digital propagation
to cut the chains of contagion

Dominique Boullier

Full professor in sociology at Sciences Po

Table of contents

Abstract

As governments face the tremendous influence of social media on personal relationships and the public sphere, they have developed a series of methods to regulate social media platforms that do not seem to adequately address the high level of disruption they have introduced. In this paper, several reasons for this inability to properly curb these harms are discussed, including a flawed understanding of the effects of platforms on media consumption, referred to in this article as "propagation processes". High frequency propagation processes on social media platforms have created a milieu of self-replicating messages that "go viral" and influence the types of media people are more likely to consume and their attention capacities. The virality and rhythm of propagation must be a core target of public policies in order to slow down all communication activities. Regulators should also take into account the role of interface design in promoting virality and invite expertise that can define standards to limit the rhythm of propagation on social media platforms. Technically and socially feasible recommendations are presented, contingent on political will to nudge all digital platforms towards legal and social responsibility. Curbing the rhythm of propagation on media platforms can change public media consumption and attention to information, through the design of new interfaces preventing any toxic rhythm of virality.

Introduction

For six years (since the Cambridge Analytica scandal broke in 2018), countries have proposed and delivered various announcements, promises, and measures regarding the regulation of social media platforms in accordance to different states, cultures, and contrasting legal systems. To what extent can we say that a real shift has been made in terms of hate speech, online harassment, content moderation, the fight against fake news, and the prevention of attention-related pathologies? Despite numerous attempts and announcements, no one seems to have found the winning regulatory combination to reformat the functioning of social networks and the behavioural norms that dominate this media sphere.

To explain this failure, which is problematic for the democratic health of nations, two series of arguments can be put forward:

1. There is an unfavourable balance of economic and political forces between all-powerful platforms and states, aggravated by an absence of political will to reformat this media architecture.

2. There are multiple proactive initiatives acting on all possible levers, but never on the very technical architecture of these platforms, the one which generates virality.

My proposal will consist of focusing on the question of virality and considering it as the key point of any reset of social networks. The theory of propagation that I have developed (Boullier, 2023) should make it possible to produce a more detailed diagnosis than the current well-intentioned discourse and to propose feasible and targeted solutions to what creates this virality. This policy brief does not claim to discuss in detail all the measures already put in place by states or international bodies, but will nevertheless not ignore them. On the other hand, this paper claims to be creative in moving away from regulationist thinking and mobilising the theory of propagations to imagine possible fundamental treatments for this chronic pathology of social media. The feasibility of the proposed measures (**displayed in bold in the text**) will be tested as much as possible, but the bias adopted here is profoundly reformist; that is to say, it takes note of the

political impossibility of radical regulation for the sanitation of social media platforms. This could lead to disappointments for certain readers but also arouse more operational initiatives with some chances of success, that is at least, what I hope. Anyway, it should trigger a conceptual renewal in the thinking of this media environment created over the past 15 years, still stuck in a classic vision of the public sphere, forged by Habermas. However, the German philosopher himself warned of "the decline of this public sphere which strove to be inclusive and which can no longer be distinguished from these semi-public spheres competing on an equal footing" (Habermas, 2023, p.115). It certainly targets, above all, the echo chamber effects, but also the drift of the media infrastructure towards subjective expressions that are neither controlled nor subject to common validity requirements. While warning of this threat against democratic systems, he hardly perceives the profound mutation (which he nevertheless compares to the printing revolution) of this "media space" into a self-replicating media milieu, favouring high-frequency propagation. It is in the light of this diagnosis, based on our theory of propagations, that we can better understand the current impasses of the regulations put in place.

Chapter 1: Why current regulations don't work

Answer: due to an unfavourable economic and political balance of power between the all-powerful platforms and the states, aggravated by a lack of political will to reformat this media architecture.

States and their agencies that want to regulate social media platforms find themselves confronted with entities that have thrived with impunity for 10 years, from 2008 to 2018. I emphasise "impunity" because many of these activities and procedures were illegal, or at least abusive, in many areas. The capture of personal data was the most prominent case, and the European General Data Protection Regulation (GDPR) made it possible to bring order only from 2018, with certain limitations, of which I only cite two:

- Checking declarations is a very long process, which then requires a considerable investment to verify the reliability of these declarations.

- The GDPR only processes personal data in the strict legal sense while the business model of the platforms is based on the exploitation of *traces of behaviour*, aggregated into profiles and recalculated using Machine Learning to be resold into "predictive products" as they are called by Shoshana Zuboff (2019). The platforms argue that this data is "neither personal nor data in the sense of legal identifiers" and therefore must be considered as outside the scope of the GDPR.

Apart from this remarkable initiative, which has become an example of taking back control of the governance of the platforms, many announcements have been made, including the possibility of dismantling them, without actions following words. By the end of 2023, 43 American states had filed an action against Meta, but we are still waiting to see how this case will progress.

However, the European Digital Services Act (DSA) and Digital Markets Act (DMA) have considered decisive acts to force platforms to behave virtuously in many areas. We will come back to this, but they have not yet produced their effects, since they only started to be implemented in 2023.

We cannot ignore the differences in pace between the technical and commercial innovation of digital platforms and the requirements of law and regulatory action. However, when action is only considered 10 years after the commercial transformation of social media, we understand that the delay is aggravated and

that reactions can no longer act against positions already acquired. The conquest of minds carried out by the American libertarians of the West Coast is, from this point of view, a very effective manoeuvre, since even liberal principles of free and undistorted competition have been abandoned in the name of competitiveness and the freedom to innovate. This could be better described, in the very terms of the libertarians, as a need to disrupt the established order and the rent-seeking behaviours. All the policies of liberal governments around the world have thus been pre-empted for 10 years by this model of radical, disruptive innovation, in defiance of the law, to the point of becoming major political axes, such as the discourse on the "start-up nation" in France. One day it will be necessary to produce a detailed history of the discursive, institutional, financial manoeuvres and networks of influence which enabled this powerful contagion of an ideology. Certainly, the so-called ideological version of the "new economy" of the late 1990s foreshadowed this state of mind, which caused the fall in the internet stock market in July 2000, a classic bubble produced from a shared belief. But the lessons have not been learned since the same laissez-faire attitude presided over all of these 2010s, even though the major financial crisis of 2008 could have encouraged caution. On the contrary, the stock market valuation of all .com firms has reached unprecedented heights of speculation and has allowed them to radically consolidate their anti-competitive positions through ever more impressive takeovers. The certainty of regular crashes in hyper-speculative areas such as virtual currencies or non-fungible tokens (NFTs) even seems to have become a constant and normal element of any strategy in these sectors.

The demand for "freedom to innovate" has above all proven to disguise a claim for a lawless zone in all areas. Let us consider the following:
- Abusive exploitation of media sources (copyright and media).
- Capture of attention by dark patterns which generate viral excesses or pathological poisonings, well known but deliberately cancelled (Facebook Files, Frances Haugen).
- All-out purchases under elimination pressure (Instagram, WhatsApp, YouTube, Waze, etc.).

- Tax fraud disguised as optimisation thanks to tax havens (tax law, ad hoc agreements to obtain de facto non-taxation, Lux Papers, Amazon).

- Lock-in strategy (Microsoft).

- Deliberate legal guerrilla warfare against local regulations (Uber).

- Capture of advertising revenue in total opacity (auction mechanisms).

- Revision of advertising placement algorithms and selection of totally opaque publications (Facebook and overvalued videos).

- Reduction of information transmitted to brands and agencies to remain the obligatory points of passage (Google).

- Work subcontracted to millions of underpaid precarious workers for moderation or validation of algorithms (Mechanical Turk from Amazon).

- Refusal to transmit personal data to allow portability and interoperability from the user's point of view (Max Schrems' judgments of the European Court of Justice).

- Submission of the data collected to the American jurisdiction (Privacy Shield).

- Closure of access to data for research despite supposedly virtuous systems (Science One) etc…

Only obvious anti-competitive practices were subject to sanctions, which became increasingly severe.

Two points deserve to be underlined to understand the omnipotence of these platforms:

1. *The advertising regime* that generates these platforms revenue has in no way been affected by the measures that have been taken so far. This allowed Meta and Google/Alphabet to capture 75% of online advertising and thus further increase their competitive power and their attractiveness to investors. This lever of action aimed at the business model has never been activated even though it is the key link in the chain which goes from users to investors, since it is the attention of both which is captured and monetised. However, this is also what explains the profound change in the nature of these platforms from 2008-2009 (YouTube, Facebook, and Twitter). All have adopted the same advertising monetisation model, implemented in the two main companies by the same person (Sheryl Sandberg, moved from Google to Facebook in 2008). However, it is as if

the regulatory authorities and the states have remained fixed on the status of social networks as they were in the 2000s between friends, during the era of web 2.0 – when the internet was social, friendly and distributed. This period was also concerning due to its non-market nature and inability to generate income. Regulating such social networks, with their admittedly excessive freedom of expression, did not seem so urgent or complicated. Their free access, however, established habits and intensive usage, which some were concerned about, particularly for the attention of the youngest users. This process was reclassified as individual user addiction, but the responsibility of the young platforms remained largely unaddressed. However, from the end of the 2000s, and with the contribution of cognitive sciences derived from captology (BJ Foggs, Stanford), the architecture of the platforms and their publication rules were revised from top to bottom to encourage attention capture and advertising monetisation. We are no longer dealing with the same companies, the same services, or the same level of economic power, even though the friendly discourse of "sharing" and "network of friends" still constitutes the ideological justification put forward. This justification works as a shield, since any attack against the platforms becomes an attack against the spirit of sharing and conviviality, which is supposed to be present even now. This storytelling, however, works to mobilise all users who are fans of these free solutions where they can reap rewards - shots of hormones of recognition - clearly visible in the vanity metrics (Boullier, 2022), since everything is measured and published.

2. *The emergence of generative AI* constitutes a factor in aggravating the situation for two reasons. First, the volume of content propagated without any reliability will soon become astronomical, since any publication can be generated automatically from a prompt summary. The poor quality of the information is not a problem since these are also machines that read, process, or calculate these publications (text, images, and videos). The self-referential circle thus created will saturate the entire cognitive space of the public and generate a disorientation much more powerful than currently. Second, all of these calculation capacities and this expertise in AI are completely monopolised by the same platforms, which can either buy existing solutions (Microsoft, with Open AI) or buy the necessary human skills at a price beyond the reach of the competition (Google, which also

acquired DeepMind in 2014). The balance of power that I describe here can only be amplified for the benefit of the platforms. The lobbying power of these platforms during the drafting of the European AI Act demonstrates this.

Chapter 2: Are the attempts by the European Commission to take control through regulation worthwhile?

Several approaches have been adopted by the European Commission to attempt to manage the take back of control of media activity on social media. Each approach corresponds to an overall classic vision of the social causes that we spontaneously attribute to these crises which seem unmanageable.

1. Denounce the bad actors and limit their power to do harm

The cause of the deterioration of social media is supposed to come primarily from a few enemies of democracy, including countries, hackers, and malicious actors of all types. It is true that it only takes a few very active and automated accounts to flood a platform with destabilising content, which in turn indicates a flaw in the control of virality. In the measures adopted, it is then a question of identifying malicious accounts and blocking them in one way or another. The key point then lies in the reporting and the speed of reaction to this reporting. Reporting is expected from users first and foremost, then from platforms, and finally from police authorities who can carry out their own surveillance, broad or targeted. This orientation poses technical problems which I will recall, but above all, it is very reductive on the conceptual level.

Research work on propaganda has existed since the 1920s and was particularly developed during the Second World War. Many forms of combating this voluntary, organised, and strategic disinformation have been tested, but their limits have since been discovered. More recent research, such as that of David Colon (2019, 2023), shows that information nuisance strategies are still very much alive but also that they have evolved considerably due to the reconfiguration of media space. Attacking malicious sources is therefore necessary but cannot be done without identifying the technical and media strategies of these sources. Their extremely distributed and versatile nature is striking and leads to a permanent race of cybersecurity forces after multiple actors who re-emerge in other forms. Worse still, the effects of polarisation, which are often described as critical in the

destruction of the climate of debate on social media, cause similar forms of behaviour to propagate in opposing camps (the camps of good and evil, to use the most simplistic and moralistic operational distinctions). The fight is therefore no longer only against identifiable sources but against deliberately toxic methods of communication, which contaminate even states, decision-makers, or communicators supposedly the most virtuous.

Furthermore, when denunciation becomes the only regime of cohabitation in these media spaces, we can see that a classic perverse effect is established. This results in the sabotage of democracy, the erosion of trust it presupposes, and in particular, the undermining of freedom of expression, which was supposed to be one of the pillars of this regime. The whistle-blowers, who are the users, find themselves immersed in these climates of hatred, violence, and disguise. Current procedures do not include any substantial means of protecting them, supporting them, or organising them collectively. We know how painful this detection work is in the moderation services of the platforms, which have relocated this operation to low-income countries, avoiding making the trauma of these operators visible. Relying on individual denunciation constitutes a very risky bet for those who practice it. This is the case at the level of interpersonal relationships of harassment, which leads to reluctance to report. The installation of procedures for collecting these reports within the platforms has proved insufficient, and their reinforcement by public services, such as Pharos in France, indicates the complexity of the matter.

It is therefore a system of autonomous and collective social control that is missing. I had the opportunity to work on similar questions of "neighbourhood watch," particularly by studying the Guardian Angels in San Francisco and Los Angeles (Boullier, 1986). While limited deviations could be observed, the advantages of this collective vigilance were remarkable since they required the organisation of collectives, the establishment of rules, and a presence on the ground as close as possible to potential incidents. An equivalent can be found in the way Wikipedia manages attempts to poison the encyclopaedia (148 administrators, 17,000 contributors on Wikipedia France) through organised attacks or more subtle manoeuvres. The encyclopaedia's self-monitoring (Cardon and Levrel, 2009) is permanent, and the reaction is often much faster

than one might think, even if it means creating revision battles that can last a few days or a few weeks. But this distributed energy is the only one capable of both maintaining a wide surveillance zone and making freedom of expression debatable and collectively controllable. We can therefore consider that the group structure on Facebook is an interesting method for distributing the power of control and self-surveillance. This could even suit the platform, which makes its responsibility (not editorial since it refuses this status specific to the media but as a moderator) distributed among several intermediaries. No proposed law or regulation fits into the social architecture of these platforms, which were originally supposed to be social "networks" (even if more accurately, the English term emphasises the status of social "media," and the European directive of 14 November 2018 speaks of "social media services" in its preamble). We could say that the imperatives of re-establishing a livable media sphere place the platforms before a clear alternative:

• Either become *social networks* again, that is to say, based on the distributed activity of participants by offering them spaces for collective regulation, with sets of moderation rules and responsibilities closer to Wikipedia. This can only be done by structuring into smaller groups, which is also found in Mastodon instances, for example.

• Or *become media*, certainly social, but with all the editorial responsibility of any media. This will require extremely strong moderation, exercised a priori as do most newspapers which still keep readers' contributions (many others decided to stop).

Their current hybrid status, further complicated by their economic imperative of advertising, does not make it possible to deal coherently with the regulation of a social media sphere of this type. We should therefore move from the idea of regulation by states, agencies, or administrations to that of a **design of regulation on these platforms** - namely, how to **prescribe possible collective self-control formats**. Legislators must then be able to **propose standard frameworks** since the experience of 15 years of life on these networks now allows us to learn lessons for survival.

2. Speed up the reaction of control agencies

All the entities that are mobilised for moderation and the suspension, banning, or legal complaint procedures that may follow, seem to agree that they need to increase their resources and in particular, their speed of reaction. It must be noted with interest, however, that during 2023, all platforms reduced their number of moderators, already notoriously insufficient in most languages and in any case, very unevenly distributed. This measure, which should lead to immediate sanctions from the control bodies, in particular under the articles of the DSA which establish an obligation of means, clearly reflects the double discourse used on the part of the platforms. This is because, at the same time, they can argue their commitment to increasingly efficient interventions of their AI systems, which are supposed to replace human operators. However, no methodologically rigorous evaluation can guarantee the reliability of such solutions.

The workforce of the platforms, their algorithmic processing capabilities, and the speed of their procedures thus become key elements in the performance of the treatment of nuisances on these networks, without questioning the origin of this obligation of responsiveness. This obligation now extends to police services, such as with the Pharos platform, and beyond to the judicial system. Some have recently proposed increasing the performance of the legal response by creating specific chambers equipped with AI to accelerate the processing of all these files (Birenbaum and Ollivier, 2023). All of these measures have the particularity of *adjusting to the rhythm of virality by considering it as an indisputable variable*. They cannot imagine reacting in any other way than a posteriori, as moderation on today's social networks does, and not on the media model already mentioned, which would authorise a priori moderation. Some kind of dogma of high frequency seems to reign over the models of transaction (such as the one click dogma), quite similar to what is happening in finance with High-Frequency Trading (HFT). HFT obliges investors (in fact the algorithms) to immediate reactions to increasingly faster market signals, without real economic significance, since transactions are authorised at an extreme fractionation level, that of the cent of a dollar, or even below, and cancelled at the end of the day for a large part of them. An entire system switches to a speculative regime which can

only operate through algorithms, without ever calling into question the social utility of this high-frequency.

This, however, is precisely the aim of regulatory authorities: to further enhance their speed processing capacities, without questioning the interest of guaranteeing such speed for increasingly short content (the reign of short formats), more and more derivations, replications of the same original piece of content (automated variations); and without any real social interest: jokes, porn, memes, fake news, insults, etc., Everything is potentially distributed with equal priority as verified and useful information for public debate or even entertainment. The acceleration, as analysed by Hartmut Rosa (2010) (who did not talk explicitly about digital since his original book was published in 2005), becomes a tyranny of high frequency that contaminates beyond specific content, to the point of wanting to accelerate all legal procedures while they require, in principle, a contradictory examination and a deliberation. The high-frequency contagion does not only destroy democracy in terms of qualities of the debate but also the quality of judicial procedures and the guarantee of a correct contradictory examination. The investments necessary for such an acceleration are glaringly socially useless, once we consider the cause of all this excitement, which is solely the viral architecture which has been built by monetised platforms without any concern for the social and political consequences of their choices.

3. Put pressure on platforms to adopt good practices

The platforms are, in fact, the central actors in these critical situations in terms of collective life, so it is not a surprise to observe that most of the measures proposed by the DSA target their responsibility. But it is here that the liberal dogma gives in to the omnipotence of these platforms and their libertarian ideology: it is in no case a question of requesting a priori approvals, inspection before placing on the market, independent testing and validation of their algorithms, all of which are common in most markets. The regulators only talk about "due diligence", reporting, and good practices.

However, the obligations of declaring risks, reporting on the methods and means implemented to address them, and the requirement to demonstrate that this is

carried out within the shortest possible deadlines (see the previous point) clash with the reality of how these platforms operate. First of all, experience should have taught us that the chances of sincere presentation of these risks and the means implemented to deal with them are very low, after nearly 15 years of impunity. Next, self-administered reporting can hardly encourage the reporting of problems that have not already been dealt with and tends to overestimate the intensity of the corrective measures taken. This is also why the DSA provides for a possible audit of the algorithms and why a unit has been set up with Commissioner T. Breton on this subject. The question of "who controls" is not secondary since it will also be necessary to explain the control procedures themselves and guarantee the autonomy of action and independence of these controllers, which would not be the case when we call on private firms.

But this inspection will have to tackle an opacity built on two levels: that of the commercial secret permanently claimed by all these platforms (even if Meta suddenly discovers the virtues of open source during the fierce competition in AI), and that of the opacity of algorithms based on connectionist AI (neural networks and deep learning, not to mention generative AI, useless in this case, but equally opaque). Industrial secrecy makes it possible to maintain opacity on recommendation algorithms, prioritisation of flows and responses, ranking, advertising placement, and the creation of profiling clusters, etc., The metrics displayed to the general public or to advertising placement agencies and brands are in fact completely artificial and uncontrollable, as Facebook revealed in 2016, announcing that it was apologising for having overvalued the viewing of videos by 80%. If a correction had been made, we had to take Facebook's word for it because no explanation of the new calculation rules had been offered! The only credible solution in such a context is a **formal and precise obligation to open data to all stakeholders**, whether regulators, advertising placement agencies, or even independent researchers. It is on this condition that a form of reflexivity could be recovered on what is really happening at the very heart of these platforms, which have, let us not forget, the capacity to organise the online lives of billions of humans to increasingly longer periods.

ARCOM, the regulatory agency for media in France, recently addressed this issue, and on this occasion, I proposed that **the media model should be applied**

(Boullier, 2023). Similar to mass media audience metrics, an independent third party carries out the measurements, **a committee of stakeholders** (including agencies and brands) decides on the advertising valuations of these audiences and sets the prices in complete transparency (without auctions), and a large part of these results are made accessible to researchers and the public. Under these conditions, we can hope to obtain fairly reliable reporting. It is a key feature of these social media platforms to track all activities in detail and be able to calculate them. It is surprising, to say the least, that this actually leads to even greater opacity than for media audience metrics. The precision of the calculation would allow for finer adjustments of advertising placements and additional measures of public interest on potential addictive effects, the propagation of disinformation, and the identification of publication strategies among various accounts.

It should be noted, in terms of political feasibility, that it is truly in the interests of brands themselves to advance on data transparency, since it is currently impossible for them to recover their data, learn from it, and use it to build a strategy. Today, platforms only learn simultaneously with their algorithms, establishing a harmful captivity – typical of a reversal of intermediation into restrictive platformisation for all stakeholders. This is what S. Zoboff calls a "division of learning," which replaces the "division of labour" of the industrial economy. I propose (Boullier, 2023) that these brands participate in a campaign in favour of **the establishment of a trusted third party responsible for measuring activity on these platforms**. This is the only way to anticipate the bursting of the "online advertising bubble" (Hwang, 2020), since no research has been able to systematically demonstrate the market gains for online advertising investments, even though the advantages in terms of visibility among financial investors remains the most likely outcome. Even more problematic, the system of sanctions takes a very long time to implement and lacks any reference point: the criterion of the percentage of sales in a given area (Europe) says nothing about the truly dissuasive nature of these potential sanctions, especially when we consider the unlimited liquidity available to these platforms. In short, we will have to wait and see, but the basic principles currently mobilised by the DSA, for instance, are hardly favourable either to platform compliance or to a structural reform of the transparency of their calculations.

4. Educate users

Regulatory approaches often recognise their limits in the face of user behaviour, which *de facto* accepts rules that undermine the quality of interactions on platforms. Thus, virality, the process of which I will explain below, relies largely on the propensity of users to replicate, share, and thereby further amplify the virality of harmful or illegal content. Sometimes it even allows platforms to justify themselves by finding a large number of allies in the public, ready to defend the *status quo* for their favourite application which provides them with the rewards they expect, for free.

Education is based on the goal and principle of individual responsibility, which is useful both morally and politically, to avoid calling into question the key players in this ecosystem (from the platforms themselves to the advertising agencies and all of the providers of social listening services, counselling to brands, etc.). The principle of individual responsibility involves educating everyone in self-control to avoid becoming trapped into the hold of social media platforms. **Managing the duration of the connection, checking the reliability of posts and their sources, exercising caution in displaying oneself, being cautious before replicating a message, controlling one's language and expressions, and having the ability to rapidly alert in the event of a problem all** fall within the range of good practices expected from users**.** It should be noted with interest that each of these expectations goes directly against the strategies of the platforms, which have designed their entire system to increase the duration of attention, encourage reactivity without control (which we call *engagement*), enhance self-preservation by learning to trigger reactions with the appropriate filters if necessary, and provoke and elicit reactions in the shortest possible times (e.g., Elon Musk's emoticons), thanks in particular to the performance of the Share and Retweet buttons. In short, once an individual has been made addicted to an application by a powerful captology design, the responsibility is placed on the user to follow the paths to self-control education in the face of an entire technical and commercial system designed to take away all spirit of control and deliberation! The ordinary user is then asked to carry out selection, filtering, and prioritisation work that they previously delegated to the media they trusted.

However, the platforms made the user leave this media system based on gatekeepers by promising total accessibility to all types of content, without filters, and then asking them to do the work of the intermediaries that they have deliberately eliminated. A wonderful operation of reversing moral imperatives, indeed!

Anyway, the platforms offer possibilities for regulation since the parameters can be constantly adjusted, although it is very difficult to see the effects on one's news feed and in particular, to test the effectiveness of the adopted filtering of advertisements or advertising content. Likewise, connection duration on each application can now be displayed. The platforms are satisfied with this and have hardly invented simple and effective tools to resolve the problems of unwanted content or poorly justified recommendations. This has led the majority of the public to consider that they must "adapt" (Stiegler, 2019), since after all, their platforms are free to access, and therefore very attractive, even if the content and the formats are nothing like what one might wish for. But there is no question of slowing down TikTok or wanting to lengthen its videos, as that would affect the very principle of the platform.

As for reporting methods, they are available, but there is no way to verify the speed of processing. In cases of harassment, this is indeed the key problem - the user remains alone in taking responsibility for their denunciation and does not have support groups or advice to deal with all the dimensions of the problem posed by certain content or relationships, as reporting is rarely sufficient. This isolation of the user constitutes one of the weaknesses of all these systems, which nevertheless come originate from the world of Web 2.0, collaborative and social but in fact totally reinvented to generate an audience of content and service consumers, who do not consider themselves in any way responsible for the social worlds they have created. Here again, it is difficult to balance the benefits of the utopian ideology of the early years of the web and the advantages of the centralised omnipotence of large, financialised platforms. Yet, it is this discourse that continues to be conveyed and which sometimes even paralyses activists who fear any attack on freedom of expression on these platforms.

I would like to now demonstrate that this dual anchoring of the thinking of social media - distributed voluntary cooperation for the common good on the one hand,

and centralised control for advertising profits on the other - are two insufficient frameworks for understanding the media milieu which has been created over the last 15 years. My theory of propagation will serve as a common thread, and the concepts of epidemiology and memetics will then be able to better account for what is happening in this milieu, explaining why it is difficult to regulate it using classical terms. The ecological connotation of the term "milieu" is deliberate, aiming to highlight that, although it is an artefact that some humans created, it has now became a way of life, a habitat within which we live, unfortunately without the capacities to make it our own as we do for our homes – a concept I term "habitèle" (Boullier, 2011,2014). I shall propose creative solutions that can be drawn from it to make this milieu liveable, without waiting for a more radical challenge, which is certainly necessary but unrealistic in the current context of digital financial capitalism. Notably, all the proactive measures that I summarised before seem to act on all possible levers, but in fact, never address the very technical architecture of these platforms, the one that generates virality.

Chapter 3: Social media platforms have created a medium for high-frequency propagation

1. The conceptual biases of current regulationist intentions

These biases can be summarised in five key points:

i. We target unwanted content that is "ready-made" and not "in the making".

ii. We target "free speech" and not "free reach".

iii. We target the nodes of the networks (the transmitters) but focus less on the structure of the networks and even less the properties and capacities of the entities which circulate.

iv. We target specific content and transmitters without understanding the ecosystem that generates virality.

v. We target disqualified content and publishers, while systemic intoxication by virality also affects supposedly virtuous content.

i. The first point obviously recalls Bruno Latour's formulation ("science ready-made", already done, vs "science in the making", in the process of being made) (Latour, 1992). Its transposition into the study of unwanted content is useful here because it allows us to measure the extent to which reporting and sanctioning bodies act afterward and only focus on what was detected and obtained great visibility. However, this does not allow us to understand the properties of the propagations that generated this success, and which should be the targets of any regulation. Let us admit that it is inevitable that unwanted content will be published by, intentionally or not, malicious entities. But understanding how, on these platforms in particular, their influence can become so great, remains the key point for acting right at the origin.

The published messages cannot be characterised one by one, but should be included in a production system, largely automated, amplified by the hierarchical properties of the platform algorithms, and subject to competition to gain the public's attention. The propagandists of Prigogine's Internet Research Agency (IRA) in St Petersburg have become "propagationists," that is to say experts in virality. They know how to vary their offerings, locations, timings, and environments, as well as their formats, tones, and connotations with a sufficiently broad objective to achieve success, namely, to disorient the audiences of democratic regimes and erode their confidence in any authority. We might then target the IRA as the issuer and consider that the work is done. However, it's well understood that such a transmitter is made up of numerous clones, and that, even worse, each Internet user can unknowingly become a relay of this disinformation. It is indeed the propagation mechanism and the environment which was created by the platforms which provides the resources for this disinformation.

ii. By focusing on certain messages and seeking to control the legality, or even the veracity, of content disseminated online, it is easy for a government agency to find itself accused of infringing upon freedom of expression. "Freedom of expression" on the Internet should be understood in the Anglo-Saxon sense of Free Speech, which is constitutional in the United States, a much broader sense than in France, for example, where freedom of expression is already framed by legal texts increasingly refined over the years. This is why it is much easier to ban

and condemn comments of an anti-Semitic nature in the French legal system of freedom of expression than in the American system (which thus sees Q Anon expressions proliferate without legal sanctions). This flaw considerably weakens any content control offensive since these platforms are globalised and can always choose the most favourable arbitration for their defence, according to the legal system of reference. For this reason, it is risky to engage in the detection and regulation of content (in the absence of a political will to transform platforms into media, which would be the most radical and coherent move) and it is better to focus on "free reach" (as opposed to free speech), which best defines the viral architecture put in place, designed as such by the platforms. It is, in fact, impossible to legally justify "free reach" which would consist of requiring the right to instantly transmit any content to any person, in any social environment. The principle of the right to equal information can in no way justify the extreme speed of propagation and the infinite extent of the propagation. However, this speed and scope are the main factors of systemic intoxication of the media milieu, thus created. Any measure that would make it possible to curb the diffusion, slow down the spread, and break the chains of contagion would, in this case, be a public health measure for the mental health of the public and not an attack on freedom of expression, since the content could always be published later (which is already the case).

It is possible to break the dogma of free reach by going beyond equipping individual users with control panels that are either too uniform or too sophisticated, the effectiveness of which cannot currently be measured (preferences, parameters, etc.). To achieve this, I proposed transposing the example of road safety policies, which address a self-regulation problem by modifying the conditions for driving across the entire road network and within the passenger compartment itself.

It would therefore be possible to **enforce that platforms must install a control panel that is permanently displayed** (a feature commonly found in video game interfaces). This control panel would indicate various metrics, such as the duration of connection to the platform, the frequency of publication per 24 hours, the frequency of reactions (likes, sharing, and commenting etc.), the average reaction speed, and the evolution of these indicators over the past few days. This

control panel would give a user the means to determine their own alert thresholds, similar to cruise control in a car. Initially, these thresholds would be optional, later becoming mandatory, with any activity exceeding the thresholds resulting in an automatic **blocking of activities on the platform for 24 hours**. This suspension would only entail postponing the publication of content and would not constitute censorship. As a result, any publication would be subject to an evaluation by the user, involving a decision-making process rather than a mere reaction. This would require prioritisation based on the publication's interest. The task that was previously done by journalists - selection and prioritisation - would now be integrated into the interface for the self-regulatory work of individual users (equivalents can be developed for groups and their administrators). Through this simple affordance (Gibson, 1979, Norman, 1999), we can **bring a constant concern for controlling their activity into the cognitive field of users**, akin to monitoring speed on the road.

It is then entirely possible for a platform, or even for the police or justice authorities, to set mandatory activity thresholds, giving rise to sanctions if they are exceeded, as is done with speed controls on the road, for example. Such measures may become necessary in certain crisis situations, in compliance with court injunctions, or once studies have demonstrated the need for this type of control for collective mental health reasons (these studies have already been conducted at Facebook but the results have never been disclosed, as in the case with Instagram). (Boullier, 2020).

iii. The focus on issuers is undoubtedly necessary for a small number of them, but this conceals the role played by the infrastructure of the platforms and the algorithmic architecture, because as Lessig (1999) used to say: "code is law". The entire internet architecture has been optimised to accelerate exchanges, reduce friction, and encourage responsiveness (called engagement), all for advertising purposes. This optimisation aims to attract investments based on the belief in the market effects obtained through these targeted placements of programmatic advertising (a practice that remains unproven but is still practiced through imitative effect). Attacking this architecture to slow down exchange and sharing, and to reduce collective intoxication with the high frequency of reaction,

is certainly targeting the economic model of these platforms, but it is not threatening the strategies of brands. On the contrary, it guarantees that brands are no longer drowned in an incessant flow of malicious content, which is damaging to their reputation.

Even more so, it is no longer appropriate to ignore the properties of messages that circulate and propagate in differentiated ways. This ecosystem created by the platforms functions as a milieu in the ecological or epidemiological sense, where entities that circulate there are in competition and can find the audience they are looking for. Certain properties increase their fitness within this milieu compared to other messages. If some of these messages manage to contaminate the minds of a large audience, it is because they have combined several factors which facilitate their virality. Thus, we need to change the paradigm to analyse the capacity of certain terms and themes to survive by mutating. This approach no longer concerns only certain transmitters but the entire ecosystem, which gives an advantage to certain features of the messages. This point is related to the question of architectures because an ecosystem like Twitter invented a Retweet button in 2009, which considerably shortens the time it takes to duplicate a message and which will favour very short messages, as proposed in the initial design of Twitter called "microblogging" in its early days. A short, catchy title, a shocking photo, and a brief message also program an easier reception and above all, a reproduction without friction (that is to say, without reading, as an Inria study has shown). By a looping effect, all senders learn to replicate these properties and the ecosystem becomes even more competitive between almost similar messages, which then requires testing other more differentiating properties. **Monitoring these mutations, enabled by the traceability of digital networks,** would make it possible to understand how the regulation of certain semiotic properties of messages could be encapsulated in the platform code to slow down propagation.

iv. Tackling specific content without understanding the production matrix means merely grabbing a few copies of a continuous flow, with no effect on their generation and virality. All of the properties that I have mentioned are in fact dictated solely by an imperative of replication speed, which then values what

circulates quickly and provokes a reaction (what we call the *engagement rate*). Propagations studies comparing fake news (identified after the fact as such) and "validated" news have shown that the veracity of the news is not in question to explain why fake news spreads faster and more widely than others (Vosoughi, 2018). Their performance is explained by their "novelty score", an index constructed by Vosoughi's team, which makes it possible to calculate their degree of dissimilarity compared to the usual flow of content received on a platform and by certain types of audiences. This novelty score accounts for the psychological effect of shock, surprise, and trigger of reactivity (to approve or disapprove, to laugh or to scream). This results from a specific design of these viral entities, other versions of which would not have been able to penetrate the capsule of attentional habit that envelops most of our informational practices. This criterion should lead to an obligation to slow down a priori the rhythm of propagations because we know that anything that shocks the most will immediately be replicated more quickly than all the other messages competing at the same time. This "priming" mechanism, which allows a signal to come to the forefront of attention, is well known in cognitive sciences. It is similar to forcing the change of routines, benchmarks, and beliefs, which constitute a key attentional principle: that of loyalty, which saves a lot of energy on the cognitive level. Conversely, the stress generated by a radically new offering also constitutes an adrenaline motivation, linked to the risk experienced in the face of such content.

Contrary to what is often asserted, social media platforms do not only live by enclosing their audiences in attentional bubbles (the filter bubble, the echo chamber), which repeat and reinforce established individual preferences. They are also permeable to any shocking intrusion, which alone explains why malicious messages manage to escape from their source environment. Not only are these platforms permeable, but their algorithms promote messages that provoke a reaction and therefore constantly disrupt habits. This is in contrast to what reading your favourite newspaper could do, whose orientations, tones and centres of interest had become familiar. This familiarity arose because these publications in the traditional media were guided by an editorial policy, a specific orientation which created a horizon of expectation (Jauss) among their audience. On social media, however, the public learns *that their horizon of expectation must be*

permanent surprise and shock, which has become a strategy adopted by certain broadcasters (which is reminiscent of Naomi Klein's shock strategy). Contrary to conventional wisdom on these topics, it is therefore necessary to address above all the contagion effects *between* social circles, which are favoured by the generalist and unregulated or established nature of the platforms. Some might be concerned to observe, for example, the American far right prospering in its own social media, as was the case with "Parler" or with "Truth Social", Trump's social media. However on the contrary, we must rejoice in this, because it is virality which then collapses by limiting exchanges within partisan communities. When Habermas (2023) laments the disappearance of this inclusive sphere, distinct from the private sphere, he deplores the loss of the principle, according to his normative method. Nevertheless, he is aware that even in the public sphere made of mass media, this ideal of discussion was almost never fulfilled since everyone selected their favourite media and stuck to it. The current fragmentation was already well advanced, and we can even question whether the myth of the construction of public opinion informed by the confrontation of publics ever had the beginnings of reality, except during the few years of state television monopoly in most countries, which remains a very brief moment. On the other hand, the porosity (a term used by Habermas) of these circles has accelerated and on a large scale, as he admits. This is why my theory of propagation accounts, above all, for this process, without freeing itself from the normative concerns of Habermas. The diagnosis of the state of public debate and democracy in this media system mutation is very similar.

v. By aiming to **reset the entire architecture of social media**, we extend the effects of these measures to all types of content, well beyond messages considered harmful, well beyond malicious sources. We must accept this general revision of the principles of propagation because intoxication does not merely consist of being affected by certain contents but, in a more systemic way, of accepting subjection to a hyper-frequency rhythm. The best content, the most reliable, the most virtuous, the most beneficial for democracy, education, and the public sphere, does not need to be formatted in accordance with such a pace of propagation. This microwave format is solely generated by advertising

imperatives for platform revenues and is analogous to the speculative activity of high frequency finance. This transformation of all public discussion or cultural dissemination into tiny elements, intended to spread like diesel microparticles without the public's knowledge, is a dead end that has endangered democracies for the past 15 years. Certainly, the movement towards the short format, short-termism, slogans, and catch phrases was already underway with the mass media and the communication and marketing agencies, which tend to colonise all areas. But the force of impact and propagation of digital platforms has changed the scale of the phenomenon, and this *change of scale* is always the key element in any *mediological mutation,* as Marshall McLuhan had already indicated.

We should therefore not worry about the loss of responsiveness of the political class, which would be caused by a possible interruption of the chains of contagion. On the contrary, it would be necessary to **prohibit any use of these platforms by political personnel** because this would free up attentional time for more substantive discusssion and would force political personnel and citizens alike to engage with more complex content. This would require longer exposure formats, which can give rise to real contradictory debates, which have nothing to do with the exchange of insults that we can sometimes experience online. Likewise, on an inter-individual scale, no supposed emergency justifies the propagation of videos between adolescents in less than three minutes to a group of 30 people. All the slowing-down processes that I propose, embedded into the very code of the platforms, would save a few minutes or a few hours, which are essential for triggering possible reports of malicious, dangerous, or misleading content. This is because the decision to replicate these messages must be restored to its status as a decision and no longer equipped to become an automatism. To do this, we need to save time, which is the equivalent of "social distancing" as was inappropriately said in France during the Covid lockdown. The social distance to be taken into account is now calculated in replication time and rate of propagation and it is possible to act on this technical, social, and economic architecture at the same time.

2. How epidemiology can help analyze the virality of social networks

I propose to reorient the regulatory policies of social networks by considering these platforms as *propagation milieus* and to borrow some elements of epidemiological methods for this. I will take care to control the scope of the analogy as much as possible, but its contribution constitutes a considerable asset for the social sciences in general and for the governance of these propagation milieus in particular. We must in some way start from a terrifying observation: we have created a media monster which functions as a self-replicating milieu. When I say "we", it's too quick a shortcut, since these networks of the 2000s from the social web were in no way programmed to become these monsters of attention capture for the benefit of a few platforms. This medium has monstrous effects because it is self-replicating. The only equivalent social system is HFT, which generates a generalised economic drift towards short-termism, the assetisation of everything, the general extension of credit, and extended opacity and risk of bubbles. From a media point of view, the drift is identical; that is to say, this apparent appendage of social media to the mass media system has transformed in 15 years into a regulator of the rhythm of publications, a distributor of advertising revenue by capturing most of it for its own benefit, and generally exciting public life. From the simple YouTuber or Instagrammer to political personnel and institutions, the entire public mind is involved in a generalised "*media warming*" (Boullier, 2020). This observation does not take into account the massive doping that the use of AI, generative in particular, will constitute for this system, which now enters into all the components of this media activity to generate the *reign of permanent fakery.* I will remain focused on the question of virality to adopt an evolutionary vision of such a system that escapes its own creators and develops its own mechanisms of contagion to all neighbouring systems. This self-replicating milieu has already partly become uncontrollable, and it will be completely uncontrollable when it is fully equipped with generative AI. There is therefore still time to act and regain control, provided that we correctly think about this media sphere as a medium for propagation.

This milieu has three qualities that are ideal for generating virality. The three criteria of Big Data are relevant here and have contributed over the past 10 years

(2012 and the victories of connectionist AI and Machine Learning at ImageNet) to this change of scale.

• *Volume* is a condition for any viral entity to be able to *replicate*: it must generate peers in quantity, and for this, the *small size* is essential because survival is no longer a problem, with the number of survivors always being sufficient to start the contamination.

• *Variety* is the other criterion of Big Data that creates favourable conditions for *virality*: any entity which propagates must present different traits and variants in large quantities, both from its emergence and then during its evolution, to test other winning solutions to contaminate suitable vehicles. However, the platforms are different, the types of messages vary, and they are replicated with variations as we see with *internet memes*. *Mutation* is the key element here and, on the internet, flawless replication is not a criterion of contagion (contrary to what some proponents of memetics say, Blackmore, 1998). The opposite is true: appropriation, reinterpretation, and variation are facilitated by the digital method of content production, which will be further amplified by AI, particularly for images.

• Finally, *Velocity* is a key element that is always underestimated or little studied by the social sciences but present in all epidemiology studies. The distances between vehicles, which are human brains, have become very short since speed is at the heart of the performance of all these networks. Contamination is therefore facilitated by this de facto density created by velocity, without almost any filter, and in a much easier way than in an urban space, for example (where one can pass from one milieu to another but by physically moving).

These three Vs of Big Data are thus good indicators to describe this milieu and understand how it becomes self-replicating, without any principle of limitation and immunity of any type, despite efforts at moderation or regulation. This technical architecture has always favoured replication speed over immunity, that is to say, network security. It is not surprising that in the absence of an efficient, immune system, the well-named computer viruses (Parikka, 2016) have thus established themselves as the main threat for all organisations, and that is without mentioning the media's immune system, which existed in the form of the media and journalists but which was overwhelmed or even itself contaminated.

This approach to virality through principles of evolution on the one hand, and through analogy with epidemiology on the other, allows us to understand the destructive effects of these viral systems. Viruses have no use for their hosts, who are only vehicles, to the point where the infection can lead to the death of this host. This is exactly what occurred with Covid-19, which indicates the automatic nature of the viral replication process, without any intention. The destruction of attentional capacities and therefore the *cognitive overflow syndrome* that we now observe, and which had already been detected and conceptualised by UTC-Costech lab researchers (Lahlou et al., 1997), is not a problem for the entities that circulate nor for the self-replicating system itself. However, it can become counterproductive, particularly for brands which are the most active carriers of this virality and who happen to have financially supported a system which favours the spread of noise to the detriment of the relevant signals associated with their message. The wild competition thus generated does not select the best: selection does not proceed in this way; it has no intention, and when we say that natural selection eliminates the weakest, we twist Darwin's thinking towards that of Spencer and its social biology. If there is selection based on superior fitness and therefore adaptation to the environment, this says nothing about the social (or biological) utility of the trait resulting from the mutation. Thus, a large number of mutations in species are completely devoid of purpose, contrary to what Lamarck wanted to believe, and are nevertheless maintained over time. Because it is *random errors ("le hasard") in transmission*, in replication, which generate this key variation to offer new opportunities of survival. As Dennett (2017) says, "no taking without mistaking". This applies to content on networks which must shock in order to propagate: everything is good to test and errors of interpretation or replication during transmission are sometimes the best chances of success and contagion of other segments of the audience. Nothing strategic here, no *belvederes* (skyhooks) which would allow an overhang but only *cranes*, cranes which use the basic elements to constitute more complex elements.

Selection conditions have often been classified into two approaches, depending on whether we think of "selection for" or "selection of" (Lenay, 2004). "Selection for" constitutes an entry from the point of arrival: we observe the success of an

entity, and we reconstruct the conditions of selection as if they had been oriented towards this finality of success. This is often the easiest entry point, but its finalist or adaptive bias can then alter the investigation. The "selection of" is more modest since it starts from an element of the observed entity, from a trait, to see its evolution in species, individuals, which allows all the variations, all the variants without a priori. But this sometimes leads to dead ends; the work is long and tedious even if the results are robust. It is not possible to eliminate one of both approaches since we want to account for any virality on social network platforms and aim to act on the chains of contagion. This is why a huge methodological investment is necessary before proposing solutions, which cannot be easily propagated recipes.

However, the most useful approach in terms of selection is that which distinguishes *R selection from K selection.* R selection (for "Rate") accounts for a virality process depending on the quantity, as we have already mentioned about the volume in Big Data. In this case, fecundity is high, size is small, and dispersal is wide, and this applies to a large number of species, including insects. In contrast, K selection (for capacity of the milieu) relies on the qualities of the entity and its replication process: in this case, the entities are large, they have a long life, and they generate a small number of descendants who are well protected and cared for. This is what we observe for mammals in general and in particular for humans. However, these two selection modes are particularly suited to different milieus: R selection, operating based on quantity, is particularly relevant in unstable and unpredictable milieux, while K selection, operating based on quality, is particularly effective in situations of stable environments, but with scarce resources and strong competition.

This distinction is particularly enlightening in matters of cultural evolution and in particular, mediology (Debray, 1991). This is because the old media regime of the mass media proceeded by K selection, which requires large stakeholders. These stakeholders have, in fact, become bigger and bigger with the entry into the competition of all the financial magnates looking for the ownership of the media securities. This regime sees fierce competition for scarce resources such as public attention and advertising investment. But this environment is (relatively) stable and allows to plan for the long term. However, it has been radically shaken

up over the past 15 years by a different ecosystem which operates on R selection, that is to say, on the proliferation of content, that of platforms. Here, there is total uncertainty about the content that can win. Success must be ensured through a mass generation of ever smaller content, which is widely dispersed on all platforms. This content even ends up contaminating the mass media and intoxicating them. When Erik Raymond (2001) published his book "The Cathedral and the Bazaar" in 1999, he accurately perceived this change in media regime, which presents different fitness conditions: the cathedral - large, durable and well-maintained, vs the bazaar – with its small, ephemeral shops that proliferate other offers. When Raymond wrote it, he did so in a spirit of disruption which enchanted the era which would become known as Web 2.0. However, he did not foresee the profound and long-term effects that the generalised financialisation of this bazaar system would allow, where certain shops now have the capacity to compete with cathedrals, without assuming any of their obligations or adhering to specific meanings or institutions. This analytical framework does not lead to specific measures in terms of regulation, it only allows us to put into perspective what has taken place over the past 15 years as a systemic change in an entire milieu, which provides prodigious opportunities for propagation to certain types of entities more than others.

Another contribution from epidemiology can be mobilised usefully to think about this viral milieu constituted by social network platforms. Since Covid-19, everyone has heard of R_0 (R zero) reproduction rate, which indicates the average number of new cases of a disease that a single infected and contagious person will generate on average in a population without any immunity (people without immunity are called susceptible people). This is an index of the start of an epidemic which can change over time, and we will then speak of R_e or "effective reproduction rate" in a population which includes healthy (Susceptible), Infected or Recovered people (SIR). Kucharski (2021) suggests breaking down the factors on which the R_0 depends into four which he calls DOTS to facilitate memorisation:

• D for duration, the length of time a person is contagious.

• O for opportunities, the average number of opportunities it has to spread the infection each day it is contagious.

• T for transmissibility, the probability that an occasion results in transmission.

• S for Susceptibility, the average sensitivity of the population.

When doing a mediology of a milieu like the one of platforms, these indicators are particularly useful for anticipating the probabilities of an item's propagation.

- *The duration* is not easily calculated, even in the case of viruses, because the incubation period does not always produce visible symptoms and it is also necessary to take into account healthy carriers, who carry the virus without being affected themselves. For the media, we rather speak of duration of *exposure*, a classic term of media studies. This exposure can be increased by repetition, which is what the propagators of toxic messages deliberately do, but also by automatism when the algorithms of the platforms boost the visibility of a post of this type by its popularity or by placing it in the trending topics for Twitter. Likewise, the automated scrolling in the fashion of TikTok allows the public to be exposed not just to a single piece of content but to similar content, as long as it has been viewed until the end (in particular, videos of choreographies, but also accidents, humour of a certain type, even problematic content). This element plays on available brain time, competing with all other content by saturation of attention time. It also creates a habituation to the continuous exposure of very brief content, which generates a non-selective cognitive posture of bewilderment that can otherwise be very satisfying. The measures to be taken to reduce the duration of exposure must not be limited to the time spent and alerts, which would make it possible to signal an excess and which certain applications now offer. We must **prohibit any automatic scrolling mechanism, any trending topics display,** and en**force the introduction of friction** into the attentional mechanism. As a result, we drastically reduce the chances of unwanted content slipping in due to the category proximity effect alone.

- *Opportunities* are among the essential factors of media virality since they depend on the number of contacts on a given platform. It has become ordinary and even considered a reputation criterion to aggregate the greatest possible number of followers and friends, whose content is then favoured in the placement of posts. However, in this way, the principle of the original social networks is completely distorted. The *real relatives* (or even friends) are only a few dozen at most, while a significant number of *contacts* are, in fact, only a form of subscription to the publications of an influencer, sometimes of modest size, but who nevertheless adopts a media strategy. Here again, we find the separation between social media and social networks. **The return to a definition of platforms as "social networks"** would make it possible to **limit the potential affiliations of relatives to 150**, which is known as Dunbar's number or "the law of 150". The anthropologist Dunbar (1992) showed the radical change in organisation once direct acquaintance is no longer possible (beyond 150 people). This principle is at the basis of the *holacracy* movement. The structure of organisations (here, its size) thus promotes loyalties and the circulation of reciprocal recognition, while exceeding the threshold leads to the propagation of a form of *anomie*. The liveable nature of social networks presupposes a form of reciprocity, mutual knowledge, and loyalty, which should require limiting these opportunities for

exchange to 150 people or accounts. This is one more way to force the public to prioritise, by adding new members to the circle on the sole condition of eliminating some others. Once we exceed this number, we enter another media dimension which becomes similar to subscribing to a media as an exposed audience and not in a reciprocity of identifiable people. In this case, it becomes justified to demand payment, to mark the change of status and entry into the classic world of media. Some influencers may deliberately seek to grow their followers and should therefore be willing to pay to attract an audience. Likewise, **this reference number, 150, should apply to groups in social networks**, which can then be subdivided into subgroups if they are very popular, but on the condition that a new group administrator takes charge of the orientation and governance of the group. **This measure radically reduces reach but increases the density of exchanges and therefore the social control of the collectives themselves**. The opportunities are then reduced, *in the same way that during a contagion we reduce the opportunities for mass gatherings.*

This equivalent form of social distancing, which works by selection, can be amplified by the use of equivalents of masks that slow down and filter exchanges. This is what Twitter did when it introduced a prevention message before any retweet, asking users if they had read the post they wanted to retweet. This single action reduces the opportunities because **prior reading creates friction in the automation installed on the platform and incorporated as a habit by users**. Twitter did this based on an observation made by researchers at Inria (the French main computer science research agency) that 60% of retweets concerned messages that had not been read (Gabielkov et al., 2016).

- *Transmissibility* depends directly on the properties of the virus, which has a greater or lesser capacity to infect all those exposed to it. It is its own power to act, its *agency,* and the precise analysis of the properties of viruses and the comparisons between variants that make it possible to see what in these properties provides a competitive advantage to certain variants (at different phases of contagion: attachment, penetration, decapsidation, replication, assembly). This analysis can be carried out on the semiotic properties of the content as soon as we undertake systematic comparison work, platform by platform, because the formatting constraints vary from one to the other. These qualities of transmissibility are, in fact, taught by the platforms themselves (Google trains Youtubers since attracting a larger audience also interests the platform) or by intermediaries who become influencers specialised in promoting their YouTube, Instagram, Facebook, TikTok, or Snapchat accounts. It is then possible to **act on these recommendations, advice, tips, etc., to discourage anything that could lead to virality during these training sessions**. **Specific formats should even be banned or moderated, as could be the case with filters on Instagram and** generators of psychological disturbances identified by Facebook. The latter, which was revealed by Frances Haugen, led to the departure of the founders of the original Instagram. **The proliferation of short formats such as stories, reels, and so on must be kept under control as well.**

This vigilance on transmissibility, just like that on opportunities, requires delving **into the specific features of messages and networks that directly influence**

the virality of content. This cannot be done without both a robust analysis and a strong political will to address content propagation structurally. Mass media regulations have evolved in this direction, even going as far as prohibiting certain image processing and requiring a clear differentiation between advertising and information content. Achieving this goal is not impossible but requires great finesse, access to control of algorithms, and mass detection methods, which AI should allow.

- *Susceptibility* is more a matter of properties of the population. A weakened person, with a fragile immune system, is more likely to be infected. In the media field, it is also possible to identify audiences who demonstrate immune weakness. The level of education in general can be an indicator, as we can consider that any training should include **education in critical thinking**. We still need to verify that this is the case and find the means to widely **disseminate media education** extended to digital media, which is often insufficiently present in schools. To this end, I inspired a digital education model in the canton of Vaud, Switzerland, with my colleagues from the EPFL Learn Laboratory (Boullier et al., 2023), which reached more than 2000 teachers at all school levels. It offers, in an inseparable way, learning the technical principles of digital technology closely coupled with education in socio-economic dimensions, the plural possibilities of delegation to machines, and the ethics of digital uses. More precisely, on platforms, **any display of popularity scores, called *vanity metrics*, weakens the immune system of an entire public and should be prohibited.** Nowadays, publication work is focused on the objective of improving scores, visibility, reputation, and capacity to generate greater responsiveness (engagement). This phenomenon was particularly visible during the urban riots of June 2023 in France when the geolocation of posts on Snapchat maps allowed each neighbourhood to compare itself and measure its notoriety based on reactions to videos published depicting mortar throwing or attacks on various buildings. As soon as imitation mechanisms are installed on platforms, they increase the susceptibility of the public, independently of the content, by favouring the most shocking content. The public then becomes hypersensitive to its image and is willing to do anything to increase its visibility, with the issue of reputation being a key point in this attention economy, just as in the financialised economy.

Conclusion

Have we truly become aware not only of the excesses of social network users and the responsibility of the platforms in these excesses, but also of the formal flaws of the ecosystem that they have created over the past 15 years? We can't even say that Google or Meta knew what they were doing. Certainly, since 2008, these firms have been willing to employ any means necessary to capture the public's attention, often with complete disregard for legality or responsibility. However, apart from this greed, which is the common lot of all digital financial capitalism and its profound contempt for the law, no one could imagine the autonomous dynamics of such a virality-driven ecosystem, which has become a self-replicating milieu and a vector of high-frequency propagation. The current state is the complete opposite of a media sphere moderated by gatekeepers, who were once mass media journalists, and also the complete opposite of the hopes for cooperation and sharing envisaged by the pioneers of the web and web 2.0. It is high time to acknowledge this, before we see this architecture irreversibly *amplified by generative AI, which could introduce us to an era of generalized fakery with self-replicating uncontrollable content.* We can regulate AI systems on one side, but if we don't take into account the milieu that they will further accelerate, we lose. We can regulate the content of social networks and urge the platforms to be a little more responsible, but if we do not understand that they themselves are overwhelmed by the monster they have created, we lose. We must resolutely tackle an overhaul, a *reset of the entire architecture of social networks*.

This reset can be radical: decoupling social networks from financialised and opaque platforms, prohibiting the economic model of advertising, favouring subscription, and forcing platforms to adopt the status of media and the editorial and legal responsibility that goes with it. However, given the prevailing liberal ideology in decision-making institutions with a semblance of power to act, it is more reasonable to propose *a reformist reset*. This should be based on a solid understanding of *the self-replicating nature of this milieu*, its virality-driven functioning, and the need to act on all the levers that can slow down the high

frequency of propagation, cutting the chains of contagion and freeing the public's minds from the influence that has taken hold.

The measures listed here have been proposed in the most concrete way possible, which easily leaves the door open for some superficial criticism. Therefore, they must serve as starting points for **committees tasked with redesigning social networks,** resulting **in specifications that are a priori binding** for any operator who would like to intervene in the European market. It remains somewhat paradoxical to note that it is the precepts and the ideological intoxication of all decision-makers by libertarianism which have created an ideal playground for the great saboteurs of democracy - multiple dictatorships - to flourish worldwide over the past 15 years. The political battle must thus confront two enemies at the same time, and any timid posture towards libertarians that only focuses on showing strength against authoritarian regimes and their destabilising practices, would at the very least be, hemiplegic, if not hypocritical or even complicit in the organised sabotage of democracy as a political regime.

List of recommendations

Here are the recommendations published along with this policy brief, which can be considered as an action plan for decision makers, provided that the conceptual framework is understood and shared, and that a true political will could inspire them:

- **Regulation must address design issues on these platforms**, **namely how to define possible collective self-control formats**. **Legislators must then be able to propose standard frameworks for algorithms and interfaces.**

- **The workforce of the platforms, their algorithmic processing capabilities, and the speed of their procedures should be made accessible and monitored for control by independent authorities.**

- **Formal and precise obligation to open data to all stakeholders.**

- **The media model must be applied: an independent third party carries out the measurements**, **where a committee of stakeholders decides and validates the prices and the strategic choices.**

- **Good practices expected from users must be trained and supported**:
    - **Controlling the duration of the connection**
    - **Checking the reliability of posts and their sources**
    - **Being cautious in displaying oneself**
    - **Being cautious before any replication of a message**
    - **Controlling one's language and expressions**
    - **Having a rapid alert system available in the event of a problem**

- **Establish a trusted third party responsible for measuring activity on these platforms**.

- **Enforce that platforms must install a control panel which is permanently displayed.**

- **Any activity exceeding a pre-defined frequency of reactions (likes, shares, comments) would automatically result in a blocking of activities on the platform for 24 hours (free speech is OK, free reach is not OK).**

- **Bring into the cognitive field of users a constant concern for controlling their activity.**

- **Monitor the mutations of messages, enabled by the traceability of digital networks, through permanent research.**

- **Reset the entire architecture of social media**.

- **Prohibit any use of these platforms (direct or indirect) by political personnel.**

- **Prohibit any automatic scrolling mechanism, any trending topics display, and enforce the introduction of friction in any publication and reaction activity.**

- **Return to a definition of platforms as "social networks" that would make it possible to limit the possible affiliations of relatives to 150 contacts.**

- **The limitation of 150 contacts should apply to groups in social networks.**

- **The limitation of 150 contacts should radically reduce reach but increase the density of exchanges and therefore the social control of the collectives themselves.**

- **Promote prior reading before publication in order to create friction in the automation installed in the platform and incorporated as a habit by users.**

- **Control the recommendations, advice, tips, etc. delivered during the training sessions of influencers, such as YouTubers etc., to discourage anything that could lead to virality.**

- **Specific formats should be banned or moderated, such as filters on Instagram.**

- **The proliferation of short formats such as stories, reels, and so on must been kept under control.**

- **Enter into the specific features of messages and networks which act directly on the virality of content.**

- **Disseminate media education and education in critical thinking.**

- **Any display of popularity scores, called *vanity metrics*, should be prohibited since it weakens the immune system of the public.**

- **Establish committees to redesign social networks which will result in specifications a priori binding for any operator.**

Bibliography

- Birenbaum et Ollivier (2023), « La déconnexion entre la rapidité de diffusion d'un contenu préjdiciable et le tempo judiciaire se révèle chaque jour un peu plus », *Le Monde*, Débats, 31 décembre 2023.
- Boullier, D. (1986), « Les Guardian Angels aux Etats-Unis », *Annales de la Recherche Urbaine*, n° 31, pp 125-136.
- Boullier, D. (2011), "Habitèle virtuelle », *Revue Urbanisme*, n° 376, Janvier-février 2011.
- Boullier, D. (2014), « Habitele: mobile technologies reshaping urban life », *URBE*, v.6, n. 1, pp.13-16.
- Boullier, D. (2020), *Comment sortir de l'emprise des réseaux sociaux*, Paris, Le Passeur éditeur.
- Boullier, D. (2022), Médiologie de la vanité en ligne, *Esprit*, n°489, Septembre.
- Boullier, D. (2023), Instituer l'accès aux données des plateformes numériques, *Revue en ligne AOC*, 20 Novembre.
- Boullier, D. (2023), *Propagations. Un nouveau paradigme pour les sciences sociales*, Armand Colin, 2023.
- Boullier D., F. Chessel-Lazzarotto, G. Liégeois, F. Mondada, D. Badoux et S. Agrebi (2023), « Un modèle pluraliste d'éducation numérique, l'expérience du canton de Vaud en Suisse », *Distances et médiations des savoirs* [En ligne], 43 | 2023
- Cardon, D. et J. Levrel (2009), « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, 154, p. 51-89.
- Colon D. (2019), *Propagande. La manipulation de masse dans le monde contemporain*, Paris, Flammarion.
- Colon, D. (2023), *La Guerre de l'information. Les États à la conquête de nos esprits*, Paris, Tallandier.
- Dennett D. C. (2017), *From Bacteria to Bach and Back. The Evolution of Minds*, Penguin Books.
- Dunbar R. I. M. (1992), "Neocortex size as a constraint on group size in primates". *Journal of Human Evolution*, 22 (6) : 469-493.
- Gabielkov M., Ramachandran A., Chaintreau A. et Legout A. (2016), "Social clicks : What and who gets read on Twitter ?", *SIGMETRICS*, 44(1), 179-192.
- Gibson J.J., (1979), *The ecological approach to visual perception*, New-York, Houghton Mifflin.
- Habermas, J. (2023), *Espace public et démocratie délibérative : un tournant*, Paris, Gallimard, 2023.
- Hwang T. (2020), *Subprime Attention Crisis, Advertising and the Time Bomb at the Hear of the Internet*, New York, Farrar, Straus and Giroux.
- Kucharski A. (2021), *Les lois de la contagion*, Paris, Dunod

- Lahlou S., Lenay C., Gueniffey Y., Zacklad M. (1997), « Le COS (cognitive overflow syndrome) », *Bulletin de l'Association pour la Recherche Cognitive*, n° 42, p. 39.

- Latour B. (1987), *Science in action: how to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard University Press.

- Lenay C. (2004), *Darwin*, Paris, Les Belles Lettres.

- Lessig L. (1999), *Code and other laws in cyberspace*, Basic Books.

- Norman, D. (1999), "Affordances, Conventions and design", *Interactions*, vol VI.3, pp 38-43.

- Parikka J. (2016), *Digital Contagions. A Media Archaeology of Computer Viruses*, New York, Peter Lang.

- Raymond E. S., (2001), *The Cathedral & the Bazaar*, O'Reilly.

- Rosa H. (2010), *Accélération*, Paris, La Découverte.

- Stiegler, B. (2019), *« Il faut s'adapter » : Sur un nouvel impératif politique*, Paris, Gallimard, coll. « NRF Essais ».

- Vosoughi S, Roy D and S Aral (2018), "The spread of true and false news online", *Science,* 359 (6380), 1146-1151.

- Zuboff S. (2019), *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York, Public Affairs.

About the author:

Dominique Boullier was trained as a sociologist and a linguist. He is a full professor of sociology at the Institut d'Etudes Politiques de Paris (Sciences Po) since 2009 (emeritus since 2022). Researcher at CEE (Centre d'Etudes Européennes et de Politique Comparée). Former director of the Social Media Lab at the Ecole Polytechnique Fédérale de Lausanne (EPFL) (2015-2019). Former scientific coordinator of the Medialab at Sciences Po with Bruno Latour (2009-2013). Former director of Lutin User Lab (Cité des Sciences, Paris (2004-2008) and Costech lab (Université de Technologie de Compiègne (1998-2005). Former entrepreneur in the 90s.

He is the author of articles and books on urban issues (L'urbanité numérique, L'Harmattan, 1999; La ville-événement, PUF, 2010), on digital technologies (Sociologie du numérique, Armand Colin, 2019, 2nd edition). His current research focuses on the technical and institutional conditions for survival in the digital world: "Comment sortir de l'emprise des réseaux sociaux" (Le Passeur éditeur, 2020), "Puissance des plateformes. Territoires et souverainetés" (Sciences Po Chaire digital, 2021), "Habiter le numérique. L'habitèle", (forthcoming) and on the conceptual and methodological mutations required for the social sciences: "Propagations. A new paradigm for the social sciences" (Armand Colin, 2023).

About the Digital, governance and sovereignty Chair:

Sciences Po's Digital, Governance and Sovereignty Chair's mission is to foster a unique forum bringing together technical companies, academia, policymakers, civil societies stakeholders, public policy incubators as well as digital regulation experts.

Hosted by the School of Public Affairs, the Chair adopts a multidisciplinary and holistic approach to research and analyze the economic, legal, social and institutional transformations brought by digital innovation. The Digital, Governance and Sovereignty Chair is chaired by **Florence G'sell**, Professor of Law at the Université de Lorraine, lecturer at the Sciences Po School of Public Affairs.

The Chair's activities are supported by:

About Project Liberty Institute:

Project Liberty Institute, 501(c)(3) founded by Frank McCourt, is bringing together technologists, academics, policymakers, civil society and citizens to build a safer, healthier tech ecosystem. The Institute has an international partner network that includes Georgetown University, Stanford University, Sciences Po, and other leading academic institutions and civic organizations; and is the steward of the Decentralized Social Networking Protocol (DSNP) which is available as a public utility to serve as the bedrock of a more equitable web and support a new era of innovation that empowers people over platforms and serves the common good.