Maël Pégny

Is Virality a Digital Concept?

Abstract: Even though virality has been widely debated in recent years, its exact relation to the digital medium is rarely explicitly discussed: Is virality just a digital version of word-of-mouth, or is it a concept of its own, necessary to understand new social phenomena induced by digital media? We try to answer this question with a particular focus on Nahon and Hemsley's definition of virality in *Going Viral* (2013). First, we try to understand viral messages among other forms of horizontal communication and their relations with cognitive techniques enabling their recording. We then move on to a detailed examination of Nahon and Hemsley's definition, trying to uncover its implicit relations to the digital medium, especially the relations with communicational costs and virality metrics. Finally, I endeavour to update their definition with an account of phenomena non-existent at the time of their publication, such as the automation of content moderation and recommendation through Machine Learning (ML). All the analysis supports the thesis that virality should be distinguished from "analogue" word-of-mouth and understood as a proper digital phenomenon, justifying a new periodisation of the history of media.

Keywords: virality, definition, digital media, periodisation, machine learning, content moderation, recommendation, ranking

It cannot be assumed that our problem, as stated in the title, is self-explanatory, so let us start by giving the reasons for this particular question. The notion of virality is part of the history of digital marketing: From its early days on[1], the concept was meant to refer to a successful digital marketing campaign, the canonical example being the Hotmail campaign. In this example, a message signalling the possibility of a free Hotmail account was attached at the bottom of each mail sent from a Hotmail account, leading to an explosion of subscriptions. It would thus seem obvious that virality is meant to describe a property of the diffusion of a given digital message.

However, many of the various definitions of this property make absolutely no reference to the digital nature of the message. In particular, the most famous and most sophisticated definition, which will be used and analysed throughout

1 The first mentions of the concept or its affiliates, such as "viral marketing", appear in the late 1990s to early 2000s on Google Scholar (see Wilson 2000, 232).

Open Access. © 2024 the author(s), published by De Gruyter. BY-NC This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
https://doi.org/10.1515/9783111311371-002

this article, i.e., that of Nahon and Hemsley (2013) in their book *Going viral*, makes no explicit reference to the digital medium where the viral message is supposed to spread.

This creates a hermeneutic ambiguity: Is the notion of virality supposed to be applied strictly to digital media, as it would describe a specifically digital phenomenon? Or is it just a synonym for the older "word-of-mouth", which marketers, politicians and intelligence services have tried to exploit long before any digital medium existed? Independently of the intentions of this or that author, or a particular terminological choice, our fundamental question is theoretical: Can we understand the viral propagation of online messages as a new phenomenon enabled by digital means of communication?

To answer this question, we start with a closer examination of the position of our problem and its methodological challenges, before discussing Nahon and Hemsley's definition of virality, its implicit relations with digital media, and the interest of complementing it with a more explicit consideration of communicational costs and virality metrics. Finally, we explain the need to update this definition to understand the impact of Machine Learning fuelled moderation and recommendation on the phenomenon of virality.

1 A Better View of our Problem: Virality among Horizontal Forms of Communication

1.1 Horizontal Forms of Communication

Let us call "peer-to-peer communication" a form of communication between individuals not acting on behalf of an institution. This form of communication is sometimes described as "horizontal" as opposed to the "vertical" communication between an institution and an individual or the masses. Let us state very clearly that we do not wish to separate vertical and horizontal communications as two genres without any overlap. It is well-known that many viral messages find their origin in influential media or institutional communication[2]. Furthermore, even when it does not derive its origins in a vertical form of communication, viral propagation can be reverberated and amplified by institutional comment, including when that comment takes

2 This point is already stated clearly by Nahon and Hemsley themselves in *Going viral* (2013).

the form of debunking[3]. Here again, many case studies have shown that the debunking of a viral rumour by an institution or mass media is often a major vector, if not the major vector of the propagation of some viral messages[4]. This paradox of debunking is made more significant by the fact that, as well-shown in Froissart (2002), some social studies of horizontal communication, such as social studies of "rumours", have been locked in an ideological denial of the role of mass media in spreading such messages. For instance, some case studies insist on classifying a given rumour as a pure form of horizontal propagation even when the role of mass media is obvious and well-documented, such as the famous case of the Orléans rumour in France, or with American World War Two rumours going back to the proliferation of German propaganda.

The same lack of demarcation is also true for the online/offline distinction. Many viral messages on social networks or other digital platforms also have an analogue life in paper publication (see Pailler and Schafer 2023) or oral conversations. When we talk of horizontal communication or digital communication in general, or viral communication in particular, we never imply that the message is spread exclusively by horizontal and/or digital means, only that those means have played a major role in the large-scale diffusion of the message.

Nevertheless, it is important to distinguish virality as a form of horizontal communication or propagation of a message from another major phenomenon of modern mass media, i.e., the reproduction, sometimes verbatim or quasi-verbatim, of the same message (paragraph, short news, or whole article) by multiple media outlets, sometimes in multiple countries and in multiple languages (see the introduction by Pinker 2020). This phenomenon has already been well-documented for British media in the eighteenth century by Will Slauter (2012) and has been a massive practice of modern media ever since, so much so that many old jokes and anecdotes present copy-and-pasting as a major journalistic practice. There is indeed in this case the massive propagation of a well-identified message, and this propagation could be seen as horizontal as the story jumps from newspaper to newspaper without a single source of authority being reproduced by all media outlets, as can be the case with the reproduction of institutional communication. However, it is relevant to the study of our phenomenon to maintain a distinction between this

3 This phenomenon is commented on several times regarding the concept of "rumour" in Pascal Froissart (2002). Different examples are also analysed there.

4 Camille Alloing and Nicolas Vanderbiest (2018) give a nice example of this in their analysis of the Twitter rumour of another terror attack in a restaurant after the Nice terror attack on July 14, 2016. They demonstrate that most re-tweets are actually critical of the rumour. Re-tweets without a conditional form or a source represent less than 10 percent of re-tweets of the rumour, which was initially spread by professional journalists.

mimetic behaviour between institutions with a horizontal communication between individuals *qua* individuals. The famous historian Marc Bloch (1999) would already complain in his short book on wartime rumours that social research should not mix up rumours spreading among soldiers and civilians and the rumours spread by the press, as those follow very different paths and constraints. This is especially true when so-called "journalistic errors" are actually a calculated decision to alter journalistic treatment to follow government propaganda or to augment the probability of achieving some sensationalist attraction of readership. When it comes to the study of communication and beliefs, it is always better by default to separate institutional and informal phenomena, and hence not to use the concept of virality to denote the propagation of a news item through "copy-and-paste" journalism. The confusion is made even more regrettable by the possibility, examined as a hypothesis by Bloch, that the interest for oral rumours was also an effect of the lack of legitimacy of the press among soldiers, as it was seen as propagating government propaganda.

With all those precautions in mind, our fundamental theoretical question can then be rephrased in the following fashion: How different are viral digital messages from older forms of peer-to-peer diffusion of a message? What are the effects of the digital medium on this form of communication? Such a question is relevant to historians who wish to describe the various forms of horizontal communication and their possible manipulations as a *longue durée* phenomenon. From a historical point of view, this may be seen as a question of periodisation: Does the irruption of the digital medium open a new era of peer-to-peer communication, or is the assumption of radical change introduced by the digital largely overblown and unwarranted? In denser terms, is virality just a digital form of word-of-mouth, or should it be considered an altogether different phenomenon, justifying the introduction of a new term? The question should be relevant at the very least for the history of politics, media, marketing (and other forms of business models based on the diffusion of a message) and the anthropology of cognitive techniques. In particular, this question should be further relevant to the long history of the institutional manipulation of peer-to-peer communication, as new forms of media may offer new technological affordances to exert influence on peer-to-peer communication. We shall see that the existence of new affordances for such influence is actually one of the main arguments for understanding virality as a new phenomenon distinct from word-of-mouth (see section 3).

Our question faces a significant risk of degenerating into byzantine quarrels: Historians are familiar with the endless bickering that frequently accompany the characterisation of a historical phenomenon as a rupture or a historical continuity. If such quarrels cannot be completely avoided, I would happily limit them to the bare minimum by following a simple methodological principle, namely, that the in-

troduction of a new concept and/or historical period should be justified by the risk that, in the absence of such a distinction between concept and/or historical periods, the observer might be blinded to the phenomena of interest. Our question about the digital nature of virality may be rephrased as such: Did digital media influence horizontal forms of communication in a manner dramatic enough to justify the delimitation of a new period in the history of those forms of communication and the use of a new term[5]? "Virality" is just a candidate term for such distinctions and the particular fate of this terminological item is not our core issue here. What matters most is the theoretical understanding of horizontal communication in digital times.

1.2 Virality in the Eye of the Beholder

I wish to address here the formidable difficulties raised by the relations between horizontal communications and the cognitive techniques used both to enable such communication, to study and influence it. To begin, it should be noted that one form of communication that is particularly favoured by digital media is the literal repetition of a content, especially through "one-click re-sharing" buttons. As is well-known concerning the anthropology of cognitive technique since at least the founding works of Jack Goody (see Goody and Watt 1963; Goody 1977; 1986), literality is typical of the world of writing. Literal word-for-word or even letter-for-letter repetition of a message is not only made easier by writing, but it only becomes something of value for a civilisation of writing. Oral societies, contrary to a persistent cliché, do not put a lot of value on literal repetition, do not learn a lot of linguistic messages by rote, and can easily consider identical two performances of a same narrative even if they vary widely from a literal perspective.

Technological affordances such as "one-click-share" buttons have made the massive literal copy and diffusion of a message an easy and virtually cost-free endeavour for many individuals. This remarkable new technological affordance allows for the literal spread of a message in peer-to-peer communication without the notorious deformation created by oral communication or re-writing. This is a considerable difference from traditional word-of-mouth that could easily be overlooked. This phenomenon obviously depends on the environment created by the digital cognitive technique. As a side note, it should also be noted that the literal repetition of the message makes the identification of a given content as viral much easier, compared to the multiple modifications of a message which can

5 For a recent and sophisticated proposition for periodisation relevant for digital media studies, see Boullier (2023).

grow so large as to make the initial message unidentifiable, something that has consequences for social sciences studying the phenomenon, but also for digital actors trying to trigger, detect and manipulate said phenomenon. This is even more important as "literal repetition" is no longer reduced, as the etymology would suggest, to the written word, but can now also include information in audio and video formats, multiplying the potential channels through which a message can be spread[6].

All those remarks plead for a deep influence of the digital medium on horizontal forms of communication. However, theoretical discussions of the definition of horizontal forms of communication, and their dependencies to a given medium, faces a formidable challenge well explained by Pascal Froissart. In his work on rumour (see reference above), he explains how "rumourality" may exist only in the eyes of the theoretical beholder. The first reason for this is ideological. As the term "rumour" can be used to destroy the legitimacy of a given message, the mere "description" of a phenomenon as a rumour may be an ideological move aimed at justifying authoritarian and elitist forms of control over information. The fight against rumours have been instrumentalised many times in history to justify government communication and censorship, as well as to support the authority of media outlets, sometimes with the further support of friendly scholars. This leads to the ideological blind spot mentioned above, when the very important role of official media in spreading rumours is systematically forgotten and the media is conveniently portrayed as a corporation of rational and professional "debunkers" and "fact-checkers" engaged in a heroic fight against the falsehoods spread by ignorant, impulsive and irrational masses.

The second bias is the well-known methodological one of scientific work, whereby some phenomenon or aspects of a phenomenon are privileged not because they are necessarily more important, but because they are easier to study. The study of horizontal forms of communication face at least three major difficulties. The first is their size, which obviously makes detailed study cumbersome. The second, especially relevant for some forms of communication such as oral rumours and interpersonal written messages, the difficulties of record collection

6 However, it should be stressed that the possibility of tracing every single communication act on digital social networks is largely a theoretical possibility. In practice, as exemplified and explained by Pailler and Schafer (2023) some viral phenomena left virtually no trace (such as email chains), social networks have various archiving policy (one of them being, no archive at all), different platforms collect different data, making cross-platform comparison very difficult, and the various and recent forms of internet archives need to make tough sampling decisions. In simple terms, viral phenomena face formidable archive collection, organisation and analysis challenges, and one should not give in to a mythology of digital platforms as a virality panopticon.

or the mere absence of record. The third is methodological, and concerns the definition of the object of diffusion, namely, what exactly is spreading when we say that a rumour, gossips, viral messages, etc., are being propagated? It is extremely common to remark that messages are constantly being reformulated, reshaped, commented and inserted in larger messages in informal, horizontal communications. That remains true even with "one-click share buttons", where the literal repetition of the original message is just a subpart of a larger message with various intents. Identifying the object of the propagation may thus be a major challenge, as the literal repetition of the exact same message is far too strict a criterium, and in the absence of an identification criterium the object of diffusion becomes utterly dissolved. This methodological challenge was implicitly present in the first experimental protocols to study rumour propagation proposed by Stern and then made famous by Kirkpatrick and Bartlett[7]. In this protocol and its variants, a subject is asked to repeat *verbatim* a text read to her by the experimenter. This attempt at repetition is written down and read to a second subject, and the experiment is thus iterated many times. The importance of text deformation, and the loss of many details, is a classic result of such experimental protocols. However, the relevance of such protocols to the study of horizontal forms of communication is dubious at best. In real life, one rarely listens to a long monologue without interruption, and one almost never attempts to learn the narrative of another person verbatim. Assimilation and reformulation of a message are just the norm in those forms of communication, and systematically favouring the literal repetition by rote reduce the propagators to intellectually passive transmitters[8]. The repetition of a written message by rote just makes the identification of

7 This example is analysed in Froissart (2002). It should be noted that writing is an essential part of the experimental protocol allowing the mere identification of the phenomenon. This could be analysed in parallel with the importance of audio recording instruments in Jack Goody's work on recitation of myths among the Camerounese LoDaaga. By Goody's admission in *The domestication of the savage mind* (1977), those instruments were decisive in the realisation that two recitations of the "same myth" according to the local people could actually diverge widely, and had nothing to do with the repetition of a message by rote. Oral messages have the particularity that they need the use of another medium to become an object of scientific study, which raises a systematic problem of distortion of the perception and conceptualisation of said message by the use of that medium.

8 Froissart makes many other relevant critical remarks, such as his criticism of the intersubstitutability of subjects in this protocol, which does not reflect the importance of social status, interpersonal relations and authority in the propagation of a message. He also insists on the fact that many rumours actually become richer in detail as they propagate instead of the impoverishment of detail that is often seen as a major characteristic of experimental protocols (see Froissart [2022, 111]). However, we only wish to insist here on the effects due to the written medium.

the object of propagation and its recording easier for the experimenter, but it does diminish the relevance of the protocol to the study of the phenomenon of interest, which takes a shape typical of oral, not written, forms of communication. In the study of horizontal forms of communication, there is thus a considerable risk that the influence of the medium is introduced by the researcher's methodological biases rather than a true part of the phenomenon itself. Thus, one can ask, how are we to avoid those pitfalls in our understanding of virality?

Virality has the first advantage of being less susceptible to an ideological instrumentalisation than the concept of "rumour". It is more descriptive than normative, and as such is not immediately incorporated in discourses aiming at legitimising or de-legitimising a given message. This does not mean that it is completely immune from such an instrumentalisation. Media outlets often caricature the internet as a communicational Wild West where misinformation spreads virally, while the heroic media fight to maintain rationality and factuality. This type of discourse is part of a legitimation effort to defend media and government communication from a potential competitor on the market of ideas and should be treated with suspicion. However, those forms of instrumentalisation did not make it into the very definition of the concept. "Viral" is employed to describe the diffusion of official government messages as well as Qanon conspiracy theories and is thus not ideologically dubious by nature.

The second and third form of biases are more relevant to the study of virality. One of the main advantages of digital messages is a spectacular increase in the faculty of recording and collecting messages. This new technological ability is a major methodological turning point in the study of horizontal forms of communications. What is more, the "one-click-share button" makes the literal repetition of messages more frequent than it would be in oral communication, making the identification of the object of propagation much easier. However, discussions about the concept of virality should remain aware that this ease in identification and collection of a text does not imply an ease in the identification of semantic items such as "information", "news", "ideas", "beliefs", "stories", "details", and so on and so forth. Here again, the study of the social science concept of "rumour" by Froissart comes with many valuable lessons for virality scholars. One of the other major drawbacks of experimental protocols over rumour propagation was (and still is) the identification of such semantic items. When one speaks of "loss of details" as a structural trend in rumour propagation, how does one delimitate a detail? How does one assess whether a detail was forgotten or considered irrelevant by some transmitter? What are the textual units of interest whose variation must be studied: exact spelling of words, words, grammatical groups of words, entire clauses? Furthermore, how does one go from the study of spelling and textual variations to more abstract, semantic items? This last problem is made even

more difficult by the ambiguity of the act of sharing a story when it comes to infer the intentions and doxastic attitudes of the story transmitter. Sharing does not imply believing in a story[9]. As we have seen above, debunking can be a major vector in the propagation of a viral story. Recording the propagation of a message is thus not the same as recording the propagation of a belief. What is more, the transmission of a story can be a subpart of many other communication acts. Questions, irony, connection with other debates and stories, commenting, interpreting, etc., are all extremely common communication acts, which cannot be guessed simply through the mere transmission of a message. This remains true in the favourable case where this transmission is literal. In the digital world, the same technological evolutions have favoured both the literal transmission of a story and the addition of comments, images, videos, sound, the insertion of a document in another document or its outright manipulation and modification. If digital media have eased the identification of the propagation of a given (literal or quasi-literal) message, it has not eased in the least the formidable hermeneutic challenges raised by the semantics of horizontal communication and its relations with doxastic attitudes.

What is more, as already mentioned in subsection 1.1, the lifecycles of messages are by no means restricted to the digital world. A message can have part of its lifecycle in the offline world, be it in paper publication, private written correspondences, posters, physical oral conversations or phone calls, public speeches, TV, radio shows and many other forms. The ease of study in the propagation of the digital forms of a message should not be confused with an ease of study of the whole lifecycle of a given message. In many cases, part of that lifecycle might be just as hard to study as it was before the advent of digital media. The danger of overestimating digital communication because it is easier to study, especially on a large scale, is still very relevant for viral messages. In any case, if virality is considered a digital phenomenon, it should never be assumed that the entire lifecycle of a viral message is digital. Virality would be by default a property of a subpart of the lifecycle of a message, and the message itself will often be a subpart of several more complex messages.

All those methodological precautions have consequences for our initial questions of conceptualisation and historical periodization. From the point of view of historiographic methodology, it is obvious that digital media represents a major turning point, because they modify the technique of source collection, especially at scale. It is standard historiographic practice to delimitate new periods not nec-

9 Conversely, not sharing does not imply not believing: the act of sharing is not a reflection of belief.

essarily because they correspond to radical social changes, but because they correspond to a period where a new type of source or methods becomes available for historiographic study. After all, the mere idea that "history begins with the invention of writing" is one such periodisation. However, it is also well-known that such a practice comes with dangers, such as the classic confusion between an evolution of the historiographic tools with a radical evolution of the underlying society, i.e., the old prejudice that societies without writing have no history (meaning "no social evolution", which is provably false) or that writing would necessarily represent a radical rupture in a given society (which has to be discussed with caution). When we wonder if viral messages belong to a new historical period, we wonder whether the underlying social phenomena of horizontal communication have been modified, not if the tools for its study have been modified.

Here we meet a classical challenge in the anthropology of writing and other cognitive techniques, which is the huge dissymmetry in sources between a social group having the cognitive technique and the one not having it. We have much more written sources on horizontal communication since we have digital means of collection, but this makes the comparison with other less easily recorded forms of horizontal communication very difficult. Such dissymmetry in sources may easily lead to an under- or overestimation of the distinctions between media and the distinctions between periods delimitated by the introduction of those media.

The final methodological challenge is the interpretation of the relation between technological affordances and actual social practices. This reflection must constantly face the double peril of technological determinism, where the social use of technology is univocally determined by the technological affordances, and a "nothing new under the social sun" position, considering the technological medium as a transparent vessel for social practices of communication. When we will single out the drop in communicational costs (see next section) as a new feature of digital media such as social networks or email accounts, we do not mean to imply that such a drop immediately determines a modification of communicational practices. It just makes it a reasonable heuristic hypothesis to assume such modification exists and to look for it. When we do insist on the new affordances for influence and manipulation of horizontal communication offered by those media, we will not mean to imply that such forms of influence and manipulation are brought into existence only by those affordances, just that the understanding of new forms of manipulation, if they exist, must consider those affordances.

The analysis of the relation between technological affordances and social practices is made even more formidable by the concomitance of other social evolutions with the apparition of new communication technology. To paraphrase Pinker (2020) in the conclusion of their book, the evolutions of norms surround-

ing "privacy" and "intimacy" might have well caused digital communicational practices as much as they have been caused by the new technologies. In so many words, the complexity of the relations between technological affordances and social practices compels us to present all our arguments in favour of virality as a new phenomenon as heuristic hypotheses, waiting for a more complete story which will most probably include other factors to account for actual practices.

To conclude, virality is a descriptive concept, capturing the shape of the diffusion of a message identified by its form. It should not pretend to capture the complex propagation of semantic items, just to be a small step in a long analysis. It should also not pretend to even capture the whole story of propagation, as messages live through several media, and scholars should not give in to the temptation to believe that the most interesting part of a phenomenon is that which is easier to record and study. Finally, studying the dependence of virality to its digital medium is not giving in to technological determinism, as technological affordances are assumed to be a strict subpart of the story of social practices. Now that we have cleared the air of possible methodological confusions, let us go back to formulate conjectures on the novelty of viral messages in the history of horizontal communication.

2 The Definitions of Virality and their Relations to Digital Media

2.1 The Implicit Relations

Let us start by quoting extensively from a couple attempts at defining the concept of "virality". Juvertson and Draper suggest that it is

> a social information flow process where many people simultaneously forward a specific information item, over a short period of time, within their social networks, and where the message spreads beyond their own networks to different, often distant networks, resulting in a sharp acceleration in the number of people who are exposed to the message (Juvertson and Draper 1997 in Nahon and Hemsley 2013, 16)[10].

10 This article is credited by Nahon and Hemsley (2013) for the coining of the term "viral marketing". It should be noted that Steve Jurvetson is neither a scholar nor a marketer or a journalist, but a venture capitalist. "Viral marketing" as a phrase does not originate from scientific analysis, but is part of the self-description of their activities by business people: as such it is an actor's category.

Denisova (2020) adds that it is "an allegory of rapid diffusion of information and ideas", while Nahon and Hemsley note the following:

> Virality is a social information flow process where many people simultaneously forward a specific information item, over a short period of time, within their social networks, and where the message spreads beyond their own [social] networks to different, often distant networks, resulting in a sharp acceleration in the number of people who are exposed to the message. Therefore, identifying and measuring virality is made on the bases of (i) the human and social aspects of information sharing from one to another; (ii) the speed of spread; (iii) the reach in terms of the number of people exposed to the content; and (iv) the reach in terms of the distance the information travels by bridging multiple networks[11]. (Nahon and Hemsley 2013, 16)

As we mentioned above, those definitions, although extremely different from one another, share a negative attribute, whereby they do not mention explicitly the digital nature of the media. There may be an implicit reference to the digital medium in Nahon and Hemsley's work, but its existence can only be established if an ambiguity of the text is solved, and one asks, are the "multiple networks" mentioned in their definition "digital networks" or a more general notion of "social network"? At face value, it seems that the phenomenon of virality, even though it may have been introduced and is still utilised massively for digital phenomena, could equally describe communication phenomena in the analogue world, and even before the advent of the modern computer.

Let me elaborate on the most famous and most sophisticate definition, that of Nahon and Hemsley. First, their definition does not rely on a single numeric value for diffusion speed or number of views, whereby a message is not viral because it reaches x users in dt amount of time[12]. By contrast, it is defined by global properties of the phenomenon. The virality is defined by the global form of the curve describing the number of individuals seeing the information through time, which should be a sigmoid, corresponding to a slow-fast-slow rhythm of diffusion. It is also defined by the power law of rate of decay.

Another important component of Nahon and Hemsley's work is their use of clusters in social networks. Clusters of social networks denote places in the network with a strong network overlap, with many individuals sharing many common acquaintances with other members of the cluster. From the viewpoint of the

11 The reader might notice that the epidemiological metaphor at the origin of the concept has virtually disappeared from Nahon and Hemsley's definition. The modalities of "contagion" from one subject to another became inessential, as the phenomenon is defined in terms of global parameters.

12 This avoids some of the objections formulated by Froissart against some mathematical modelisations of "rumour", that were dependent on an arbitrary numeric threshold.

sociology of ideas, it is frequently a place of belief overlap. They distinguish between clusters of strong ties (friends and family, people we see in our everyday life) and clusters of weak ties (indirect knowledge, someone you lost contact with). They make two fundamental assertions: i) social networks augment the number of weak ties that a person may have, and ii) new knowledge comes from weak ties. Viral content typically saturates a local cluster before jumping to another cluster through a weak tie.

Here again, the interpretation of Nahon and Hemsley's work is plagued by a slight ambiguity. Even if it has become common practice to use the phrase "social networks" to denote only "digital social networks", this could be made more explicit. However, if we make the not so bold assumption that this is also the case here, then Nahon and Hemsley's definition of virality contains an implicit reference to the digital nature of the medium. Even more noteworthy, they argue that digital social networks had the effects of augmenting the number of weak ties individuals may have. This clearly seems to be a new effect of the digital medium, and it is relevant for the understanding of virality. If social networks augment the number of weak ties, and if messages become viral by jumping from one local cluster of strong ties to another via weak ties, this would mean that the digital medium has created new opportunities for messages to turn viral. This points to the possibility that, even though they did not make it perfectly explicit, Nahon and Hemsley were actually trying to understand the specific impact of the digital medium on the phenomenon of virality.

As argued by Nahon and Hemsley, this definition has the advantage of offering a clear-cut distinction from other forms of peer-to-peer diffusion of a message, which are all familiar to digital media scholars, such as popular messages, memes, cascades, or word-of-mouth[13]. A popular message does not need a sigmoid curve, as some messages remain seen by a large number of individuals for a long time, sometimes plateauing at a given interval of values for years. The concept of "meme" is not meant to capture a single information event and the modalities of its spread. Researchers in memetics, to quote Nahon and Hemsley, "are typically more focused on understanding the transition from the original content to its derivatives, comparing memes, or looking for common elements" (2013, 43). Memes are thus a more abstract unit of culture than a viral message. As for information cascades, they are, to quote Nahon and Hemsley again, "a concept used to explain why people imitate other people's behaviours". Information cascades are supposed to be part of herd

13 One of the main explicit motivations of Nahon and Hemsley's definition is to enable a distinction with other forms of "information flows". They explicitly regret the assimilation of virality to a simple "digital word-of-mouth" by a subpart of the literature (2013, 43).

behaviour, when an individual imitates another individual's behaviour on their authority or assuming that they have a good reason to behave as they do, such as people going to a restaurant because it is popular, or joining a line supposing the individuals in line must be waiting for something interesting. Unlike virality, this phenomenon does not need to reach a great number of individuals or to unfold quickly. What is more, there is no need to assume that messages go viral only when people share a message because other people are sharing this message. This would be a rather simplistic hypothesis, ignoring the literature indicative of selection in sharing decisions and the great variety of reasons for sharing (see Nahon and Hemsley 2013, 45). Word-of-mouth does not presuppose the speed, reach-by-numbers and reach-by-networks which are key elements of the definition of virality. What is more, word-of-mouth presupposes that individuals engage in physical conversations, while "virality must employ the many-to-many, mass-personal communication we described above". This last passage, added to the mention that virality is a "network phenomenon" as opposed to word-of-mouth, may be the clearest indication that despite very slight ambiguities in their phrasing, Nahon and Hemsley fully understand virality as a digital phenomenon necessitating the use of digital networks.

2.2 The Importance of Communicational Costs and System Piggybacking

If some features of digital communication can easily be read in Nahon and Hemsley's definition with only a slight dose of intellectual charity, some other features of digital communication remain entirely implicit, even though they may actually contribute to Nahon and Hemsley's theoretical effort:

1. The first is that digital media such as e-mail accounts and digital networks are prodigious enablers of one-to-many communication. Circumstances such as public speeches and sending a copy of a letter to multiple recipients exist in the analogue world, but they are few and far between, demand considerable effort and are submitted to strong constraints of time, space, and costs. The presence on a digital network, on the contrary, makes the communication of a message to hundreds if not thousands of individuals, sometimes spread through many different countries, an effortless endeavour. This greatly facilitates the exponential diffusion of a message.
2. The relative speed of diffusion and fast decay that is characteristic of viral messages can be seen as a consequence of three more fundamental features. The first is the exponential diffusion facilitated by the digital medium that we just commented on. The other two are features of communications that

are not directly dictated by the digital medium, but are rather descriptive properties of their common use. The second property is the short time-interval of resharing. When a viral message is spread, it is typically in an interval going from a couple seconds after reception to a couple days (see Nahon and Hemsley 2013). This implies that if an exponential diffusion is to happen, it is to happen in a short interval of time, hence the relative speed of diffusion of viral messages. The third and final feature is the low frequency of re-sharing. Users of digital media typically share a viral information once, not several times (see Pailler and Schafer 2023). This implies that viral messages are bound to have a fast decay of their number of viewers-resharers. As the exponential function has the decisive property to quickly outrun the world population, without resharing by the same individuals the propagation of a viral message is bound to decay rapidly.

3. Finally, it can be said that digital social networks and other forms of digital media augment the reach of the message, understood both as the number of viewers and the geographical area in which it will be spread. The absence of any mention of the physical space may be revealing of the implicit reference to digital media in Nahon and Hemsley's work. Communication through large distances has become so easy to be considered unproblematic, and Nahon and Hemsley use purely informational metrics to characterize virality. This is not to say that long-range diffusion of a given message is a new phenomenon. As noted by Froissart (2002b), the diffusion of images through very long distances is an old phenomenon. However, the gentle reader may be too young to remember that there used to be a time where chatting online with an individual in the Philippines while living in Europe was considered amazing, and where a long-distance phone call was a very costly endeavour that had to be kept exceedingly short and infrequent. It is a testament to the transformation of communication by digital media that those operations now seem to be effortless and almost costless.

The reader may have noticed that consideration of communicational costs is essential to all the remarks just made. I use here a notion of cost that is not purely financial but denotes all relevant resources that may be spent in communication, such as time, energy, the cognitive effort of learning necessary skills and conceiving the message and its diffusion. When analysing the history of media, it is not sufficient to consider whether a given form of communication exists, but how costly it is. A radical drop in costs may similarly alter the frequency of that form of communication, the population that uses it, the type of content that can be spread through that mode of communication, the business models developing around it, and so many other social phenomena of interest. What may be the

most impactful in the digital medium is not necessarily the creation of completely new forms of communication, but a radical decrease in costs for some of those forms[14].

Another singularity of virality as opposed to word-of-mouth from the viewpoint of communicational costs was noted very early by Ralph F. Wilson. Virality supposes to piggyback the resources offered by the digital platform. A content spreading massively and rapidly through email accounts or on a digital social network consumes a significant amount of resources in terms of computing power, memory on the company's servers or bandwidth. A government, marketing agency or activist trying to launch a viral message is effectively piggybacking on another actor's resources, thus transferring the costs of message reproduction and diffusion that they should otherwise have assumed on their own. This modifies the economy of peer-to-peer message diffusion. It also implies that the institution running the digital forum must have an interest in being the place where at least some viral messages are spread, because otherwise it would see a significant share of its resources instrumentalised by other actors without reaping any reward. From a strategic perspective, virality on modern digital fora depend on a tacit confluence of interests between actors having an advantage in promoting some viral messages and digital fora having an interest in being the place where virality happens. This implicit two-player economic game was absent when an intelligence service or marketing agency was trying to launch and manipulate a rumour.

Before we come back to this topic in detail in section 3, it is essential to understand some major political phenomenon such as the difficulty of obtaining from private companies any moderation of nefarious viral messages, or even simply to get them to stop recommending such content. It is easy to manipulate a system encouraging virality even without detailed knowledge of moderation systems. A coordinated surge of sharing created by fake accounts and bots is a well-known and widely used strategy. If digital actors can manage to detect such attempt at viral manipulation, they may also show some obvious ill will for doing so, as virality is part of their own business model. For instance, Facebook's ill will towards moderate political and medical misinformation has become blatant and infamous (Frenkel and Kang 2021), and even Pre-Musk Twitter was criticised for its slow reaction to some misinformation and its maintenance of the easily hackable Trends (see Ohlheiser 2022). Obtaining moderation of viral misinformation from digital companies profiting from virality has thus turned into a major political problem.

14 This importance of costs is well seen by Roy Pinker (2020) in the conclusion of their book.

Understanding the exact problems raised by this incorporation of virality in business models is a great analytical challenge beyond the scope of this chapter. After all, spreading misinformation for profit is a problem as old as the for-profit mass media, and the term "yellow journalism" was created for some of the first newspaper business empires in the US in the late nineteenth century, especially under William Randolph Hearst, as some media outlets became infamous for their desire to spread dubious information and outright falsehoods in their cynical quest for the next buzz. It is thus slightly unnerving to hear journalists discuss "fake news" and "misinformation" as if this was a brand-new phenomenon created by "the internet" or "social networks". If such superficial analysis can trigger a healthy sceptical reaction before claims of radical novelty[15], it should not blind us to the possibility of novelty introduced by virality in the economics of for-profit media. As has been noted many times, the click-bait-industry is based on the monetisation of an attention span of a couple of seconds, a feat inaccessible to previous media industries (Venturini 2019). In that ultra short attention span, actions such as following a link or re-sharing a message, typical of viral propagation, are major targets of monetisation. As we will see below, the expansion of Machine Learning, especially predictive models of click-and-share, gives through customization greater exposure to information favouring virality.

2.3 Virality Metrics, or the Measure of the Phenomenon as Part of the Phenomenon

Before we move on to more advanced technological features such as the use of ML, other technical features typical of the digital medium should be mentioned as part and parcel of the virality phenomenon. One of them is the computation and display on user interfaces of virality metrics. With virality, it is not an exaggeration to say the measure of the phenomenon is part of the phenomenon itself, as virality metrics are essential to the platforms and researchers' perception and reaction to viral phenomena and have a self-confirmatory effect.

Those metrics are both used by the managers of digital fora to monitor activity and by the users and content creators, encouraging re-sharing and guiding the strategic creation of content. This again depends on digital technological affordances, such as the ability to quantify and measure activity at scale and in real time. Hashtags could also be analysed as a form of self-identification of the viral message as such, which could be a completely new practice. A new system of so-

15 An example of this healthy scepticism can be found in Pinker (2020).

cial norms creates new attempts at "hacking" those norms, namely to instrumentalise and manipulate the norms. New forms of hacking on digital media are thus a good indicator that those media are modifying the norms of communication. Some practices explicitly target the underlying mechanism quantifying and promoting virality, such as "trend piggybacking", when a user adds a popular hashtag to a video which may not have a direct relation to this hashtag in order to increase its chances of going viral. Other users go as far as using fake accounts to boost the number of re-shares of a given post (see Elmas et al. 2023).

It should be noted that such forms of "virality hacking" – and also research – are made harder by the secrecy surrounding some virality metrics. For instance, until very recently, Twitter did not publicly share its virality metrics even on the API dedicated to researchers (see Ling et al. 2022). This constrains researchers to using indirect metrics of virality such as number of likes or number of views (for the latter, only when videos do not start automatically, as it corrupts this metric as an indicator of virality). Virality is thus part and parcel of a wider phenomenon of digital media, namely the privatisation and opacification of norms governing public speech (see references to "shadowbanning" below and in Grison's chapter).

3 Updating Nahon and Hemsley's Definition: The Importance of Machine Learning in Virality and its Control

There is another factor in need of consideration in Nahon and Hemsley's work, namely, its age. The book containing the full exposition of their theory was published in 2013. For all their brilliance, they could not see into the future. I will argue that what has happened since then in communication technology is of great relevance to the analysis of viral phenomena and should be used to update their definition and deepen our reflection on the role of digital medium.

The most remarkable development for virality since the publication of *Going viral* is probably the advent of ML as a technology to control the flow of information on social networks and other platforms. For instance, Meta (formerly known as Facebook) started to use ML on a massive scale for content recommendation and ranking from 2016 on (see Hao 2021b), and many other players have followed suit, making recommendation and ranking one of the great industrial use cases of modern ML (see Portugal et al. 2018; Sharda and Josan 2021; Da'U and Salim

2020[16]). This cannot be without an effect on the phenomenon of virality, as it implies a centralised control of the information to which the user is exposed. It can thus be stated that ML created new, automated modalities of control over the phenomenon of virality.

Furthermore, recommendation and ranking are not the only use cases of ML that are relevant to the phenomenon of virality. Major players such as Meta-Facebook widely use predictive ML to estimate the probability of clicking and sharing, and those models play a major role in their strategy of engagement maximisation[17]. Finally, content moderation has also been hugely automated in the last couple of years with ML models[18]. While recommendation and ranking help promote content and thus increase its chance of becoming viral, content moderation may suppress a content altogether or reduce its visibility to other users in a variety of ways, diminishing its likelihood of going viral or effectively reducing it to zero. The means at the disposal of digital platform managers range from account, message, search result or recommendation suppression to flagging a message or account as suspicious, creating frictions of sharing such as asking you to read before you share[19], or surreptitiously reducing the number of contacts in a network who can see the message ("shadowbanning")[20].

16 For an example of the type of research conducted to maximize engagement with social media marketing, see Lee et al. (2018).

17 Meta publishes guidelines for content creator on how to maximise engagement for their content, for instance this 2022 publication: Meta, A creator's guide to growth: how to get your content seen on Facebook, September 20, 2022, https://www.facebook.com/creators/how-to-get-your-content-seen-on-facebook. For examples of Machine Learning research on this topic, see Zhao et al. (2018); Zou et al. (2019).

18 For a recent review of the Machine Learning literature on content detection and moderation, with an optimistic take on automation, see Gongane et al. (2022); Androcec (2020). For an introduction to algorithmic moderation with a discussion of its ethical and political challenges, see Gorwa et al. (2020).

19 As remarked by Frances Haugen in her Senatorial testimony (see Facebook whistleblower Frances Haugen testifies before Congress. CBS News, https://www.youtube.com/watch?v=juZEkeTjTRY, min. 1'08), it is known that a measure as simple as asking users to click on a link before they share can significantly slow the spread of misinformation.

20 This point is explicitly made in the very title of Gillepsie (2022). The exact definition of shadowbanning is a controversial point, and this controversy itself illustrates the complexity of moderation practices. For instance, does shadowbanning mean that your content is absolutely invisible to all other users, or that you are no longer accessible by the search function? Several platforms have unsurprisingly vehemently denied the very existence of shadowbanning. For these questions, see Savolainen (2022). The paper also makes the intriguing point that beyond the mere existence of such and such a moderation practice, the deep opacity of the rules that public discourse is submitted to on digital platforms, and the impossibility for users to even guess them in a consistent fashion are disturbing problems of platform governance.

It may be beneficial not only to stop conceiving recommendation and moderation as all-or-nothing businesses, but also to stop opposing recommendation and moderation in a dichotomy to conceive of them as two poles in a spectrum of content propagation control. If one theoretical pole is the complete and definitive suppression of a message before it can be seen by anyone, then the other would be the immediate, uncommented and literal sharing with the entire digital platform. Recommendation and moderation effectively move between those two poles to modulate the exposure of a given message. ML has thus provided managers of digital platforms with powerful and radically new affordances to control the phenomenon of virality using a variety of means, with a precision and power unknown to the various forms of control related to word-of-mouth.

Before we elaborate on this last point, it is important to not give in to a fantasy of absolute information control through ML. Recommendation and moderation are challenging tasks with a substantial error rate, and unexpected behaviours of systems sometimes thwart the efforts of digital platforms to control information propagation while respecting current legislation, their own Terms of Use and their business and brand strategies. For instance, during her testimony as a Facebook whistle-blower before a Senate committee[21] on October 2021, Frances Haugen insisted on the various difficulties met by the moderation systems of the tech giant, especially for non-English languages, and the tremendous complexity created by the interaction of many ML models on the platform[22]. Those mistakes hark back to the infamous opacity of those large ML systems, the behaviour of which remains to this day hard to understand and control in all its details.

However, those limitations of ML information control should not be used to deny the novelty of the phenomena unfolding in front of our eyes. First, the occasional lack of success of a form of control does not necessarily diminish its nov-

21 https://www.youtube.com/watch?v=GOnpVQnv5Cw.

22 See Hao (2021a). Haugen's public senate testimony is a document well-worth consulting on its own. It touches many topics, including Meta's negative impact on children's mental health, its amplification of divisive political content and misinformation, its lack of reaction against its instrumentalisation during the massacres in Myanmar, but for our topic, it is worthwhile to hear her explain that "anger-driven virality" is a conscious design decision, and need not be the price to pay of sharing some pictures with your friends. The overallocation of "integrity spendings" to English-speaking users is mentioned at the end of minute 57'-beginning of min. 58'. This is not only due to the availability of English language data, but also to the market importance of the language. One may also mention the lack of training on mixed language data, such as "Hinglish" or the mixture of English and Hindi commonly found on Indian social networks. The withdrawal of some virality-induced features during the American elections because of their known dangers to the public sphere, and their re-instalment for growth purposes after that, is mentioned at min. 1'01.

elty, and second, it is enough that those systems are able to impact information propagation at scale to constitute a new phenomenon of the greatest importance, even if this impact does not take the exact form wished by the system's designers. It should also be underlined that according to Haugen the use of ML is not always motivated by a belief in its impeccable efficiency and safety, but by a belief that is automation potential will help scalability and hence company growth.

When we consider the effects of ML systems on virality, it becomes apparent that we are talking about a digitally native phenomenon. There is no automated recommendation system or content moderation system in the analogue world[23]. This makes virality a digital phenomenon not only because it takes place on digital platforms, but because those digital platforms use digital means to control this phenomenon[24]. This constitutes an historically important feature, as it radically distinguishes virality from word-of-mouth. No government and no marketing agency ever had the power to control every peer-to-peer interaction in the street to favour or suppress a given message. The new situation on digital fora would be analogous to a central power being able not only to record what every individual is saying in real time, but also to mute an individual or to limit the number of persons who can hear her when she spreads a message disapproved of by the central power, while also being able, at least in principle, to augment the audience and exposure of all, and not just some, individuals spreading an approved message.

Such a gigantic power could only be enhanced by its surreptitious character. If our imaginary power could control peer-to-peer communication without the knowledge of individuals, this would certainly reduce the probability of protest and resistance. This again is the case on modern digital fora, as both recommendation and moderation remain largely invisible to individual actors. If more and more tech savvy actors are definitely aware of this phenomenon and try to circumvent it in their favour, thus creating a new phenomenon of interest in the process (see examples below), many individuals remain ignorant either of the phenomenon itself or of its exact modalities. I have already insisted elsewhere on the necessity of taking into account this surreptitious character of digital information control as one of its most distinctive features (see Pégny 2024). The ability to exert influence on information propagation at scale while maintaining relative

23 For more on this topic, see for instance Krafft and Donovan (2020).

24 This use of digital forms of control is not exclusive, as platforms still use an army of human moderators. The study of virality cannot be complete without an analysis of the guidelines given to those moderators and their concrete work conditions, as can be found in Roberts (2019). However, we do not aim at such a complete study here, only to explore what makes it typically digital.

secrecy is definitely an original phenomenon dependent on digital means. It would have been hard to imagine police officers intervening in every street conversation to control word-of-mouth messaging while pretending to maintain the operation in secret. Such a combination of detailed control at scale and relative discretion is dependent on the technological affordances offered by digital fora, and ML recommendation and moderation are part of those affordances.

The definition of content moderation as a technical task has in itself many normative effects. First, it is crucial to underline the ill-defined and intrinsically complex nature of content moderation. This cannot be a well-defined task, not only because "appropriate content" is a value-laden, culture-dependent, and controversial issue but because it concatenates under the same name different concepts with vastly different identification and demarcation issues for both the law and computer science: harassment, abusive language, threats, offensive language, slandering, misinformation, etc. All those different forms of undesirable content only share the negative property that they are undesirable on a given platform. What is more, the property of being undesirable can also be platform dependent. A dance video might be inappropriate on a professional network platform, but perfectly appropriate elsewhere (see Gongane et al. 2022). Finally, each of those issues is notoriously difficult in itself, and it is not to be excluded that their simultaneous implementation on the same platform could create a complex interaction between issues. What is to be done, for instance, if a social media post can be seen both as misinformation and a threat of violence, and those two problems are identified by two different subsystems subjugated to different rules? The pretence of obtaining a technical solution to "the problem of content moderation" is thus in itself a major source of conceptual confusion, as it hides the plurality of problems at hand.

Before concluding on this, I would like to add a couple more specific issues created by the use of ML techniques for automated content moderation[25]. The problems of this "unification through technologization" of several moderation challenges are compounded by the banalisation of naïve quantification of social phenomena. For instance, the review paper by Gongane et al. referenced above quotes quantities and percentages of "detrimental content" and "hate speech"

25 The reader may have noticed that I have not, and will not, try to apply the same analysis to recommendation and ranking, even though this would be absolutely necessary for a complete analysis of the effect of digitalisation on virality. This is only due to editorial limitations in space, and in intrinsic limitations in scope of this chapter. I am here just trying to convince the reader of the relevance of taking into consideration the properly digital nature of virality by examining the impact of the automated means of virality control: I am not trying to give a complete description of the effects of those means on virality, which would be so much more ambitious.

without any definition or methodology, as if the measures of such phenomena were not controversial and fraught with methodological issues. The authors go on to make very strange affirmations, such as the rarity of uncertain cases of hate speech, or a classification of satirical news as fake news, without any comment or methodological precaution. Such a naïve approach to delicate social issues creates a high risk of "phenomenon creation" as well as maintaining ignorance about the real phenomena of abuse. If there is nothing intrinsically wrong with applying ML techniques to moderation tasks, the risks of such naïve use of those techniques are very real.

Be that as it may, the training of ML systems for the identification of inappropriate content supposes the conception of a training dataset by a group of data labellers. Only a vast amount of examples labelled as appropriate or inappropriate will enable the system to pick up statistical features enabling successful identification. This entails that the group of data labellers must be given a definition of what inappropriate content is, and a set of instructions to decide on the multitude of cases they will have to face. Even if we set aside the considerable difficulties of definition that we just mentioned, the use of a large group of data labellers creates its own difficulties of consistency and quality control caused by the notoriously murky and controversial application of rules to cases of content moderation[26]. This new modality of identification induced by the choice of ML methodology both inherits classical issues of man-made moderation and adds its own peculiarities.

For instance, a new issue introduced into content moderation is due to a current particularity of NLP (Natural Language Processing) models, that is their inability to handle emojis, GIFs and other visual symbols. Those models are trained on a normalised language where those symbols are simply suppressed, which may cause issues when those symbols have a crucial modifier role in the whole message. This situation may evolve quickly as models become more and more multi-modal, but it is for the time being a feature worth knowing about among social scientists studying communication practices on digital media.

What is more, it is unclear in the current state of the art whether ML systems trained to identify "inappropriate content" are learning robust features of the underlying phenomena – for instance, what makes a threat out of a given statement – or if they are learning proxy signals commonly associated with (what a given group of people think according to a given set of definitions and instructions is) inappropriate content. This is a generic problem of current ML, as it

26 See Gongane et al. (2022) for some details on technical quality control measures. For more on the extreme complexity of human moderation, see the book by Roberts mentioned above. *The MIT Tech Review* has also published extensively on the topic in the last years.

lacks any intrinsic methodological warranty that the features identified are robust (for a short, entry-level introduction to this gigantic issue, see Wang 2023).

The automation of content moderation on vast platforms with an international user base also raises deep issues of cultural sensitivity and pluralism. The technical system may lead to a uniform application of norms throughout the platform, if only to simplify the issues at hand, which may be conceived of as a surreptitious imposition of cultural norms on certain countries. Users of Facebook have sometimes complained about the imposition of the American taboo against female breast nudity on the platform (see, for instance, Demopoulos 2023).

To recap, content moderation is deeply influenced by its position as a technical problem, and its particular solution through ML systems. Such an influence is naturally transmitted to virality. However, another relation between virality and content moderation can be found in Frances Haugen's testimony. She has insisted on the deep tensions between content moderation, its possible automation, and the will to promote viral content. Meta-Facebook promised that AI systems would be able to detect and eliminate toxic content that would be promoted by their engagement-based ranking systems (see min. 1′07′ in the reference above). As a matter of fact, they experimented with such a strategy, and internally reached the conclusion that the detection of toxic content was performing too poorly to amount to a satisfying solution. Haugen herself defends the position that making the platform less viral is the only way to diminish its toxicity. especially when she has defended the position that only a less viral network could be well moderated. This would mean that the commercial quest of virality and content moderation are in structural tension, and automated content moderation is just a *pis-aller* after the management of the platform sacrificed the genuine moderation of their content on the altar of virality (and profit). This is a very interesting idea in its own right, but its examination is beyond the scope of this work.

In simple terms, digital technological affordances enable an extreme centralisation of control on the peer-to-peer diffusion of messages at an unprecedented level of detail, and it constitutes without any shadow of a doubt a new historical situation that no political analysis of modern media can afford to ignore. If the central manipulation of seemingly horizontal, peer-to-peer communication is by no means a new phenomenon, digital media offer new means of manipulation that warrant an autonomous account, including understanding its failures and unforeseen effects. To wrap it up in a provocative formula: Virality is a new form of peer-to-peer propagation because it can be influenced in new (digital) ways.

Conclusion

We have pleaded for a heuristic understanding of virality as a new form of horizontal communication well-distinguished from word-of-mouth. Methodological precautions should wrap this assertion in a fair amount of humility. Virality is a descriptive concept which does not pretend to model the propagation of ideas or beliefs. It can only constitute part of a story of propagation going through multiple channels and should not be given an exaggerated importance because it is easier to study than oral communication. Finally, its dependence on the digital medium does not imply that the whole story of social practices surrounding virality should be reduced to a technological determination.

However, we have seen that digital media present significant technological affordances for horizontal communication which should not be downplayed. Digital media have caused a radical drop in the cost of some forms of communication, especially the one-to-many, literal, fast and long-distance propagation of a message. Virality metrics are also a feature typical of the digital world, absent from analogue word-of-mouth, which create a self-confirmatory feedback loop. The measure of the phenomenon, and its manipulation, is part of the phenomenon.

Through ML recommendation and moderation, new opportunities of centralised control of peer-to-peer communication, even if their exact effects may be hard to master, have arisen, which is a new phenomenon from a technological and media perspective. What is more, virality has created a new economy of piggybacking where content creators and propagators aiming at virality harness the communicational resources offered, oftentimes for free, by another actor, who must consequently develop a business interest in being the place where virality happens. Furthermore, those actors must develop an understanding of the automated means of influence on information propagation to circumvent or instrumentalise them to their advantage.

ML-based content recommendation and moderation blurs the distinction between horizontal and vertical communication in novel ways. Older forms of institutional manipulation, such as rumour propagation or astroturfing, arguably aimed at blurring such a distinction by creating or harnessing a seemingly horizontal phenomenon submitted to centralised design and manipulation. However, recommendation and moderation enable a new form of confusion of genres, as every step of content propagation can be controlled by recommendation and moderation algorithms, a level of granularity never achievable before. This is arguably one of the reasons why the responsibility of a platform for content is so hard to define legally, as modern platforms enable deep propagation control without actual content creation.

All of this pleads for an understanding of virality not as a simple "digital word-of-mouth", but as a new form of horizontal communication, starting in the late 1990s and taking an important development and turn in the mid-2010s with the advent of ML-fuelled content propagation control. This should be read as a heuristic statement: A lot more work needs to be done to better understand the interplay of media and communicational practices in the three last decades and compare it with similar studies of other periods and media.

References

Alloing, Camille, and Nicolas Vanderbiest. "La fabrique des rumeurs numériques. Comment la fausse information circule sur Twitter ?" *Le Temps des médias* 1 (2018: 105–123. https://hal.science/hal-01712206/document.

Androcec, Darko. "Machine learning methods for toxic comment classification: a systematic review." *Acta Universitatis Sapientiae*, *Informatica* 12, no. 2 (2020): 205–216.

Bloch, Marc. *Réflexions d'un historien sur les fausses nouvelles de la guerre*. Paris: Éditions Allia, 1999.

Boullier, Dominique. *Propagations. Un nouveau paradigme pour les sciences sociales.* Malakoff: Armand Colin, 2023.

Da'U, Aminu, and Naomie Salim. "Recommendation system based on deep learning methods: a systematic review and new directions." *Artificial Intelligence Review* 53, no. 4 (2020): 2709–2748.

Demopoulos, Alaina. "Free the Nipple: Facebook and Instagram told to overhaul ban on bare breasts." *The Guardian*. 18 January 2023.

Denisova, Anastasia. "How to define 'viral' for media studies?." *Westminster Papers in Communication and Culture*, vol. 15, no 1 (2020).

Elmas, Tuğrulcan, Selim, Styephane, and Célia Houssiaux. "Measuring and Detecting Virality on Social Media: The Case of Twitter's Viral Tweets Topic." In *Companion Proceedings of the ACM Web Conference 2023*, 314–317. New York: ACM, 2023.

Froissart, Pascal. *La rumeur. Histoire et fantasmes*. Paris: Belin, 2002.

Froissart, Pascal. "Les images rumorales. Une nouvelle imagerie populaire sur Internet." *Médiamorphoses* 5, no. 1 (2002b): 27–35.

Froissart, Pascal. "L'invention de la lutte contre les rumeurs." *Le Temps des médias*, no. 1 (2022): 223–240.

Frenkel, Sheera, and Célia Kang. *An ugly truth: Inside Facebook's battle for domination*. London: Hachette UK, 2021.

Gillespie, Tarleton. "Do not recommend? Reduction as a form of content moderation." *Social Media+ Society* 8, no. 3 (2022). https://doi.org/10.1177/2056305122111755.

Gongane, Vaishali U., Mousami V. Munot, and Alwin D. Anuse. "Detection and moderation of detrimental content on social media platforms: current status and future directions." *Social Network Analysis and Mining* 12, no. 129 (2022). https://doi.org/10.1007/s13278-022-00951-3.

Goody, Jack, and Ian Watt. "The consequences of literacy." *Comparative studies in society and history* 5, no. 3 (1963): 304–345.

Goody, Jack. *The domestication of the savage mind*. Cambridge, MA: Cambridge University Press, 1977.

Goody, Jack. *The logic of writing and the organization of society*. Cambridge, MA: Cambridge University Press, 1986.

Gorwa, Robert, Binns, Reuben, and Christian Katzenbach. “Algorithmic content moderation: Technical and political challenges in the automation of platform governance.” *Big Data & Society*, vol. 7, no 1 (2020). https://journals.sagepub.com/doi/full/10.1177/2053951719897945.

Hao, Karen, “The Facebook whistleblower says its algorithms are dangerous. Here’s why.” *MIT Tech. Review.* 5 October 2021.

Hao, Karen, “How Facebook Got Addicted to Spreading Misinformation.” *MIT Tech Review.* 11 March 2021.

Jurvetson, Steve, and Tim Draper. “Viral marketing.” *Netscape M-Files* 1, no. 1 (1997): 1–20.

Krafft, Peaks M., and Joan Donovan. “Disinformation by design: The use of evidence collages and platform filtering in a media manipulation campaign.” *Political Communication* 37, no. 2 (2020): 194–214.

Lee, Dokyun, Kartik Hosanagar, and Harikesh S. Nair. “Advertising content and consumer engagement on social media: Evidence from Facebook.” *Management Science* 64, no. 11 (2018): 5105–5131.

Ling, Chen, Blackburn, Jérémy, and Emiliano De Cristofaro, *et al.* “Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos.” In *Proceedings of the 14th ACM Web Science Conference 2022*, 164–173. New York: ACM, 2022.

Nahon, Karine, and Jeff Hemsley. *Going viral*. Cambridge, UK: Polity Press, 2013.

Ohlheiser, A.W., “Why Twitter Still Has Those Terrible Trends,” *MIT Tech Review*. 28 July 2022.

Pailler, Fred, and Valérie Schafer. “Keep Calm and Stay Focused: Historicising and Intertwining Scales and Temporalities of Online Virality.” In *Zoomland. Exploring Scale in Digital History and Humanities*, edited by Florentina Armaselu and Andreas Fickers, 119–150. Berlin: De Gruyter, 2023. https://www.degruyter.com/document/doi/10.1515/9783111317779-006/html?lang=en.

Pégny, Maël. *Introduction à l’éthique des algorithmes (et de l’IA*). Paris: J. Vrin, 2024 (forthcoming).

Pinker, Roy. *Fake news et viralité avant Internet. Les lapins du Père-Lachaise et autres légendes médiatiques*. Paris: CNRS Éditions, 2020.

Portugal, Ivens, Alencar, Paulo, and Donald Cowan. “The use of machine learning algorithms in recommender systems: A systematic review.” *Expert Systems with Applications* 97 (2018): 205–227.

Roberts, Sarah T. *Behind the Screen. Content Moderation in the Shadows of Social Media.* New Haven and London: Yale University Press, 2019.

Savolainen, Laura. “The shadow banning controversy: perceived governance and algorithmic folklore.” *Media, Culture & Society* 44, no. 6 (2022): 1091–1109.

Sharda, Shreya, and Gurpreet S. Josan. “Machine Learning Based Recommendation System: A Review.” *International Journal of Next-Generation Computing* 12, no 2 (2021). https://doi.org/10.47164/ijngc.v12i2.20.

Slauter, Will, “Le paragraphe mobile. Circulation et transformation des informations dans le monde atlantique du XVIIIe siècle.” *Annales* 67, no. 2 (2012): 363–389.

Venturini, Tommaso. “From fake to junk news: The data politics of online virality.” In *Data politics*, edited by Didier Bigo, Engin Isin and Evelyn Ruppert, 123–144. London: Routledge, 2019.

Wang, Yifei. “Robustness and Reliability of Machine Learning Systems: A Comprehensive Review.” *Eng OA* 1, no. 2 (2023): 90–95.

Wilson, Ralph F. “The six simple principles of viral marketing.” *Web marketing today* 70, no. 1 (2000). https://www.practicalecommerce.com/viral-principles

Zhao, Qian, Maxwell F. Harper, Gediminas Adomavicius, et al. “Explicit or implicit feedback? Engagement or satisfaction? A field experiment on machine-learning-based recommender systems.” In *Proceedings of the 3rd Annual ACM Symposium on Applied Computing*, 1331–1340. New York: ACM, 2018.

Zou, Lixin, Long Xia, Zhuoye Ding, et al. “Reinforcement learning to optimize long-term user engagement in recommender systems.” In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2810–2818. New York: ACM, 2019.