_ _

Junk news bubbles modelling the rise and fall of attention in online arenas

_ _

Castaldo, Maria; Venturini, Tommaso; Frasca, Paolo; Gargiulo, Floriana

Article

Junk news bubbles modelling the rise and fall of attention in online arenas

Maria Castaldo
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France

Tommaso Venturini [ID]
Centre for Internet and Society, CNRS, France

Paolo Frasca
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France

Floriana Gargiulo
GEMASS, CNRS, France

Abstract
In this article, we present a type of media disorder which we call 'junk news bubbles' and which derives from the effort invested by online platforms and their users to identify and circulate contents with rising popularity. Such emphasis on trending matters, we claim, can have two detrimental effects on public debates: first, it shortens the amount of time available to discuss each matter and second, it increases the ephemeral concentration of media attention. We provide a formal description of the dynamic of junk news bubbles, through a mathematical exploration of the famous 'public arenas model' developed by Hilgartner and Bosk in 1988. Our objective is to describe the dynamics of the junk news bubbles as precisely as possible to facilitate its further investigation with empirical data.

Keywords
Attention, attention cycles, attention economy, fake news, modelling, online platforms

Corresponding author:
Tommaso Venturini, Centre for Internet and Society, CNRS, rue Pouchet, Paris 75794, France.
Email: tomm.venturini@gmail.com

Conceptualizing junk news bubbles

Much has been written in the last years about online media and the threat of 'selective exposure' (Sears and Freedman, 1967), 'echo chambers' (Garrett, 2009) and 'filter bubbles' (Pariser, 2011). In various ways, all these notions point to the risk that the growing availability of information and the perfecting of filtering and recommendation technologies may create a 'situation in which thousands or perhaps millions or even tens of millions of people are mainly listening to louder echoes of their own voices' (Sunstein, 2001: 16). Reviving a long tradition of homophily and segregation models (Schelling, 1971), this idea has sparked much interest in computational sociology (cf., among others, Barberá et al., 2015; Colleoni et al., 2014; Geschke et al., 2019; Quattrociocchi et al., 2016) despite evidence that its impacts may be overestimated (cf. Bakshy et al., 2015; Boxell et al., 2017; Dubois and Blank, 2018; Flaxman et al., 2016).

Less attention and computational efforts have been dedicated to a different type of media bubbles, whose danger comes not from the fragmentation, but from the ephemeral concentration of public attention. In a previous article (Venturini, 2019), we proposed the term 'junk news' to expand the notion of 'fake news' beyond its excessive focus on deceitfulness (cf., among others, Benkler et al., 2018; Gray et al., 2020; Zuckerman, 2017). Besides false contents disguised as mainstream news and explicitly directed at deceiving their receivers, media scholars should worry about the avalanche of memes, click-baits, trolling provocations and other forms of ephemeral distractions that prevent online audiences from engaging in a thoughtful public debate. This type of 'information disorders' (Wardle and Derakhshan, 2017) cannot be defined on the basis of its content or its style (as in manual and automated fact-checking, cf. Graves, 2018 and Ciampaglia et al., 2015) nor on the basis of the way in which it polarizes the debate (as in echo chambers and filter bubbles). Instead, ephemeral distractions are distinguished by a feature often neglected when considering online misinformation – their temporal profile. Like junk food, junk contents tantalize without ever satiating. Like echo chambers, they can be described as bubbles, although not the sense of a secluded informational space, but rather as market bubbles: speculative waves that destroy rather than create richness in public debate.

If the dangers of junk news bubbles are neglected by current media research, it is not because they are lesser than that of fake news and filter bubbles, but because this particular information disorder still lacks the precise conceptualization that supports the research on fact-checking and polarization. To outline such a conceptualization, we propose to define junk news bubbles as an *adverse media dynamic in which a large share of public attention is captured by items that are incapable of sustaining it for a long time*. Both elements of this definition are crucial. Popular stories and even viral contents are not necessarily junk news, no matter how quickly and largely they spread in online networks (Jenkins et al., 2013). To qualify as junk news bubbles, contents must fade away as quickly as they rose, so that they distract public debate rather than nourishing it. Note that our definition is agnostic about the quality of junk contents. Even a patent piece of misinformation such as the infamous claim that 'Brexit will make available 350 million pounds per week for the NHS' plastered onto a red bus during the UK-EU referendum campaign, can end up generating productive discussions if it sticks long enough in the

public debate (Marres, 2018). Vice versa, newsworthy stories cannot contribute to democratic conversation if they are too quickly pushed out of the public agenda. In other words, the notion of junk news bubbles applies less to specific pieces of content, than to a general acceleration of online attention cycles.

Central in the '70s and '80s, the question of 'attention cycles' (Downs, 1972) has lost steam in current media research because of the advent of digital technologies and the extension of the media system that they brought with them. Because of this extension, the question of the occupation of public debate has begun to be formulated in spatial rather than in temporal terms (i.e. where something is discussed rather than when). Temporal dynamics, however, remains crucial for, as in the words of McLuhan, 'the "message" of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs . . . amplif[ying] or accelerate[ing] existing processes' (McLuhan, 1964: 8). As noted by scholars working on the attention economy (Crogan and Kinsley, 2012; Lanham, 20066; Terranova, 2012), digital technologies are particularly inclined to amplify 'media hypes' (Vasterman, 2005) and to concentrate public attention on widespread but ephemeral trends.

Our argument echoes McCombs' (2005) plea not to desert the research on 'agenda setting' (McCombs, 2004; McCombs and Shaw, 1972) on the grounds of online attention being scattered in a myriad of different channels each with its own agenda. Precisely because of this overabundance of sources, digital media have since their inception channelled their flows through increasingly sophisticated 'Engines of Order' (Rieder, 2020) constantly ranking – and updating the rank – of their contents (Cardon et al., 2018; Cardon, 2015). This is particularly true of social media platforms, whose size could not be sustained without infrastructures to sieve through millions of contents produced every day and whose business model as marketplaces of ideas entails a relentless circulation of new trending topics (Webster, 2014). This unremitting production of trends is an inherent feature of recommendation algorithms, as candidly admitted by YouTube engineers: 'in addition to the first-order effect of simply recommending new videos that users want to watch, there is a critical secondary phenomenon of boot-strapping and propagating viral content' (Covington et al., 2016: 193, see also Zhao et al, 2019). Algorithms, however, are not the only component of online attention infrastructures and not the only actors at play. A similar push for trendiness comes from the interaction between the practices of social media users, which are increasingly driven by micro-celebrity strategies (Khamis et al., 2017; Marwick and Boyd, 2011), and the graphic interface of platforms and apps, which are partial to ephemeral visibility and vanity metrics (Rogers, 2018).

The term 'junk news bubbles', thus, designates less a specific type of content (as in the case of misinformation or 'fake news') than a general tendency of media systems to privilege trendy items and thus create an extremely turbulent and superficial public debate. False and low-quality stories are the type of content that thrives most directly on the incentives of trendiness, but to some extent, all digital conversations are influenced by this acceleration. In the article where we first proposed the term 'junk news' (Venturini, 2019), we discuss in more detail the political economy of online acceleration and the way in which it influences the circulation of news. In this article, our goal is to provide a *formal* description of these attention dynamics to encourage their further empirical study. With a few remarkable exceptions (see in particular Leskovec et al., 2009 and

Lorenz-Spreen et al., 2019), no large-scale research has been devoted to attention cycles, despite the growing availability of traces produced by digital media (Latour et al., 2012; Lazer and et al, 2009; Venturini et al., 2015).

To facilitate such research, we propose a mathematical formalization of one of the most influential accounts of attention dynamics: the 'public arenas model' introduced in 1988 by Stephen Hilgartner and Charles Bosk. Despite its clarity and insightfulness, H&B's framework has so far found no mathematical formalization for its complexity and lack of formal description. In this article, we streamline H&B's model focussing on the rise and fall of attention matters (and ignoring the linkages across different arenas and the actors within each arena). Doing so, we propose a ready-to-test (*prêt-à-tester*) version of H&B's model hoping that it will encourage further empirical investigation on junk news bubbles. Ours is a case of 'toy model' (Reutlinger et al., 2018), whose function is not to be applied or fitted to empirical data nor to offer an accurate description of the phenomenon that it presents, but to help in defining it and setting the conceptual bases for its future study.

Model description

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

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

arena). Examples could be different videos in a given YouTube channel or different threads in a given Reddit subreddit. To be sure, we are not promising that our model will fit all media debate, but inviting scholars to test it empirically on different phenomena to determine to which it can be fruitfully applied.

(b) The second ingredient of our model is two **competition mechanisms** that favour some attention matters over others. The four different 'principles of selection' distinguished by H&B find in our model a formalization in two main mechanisms:

- *Exogenous influences*. Three of the four 'principles of selection' distinguished by H&B, 'drama' (pp. 61–62), 'culture and politics' (H&B, p. 64) and 'organizational characteristics' (pp. 65, 66) are rendered in a deliberately coarse way in our model. The dramatic value of attention matters as well as the way in which they resonate with the general culture or with the specific organization of the medium are important, but their influence falls outside the self-induced media dynamics that constitute the focus of our model. In our formalization, the influence of these features is thus rendered as a noise which randomly increases or decreases the visibility of each item at each iteration. This solution allows for accounting for this type of influence (and to explore the effect of its variation) under the assumption that its specific nature does not affect the dynamic of junk news bubbles.
- *Endogenous trending*. The last selection principle identified by H&B, 'novelty and saturation', is crucial to our model. At each iteration, the model increases or decreases the visibility of each matter, repeating its previous variation, *multiplied by a parameter that accelerates or decelerates such variation*. The model therefore rewards rising items and penalizes declining ones. This mechanism works as a *Matthew effect* (Merton, 1968; Newman, 2001), but a *dynamic one* which rewards not the most visible matters, but the ones that have increased the most since the previous iteration. This boosting of trendiness is consistent with the way in which online platforms 'emphasiz[e] novelty and timeliness . . . [by] identifying unprecedented surges of activity' and 'reward[ing] popularity with visibility' (Gillespie, 2016: 55, 60). Such partiality for trendiness is characteristic of both social media and their users, in a sociotechnical loop in which the visibility granted by platform algorithms both *depends on* and *is influenced by* the number of views generated by different contents.

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

- *Lower boundary: exclusion of negative visibility*. Because it is impossible to conceptualize such a thing as negative attention, when noise or acceleration push the visibility of a matter of attention below zero, the item is removed from the arena and replaced with a new one with null initial visibility. Because a new attention

matter can enter the arena only when an old one leaves it, the number of items in the model remains fixed (but some items can have visibility equal to zero).

- *Upper boundary: saturation of the attention capacity*. After having applied noise and acceleration and corrected for negative attention, the model divides the potential visibility of each item by the sum of the potential visibilities of all items. This normalization makes sure that the sum of all computed visibilities remains equal to one. This boundary implements a key ingredient of H&B framework, the idea that each debate arena has a **fixed attention capacity** (or 'carrying capacity', in H&B terms). The fixity of the global 'carrying capacity' is crucial to ensure that our model does not converge to a trivial winner-takes-all equilibrium. While raising attention matters are pushed to an increasing visibility by their trendiness, they all end up reaching a point where they exhausted their potential for growth, begin to slow down and are penalized by competition mechanisms.

The inelasticity of attention capacity also ensures that the visibility gained by one matter of attention is always lost by some other so that 'the ascendance of one social problem will . . . be accompanied by the decline of one or more others' (H&B, p. 61). While we are, of course, aware that public attention fluctuates with circadian and professional rhythms, we believe that these cyclical fluctuations can be discounted for the sake of simplicity. Following H&B, we think that good reasons for a fixed attention capacity can be found in the limited staging capacity of media ('the prime space and prime time for presenting problems publicly are quite limited', p. 59) and, more importantly, in the limited capacity of the public to attend to public ('members of the public are limited not only by the amount of time and money they can devote to social issues, but also by the amount of "surplus compassion" they can muster for causes beyond the usual immediate concerns', p. 59). This second element is crucial to understand why the assumption of a limited carrying capacity remains relevant for online media even if digital technologies removed most of the barriers of conventional news gatekeeping (Shoemaker and Vos, 2009). As noted by Ray Maratea (2008) in relation to the visibility in the blogosphere:

> "*Although the Internet may provide an infinite carrying capacity and make the claims-making process more efficient, it cannot resolve the problem that audiences have limited amounts of time and attention to focus on various social problems. Furthermore, the blogosphere has developed according to a hierarchical structure, meaning readers and traditional journalists largely concentrate their attention on a relatively small number of well-known blogs. In short, the development of the blogosphere has the potential to dynamically change the claims-making process, but Hilgartner and Bosk's public arenas model remains vital to understanding the rise and fall of social problems in the new media age.*" (p. 140)

In fact, by eroding the boundaries between news and entertainment (Prior, 2005), digital convergence may have created an even harsher 'all-out war for the time of an audience that has more choices than at any point in history' (Klein, 2020: 279). While others (see, e.g. Cinelli et al., 2019) takes this competition as a reason for selective exposure and filter bubbles, we believe that they are a crucial ingredient of ephemeral concentration and fake news bubbles. According to the limited capacity model of mediated message processing proposed by Annie Lang (2000), media users can deal with an overabundance

of media stimuli by switching to an entertainment – rather than understanding – mode, saving their limited cognitive resources by interacting more superficially with the medium ('running on automatic', Lang, 2000: 53).

Model formulation and parameters

(a) We call x_i each item of our **population of matters of attention**, with $i = 1, ..., n$, where n is the maximum number of items in the population. We call 'visibility' or π_i^t the share of attention captured by x_i at time t. By a mechanism explained below, at each timestep, the sum of π_i^t for all i is fixed and equal to one. This allows to interpret each π_i^t as the percentage of the total attention captured by each item i at time t

(b) We model the two **competition mechanisms** as follows:
- *Endogenous trending.* At every timestep $t+1$, the visibility π_i^{t+1} of each item i is modified by adding to its current visibility π_i^t a term which repeats its previous variation (i.e. $\Delta\pi = \pi_i^t - \pi_i^{t-1}$) multiplied by a positive factor α, which could be interpreted as a boost of trendiness.
- *Exogenous influences.* In our formalization, we render all external influences on media dynamics as a noise ε_i^t which increases or decreases the visibility of item i randomly at timestep t. The noise ε_i^t is a realization of a normal distribution
- $N(0, \dfrac{1}{c * n^2})$ with mean = 0 and standard deviation = $\dfrac{1}{\sqrt{cn}}$

where c is a positive parameter. We can, therefore, write the potential visibility of each item after the iteration p_i^{t+1} as the output of the two above mechanisms as follows:

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

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

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

$$\tilde{p}_i^{t+1} = max(0, p_i^{t+1}) \tag{2}$$

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

$$\pi_i^{t+1} = \frac{\tilde{p}_i^{t+1}}{\sum_j \tilde{p}_j^{t+1}} \tag{3}$$

Initialization. At the first step of the model, the visibility of every i (i.e. π_i^1) is initialized with a random numbers drawn from a uniform distribution between 0 and 1 and

normalized to satisfy the constraint $\sum_i \pi^1{}_i = 1$. At the second step, the visibility every i

(i.e. π_i^2) is obtained by adding to π_i^1 a noise $\varepsilon^t{}_i$ drawn from the normal distribution

$N(0, \dfrac{1}{c * n^2})$ and normalizing. After the first two steps, the dynamics is self-sustained by

evaluating equations (1), (2) and (3) at each iteration.

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

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

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

Model results and discussion

Despite its simplicity, our model is able to generate patterns comparable with the empirical observations of media systems (Leskovec et al., 2009; Lorenz-Spreen, 2019). In particular, our formalization supports the H&B intuition that the 'shifting waves of social problems' (H&B, p. 67) typical of media attention cycle can be explained by the interaction between the push of trendiness and the saturation of the carrying capacity:

> if we explore these complex linkages, we find a huge number of positive feedback loops, 'engines', that drive the growth of particular problems. Growth is constrained, however, by the negative feedback produced by the finite carrying capacities of the public arenas, by competition among problems for attention, and by the need for continuous novel drama to sustain growth..

Previous studies (Cattuto et al., 2007; Gonçalves et al., 2011; Weng et al., 2012) considered the role of users' limited attention in media competition assigning users a maximal number of possible interactions (a sort of Dunbar number for individual attention). In most of these models, the fall of popularity is obtained forcing an aging process of media items through an explicit time decay term. This aging process, however, is difficult to justify
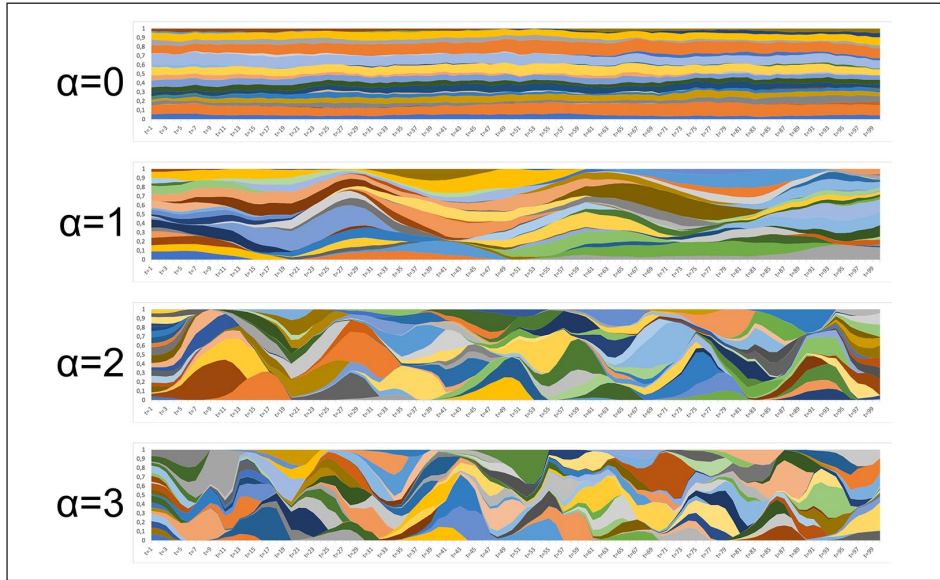
Figure 1. Evolution of our model for trendiness boost $= 0$, 1, 2 and 3 (with $n = 20$ and $c = 12$). Each colour area corresponds to the attention received by an item. Only the first 100 iterations are shown as the shape of the curves does not change in further iterations.

theoretically and empirically. One of the most original aspects of our model is that it dispenses with this aging process: items' popularity decays naturally through the interplay between the trendiness and the saturation of the overall attention capacity.

The comparison between the graphs in Figure 1 suggests that, as the boost of trendiness grows, the rise and fall of attention matters becomes steeper. This relation can be tested by computing the mean steepness of attention curves (the absolute increase or decrease by unit of time) and observing that it increases monotonically with the increase of alpha before reaching a plateau (probably due to the upper and lower constraints on the state and to the impossibility of compressing the width of curve beyond a certain point).

Figure 2 confirms that the relation between the steepness of attention curve and trendiness boost is not substantially transformed by the other parameters of our model. The number of attention matters and the importance of exogenous influences shift the position of the curve, but do not change its shape. Also, because both n and c affect the curve in the same way, only n will be explored in the next figures.

Considering together Figures 1 and 2, it is also interesting to notice that trendiness boost increases rise-and-fall steepness by affecting both dimensions of the media cycle: the *height* of attention curves and their *width*. This suggests that junk news bubbles can combine features that may appear contradictory.

- Regardless of the number of items or the level of noise, the stronger is trendiness boost the shorter is the lifecycle of individual attention matters (Figure 3a).
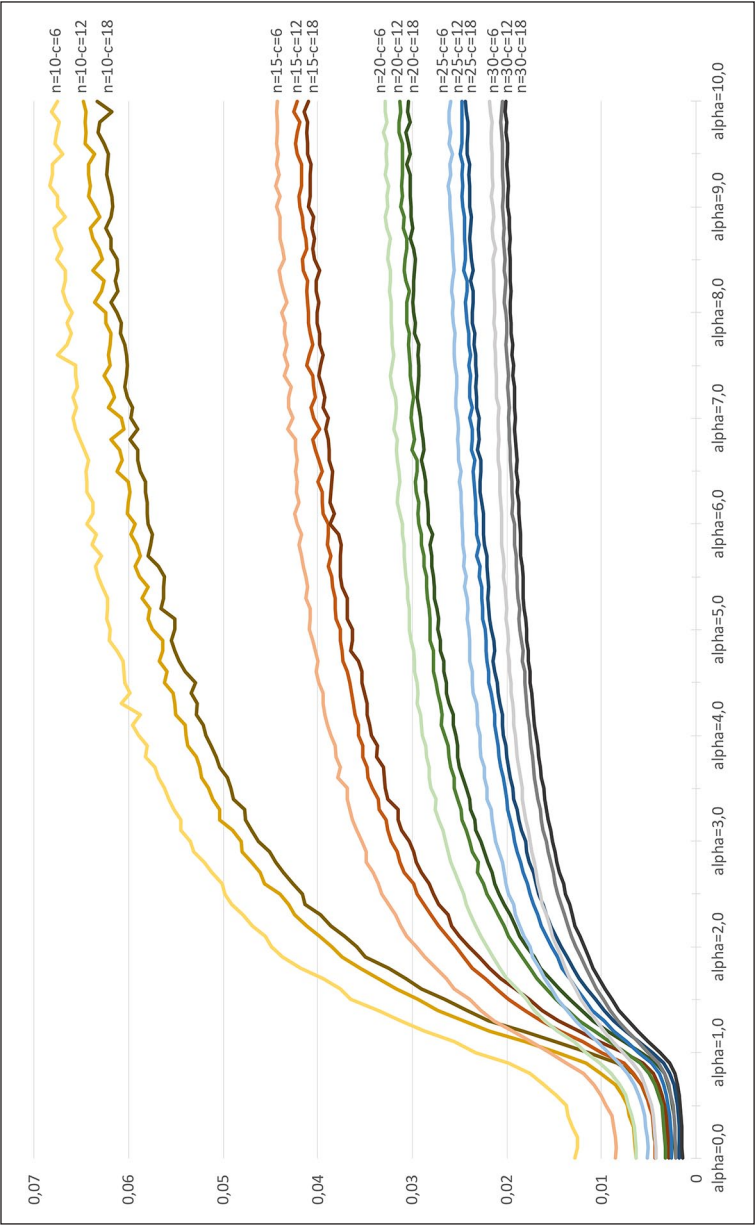
Figure 2. Mean slope of attention curves as function of the trendiness boost (for different values of *n* and *c*).
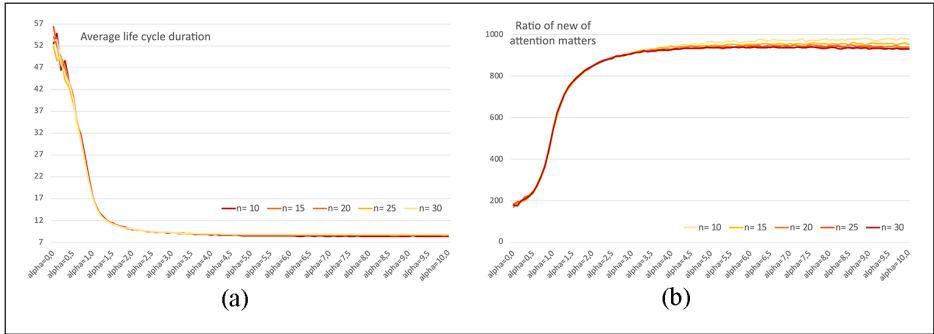
Figure 3. (left) Mean length of attention matters' life cycle and (right) ratio of new attention matters entering the model in its first 10,000 iterations, at the variation of trendiness boost (for different values of *n* and with *c* set to 12).
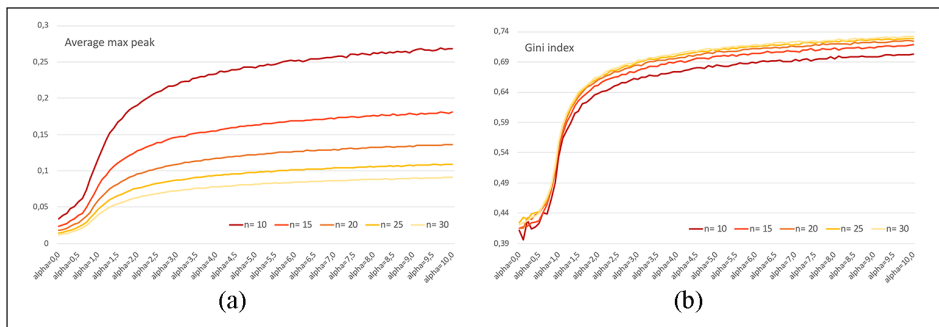

Figure 4. (left) Mean height of the attention curve peaks and (right) Gini index of attention concentration at each iteration of the model, at the variation of trendiness boost (for different values of *n* and with *c* set to 12).

Remarkably, this is true for all attention matters: even items that reach very high levels of visibility end up falling as quickly as they rose. As a consequence of the shortening of attention weaves, a higher number of matters enter and exit the arena (Figure 3b). This may contribute to making platforms more attractive by increasing the dynamism of their offer of information and entertainment.

- On the other hand, higher trendiness boost increases the maximum visibility reached by attention matters (Figure 4a) and, most importantly, amplifies the difference between successful and unsuccessful attention matters, creating a situation in which, at each iteration, most of the available attention is captured by a minority of over-visible items (Figure 4b). There is a huge 'population' of potential problems-putative situations and conditions that could be conceived of as problems. This population, however, is highly stratified. An extremely small fraction grows into social problems with 'celebrity' status . . . [while] the vast

majority of these putative conditions remain outside or on the extreme edge of public consciousness' (H&B, p. 57).

Some empirical examples

Being a *toy model*, our mathematical formalization is not meant to serve as a simulation of a real-world dynamic or to be fit to empirical data. Still, simplistic as it is, our toy model does capture some features of the current media system as confirmed, for example, by the results of an empirical investigation of *Accelerating Dynamics of Collective Attention* by Lorenz-Spreen et al. (2019). Comparing different media and over several years, the authors found that 'the attention associated with individual topics rises and falls with increasing gradients' and 'the shifts of collective attention between topics occur more frequently' (p. 1763) suggesting a general increase of trendiness's importance in the media system. They also find that this tendency is stronger for media, such as Twitter, Google and Reddit and less pronounced for Wikipedia and the scientific literature, suggesting that the arenas that are more sensitive to trendiness are also the ones that are most exposed to junk news bubbles.

To find other examples of our model's insights, we explored the dynamics of visibility of the French vlogosphere. While the most obvious (and probably the best) operationalization of H&B model would be to consider the whole French YouTube as an arena for the competition of different thematic issues, this would require a huge data collection and automatic identification of themes which exceed the scope of this article. We thus decided to exploit the *fractal* nature of media systems and to consider each YouTube channel as an arena and each video as an attention matter. As the competition between video is likely to take place across channels rather than within them, this operationalization is far from ideal but, as we will see, sufficient to exemplify our argument.

Starting on 9 December 2019, we recorded the number of views collected every hour by each video published by about 1000 YouTube channels active in the French public debate (selection was made by iterating expert review and snowball sampling). For this article, we limit our analysis to the data collected until 14 March 2020 (to exclude the coronavirus disease 2019 lockdown period where we observed slightly different attention dynamics, Castaldo et al., 2020). Since we are exploring in-channel competition, we also restrict our analysis to channels having multiple active videos at the same time, thus focussing on the 60 most active channels on our corpus. In this subcorpus, the most active channel is 'Europe 1' with 4,291 videos published in the 3 months of collection and the least active is 'Charente Libre' with 216 videos. The most popular channel is 'France 24' with 731,950 total views and the least popular 'L'Opinion' with 1552 total views. For each of these 60 channels, we calculated the same two metrics computed for in our simulation above: the average life cycle of the videos (defined as the average number of hours necessary for a video in the channel to reach the 95% of the total views collected in the first week) and the average Gini index of hourly concentration.

The scatter plot in Figure 5 confirms the main insight of our model: the fact that short life cycles and high concentration go hand in hand in the distribution of media attention. The distribution on the diagonal of the scatter plot suggests a sort of continuum that goes from channels that publish few videos with a relatively longer life expectancy and a less
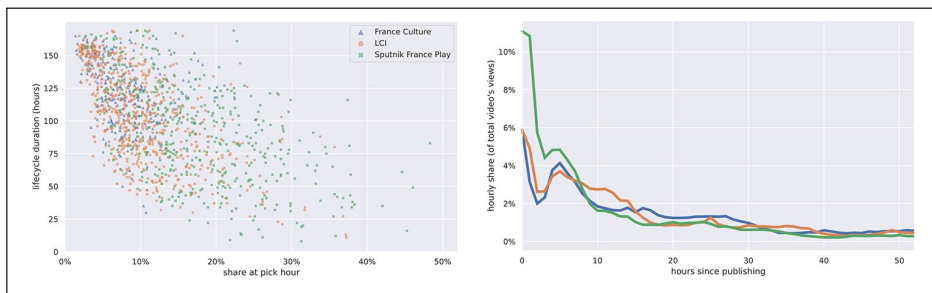
Figure 5. Scatterplot of the average concentration and average life cycle of the videos of 60 YouTube channels highly active in the French political vlogosphere. The size of the dots corresponds to the total number of views collected by the channel and the colour corresponds to the number of published videos (going from dark red for fewer videos, to orange, yellow, green and violet for more videos).

skewed distribution of attention (top-left), to channels that publishes videos at a much more rapid pace knowing that most will spike and die very quickly, but hoping that some of them will stick (bottom-right). It is tempting to interpret such continuum as a quality VS click-bait, even though many exceptions are manifest in the diagram (e.g. France24, Europe1 and TV5Monde, are news channels of established quality and yet all find themselves in the bottom-right corner because of their practices of publishing multiple videos a day).

The scatterplot also allows us to select three chains (boxed in Figure 5) with a similar number of videos and total views, but with very different life spans and concentrations, which we can use as examples of three different attention regimes:

1. 'France Culture' (276 videos, 34,694 views) with long life cycles and low concentration, is a renowned public radio, appreciated for the quality of its contents and the seriousness of its debate.
2. 'LCI' or 'La Chaîne Info' (738 videos, 54,198 views) with medium life cycles and medium concentration, is an established live-news broadcasting service by French public television channel.
3. 'Sputnik France' (466 videos, 32,694 views) with short life cycles and high concentration, is a Russian state-owned news agency, often criticized for spreading misinformation and low-quality stories.

To highlight the difference between the three attention regimes, Figure 6 (left) plots all of their videos for the duration of their lifecycle (the number of hours needed to reach

Figure 6. Left: Scatterplot of the concentration (share at peak hours) and average life cycle of all the videos published by three selected channels (France Culture, LCI and Sputnik France Play). Right: Average temporal profile of the videos of the three channels.

the 95% of the total views collected in the first week) and the share of visibility over the channel that they obtained at their peak hour (the hour in which they collected most of their views). At their peak hour, videos published by France Culture rarely captured over 10% of the total attention for the channel, but many of them remain alive through most of their first week. The videos by Sputnik France often reach shares of 20% or more, but rarely survive after the first two or three days from their publication. The videos from La Chaîne InfoLa Chaîne Info (LCI) are somewhere in between. The difference is clearer when observing the average trajectory of visibility for the videos of the three channels for the first 2 days after their publication (Figure 6 right). While the average Sputnik's video makes most of its views in its first hours and then declines extremely quickly, the average France Culture's video has a slower start, but a thicker visibility tail (LCI presents once again an intermediate profile).

Besides focussing on individual channels, it would be interesting to investigate whether, over a longer period of time, it is possible to observe a general shifting down and to the right of the media system as a whole. Unfortunately, the data at our disposal do not yet allow such an inquiry.

Conclusion

Taken alone, none of the consequences of junk news bubbles highlighted by our model is particularly surprising: being an acceleration, trendiness predictably shortens the lifespan of attention matters and, being a positive feedback, it increases their maximum visibility. Their combination, however, is remarkable as it creates a *shoaling* of attention waves which reduces the width *and* increases the height of attention curves. Debate arenas characterized by a stronger focus on trendiness may therefore end up displaying a syncopated rhythm of attention that is at the same time *increasingly dispersed and increasingly concentrated* (as one can easily observe, for example, in YouTube channels or subreddits devoted to buzzing news, memes and viral contents). Junk news bubbles are characterized by the same attention skewness of Boydstun et al.'s (2014) 'media storms', but not by the same persistence in time. As such, they are particularly worrying because they take attention away from other discussions (because of their skewness),

without producing the heightened public awareness created by media storm (because of their ephemerality).

While evidence exists that the syncopation described by our model can be found in numerous online platforms (Bandari et al., 2015; Castillo et al., 2014; Crane and Sornette, 2008; Wu and Huberman, 2007; Yang and Leskovec, 2011), little empirical research has been carried out on the consequences of such attention regime. The risks of distraction related to screens and online media have been decried at the individual and cultural level (cf. for instance, Citton, 2014; Crawford, 2015; Goldhaber, 1997; Hassan, 2011), but hardly investigated through the records increasingly made available by digital platforms themselves. This article hopes to facilitate such line research by providing a formal description of a distracted attention regime: a situation in which attention waves becomes both higher and narrower and in which public debate is trapped in a continual succession of hot button issues.

Such a situation, arguably, is not particularly propitious to quality. While our model defines junk news bubbles independently from their content value, we suspect this attention regime to be associated with misinformation and poor quality. A regime in which visibility is granted and withdrawn with great rapidity unsurprisingly favours *click-baity* content designed to catch the attention more than to retain it. This observation may explain why, in political discourse, traditional propaganda is increasingly replaced by political trolling aimed at drowning opponents' discourse in noise (Flores-Saviaga et al., 2018; Jack, 2017) or simply to monetize political outrage (Braun and Eklund, 2019).

Being a simplified formalization of a relatively abstract framework, our mathematical model does not allow substantial claims about actual attention dynamics. It allows, however, to advance a precise hypothesis about the junk news bubbles and their detrimental effects on public debate: the fascination with trendiness of digital platforms and their users may create an over-accelerated public debate in which a disproportionate share of media attention is captured by matters which are incapable to sustain it. As the shoaling of sea waves is associated with the entering in shallower waters, so junk news bubbles may be associated with a shallower public debate, a risk that raises concerns and normative implications different from those associated with filter bubbles and fake news. While the latter can be (and has been) addressed by tweaking the recommendation algorithms to favour mainstream sources of information, this solution does not necessarily solve the problem highlighted in this article.

Describing an attention regime that is increasingly pervasive in online media, junk news bubbles cannot be fought by censoring specific content or specific sources, but demands a deep restructuring of the system of incentives that characterize digital communication. Until social media will obtain the largest share of their profits from selling metrics of shallow and ephemeral engagement (e.g. impressions, views, clicks and shares) and until their users will be rewarded according to the same metrics, little are the chances to avoid dynamics of hyper-acceleration. This does not mean that all content producers will play the game of click-bait and junk news – think of the many amazing works produced on YouTube by both mainstream and native creators (Burgess and Green, 2009; Snickars and Vonderau, 2009) – nor that online platforms can only promote superficial forms of engagement – think of how Twitter has been invested by all sorts of political activists (Gerbaudo, 2012). It does mean, however, that in the old opposition

between a distracted public opinion (Lippmann, 1922, 1927) and engaged public inquiry (Dewey, 1927), accelerated attention regimes stack the odds against the latter.

Funding

ORCID iD

Tommaso Venturini https://orcid.org/0000-0003-0004-5308

References

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.

Bandari R, Asur S and Huberman BA (2015) The pulse of news in social media: forecasting popularity. In: *Proceedings of the sixth international AAAI conference on weblogs and social media*. Available at: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view-File/4646/4963

Barberá P, Jost JT, Nagler J, et al. (2015) Tweeting from left to right: is online political communication more than an echo chamber? *Psychological Science* 26(10): 1531–1542.

Benkler Y, Faris R and Roberts H (2018) *Network Propaganda*. Oxford: Oxford University Press.

Boxell L, Gentzkow M and Shapiro J (2017) *Is the Internet Causing Political Polarization? Evidence from Demographics*. Cambridge, MA. Available at: http://www.nber.org/papers/w23258.pdf

Boydstun AE, Hardy A and Walgrave S (2014) Two faces of media attention: media storm versus non-storm coverage. *Political Communication* 31(4): 509–531.

Braun JA and Eklund JL (2019) Fake news, real money: ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism* 7(1): 1–21.

Burgess J and Green J (2009) *Journal of Chemical Information and Modeling Youtube: Online Video and Participatory Culture*. Cambridge: Polity.

Cardon D (2015) *A Quoi Rêvent Les Algorithmes. Nos Vies à l'heure: Nos Vies à l'heure Des Big Data*. Paris: Seuil.

Cardon D, Cointet JP and Mazières A (2018) Neurons spike back. *Réseaux N* 211(5): 173–220.

Castaldo M, Venturini T, Frasca P, et al. (2020) The rhythms of the night: increase in online night activity and emotional resilience during the Covid-19 lockdown. 1–15. Available at: http://arxiv.org/abs/2007.09353

Castillo C, El-Haddad M, Pfeffer J, et al. (2014) Characterizing the life cycle of online news stories using social media reactions. In: *Proceedings of the ACM conference on computer supported cooperative work, CSCW*, pp. 211–223. Available at: https://chato.cl/papers/castillo_elhaddad_pfeffer_stempeck_2013_news_lifecycle.pdf

Cattuto C, Loreto V and Pietronero L (2007) Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences* 104(5): 1461–1464.

Ciampaglia GL, Shiralkar P, Rocha LM, et al. (2015) Computational fact checking from knowledge networks. *PLOS ONE* 10(10): e0141938. DOI: 10.1371/journal.pone.0141938.

Cinelli M, Brugnoli E, Schmidt AL, et al. (2019) Selective exposure shapes the Facebook news diet. Epub ahead of print 13 March. DOI: 10.1371/journal.pone.0229129.

Citton Y (2014) *Pour Une Écologie de l'attention*. Paris: Seuil.

Colleoni E, Rozza A and Arvidsson A (2014) Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication* 64(2): 317–332.

Covington P, Adams J and Sargin E (2016) Deep neural networks for YouTube recommendations. *Proceedings of the 10th ACM conference on recommender systems – RecSys'16*, pp. 191–198. Available at: http://dl.acm.org/citation.cfm?doid=2959100.2959190.

Crane R and Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America* 105(41): 15649–15653.

Crawford MB (2015) *The World Beyond Your Head: On Becoming an Individual in an Age of Distraction*. New York: Farrar, Straus and Giroux.

Crogan P and Kinsley S (2012) Paying attention: toward a critique of the attention economy. *Culture Machine* 13: 1–29.

Dewey J (1927) *The Public and Its Problems*. Athens, OH: Ohio University Press.

Downs A (1972) Up and down with ecology: the 'issue-attention cycle'. *The Public Interest* 28: 38–50.

Dubois E and Blank G (2018) The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information Communication and Society* 21(5): 729–745.

Flaxman S, Goel S and Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80(S1): 298–320.

Flores-Saviaga C, Keegan BC and Savage S (2018) Mobilizing the trump train: understanding collective action in a political trolling community. *ICWSM*: 82–91. Available at: http://arxiv.org/abs/1806.00429

Garrett RK (2009) Echo chambers online? Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication* 14(2): 265–285.

Gerbaudo P (2012) *Tweets and the Streets: Social Media and Contemporary Activism*. London: Pluto Books.

Geschke D, Lorenz J and Holtz P (2019) The triple-filter bubble: using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology* 58(1): 129–149.

Gillespie T (2016) #trendingistrending when algorithms become culture. In: Seyfert R and Roberge J (eds) *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*. London and New York: Routledgepp. 52–75.

Goldhaber MH (1997) The attention economy and the net. *First Monday* 2(4). Available at: https://firstmonday.org/article/view/519/440 (accessed 9 October 2010).

Gonçalves B, Perra N and Vespignani A (2011) Modeling users' activity on Twitter networks: validation of dunbar's number. *PLoS One* 6(8): e22656.

Gray J, Bounegru L and Venturini T (2020) The infrastructural uncanny: 'Fake News' and the fabric of the web as matter of concern. *New Media & Society*.

Graves D (2018) *Understanding the Promise and Limits of Automated Fact-checking*. Reuters Institute for the Study of Journalism

Hassan R (2011) *The Age of Distraction: Reading, Writing, and Politics in a High-speed Networked Economy*. Abingdon: Taylor & Francis.

Hilgartner S and Bosk CL (1988) The rise and fall of social problems: a public arenas model. *American Journal of Sociology* 94(1): 53–78.

Jack C (2017) What's propaganda got to do with it? *Points*. Available at: https://points.datasociety.net/whats-propaganda-got-to-do-with-it-5b88d78c3282

Jenkins H, Ford S and Green J (2013) *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: New York University Press.

Khamis S, Ang L and Welling R (2017) Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies* 8(2): 191–208.

Klein E (January 28, 2020) *Why We're Polarized*. Avid Reader Press, p. 203. ISBN 9781476700397.

Lang A (2000) The limited capacity model of mediated message processing. *Journal of Communication* 50(1): 46–70.

Lanham RA (2006) *The Economics of Attention: Style and Substance in the Age of Information*. Chicago, IL: University of Chicago Press.

Latour B, Jensen P, Venturini T, et al. (2012) 'The whole is always smaller than its parts': a digital test of gabriel tardes' monads. *The British Journal of Sociology* 63(4): 590–615.

Lazer D, Pentland A, Adamic L, et al. (2009) Life in the network: the coming age of computational social science. *Science* 323(5915): 721–723.

Leskovec J, Backstrom L and Kleinberg J (2009) Meme-tracking and the dynamics of the news cycle. In: *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 497–506. Paris, France: ACM. Available at: http://portal.acm.org/citation.cfm?id=1557019.1557077 (accessed 2 October 2010). 497–506.

Lippmann W (1922) *Public Opinion*. New York: Simon & Schuster.

Lippmann W (1927) *The Phantom Public*. New York: The Macmillan Company.

Lorenz-Spreen P, Mønsted BM, Hövel P, et al. (2019) Accelerating dynamics of collective attention. *Nature Communications* 10(1): 1759.

McCombs M (2004) *Opinion, Setting the Agenda: Mass Media and Public*. Cambridge: Polity.

McCombs M (2005) A look at agenda-setting: past, present and future. *Journalism Studies* 6(4): 543–557.

McCombs ME and Shaw DL (1972) The agenda-setting function of mass media. *Public Opinion Quarterly* 36(2): 176.

McLuhan M (1964) *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.

Maratea R (2008) The E-rise and fall of social problems: the blogosphere as a public arena. *Social Problems* 55(1): 139–160.

Marres N (2018) Why we can't have our facts back. *Engaging Science, Technology, and Society* 4: 423.

Marwick A and Boyd D (2011) To see and be seen: celebrity practice on Twitter. *Convergence* 17(2): 139–158.

Merton RK (1968) The Matthew effect in science. *Science* 159(3810): 56–63.

Newman ME (2001) Clustering and preferential attachment in growing networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 64: 025102.

Pariser E (2011) *The Filter Bubble. What the Internet Is Hiding from You*. New York: Penguin Press.

Prior M (2005) News vs. Entertainment: how increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science* 49(3): 577–592.

Quattrociocchi W, Scala A and Sunstein CR (2016) Echo chambers on Facebook. *SSRN*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110

Reutlinger A, Hangleiter D and Hartmann S (2018) Understanding (with) toy models. *The British Journal for the Philosophy of Science* 69(4): 1069–1099.

Rieder B (2020) *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam: University Press.

Rogers R (2018) Otherwise engaged: social media from vanity metrics to critical analytics. *International Journal of Communication* 12(732942): 450–472.

Schelling T (1971) Dynamic models of segregation. *Journal of Mathematical Sociology* 1: 143–186.

Sears DO and Freedman JL (1967) Selective exposure to information: a critical review. *The Public Opinion Quarterly* 31(2): 194–213.

Shoemaker PJ and Vos T (2009) *Gatekeeping Theory*. New York: Routledge.

Snickars P and Vonderau P (2009) *The YouTube Reader*. New York: Columbia University Press.

Sunstein C (2001) *Republic.com*. Princeton: University Press.

Terranova T (2012) Attention, economy and the brain. *Culture Machine* 13: 1–19.

Vasterman PLM (2005) Media-hype: self-reinforcing news waves, journalistic standards and the construction of social problems. *European Journal of Communication* 20(4): 508–530.

Venturini T (2019) From fake to junk news, the data politics of online virality. In Didier Bigo, Engin Isin, and Evelyn Ruppert (eds) *Data Politics: Worlds, Subjects, Rights*. London: Routledge, pp. 123–144. Available at: http://www.tommasoventurini.it/wp/wp-content/uploads/2018/10/Venturini_FromFakeToJunkNews.pdf

Venturini T, Jensen P and Latour B (2015) Fill in the gap: a new alliance for social and natural sciences. *Journal of Artificial Societies and Social Simulation* 18(2): 11.

Wardle C and Derakhshan H (2017) *Information disorder: toward an interdisciplinary framework for research and policymaking*. Report to the Council of Europe. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Webster J (2014) *The Marketplace of Attention*. Cambridge, MA: MIT Press.

Weng L, Flammini A, Vespignani A, et al. (2012) Competition among memes in a world with limited attention. *Scientific Reports* 2(1): 335.

Wu F and Huberman BA (2007) Novelty and collective attention. *Proceedings of the National Academy of Sciences of the United States of America* 104(45): 17599–17601.

Yang J and Leskovec J (2011) Patterns of temporal variation in online media. In: *Proceedings of the 4th ACM international conference on web search and data mining, WSDM 2011*, pp. 177–186. Available at: https://cs.stanford.edu/~jure/pubs/memeshapes-wsdm11.pdf

Zhao Z, et al (2019) Recommending what video to watch next: a multitask ranking system. In: *RecSys 2019 – 13th ACM conference on recommender systems*, pp. 43–51. Available at: https://www.researchgate.net/publication/335771069_Recommending_what_video_to_watch_next_a_multitask_ranking_system

Zuckerman E (2017) Stop saying 'Fake news'. It's not helping. Available at: http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/ (accessed 8 September 2017).

Author biographies

Maria Castaldo is a CNRS PhD student. Her interdisciplinary PhD etitled "System for social network disorders" is settled in-between Mathematical Modelling, Digital Methods and Data Analysis.

Tommaso Venturini is a CNRS permanent researcher. He is an associate researcher of Inria and of the Médialab of Sciences Po Paris and a founding member of the Public Data Lab. His research focuses on Digital Methods, STS and Visual Network Analysis. He teaches Controversy Mapping, Data Journalism, Data Activism and Information Design at graduate and undergraduate level.

Paolo Frasca is a CNRS Researcher based in Grenoble, France. He is an associate researcher of CNR-IEIIT, Turin, Italy. Paolo's research lies at the crossroad of network science and mathematical control theory, with applications to transportation and social networks.

Floriana Gargiulo is a CNRS permanent researcher, at the GEMASS laboratory (Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne). Her profile is inscribed in the framework of Computational Social Science. Using methods from data science, complex networks and modelling, her research is focused on digital collective behaviors, opinion formation and science of science.