PLOS ONE

Modeling opinion polarization on social media: Application to Covid-19 vaccination hesitancy in Italy

Jonathan Franceschi[1]*, Lorenzo Pareschi[2], Elena Bellodi[3], Marco Gavanelli[3], Marco Bresadola[4]

1 Department of Mathematics "F. Casorati", University of Pavia, Pavia, Italy, 2 Department of Mathematics and Computer Science, University of Ferrara, Ferrara, Italy, 3 Department of Engineering, University of Ferrara, Ferrara, Italy, 4 Department of Humanities, University of Ferrara, Ferrara, Italy

* jonathan.franceschi01@universitadipavia.it

Abstract

The SARS-CoV-2 pandemic reminded us how vaccination can be a divisive topic on which the public conversation is permeated by misleading claims, and thoughts tend to polarize, especially on online social networks. In this work, motivated by recent natural language processing techniques to systematically extract and quantify opinions from text messages, we present a differential framework for bivariate opinion formation dynamics that is coupled with a compartmental model for fake news dissemination. Thanks to a mean-field analysis we demonstrate that the resulting Fokker-Planck system permits to reproduce bimodal distributions of opinions as observed in polarization dynamics. The model is then applied to sentiment analysis data from social media platforms in Italy, in order to analyze the evolution of opinions about Covid-19 vaccination. We show through numerical simulations that the model is capable to describe correctly the formation of the bimodal opinion structure observed in the vaccine-hesitant dataset, which is witness of the known polarization effects that happen within closed online communities.

1 Introduction

Vaccination coverage, globally, has been at its highest levels for the last decades, with the notable exception of measles and diphtheria [1]. Measles outbreaks in particular [2] have raised concern in the Western public since they were about a disease that vaccination and treatments had reduced to a condition of rarity. The SARS-CoV-2 pandemic has been one compelling reason to rethink vaccination as an effective medical practice to prevent the spreading of diseases, especially in relation to the massive media coverage of the topic. The issue, in this case, is that a polarizing debate could exacerbate *vaccination hesitancy*, i.e., the reluctance in getting vaccinated (see [3–6], but also the recent [7–9]), with potentially dangerous implications for healthcare [10–13]. Moreover, unstable contexts like this one are more likely to develop irreparable fractures when misinformation is disseminated among people, and a positive reinforcement loop clusters the audience into isolated groups (the so-called *echo chambers*) where the only information shared is the one aligned with the majority point of view [5, 14–16]. Therefore,

the need for the policymaker to be able to take informed decisions based on the understanding of the directions of the debate evolution is critical.

Vaccines are no stranger issue to mathematical literature, too [17, 18], especially the game-theoretical one. The *free-ride effect* of the portion of population not taking the possible risks associated with vaccination to enjoy the benefits given by the herd immunity are well known [19, 20]. So are opinions [21–25], especially in the context of multi-agent systems and kinetic theory more generally [26–32]. Within the same field, works have been proposed recently that borrow from the classical compartmental framework of epidemiological theory [33, 34], both for the spread of diseases and misinformation [35–39].

In this paper, we build on these elements to present a differential model, based on a mean-field description of agent dynamics, for the evolution of opinions in the presence of fake news spreaders. Although the model is mainly applied to the spread of fake news on social networks regarding the hesitancy to Covid-19 vaccination, its structure easily finds application in more general contexts where polarization of opinions within closed online communities is observed.

More in details, following the seminal paper [36], we consider a multi-agent population with a structure, where the modeling of fake news dissemination is managed via a set of compartments. In this setting, where the fake-news is treated as the spreading of a virus, the underlying variable that is shared by each agent characterizes a bivariate opinion distribution, that takes into account both positive and negative opinions about a given topic. This last aspect is crucial with a view to aligning the model with experimental data from *sentiment analysis* carried out on social media platforms, such as blogs and social networks. In such a situation, each opinion is inherently two-dimensional, as it classifies the polarity of a given text according to which level the opinion expressed is positive or negative [40–44].

We emphasize that, unlike [30, 36, 45], our starting point is a system of stochastic differential equations (SDEs) for the dynamics of opinions and not a binary interaction dynamics leading to a Boltzmann-type equation in the limit of a large number of agents. In fact, we are interested in modeling a situation in which agents interact simultaneously with the entire population, a scenario typical of group chats in instant messaging. As a consequence, the resulting model can be analyzed directly thanks to its mean-field approximation that permits to compute explicitly the steady states of the system without resorting to the quasi-invariant opinion approximation. The equilibrium states, in contrast with the classical case [30, 46], are characterized by a superposition of Beta distributions that give rise to bimodal shapes, i.e., individuals' thoughts polarizing around different extreme positions, with a certain absence of compromise, in agreement with those observed opinion polarization effects in closed communities.

The model is then interfaced with available data concerning Covid-19 vaccination in Italy from the popular messaging app Telegram; one of its features is the possibility of having large online group chats focused on a topic of choice. They effectively form closed communities, where conversations experiment a low degree of noise: they are therefore ideal to analyze the evolution of sentiments about a certain subject. Numerical simulations show the model's ability to interface correctly with the data extracted using NLP techniques and to describe the polarization phenomenon over time very well.

The rest of the manuscript is organized as follows: in Section 3, we present the stochastic differential model for opinion-formation processes characterized by two-dimensional vectors, which in the mean-field limit is approximated by a Fokker-Planck equation that allows us to compute steady states for marginals in explicitly solvable special cases. Next, in Section 4, we merge the model with a compartmental framework to take into account the potential spread of misinformation which can act as a catalyst for the polarization. In Section 5 we present the social media dataset and compare the evolution predicted by the model to the data one. Finally, in the last section some final considerations and concluding remarks are reported.

2 Materials and methods

All records from our datasets are posts from open, freely accessible, Telegram Italian chats, all of which are related to the topics of Covid-19 or vaccination:

- Io Non Mi Vaccino Chat

- Vittime vaccino Covid in Italia

- Mamme dissociate d'Italia/No vaccini COVID su bambini

- Singles italiani NON vaccinati

- COMBATTENTI NO BOOSTER—NO TERZA DOSE—NO VAC—NO GREEN PASS

- Personale Scuola—No Green Pass—No Booster Vax

We collected a total of 4077 posts from six different chats, from August 20th, 2021 to February 27th, 2022, which we anonymized and aggregated in order to obtain a unique, larger cluster. The collection method complied with the terms and conditions for the source of the data: in the Telegram Desktop version the "Export Chat history" functionality was used to download the data of interest from each chat in JSON format.

3 Mean-field models of bivariate opinion formation

When modelling the dynamics of opinions within individuals from a mathematical point of view, several approaches based on multi-agent interactions at various levels are possible [14, 26, 45, 47–50]. It is customary to set the interval $[-1, 1] \subseteq \mathbb{R}$ as a natural space for the variable w representing the opinion, intending that radical positions are assumed as the absolute value $|w|$ approaches 1, while neutral ones are assumed near 0. This choice embeds opinions as a continuous spectrum between positive and negative convictions and allows for a relatively simple description as a one-dimensional variable.

Here, we are setting ourselves in a subspace of the plane to better interpret the inherently two-dimensional nature of the description of opinions given by *natural language processing* (NLP) techniques like sentiment analysis, which assign scores based on how much a certain thought can be perceived as positive or negative, so that each record is associated with a pair of scores. Although a multivariate model for opinion requires greater care to devise it and to perform computations, it also gives us more coherent informations when aligning the model with data.

Of course, one can simplify the problem by projecting the two-dimensional opinion space $[0, 1]^2$ onto a simpler one-dimensional space $[-1, 1]$. There are many ways to achieve this. Besides the previously mentioned mapping of positive and negative opinions in $[0, 1]$ and $[-1, 0]$, one can also achieve a one-dimensional mapping by considering the difference between positive and negative opinions in each output of the NLP algorithm. However, these one-dimensional embeddings, in addition to losing the correlation effects between positive and negative opinions, show limitations when it comes to polarization. For instance, if you choose to subtract the terms, a resulting neutral opinion close to zero is obtained both in the case of two bold contrasting statements and for a general mild thought.

The other main modeling choice is the use of SDEs instead of other alternatives, such as binary interactions described according to a broader interpretation of particle dynamics typical of statistical mechanics [30]. Here we are interested in modeling a situation in which agents interact simultaneously with the whole population at all times: this is the typical scenario of group chats within instant messaging applications.

### 3.1 A multi-agent stochastic differential model

We consider a population of N indistinguishable agents: at the instant $t \geq 0$, the i-th agent possesses an opinion expressed as a pair of sentiments (one positive and one negative), indicated by the vector $\mathbf{W}_t^i = (W_{+,t}^i, W_{-,t}^i) \in [0,1]^2$. The variables $W_{\pm,t}^t$ can be thought as the intensities of those sentiments, so that when they are close to zero, they express a mild opinion, while when they are close to one they translates into fierce ones. In Fig 1 we sketch how the model works.

The continuous time evolution of the pair $\mathbf{W}_t^i$ can be expressed via a stochastic differential system of the general form

$$\begin{cases} dW_{+,t}^i = \dfrac{1}{N}\sum_{j=1}^N \Lambda^+(\mathbf{W}_t^i, \mathbf{W}_t^j)(W_{+,t}^j - W_{+,t}^i)dt + \sigma_+ D^+(\mathbf{W}_t^i)dB_{+,t}^i \\[3mm] dW_{-,t}^i = \dfrac{1}{N}\sum_{j=1}^N \Lambda^-(\mathbf{W}_t^i, \mathbf{W}_t^j)(W_{-,t}^j - W_{-,t}^i)dt + \sigma_- D^-(\mathbf{W}_t^i)dB_{-,t}^i \end{cases} \tag{1}$$

where $\Lambda^\pm(\mathbf{W}_t^i, \mathbf{W}_t^j)$ are nonnegative functions characterizing the rate towards compromise when two agents i and j interact. Then $\sigma_\pm$ are positive constant diffusion coefficients, while the nonnegative functions $D^\pm(\mathbf{W}_t^i)$ represent the local incidence of the diffusion effects due to self-thinking of agent i. The latter functions usually vanish at the boundary of $[0,1]^2$ so that people at extreme positions are less subject to noise effects. Finally, $dB_{\pm,t}^i$ are independent one-dimensional Brownian motions to take into account the random nature of social interactions. In Fig 2 we show an example of simulation of model (1) for demonstration purposes.

It should be noted that the model introduced in (1), while inspired by the classical works by Hegselmann and Krause [51], Deffuant et al. [47], and Sznajd-Weron et al. [48], has some important differences due to the presence of a noise term in the interactions and the continuous-time description of the opinion formation process. In particular, due to the presence of noise, the dynamics may give rise to an inadmissible opinion vector. For this reason, the
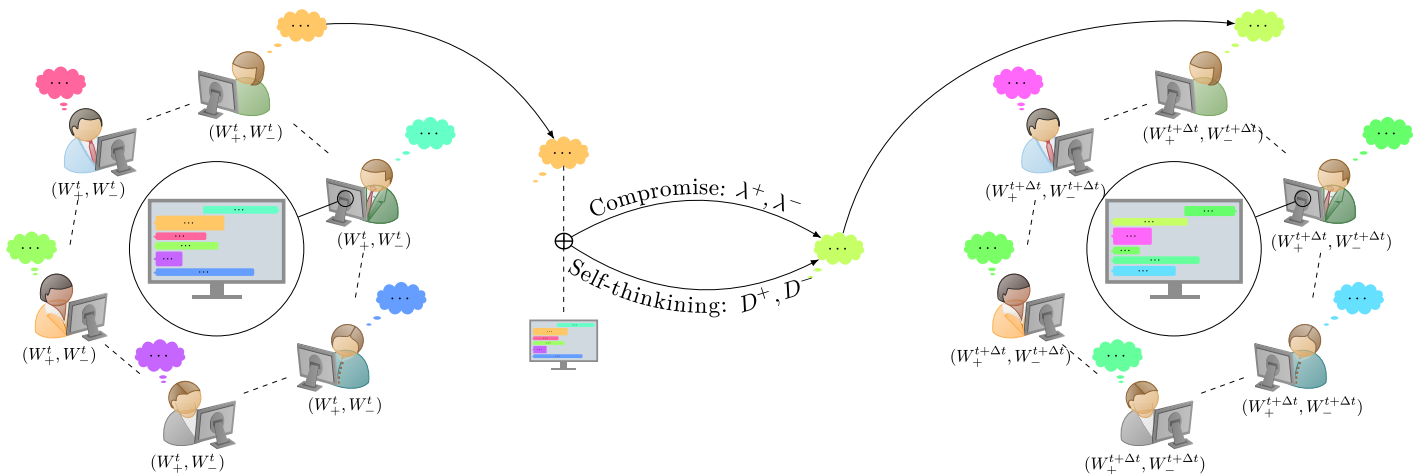
Fig 1. Opinions expressed by users in a group chat (represented here by different colors) are extrapolated by the messages via NLP and are denoted by a pair of continuous, time-dependent, real values $(W_t^+, W_t^-) \in [0,1]^2$: One for how positive the opinion is and one for how negative it is, respectively. The dynamics is then characterized by the functions λ^+, λ^- that define the compromise process and the functions D^+, D^- that formalize individual self-thinking.
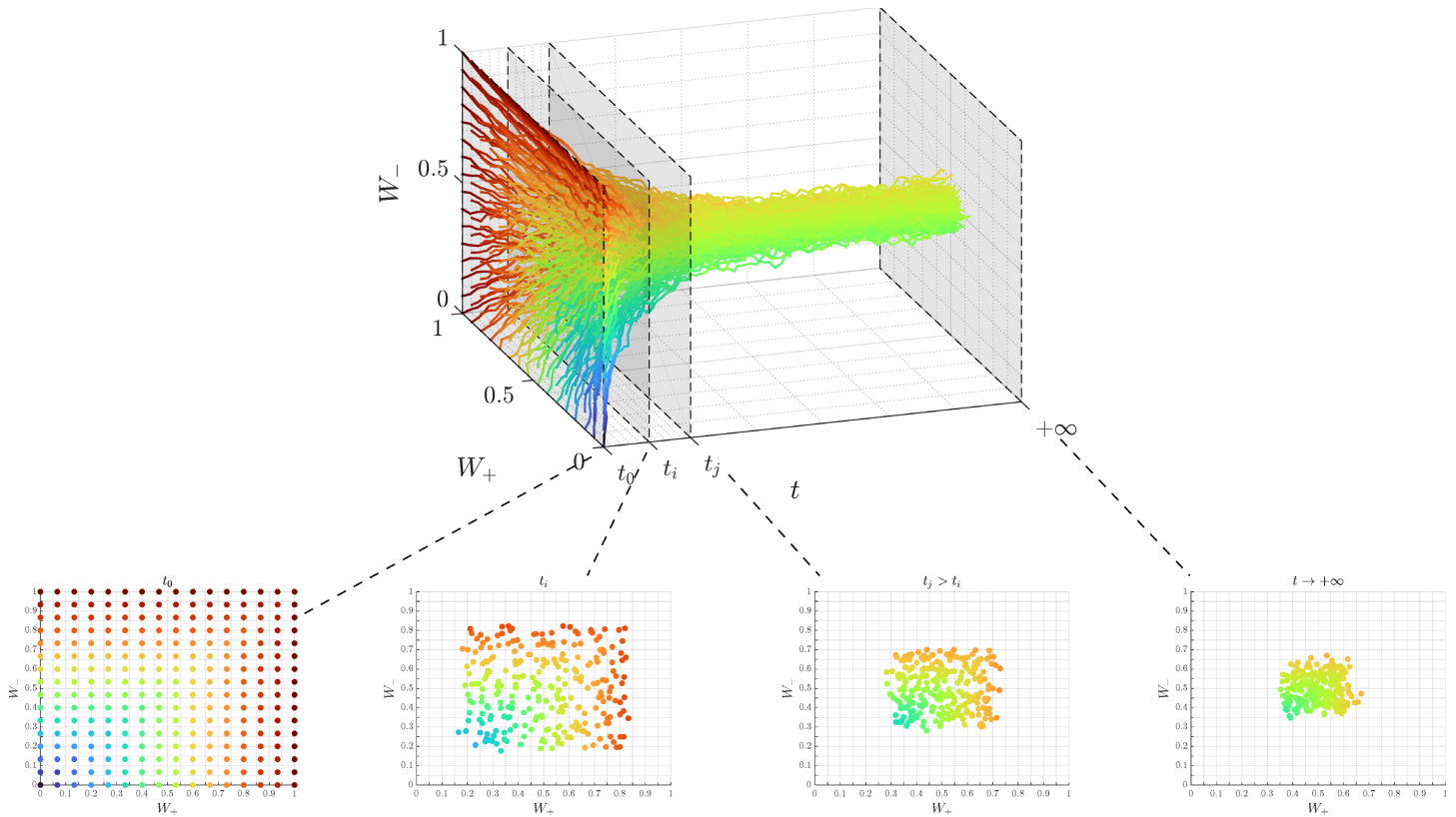
Fig 2. Simulation of model (1) using Euler-Maruyama scheme using $N = 256$ agents with $\lambda^+ = \lambda^- = \sigma_+ = \sigma_- = 0.05$. Here colors represent opinion's intensity measured as $\|\mathbf{W}_t\|_\infty$. Opinions initially are drawn from a uniform distribution and then, due to the compromise dynamics, concentrate toward the center of the unit square.

model is supplemented with appropriate boundary conditions that constrain the opinion vector within the domain $[0, 1]^2$.

A typical example of compromise functions is represented by the bounded confidence model [51] where agents interact only if their opinions differ (component-wise) no more than a certain confidence level $\Delta_\pm \in [0, 1]$

$$\Lambda^\pm(\mathbf{W}_t^i, \mathbf{W}_t^j) = \lambda^\pm \Psi(|W_{\pm,t}^i - W_{\pm,t}^j| \leq \Delta_\pm)$$

with $\Psi(\cdot)$ the indicator function and $\lambda^\pm$ suitable positive constants. Further in this paper, we will restrict to the simplified situation where $\Delta_\pm = 1$ (which means that agents can interact with everyone else) and thus $\Lambda^\pm(\mathbf{W}_t^i, \mathbf{W}_t^j) = \lambda^\pm$ represent the alignment strengths towards the current mean positive and negative opinions of the population

$$M_{+,t}^N = \frac{1}{N}\sum_{j=1}^N W_{t,+}^j, \qquad M_{-,t}^N = \frac{1}{N}\sum_{j=1}^N W_{t,-}^j. \tag{2}$$

Next, we introduce the empirical measure

$$f^N(\mathbf{w}, t) := \frac{1}{N}\sum_{i=1}^N \delta(\mathbf{w} - \mathbf{W}_t^i), \tag{3}$$

where $\mathbf{w} = (w_+, w_-) \in [0, 1]^2$ and $\delta(\cdot)$ is the Dirac delta function, which counts how many agents share the same pair of opinions at time $t \geq 0$. Our main goal is to analyze the evolution of the empirical measure of the system, especially when the number of individuals in the population grows large. The advantage of resorting to a limit procedure in order to study the mean-field version of system (1) is that under simplifying hypotheses it is possible to compute the stationary state of quantities of interest concerning the system.

3.2 Mean-field limit and Fokker-Planck approximation

A classical approach to formally analyze the behavior of the system when the number N of agents in the population is large, is to consider the N-particle probability density [52, 53]

$$f^{(N)}(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N, t)$$

and the associated first marginal

$$f_1^{(N)}(\mathbf{W}_t^1, t) = \int_{[0,1]^{2N-2}} f^{(N)}(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N, t) \, d\mathbf{W}_t^2, \ldots, d\mathbf{W}_t^N.$$

and make the so-called *propagation of chaos assumption* on the marginals. More specifically, we assume that $f^{(N)} \approx f^{\otimes N}$ for $N \gg 1$, i.e., the random vectors $\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N$ are approximately independently $f(\mathbf{w}, t)$-distributed.

In this case, for $N \gg 1$ we can write

$$f^N(\mathbf{w}, t) \approx f(\mathbf{w}, t), \qquad (M_{+,t}^N, M_{-,t}^N) \approx (m^+(t), m^-(t)) = \int_{[0,1]^2} f(\mathbf{w}, t)\mathbf{w} \, d\mathbf{w}, \qquad (4)$$

due to the law of large numbers. Consequently, the SDE model (1) becomes independent of $j \neq i$ and we obtain a so-called *Mc-Kean nonlinear process* (see, e.g., [54]) which, in the simplified situation where $\lambda^\pm$ are non negative constants, reads

$$\begin{cases} dW_t^+ = \lambda^+ (m^+(t) - W_t^+)dt + \sigma_+ D^+(\mathbf{W}_t)dB_t^+, \\ dW_t^- = \lambda^- (m^-(t) - W_t^-)dt + \sigma_- D^-(\mathbf{W}_t)dB_t^-, \end{cases} \qquad (5)$$

with $f := \mathrm{law}(\mathbf{W}_t)$, i.e., the measure induced by the process $\mathbf{W}_t$. The above system may be equivalently expressed by a nonlinear Fokker-Planck equation of the form [52, 53, 55]

$$\begin{aligned} \frac{\partial}{\partial t} f(\mathbf{w}, t) &= \lambda^+ \frac{\partial}{\partial w_+} [(w_+ - m^+(t))f(\mathbf{w}, t)] + \lambda^- \frac{\partial}{\partial w_-} [(w_- - m^-(t))f(\mathbf{w}, t)] \\ &+ \frac{\sigma_+^2}{2} \frac{\partial^2}{\partial w_+^2} \left(D^+(\mathbf{w})^2 f(\mathbf{w}, t) \right) + \frac{\sigma_-^2}{2} \frac{\partial^2}{\partial w_-^2} \left(D^-(\mathbf{w})^2 f(\mathbf{w}, t) \right). \end{aligned} \qquad (6)$$

Eq (6) needs to be complemented with suitable no-flux boundary conditions that guarantee $f(\mathbf{w}, t)$ to be compactly supported in $[0, 1]^2$

$$\lambda^+ (w_+ - m^+(t))f(\mathbf{w}, t) + \frac{\sigma_+^2}{2} \frac{\partial}{\partial w_+} \left(D^+(\mathbf{w})^2 f(\mathbf{w}, t) \right) = 0, \quad \text{on } w_+ = 0, 1$$

$$\lambda^- (w_- - m^-(t))f(\mathbf{w}, t) + \frac{\sigma_-^2}{2} \frac{\partial}{\partial w_-} \left(D^-(\mathbf{w})^2 f(\mathbf{w}, t) \right) = 0, \quad \text{on } w_- = 0, 1. \qquad (7)$$

Thanks to the above conditions we can introduce the normalization assumption

$$\int_{[0,1]^2} f(\mathbf{w}, t) d\mathbf{w} = 1, \quad \forall\, t \geq 0.$$

We refer to [53], and the references therein, for rigorous results concerning the mean-field limit of stochastic particle system of type (1).

Note that, if in addition, at the boundary of $[0, 1]^2$ we have

$$D^+(\mathbf{w})^2 f(\mathbf{w}, t) = D^-(\mathbf{w})^2 f(\mathbf{w}, t) = 0, \tag{8}$$

integrating by parts and using the no-flux boundary conditions (7), we have conservation of the mean opinion

$$\frac{d\,\mathbf{m}(t)}{dt} = \int_{[0,1]^2} \frac{\partial}{\partial t} f(\mathbf{w}, t) \mathbf{w}\, d\mathbf{w} = 0.$$

If we now define the variances of the variables w^+ and w^- as

$$V^+(t) = \int_{[0,1]^2} f(\mathbf{w}, t)(w^+ - m^+)^2\, d\mathbf{w}, \quad V^-(t) = \int_{[0,1]^2} f(\mathbf{w}, t)(w^- - m^-)^2\, d\mathbf{w},$$

we have

$$\frac{dV^\pm(t)}{dt} = -2\lambda^\pm V^\pm(t) + \sigma_\pm^2 \int_{[0,1]^2} D^\pm(\mathbf{w})^2 f(\mathbf{w}, t)\, d\mathbf{w}. \tag{9}$$

This shows that the particular choice of the functions $D^\pm(\mathbf{w})$ influences the behavior of the variance and so the convergence to equilibrium of the Fokker-Planck equation (6).

3.3 Equilibrium states for the marginal densities

The the functions $D^\pm(\mathbf{w})$ characterizing the local effect of diffusion, and thus the individual behavior of agents, turn out to be essential for the purpose of studying the equilibrium states of the system. For example taking

$$D^\pm = |w^\pm - m^\pm|,$$

where agents tends to reduce self-thinking as their opinion is close to the average, from (9) we get the uniform decay of the variances as soon as $2\lambda^\pm > \sigma_\pm^2$. In this case the long time behavior is characterized by a Dirac delta function $f^\infty(\mathbf{w}) = \delta(\mathbf{w} - \mathbf{m})$ where all agents are concentrated on the same opinion.

In the rest of the paper, we assume that opinions close to zero and one are less prone to random opinion effects, in the sense that both very moderate and more extreme individuals in expressing opinions have less freedom to change opinion, since they are already positioned in an extremal state. This assumption turns out to be essential in order to derive steady states in agreement with the experimental data, and differs from classical one-dimensional opinion models where individuals with an opinion around zero are assumed to be hesitant and so mostly prone to the effect of diffusion.

To this aim, we consider the local diffusion function to be such that $D^+(\mathbf{w}) = D(w_+)$ and $D^-(\mathbf{w}) = D(w_-)$ with

$$D(w) = \sqrt{w(1 - w)}, \tag{10}$$

so that it vanishes in 0 and 1. This assumption, if the solution $f(\mathbf{w}, t)$ is sufficiently regular,

guarantees conditions (8) at the boundary and therefore the mean opinion is independent from time. We refer to [30] for other admissible choices leading to interesting steady states. As we will see, thanks to (10) we are able to compute explicitly the steady state for the marginal densities.

Indeed, let us integrate directly system (6) with respect to the negative opinion w_-, so that we have

$$
\begin{aligned}
\frac{\partial}{\partial t} \int_0^1 f(\mathbf{w}, t)\, dw_- \;=\;& \lambda^+ \frac{\partial}{\partial w_+} \left[(w_+ - m^+) \int_0^1 f(\mathbf{w}, t)\, dw_- \right] \\
&+ \lambda^- (w_- - m^-) f(\mathbf{w}, t)|_0^1 \\
&+ \frac{\sigma_+^2}{2} \frac{\partial^2}{\partial w_+^2} \left[w_+(1 - w_+) \int_0^1 f(\mathbf{w}, t)\, dw_- \right] \\
&+ \frac{\sigma_-^2}{2} \frac{\partial}{\partial w_-} [w_-(1 - w_-) f(\mathbf{w}, t)]|_0^1 .
\end{aligned}
\tag{11}
$$

Thanks to the boundary conditions in (7) we have the simplification

$$
\frac{\partial}{\partial t} g(w_+, t) = \lambda^+ \frac{\partial}{\partial w_+} \left[(w_+ - m^+) g(w_+, t) \right] + \frac{\sigma_+^2}{2} \frac{\partial^2}{\partial w_+^2} \left[w_+(1 - w_+) g(w_+, t) \right],
\tag{12}
$$

where we denote the marginal density of the positive opinion as

$$
g(w_+, t) = \int_0^1 f(\mathbf{w}, t)\, dw_- .
$$

We can now compute the stationary solution $g^\infty(w_+)$ by observing that using the boundary conditions (7) it satisfies

$$
\lambda^+ \left[(w_+ - m^+) g^\infty(w_+) \right] + \frac{\sigma_+^2}{2} \frac{\partial}{\partial w_+} \left[w_+(1 - w_+) g^\infty(w_+) \right] = 0.
$$

Thus, g^∞ is computed explicitly as [30]

$$
g^\infty(w_+) = C_+ w_+^{m^+/\mu_+ - 1} (1 - w_+)^{(1 - m^+)/\mu_+ - 1},
\tag{13}
$$

where $\mu_+ := \lambda^+/\sigma_+^2$, $m_+ \in (0, 1)$ and C_+ is a normalization constant which depends on all parameters appearing in (13). Eq (13) represents a Beta distribution of the form Beta($w; a, b$) with $a = m^+/\mu_+$ and $b = (1 - m^+)/\mu_+$ (see Fig 3). Note also that since $a, b > 0$ condition (8) is always guaranteed.

Analogous calculations can be performed to obtain a closed expression for the steady state of the marginal density of the negative opinion $h(w_-, t)$ which reads

$$
h^\infty(w_-) = C_- w_-^{m^-/\mu_- - 1} (1 - w_-)^{(1 - m^-)/\mu_- - 1},
\tag{14}
$$

where now $\mu_- := \lambda^-/\sigma_-^2$, $m_- \in (0, 1)$ and C_- is a normalization constant. In Fig 3 are reported examples of various stationary marginal opinion distributions.

One of their interesting properties is that they are flexible enough to give rise to several different shapes, including some that are unbounded near the ends of the support. This is representative of extreme polarization phenomena in which the vast majority of the population shares extreme ideas. Beta distributions are also *unimodal*, i.e., agents well described by a Beta tend to aggregate around a certain, unique, value, where this tendency depends on the variance
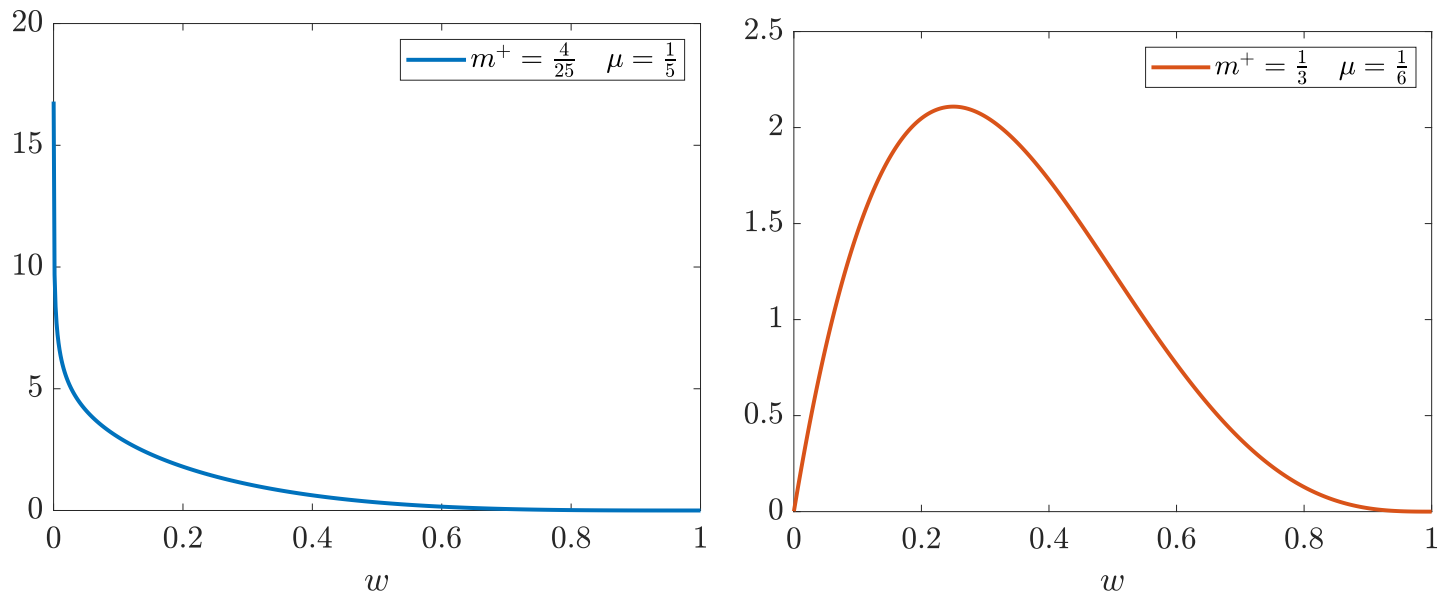
Fig 3. Examples of stationary marginal opinion distributions $g^{\infty}(w_+)$ characterized by Beta(w_+; a, b) functions obtained with different choices of m^+ and μ. Left: $m^+ = 4/25$, $\mu = 1/5$, corresponding to $a = 0.8$, $b = 4.2$. We can see that the function tends to infinity as we approach the left boundary. Right: $m^+ = 1/3$, $\mu = 1/6$, corresponding to $a = 2$, $b = 4$, with a unimodal structure.

of the distribution. However, this implies that a similar model would not be able to accurately describe different kinds of polarization, the ones that are local to certain population subsets and that may have a *bimodal* structure.

These latter phenomena are the ones more commonly associated with the formation of separate clusters within communities after the population has been exposed to fake news: misinformation exacerbates underlying radical opinions in a group of individuals, who then progressively proceed to discard any other belief. While this happens broadly, at different levels, (the phenomenon of *echo chambers*), the less moderate one are often the one that raise more concern [56–61].

Therefore, the model should also take into account the effects brought by the dissemination of fake news that also lead to changes of the average opinions within subgroups of individuals. This will be explored in the next section.

4 Merging opinion formation with fake news dissemination

The full model focuses again on a structured population where N indistinguishable agents all share a vector-valued variable $\mathbf{w} \in [0, 1]^2$. In the current setting, concurrent to the opinion formation process is also spread of misinformation, whose dynamics can be fruitfully approached through the compartmental framework typical of epidemiology [33, 34, 36].

4.1 Defining fake news

Defining what fake news is and why it is a phenomenon deserving its own category (think for instance to other classifications of lies, e.g., scams, hoaxes, urban legends and etc.) is itself challenging. Our approach will be to consider fake news any piece of information whose *initial* diffusion is made with the purpose of mislead people intentionally. The word 'initial' here is key, because fake news is most often spread by people who do not know (or care) it is false (see Fig 4). This has been linked to the concept of *post-truth* and explored also from a philosophical
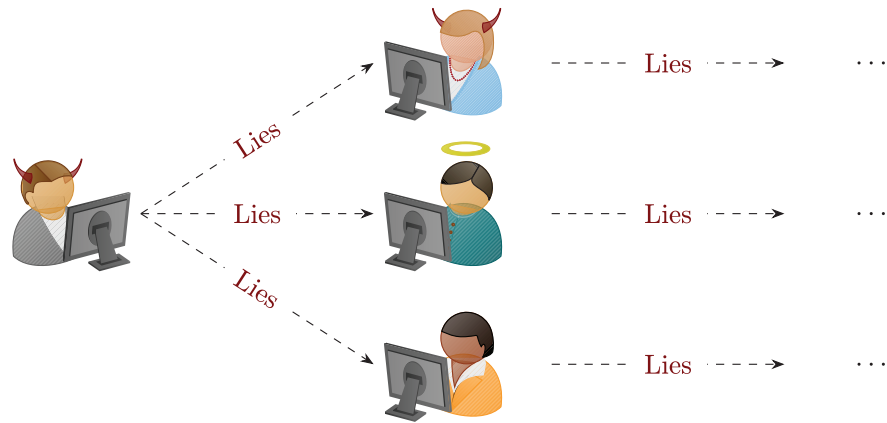
Fig 4. Defining fake news: here the evil look of the person on the left symbolizes the purpose of voluntarily mislead others with false information; whereas the first recipients of the news spread it, in turn, either animated by the same will (evil agent, top), by the desire of sharing helpful or otherwise legitimate information (angelic agent, middle) or finally guided by no specific goal (neutral agent, bottom).

https://doi.org/10.1371/journal.pone.0291993.g004

point of view [62]; see [36] and the references therein for a recent overview on the different approaches for fake news detection.

4.2 A compartmental model for fake news spreading

For what concerns the dissemination of fake news, we assume that the people within the community can be divided into four disjoint compartments: susceptible (or ignorant), exposed (or incubator), infected (or spreader) and recovered (or stifler). Concerning the nomenclature, we opted for the epidemiological convention here to adhere to the previous works [36, 37]. We refer to [36] and the references therein for other popular choices.

Then we use suitable differential equations to describe the way individuals change compartment. To each compartment will be assigned its initial as identifying letter, so that we will refer to them as the set $\mathcal{C} := \{S, E, I, R\}$. We shall therefore study the evolution of the opinion's distribution of the agents in each compartment, noted, respectively, by $f_S = f_S(\mathbf{w}, t)$, $f_E = f_E(\mathbf{w}, t)$, $f_I = f_I(\mathbf{w}, t)$, and $f_R = f_R(\mathbf{w}, t)$.

Like before, we restrict ourselves to consider a reduced time-span, during which we can assume that the population is fixed, i.e., nobody enters or leaves it; This choice is based on the average lifespan of fake news. Thus, we set the overall opinion distribution as a probability density for all $t \geq 0$, i.e.,

$$\int_{[0, 1]} \sum_{J \in \mathcal{C}} f_J(\mathbf{w}, t) \, d\mathbf{w} = 1, \quad t > 0.$$

The quantities in the first column of

$$g_J(w_+, t) = \int_{[0, 1]} f_J(\mathbf{w}, t) \, dw_- \qquad m_J^+(t) = \frac{1}{\rho_J(t)} \int_{[0, 1]^2} w_- f_J(\mathbf{w}, t) \, d\mathbf{w},$$

$$h_J(w_-, t) = \int_{[0, 1]} f_J(\mathbf{w}, t) \, dw_+ \qquad m_J^-(t) = \frac{1}{\rho_J(t)} \int_{[0, 1]^2} w_+ f_J(\mathbf{w}, t) \, d\mathbf{w}$$

denote the marginals densities, i.e., the fractions of the population that belongs to compartment $J \in \mathcal{C}$ with positive and negative opinion, respectively, at time $t \geq 0$, while in the second column we denote the mean relative to the positive and to the negative opinion, respectively. Finally,

$$\rho_J(t) = \int_{[0,1]^2} f_J(\mathbf{w}, t) \, d\mathbf{w}$$

is the total mass fraction of agents in the compartment J.

When the fake-news dynamic is independent from the opinion of individuals it follows the simple system of ordinary differential equations [36]

$$\begin{cases} \dfrac{d\rho_S(t)}{dt} = -\beta\rho_S(t)\rho_I(t) + (1-\alpha)\gamma\rho_I(t) \\[2mm] \dfrac{d\rho_E(t)}{dt} = \beta\rho_S(t)\rho_I(t) - \zeta\rho_E(t) \\[2mm] \dfrac{d\rho_I(t)}{dt} = (1-\eta)\zeta\rho_E(t) - \gamma\rho_I(t) \\[2mm] \dfrac{d\rho_R(t)}{dt} = \eta\zeta\rho_E(t) + \alpha\gamma\rho_I(t) \end{cases} \tag{15}$$

with $\rho_S(t) + \rho_E(t) + \rho_I(t) + \rho_R(t) = 1$. We will refer to it as a SEIR model for fake-news spreading. Basically, susceptible agents get exposed at a rate that is proportional to the probability of them interacting with an active spreader (i.e., an infectious individual). Once they are exposed, they wait for an average time $1/\zeta$ and start disseminate the fake news with probability $1 - \eta$. After an average time $1/\gamma$, they stop doing so and are removed permanently from the dynamics with probability α. A schematic depiction of the dynamics is showed in Fig 5, whereas in Fig 6 an example of evolution within a closed population is sketched. For simplicity, we assume $\eta = 0$, $\alpha = 1$ so that the exposed individuals, after the latency period, always start to spread the fake-news and after an average time spreaders are permanently removed from the dynamic.
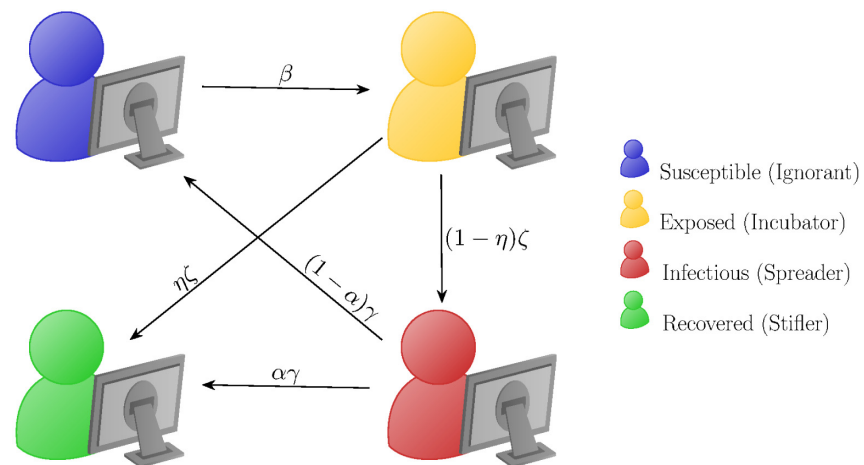
Fig 5. Compartmental dynamic and parameters in the fake news SEIR model (15): People get exposed to fake news with a contact rate β with infected individuals, after a latency period $1/\zeta$ with probability $1 - \eta$ they start spreading it until they finally stop after an average time $1/\gamma$ and become 'immunized' or uninterested in it, thus removing themselves from the dissemination dynamics with probability α.

Susceptible (Ignorant)   Exposed (Incubator)   Infectious (Spreader)   Recovered (Stifler)
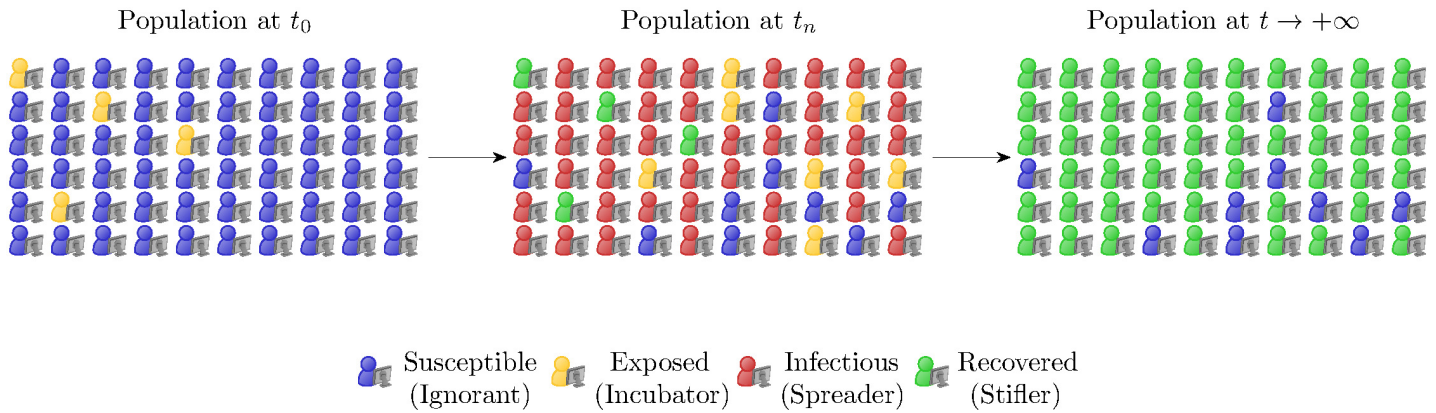
Fig 6. Evolution in time of the dissemination of fake news within a population for model (15). For large times the fake-news infection disappears and the population is composed only by susceptible and recovered individuals.

If we combine the dissemination dynamics with the opinion formation process described in the previous section we obtain the following mean-field model

$$
\begin{aligned}
\frac{\partial f_S(\mathbf{w},t)}{\partial t} &= -K(f_S, f_I)(\mathbf{w},t) + \lambda_S^+ \frac{\partial}{\partial w_+}\left[(w_+ - m_+(t))f_S(\mathbf{w},t)\right] \\
&+ \lambda_S^- \frac{\partial}{\partial w_-}\left[(w_- - m_-(t))f_S(\mathbf{w},t)\right] \\
&+ \frac{\sigma_{+,S}^2}{2}\frac{\partial^2}{\partial w_+^2}\left(D(w_+)^2 f_S(\mathbf{w},t)\right) + \frac{\sigma_{-,S}^2}{2}\frac{\partial^2}{\partial w_-^2}\left(D(w_-)^2 f_S(\mathbf{w},t)\right),
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
\frac{\partial f_E(\mathbf{w},t)}{\partial t} &= K(f_S, f_I)(\mathbf{w},t) - \zeta(\mathbf{w})f_E(\mathbf{w},t) + \lambda_E^+ \frac{\partial}{\partial w_+}\left[(w_+ - m_+(t))f_E(\mathbf{w},t)\right] \\
&+ \lambda_E^- \frac{\partial}{\partial w_-}\left[(w_- - m_-(t))f_E(\mathbf{w},t)\right] + \frac{\sigma_{+,E}^2}{2}\frac{\partial^2}{\partial w_+^2}\left(D(w_+)^2 f_E(\mathbf{w},t)\right) \\
&+ \frac{\sigma_{-,E}^2}{2}\frac{\partial^2}{\partial w_-^2}\left(D(w_-)^2 f_E(\mathbf{w},t)\right),
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\frac{\partial f_I(\mathbf{w},t)}{\partial t} &= \zeta(\mathbf{w})f_E(\mathbf{w},t) - \gamma(\mathbf{w})f_I(\mathbf{w},t) + \lambda_I^+ \frac{\partial}{\partial w_+}\left[(w_+ - m_+(t))f_I(\mathbf{w},t)\right] \\
&+ \lambda_I^- \frac{\partial}{\partial w_-}\left[(w_- - m_-(t))f_I(\mathbf{w},t)\right] + \frac{\sigma_{+,I}^2}{2}\frac{\partial^2}{\partial w_+^2}\left(D(w_+)^2 f_I(\mathbf{w},t)\right) \\
&+ \frac{\sigma_{-,I}^2}{2}\frac{\partial^2}{\partial w_-^2}\left(D(w_-)^2 f_I(\mathbf{w},t)\right),
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
\frac{\partial f_R(\mathbf{w},t)}{\partial t} &= \gamma(\mathbf{w})f_I(\mathbf{w},t) + \lambda_R^+ \frac{\partial}{\partial w_+}\left[(w_+ - m_+(t))f_R(\mathbf{w},t)\right] \\
&+ \lambda_R^- \frac{\partial}{\partial w_-}\left[(w_- - m_-(t))f_R(\mathbf{w},t)\right] \\
&+ \frac{\sigma_{+,R}^2}{2}\frac{\partial^2}{\partial w_+^2}\left(D(w_+)^2 f_R(\mathbf{w},t)\right) + \frac{\sigma_{-,R}^2}{2}\frac{\partial^2}{\partial w_-^2}\left(D(w_-)^2 f_R(\mathbf{w},t)\right),
\end{aligned}
\tag{19}
$$

where

$$m_+ = \sum_{J \in \mathcal{C}} \rho_J m_J^+, \qquad m_- = \sum_{J \in \mathcal{C}} \rho_J m_J^-. \tag{20}$$

The functional

$$K(f_S, f_I)(\mathbf{w}, t) = f_S(\mathbf{w}, t) \int_{[0,1]^2} \kappa(\mathbf{w}*) f_I(\mathbf{w}*, t) \, d\mathbf{w}* \tag{21}$$

is the local incidence rate of interactions between susceptible and infectious individuals, where $\kappa(\mathbf{w})$ is a contact function which measures the impact of the opinion in the dissemination of fake-news. A simplifying assumption is that $\kappa(\cdot, \cdot)$ is separable in the two variables, i.e., $\kappa(\mathbf{w}) = \beta k(w_+)\bar{k}(w_-)$, with $\beta > 0$ a constant. That way, we can think the impact of the strength of positive sentiment acts independently by the one of the strengths of the negative sentiment and we can decouple their evolution. A choice of particular interest would be one in which $\kappa(\cdot)$ is a function of the sole variable w_+ (respectively, w_-).

System (16)–(19) needs to be complemented by the no flux boundary conditions for all $J \in \mathcal{C}$

$$\lambda_J^+\left(w_+ - m_+(t)\right)f_J(\mathbf{w}, t) + \frac{\sigma_{+J}^2}{2}\frac{\partial}{\partial w_+}\left(D(w_+)^2 f_J(\mathbf{w}, t)\right) = 0, \quad \text{on } w_+ = 0, 1,$$
$$\lambda_J^-\left(w_- - m_-(t)\right)f_J(\mathbf{w}, t) +, \frac{\sigma_{-J}^2}{2}\frac{\partial}{\partial w_-}\left(D(w_-)^2 f_J(\mathbf{w}, t)\right) = 0, \quad \text{on } w_- = 0, 1. \tag{22}$$

Note that, if the alignment rates $\lambda_J^\pm = \lambda^\pm$ independent from $J \in \mathcal{C}$, as a consequence of the above boundary conditions and the choice of the diffusion function (10), the quantities $m_+(t)$ and $m_-(t)$ are conserved in time.

4.3 Stationary marginal densities

If in system (16)–(19) we choose a constant function $\kappa(\cdot, \cdot) \equiv \beta > 0$ as well as constant epidemiological parameters we obtain again system (15) with $\alpha = \eta = 1$ by integrating in the variable $\mathbf{w}$. As a consequence, classical results in epidemiology [33] guarantee that when $t \to \infty$ the fake-news spreading vanishes and we have both $\rho_E(t) \to 0$ and $\rho_I(t) \to 0$. Moreover, $\rho_S(t) \to \rho_S^\infty$ and $\rho_R(t) \to \rho_R^\infty = 1 - \rho_S^\infty$ where ρ_S^∞ solves

$$\log\left(\frac{\rho_S}{\rho_S(0)}\right) = \frac{\beta}{\gamma}(1 - \rho_S^\infty). \tag{23}$$

Let us denote with m_+^∞, m_-^∞ the large time behavior Similarly the evolutions of the first

moments are obtained by integrating in **w** after multiplication for $w_\pm$ to obtain

$$
\begin{cases}
\dfrac{d}{dt}(\rho_S(t)m_S^\pm(t)) = -\beta\rho_S(t)m_S^\pm(t)\rho_I(t) - \lambda_S^\pm\rho_S(m_S^\pm - m_\pm) \\[2mm]
\dfrac{d}{dt}(\rho_E(t)m_E^\pm(t)) = \beta\rho_S(t)m_S^\pm(t)\rho_I(t) - \zeta\rho_E(t) - \lambda_E^\pm\rho_E(m_E^\pm - m_\pm) \\[2mm]
\dfrac{d}{dt}(\rho_I(t)m_I^\pm(t)) = \zeta\rho_E(t)m_E^\pm(t) - \gamma\rho_I(t)m_I^\pm(t) - \lambda_I^\pm\rho_I(m_I^\pm - m_\pm) \\[2mm]
\dfrac{d}{dt}(\rho_R(t)m_R^\pm(t)) = \gamma\rho_I(t)m_I^\pm(t) - \lambda_R^\pm\rho_R(m_R^\pm - m_\pm).
\end{cases}
\tag{24}
$$

So that for large times we have $m_\pm(t) \to m_\pm^\infty$ and $m_S^\pm = m_R^\pm = m_\pm^\infty$.

Then, if we integrate (16)–(19) with respect to w_- in the same way we did for equation (6), using the boundary conditions (22) we have that at the stationary state the marginals for the positive opinion satisfy

$$
\frac{\partial}{\partial w_+}\left[(\lambda_S^+ w_+ - m_+^\infty)g_S^\infty(w_+)\right] + \frac{\sigma_{+,S}^2}{2}\frac{\partial^2}{\partial w_+^2}\left[w_+(1-w_+)g_S^\infty(w_+)\right] = 0,
$$
$$
\frac{\partial}{\partial w_+}\left[(\lambda_R^+ w_+ - m_+^\infty)g_R^\infty(w_+)\right] + \frac{\sigma_{+,R}^2}{2}\frac{\partial^2}{\partial w_+^2}\left[w_+(1-w_+)g_R^\infty(w_+)\right] = 0,
$$

which provide the stationary distributions

$$
\begin{aligned}
g_S^\infty(w_+) &= \rho_S^\infty C_S^+ w_+^{m_+^\infty/\mu_S^+ - 1}(1-w_+)^{(1-m_+^\infty)/\mu_S^+ - 1}, \\
g_R^\infty(w_+) &= (1-\rho_S^\infty)C_R^+ w_+^{m_+^\infty/\mu_R^+ - 1}(1-w_+)^{(1-m_+^\infty)/\mu_R^+ - 1}
\end{aligned}
\tag{25}
$$

where $\mu_S^+ = \lambda_S^+/\sigma_{+,S}^2$, $\mu_R^+ = \lambda_R^+/\sigma_{+,R}^2$ and C_S^+, C_R^+ are normalization constants.

Introducing analogous hypotheses, we obtain the same result for the total marginal density of negative opinions

$$
\begin{aligned}
h_S^\infty(w_-) &= \rho_S^\infty C_S^- w_-^{m_-^\infty/\mu_S^- - 1}(1-w_-)^{(1-m_-^\infty)/\mu_S^- - 1}, \\
h_R^\infty(w_-) &= (1-\rho_S^\infty)C_R^- w_-^{m_-^\infty/\mu_R^- - 1}(1-w_-)^{(1-m_+^\infty)/\mu_R^- - 1}
\end{aligned}
\tag{26}
$$

with $\mu_S^- = \lambda_S^-/\sigma_{-,S}^2$, $\mu_R^- = \lambda_R^-/\sigma_{-,R}^2$ and C_S^-, C_R^- normalization constants.

This means that, depending on the given regime of parameters, since the total marginal density for the positive or negative opinions is the mixture of two Beta distributions, it can be unimodal or bimodal (see Fig 7).

5 Application to Covid-19 vaccination hesitancy

As we mentioned in the introduction, the proposed model is suitable to describe the evolution of opinions within closed online communities where data are collected through a suitable use of NLP techniques like sentiment analysis. Here, we present the particular case of selected groups of people which shared preoccupation for the Italian vaccination campaign in response to the SARS-CoV-2 pandemic [63]. More in details, they all share, to a certain extent, *vaccination hesitancy*.
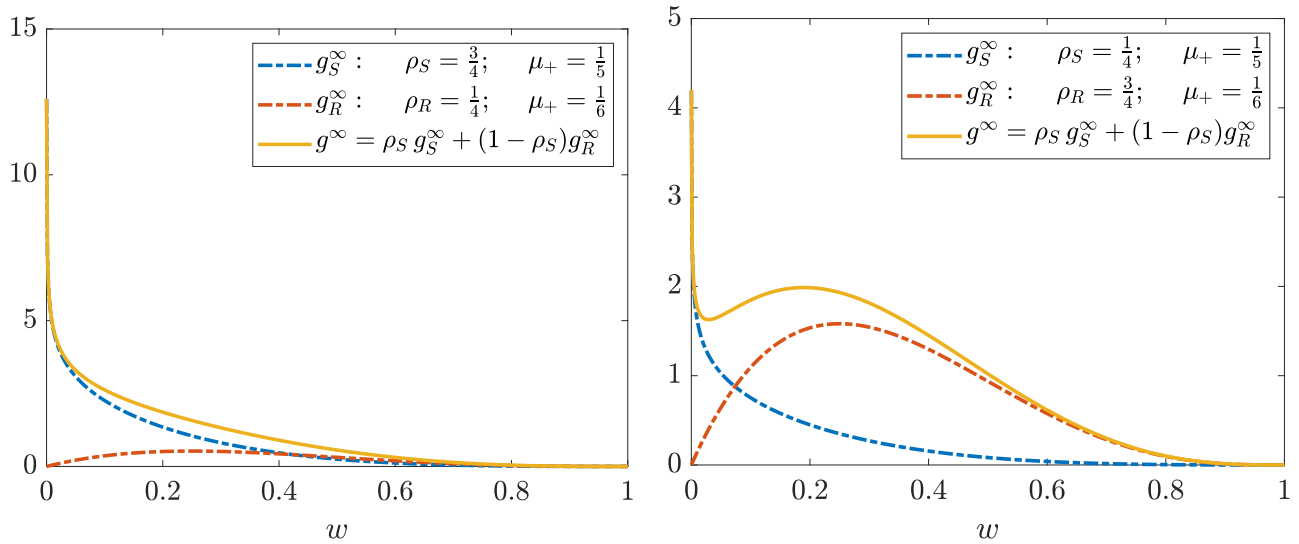
Fig 7. Examples of stationary solutions obtained from (25) with different choices of parameters ρ_S, ρ_R, m_+, λ_S^+, λ_R^+, $\sigma_{+,S}$ and $\sigma_{-,S}$. We can see that the resulting function can have both a unimodal character (left) or a bimodal behavior (right).

5.1 Data collection

We comment here about the data records that we outlined in the Methods' Section. All chat posts focus on vaccination, from various perspectives and scopes, which refer to different social groups the users belong to.

References to conspiracy theories, plain misinformation, rage bursts and mockery, all mix with sincere pleas of explanations on vaccination as well as other measures employed by the former Italian government to combat and contain the effect of the SARS-CoV-2 pandemic on both the national healthcare and economy.

Using techniques proper of the framework of *sentiment analysis* [40–44], we assigned to each post a pair $(\bar{w}_+, \bar{w}_-) \in [0, 1]^2$ of scores, respectively positive and negative which reflect how good or bad the opinion of the user might be in that instant of time.

Scores are assigned in an automatic fashion by SentIta [44] a pre-trained model which analyzes the content of each post. SentIta is a Bidirectional LSTM-Convolutional Neural Network with two output signals ranging between 0 and 1, which performs sentiment analysis on Italian texts (no translation was required). The model is released as an open-source Python library (https://nicgian.github.io/Sentita/, last accessed on June 1, 2023), which we directly applied to our dataset. The model has been trained and tested on Sentipolc2016 [64] and ABSITA2018 [65] datasets for a total of 15,000 positively and negatively labelled Italian sentences plus 90,000 Wikipedia sentences automatically labelled as neutral. The following is a post from our dataset which was evaluated as a score pair of (0.055934787, 0.820981), with the English translation aside.

Ma tanto anche se a queste persone diciamo non fatevi più altre dosi che vi fanno male, non ci ascoltano. Ormai per loro siamo noi i cattivi, e non quelli che veramente sono i cattivi. Ormai non hanno in mente altro. Pensano solo alla prossima dose e ai cattivi no vax.

Still even if we tell this people don't take any more doses, that they hurt you, they won't listen. At this point to them it's us the bad guys and not the ones that are the bad guys for real. By now they have nothing else in mind. They just think about the next dose and to the bad no-vaxs.

The following one instead scored a pair (0.31506833, 0.69410014).

Io di Paragone non mi fido. . . cmq io la mia battaglia la faccio qui. E siamo davvero na marea. Non so come e in Italia ma qui pian piano la gente si sta svegliando, anche i vaccinati si stanno unendo a noi.

I don't trust Paragone [former Italian politician, authors' note]. . . however I fight my battle here. And we really are a ton. I don't know how and in Italy but here slowly but surely people are waking up, even the vaccinated are joining us.

As a last example, we report one of the few posts that were originally written in English (score of (0.028655171, 0.045119375)).

truth revealed: FAUCI just confess on a live stream with Mark Zuckerberg that vaccination actually may cause the problem.

The last post is revealing of some of the issues involved in using software-based sentiment analysis techniques: the form may be neutral and plain, but its content arguably is. Also the small caps text is typical of the sensationalist tones affine with conspiracy theories and fake news in general.
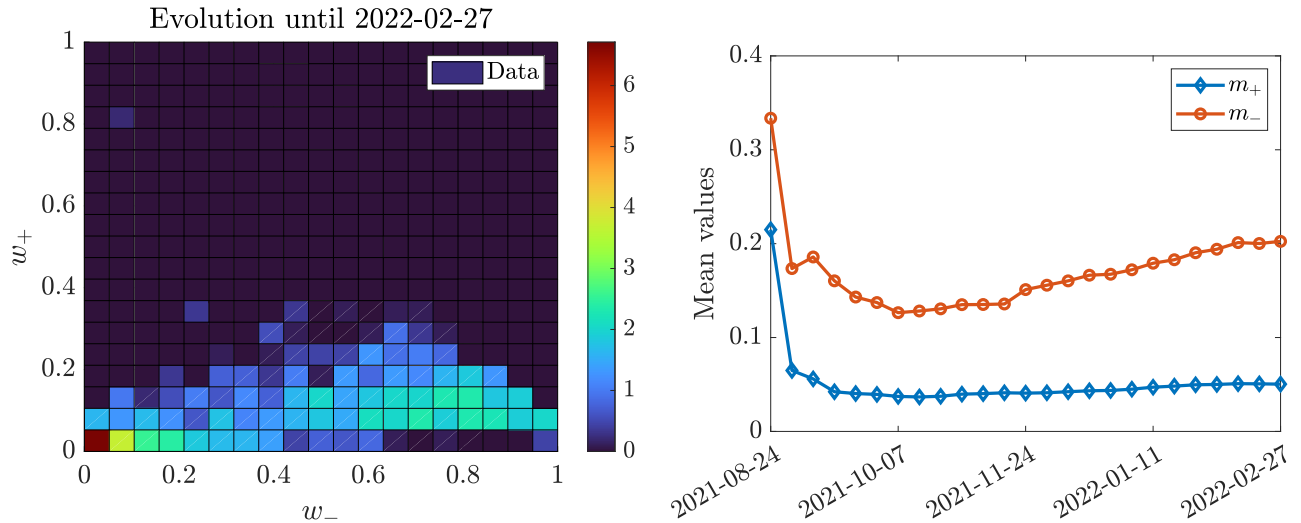
In spite of sentiment analysis having become spread both in academic and corporate works [43, 66, 67], its evaluation is not free from risk: since NLP is a relatively young discipline which faces lots of challenging tasks, there are no current one-solves-all approaches for parsing human-produced syntax in a robust way. Besides, online chats might not be the most suited environment for unambiguous, error-free communication, not to mention the use of non-verbal means, such as non-plain-text characters (emojis, for instance) to express emotions and concepts which necessarily would go undetected by a not instructed software. Hence, the evaluation of the records in our datasets comes with inherent uncertainties. Here we do not try to quantify these uncertainties, we refer to [68] for related approaches to uncertain data in compartmental models.

In the following, when interfacing data with our model, we always considered aggregate data, i.e., scores gathered for posts from every group chat combined into a unique dataset. Moreover, if not otherwise specified, we always discretized the dataset into a grid of 20×20 bins.

The main peculiarity of the dataset is depicted in Fig 8: at the end of the evolution period, a significant concentration of people with strong negative opinion and essentially neutral positive opinion has formed. This is precisely the kind of clustering polarization that we mentioned in previous sections: here the dataset is showing a clear instance of bimodal distribution. Moreover, as the evolution of the positive and negative mean opinions shows, this bimodal distribution is the outcome of a polarizing trend across the population, which, interestingly, involves only the negative opinions. Let us focus on the marginal of the negative opinion at the final time snapshot. We make the *ansatz* that it can be approximated well by a suitable convex combination of two Beta distributions and we also assume that the distribution of data at the last time snapshot is the equilibrium of our dynamics. Then, we leverage our knowledge of the analytical expression for the steady state of the marginals (26) to solve the optimization problem

$$\min_{\substack{m_{-,S}^\infty,\ m_{-,S}^\infty, \\ \lambda_{-,S},\ \lambda_{-,R}, \\ \sigma_{-,S},\ \sigma_{-,R}, \\ \rho_S^\infty}} \left\| \rho_S^\infty h_S^\infty\left(w_-; m_{-,S}^\infty, \frac{\lambda_{-,S}}{\sigma_{-,S}^2}\right) + (1-\rho_S^\infty)h_R^\infty\left(w_-; m_{-,R}^\infty, \frac{\lambda_{-,R}}{\sigma_{-,R}^2}\right) - h^\infty(w_-) \right\|_2, \quad (27)$$

Fig 8. Left: final time snapshot of the dataset. The base 2-logarithm of the data is used for coloring in order to better appreciate the differences in magnitude. Two main concentration regions are clearly visible: the stronger one, around the origin, with the highest peak and the lowest local variance, and the weaker one, in the bottom right region of the surface plot. Right: time evolution of mean positive and negative opinion. While the mean positive opinion quickly converges toward an equilibrium point, the negative one presents an increasing trend after a brief decreasing phase.

https://doi.org/10.1371/journal.pone.0291993.g008

where $h^\infty(\,\cdot\,)$ is the marginal distribution of the negative opinion extrapolated from the last recorded time snapshots, that is, we are fitting our parameters directly on the marginal distributions of the negative sentiments, which is the one with most features in our dataset. Problem (27) is solved via standard nonlinear least squares algorithm; we report in Table 1 the results we obtained from the fitting procedure and in Fig 9 the plot we obtain.

The positive parameters were then established solving the analogous minimization problem

$$\min_{\substack{m_{+,S}^\infty,\ m_{+,S}^\infty, \\ \lambda_{+,S},\ \lambda_{+,R}, \\ \sigma_{+,S},\ \sigma_{+,R}, \\ \rho_S^\infty}} \left\| \rho_S^\infty g_S^\infty\left(w_-; m_{+,S}^\infty, \frac{\lambda_{+,S}}{\sigma_{+,S}^2}\right) + (1 - \rho_S^\infty)g_R^\infty\left(w_-; m_{+,R}^\infty, \frac{\lambda_{+,R}}{\sigma_{+,R}^2}\right) - g^\infty(w_-) \right\|_2,$$
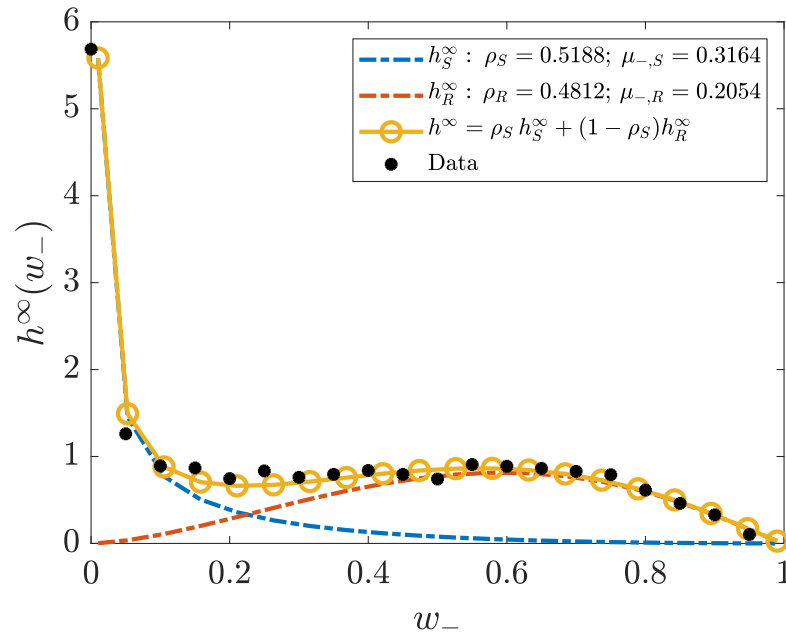
(28)

where this time $g^\infty(\,\cdot\,)$ is the marginal distribution of the positive opinion from our dataset; the result is reported in Table 1.

Finally, once problem (27) is solved, we can also set the parameters responsible for the compartmental dynamics: we fixed $\zeta = \gamma = 1$ to consider the average lifespan of a piece of fake news to be around 24 hours. The initial state we considered was $\rho_S(0) = 0.9$, $\rho_E(0) = 0.05$ and $\rho_I = \rho_R = 0.025$. We chose to have nonzero fractions of mass for the compartments of active spreaders and for the one of removed individuals to take into account exposure to fake news previous to the group-chat-related dynamics. Then, we used relation (23) along with

Table 1. Fitting parameters obtained as solution of minimization problems (27) and (28).

Opinion-related parameters						
ρ_S	$m_{-,S}^\infty$	$\lambda_{-,S}$	$\sigma_{-,S}$	$m_{-,R}^\infty$	$\lambda_{-,R}$	$\sigma_{-,R}$
0.5188	0.0793	0.0475	0.3871	0.5574	0.0063	0.1756
	$m_{+,S}^\infty$	$\lambda_{+,S}$	$\sigma_{+,S}$	$m_{+,R}^\infty$	$\lambda_{+,R}$	$\sigma_{+,R}$
	0.0400	0.0412	0.4658	0.1375	0.0126	0.3228

https://doi.org/10.1371/journal.pone.0291993.t001

Fig 9. Fitting of equilibrium data: marginal of the negative opinion with a convex combination of the marginal distribution for susceptible and recovered individuals, respectively. The fitting parameters are shown and for ease of retrieval are also reported in Table 1.

https://doi.org/10.1371/journal.pone.0291993.g009

our estimate of ρ_S^∞ in Table 1 to determine the value of the basic reproduction number and therefore of $\beta = 1.21$.

5.2 Simulation results

The next section is devoted to compare the evolution provided by the model (16)–(19) and the one provided by the data. The model has been calibrated with the parameters in Table 1. For what concerns the epidemiological coefficients, they were chosen in order to achieve total masses at equilibrium that were compatible with the quantity ρ_S in Table 1, which gives the mass fractions of the two Beta distributions that concur to provide the steady state for the marginal negative density.

Computing the numerical evolution of the model requires a careful discretization of the system in order to keep high accuracy when describing the stationary solutions. To this aim we adopt the steady state preserving approach devised for Fokker-Planck equations in [69] by extending it to systems in the multidimensional case.

To this aim, system (16)–(19) was split both in time and opinion space. To describe the splitting, let us rewrite the system as follows

$$\partial_t \mathbf{f}(\mathbf{w},t) = \mathbf{F}_+[\mathbf{f}(\mathbf{w},t)](w_+,t) + \mathbf{F}_-[\mathbf{f}(\mathbf{w},t)](w_-,t) + \mathbf{P}[\mathbf{f}(\mathbf{w},t)](\mathbf{w},t),$$

where the bold operators are vector valued such as

$$\mathbf{F}_+[\mathbf{f}(\mathbf{w},t)](w_+,t) = \left[\frac{\partial}{\partial w_+}[(\lambda_J^+ w_+ - m_+(t))f_J(\mathbf{w},t)] + \frac{\sigma_{+,J}^2}{2}\frac{\partial^2}{\partial w_+^2}\left(D(w_+)^2 f_J(\mathbf{w},t)\right) \right]_J, J \in \mathcal{C}$$

$$\mathbf{F}_-[\mathbf{f}(\mathbf{w},t)](w_-,t) = \left[\frac{\partial}{\partial w_-}[(\lambda_J^- w_- - m_-(t))f_J(\mathbf{w},t)] + \frac{\sigma_{-,J}^2}{2}\frac{\partial^2}{\partial w_-^2}\left(D(w_-)^2 f_J(\mathbf{w},t)\right) \right]_J, J \in C$$

$$\mathbf{P}[\mathbf{f}(\mathbf{w},t)](\mathbf{w},t) = [-K(f_S,f_I); K(f_S,f_I) - \zeta f_E; \zeta f_E - \gamma f_I; \gamma f_I](\mathbf{w},t).$$

Then, if we discretize the time domain with a time step of size $\Delta t > 0$ and we denote by $\mathbf{f}^n(\mathbf{w})$ an approximation of $\mathbf{f}(\mathbf{w}, n\Delta t)$, the (first-order) time splitting method consists in solving in the time interval $[0, \Delta t]$ the following sequence of problems

$$\text{Evolve positive opinions} \;\Rightarrow\; \begin{cases} \dfrac{\partial \mathbf{f}^\dagger}{\partial t} &= \mathbf{F}_+[\mathbf{f}^\dagger], \\[2mm] \mathbf{f}^\dagger(\mathbf{w}, 0) &= \mathbf{f}^n(\mathbf{w}), \end{cases} \tag{29}$$

$$\text{Evolve negative opinions} \;\Rightarrow\; \begin{cases} \dfrac{\partial \mathbf{f}^{\dagger\dagger}}{\partial t} &= \mathbf{F}_-[\mathbf{f}^{\dagger\dagger}], \\[2mm] \mathbf{f}^{\dagger\dagger}(\mathbf{w}, 0) &= \mathbf{f}^\dagger(\mathbf{w}, \Delta t), \end{cases} \tag{30}$$

$$\text{Evolve fake} - \text{news spreading} \;\Rightarrow\; \begin{cases} \dfrac{\partial \mathbf{f}^{\dagger\dagger\dagger}}{\partial t} &= \mathbf{P}[\mathbf{f}^{\dagger\dagger\dagger}], \\[2mm] \mathbf{f}^{\dagger\dagger\dagger}(\mathbf{w}, 0) &= \mathbf{f}^{\dagger\dagger}(\mathbf{w}, 2\Delta t), \end{cases} \tag{31}$$

and finally set $\mathbf{f}^{n+1}(\mathbf{w}) = \mathbf{f}^{\dagger\dagger\dagger}(\mathbf{w}, \Delta t)$. Higher order splitting can be constructed as well (see [69] and the references therein).

Each one-dimensional opinion direction was discretized using the structure-preserving scheme in [69] that for completeness is outlined below.

We want to solve a nonlinear Fokker-Planck equation of the form

$$\begin{cases} \partial_t f(w, t) &= \partial_w \mathcal{F}[f](w, t) \\ f(w, 0) &= f_0(w), \end{cases}$$

where $t \geq 0$, $w \in \Omega \subseteq \mathbb{R}$ and $\mathcal{F}[\,\cdot\,]$ is the so-called flux

$$\mathcal{F}[f](w, t) = (\mathcal{B}[f](w, t) + \partial_w D(w))f(w, t) + D(w)\partial_w f(w, t),$$

with $\mathcal{B}[\,\cdot\,]$ is a bounded operator describing aggregation dynamics and $D(\cdot)$ is a diffusion function. Now we introduce a uniform mesh $w_i \in \Omega$, so that $w_{i+1} - w_i = \Delta w$. We use the subscript notation $\mathcal{F}_{i\pm 1/2}(t)$ to denote an approximation of the fluxes $\mathcal{F}[f](w_i \pm \Delta w/2, t)$. Let us consider the discretization

$$\frac{d}{dt} f_i(t) = \frac{\mathcal{F}_{i+1/2}(t) - \mathcal{F}_{i-1/2}(t)}{\Delta w},$$

where $f_i(t)$ approximates $f(w_i, t)$. The numerical flux $\mathcal{F}_{i\pm 1/2}$ is calculated as

$$\begin{cases} \mathcal{F}_{i+1/2} &= \tilde{\mathcal{C}}_{i+1/2} \tilde{f}_{i+1/2} + D_{i+1/2} \dfrac{f_{i+1/2} - f_i}{\Delta w} \\[3mm] \mathcal{F}_{i-1/2} &= \tilde{\mathcal{C}}_{i-1/2} \tilde{f}_{i-1/2} + D_{i-1/2} \dfrac{f_i - f_{i-1/2}}{\Delta w}, \end{cases}$$

with the notation $D_{i\pm 1/2} = D(w_{i\pm 1/2})$. Finally (omitting for simplicity the terms with subscript
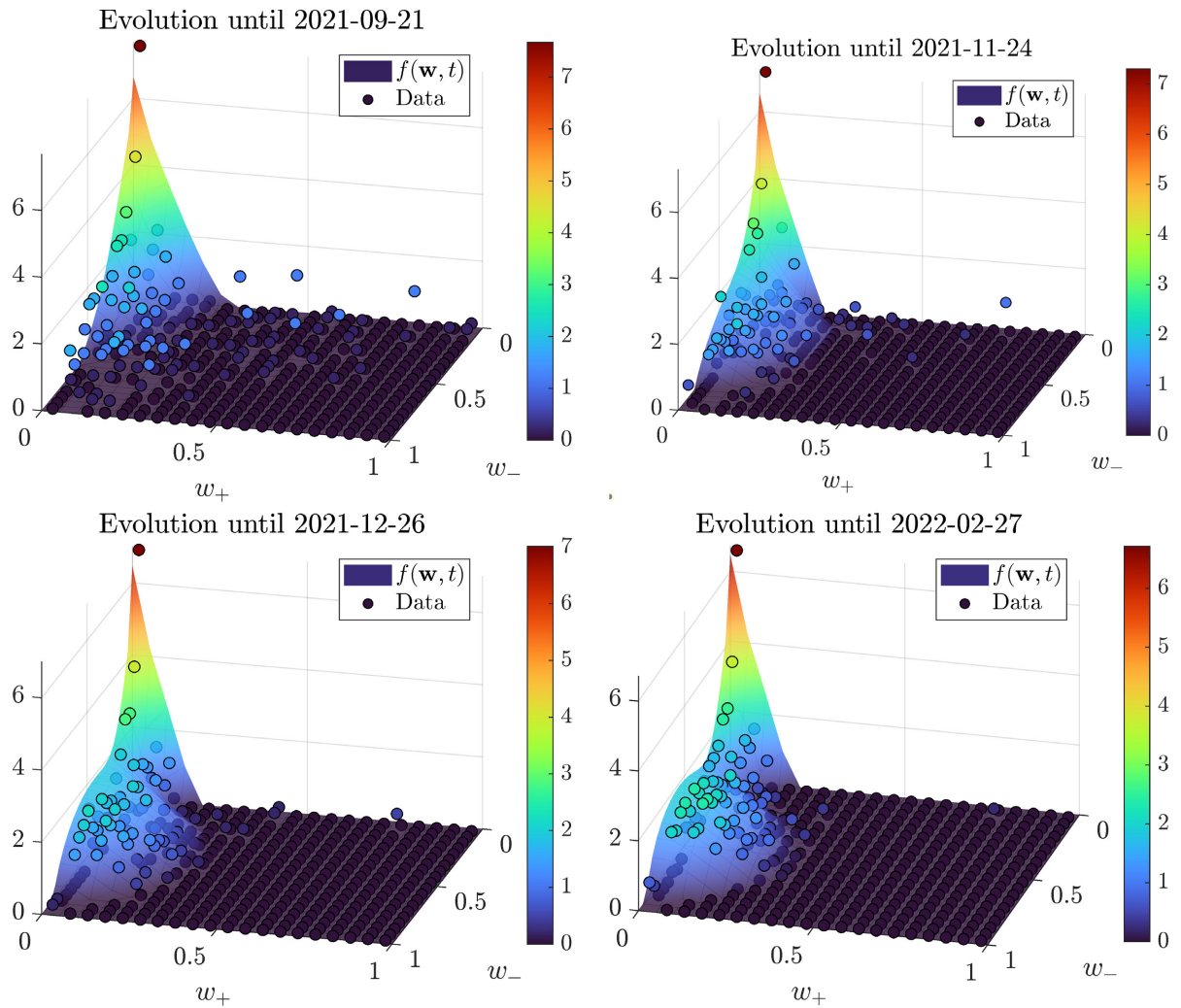
Fig 10. 3D snapshots of dataset time series and model evolution. Again, base 2-logarithm of values is used for coloring. The warmer colors in the bottom left quadrant show the increasing polarization effect around negative opinions.

$i-1/2$, which are of analogous writing), we define

$$\tilde{f}_{i+1/2} = (1 - \delta_{i+1/2})f_{i+1} + \delta_{i+1/2}f_i,$$

$$\delta_{i+1/2} = \frac{1}{\lambda_{i+1/2}} + \frac{1}{1 - \exp(\lambda_{i+1/2})},$$

$$\lambda_{i+1/2} = \frac{\Delta w \tilde{\mathcal{C}}_{i+1/2}}{D_{i+1/2}},$$

$$\tilde{\mathcal{C}}_{i+1/2} = \frac{D_{i+1/2}}{\Delta w} \int_{w_i}^{w_{i+1}} \frac{\mathcal{B}[f](w,t) + \partial_w D(w)}{D(w)} \, dw.$$

In the numerical simulations we used a coarse mesh of 20 points, whereas the time integration of the opinion evolution was computed with a semi-implicit scheme where the time step Δt
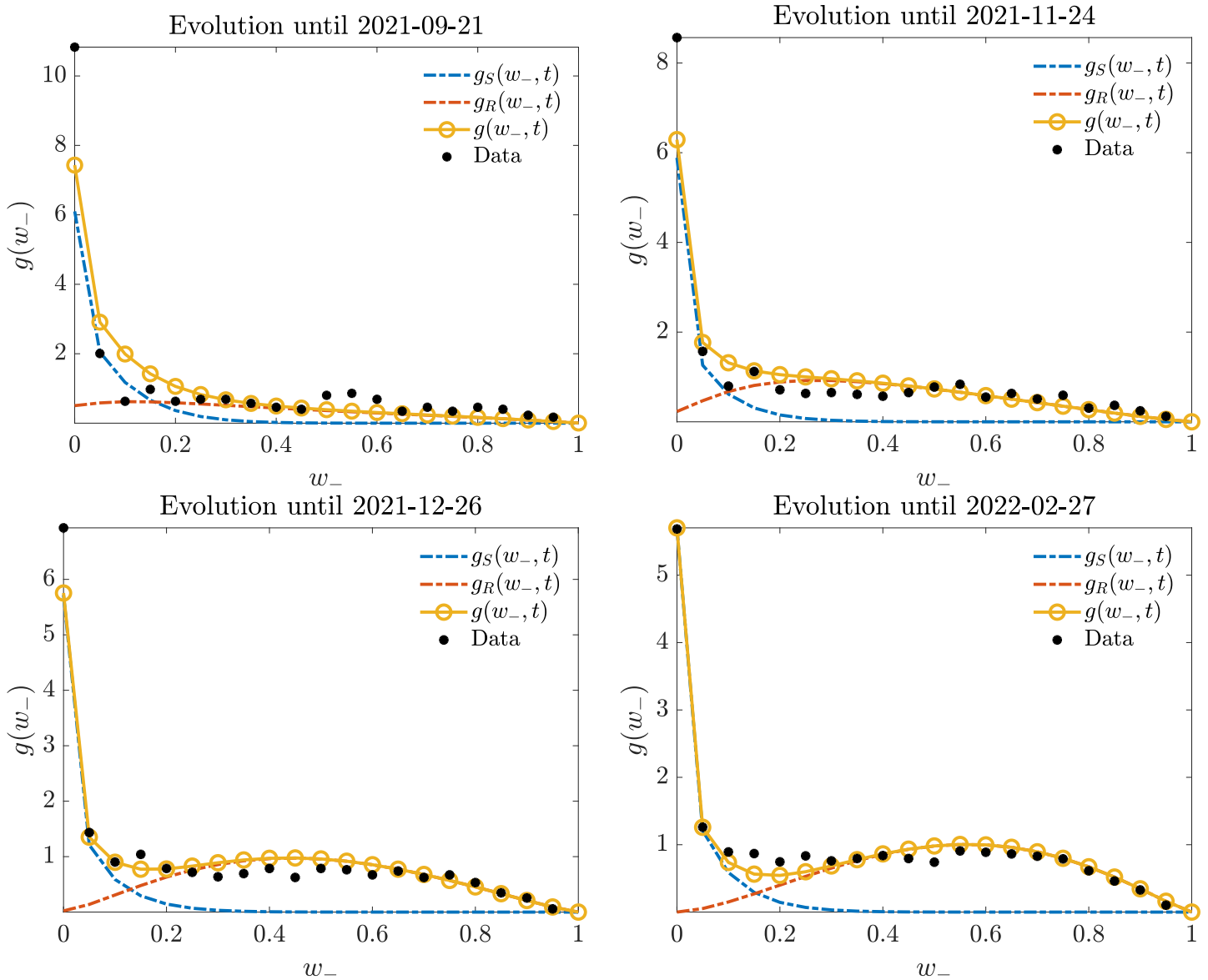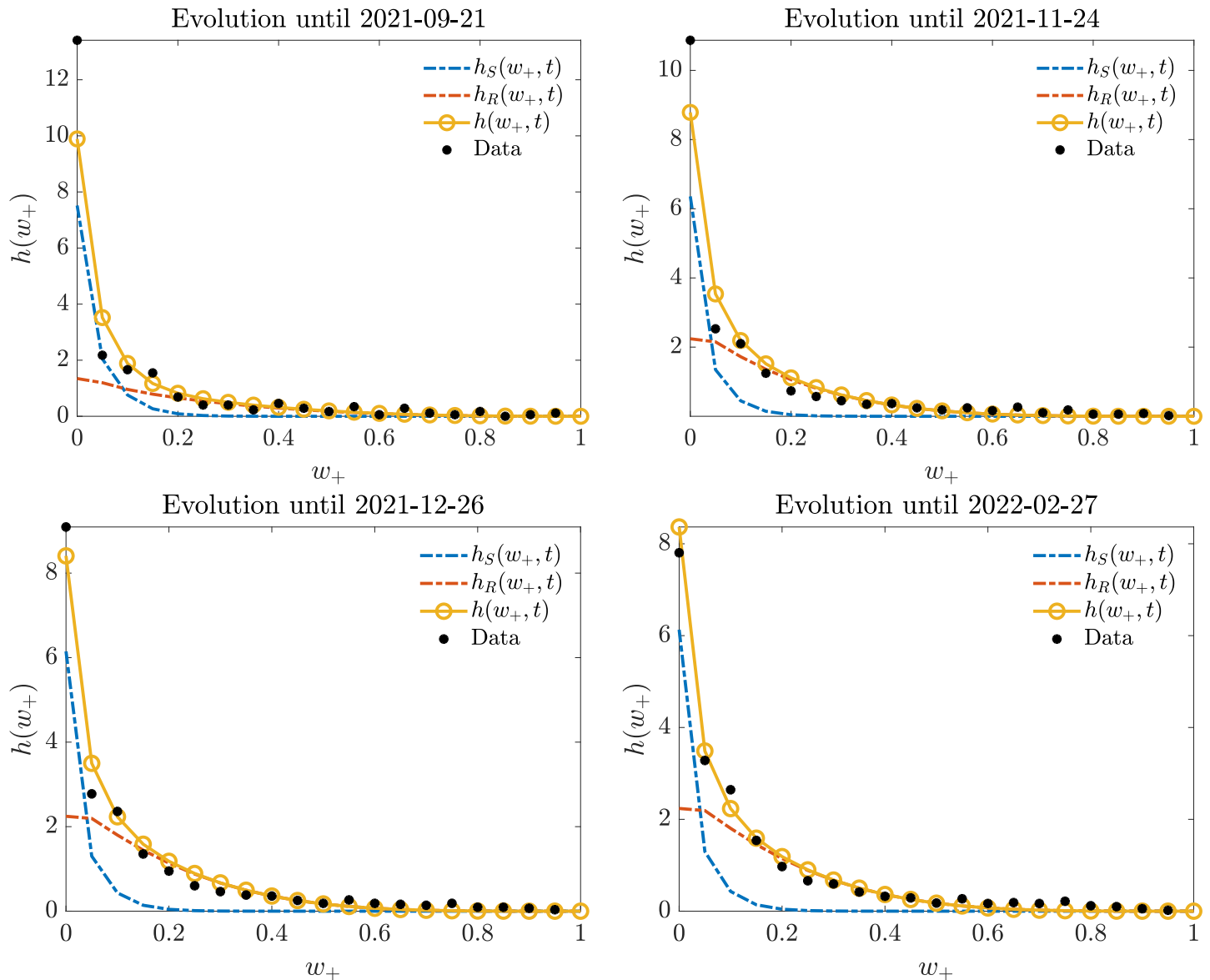
Fig 11. Evolution of the marginal density for the negative opinion: It is clear the emergence of a peak around the value of $w_- = 0.6$, i.e., a polarizing effect toward negative sentiments.

was chosen to be $O(\Delta w_+)$, with $\Delta w_+ = \Delta w_-$ being the steps for the 2D opinion domain, whereas the epidemic exchange portion was integrated through a simple explicit method.

In Fig 10 we report the comparison. Here, we can see that concentration around the origin starts very early, followed by a portion of data-points gathering towards a neighborhood of $(0.25, 0.6)$ to form a peak later on. This polarization trend concerns only the negative opinions, as testified by the evolution of the marginal showed in Fig 11; while the mixture of equilibria for the positive marginal keeps substantially the same profile of unimodal decrease (Fig 12). Overall, we can see that the model is capable of correctly identifying the formation and evolution of both unimodal and bimodal trends happening at the same time in the two-dimensional

Fig 12. Evolution of the marginal density for the positive opinion: Even if the total distribution appears to be unimodal, it is still the sum of two distinct profiles.

evolution. To sum up, the model can accurately predict the polarization process towards negative extreme shown in our dataset.

6 Conclusion

Mathematical modeling of fake-news spreading is a particularly timely and challenging topic, involving numerous areas of research with strong social impacts. In this paper we focused on opinion formation processes within closed communities in presence of spreaders of misinformation. Inspired by a real case study from social data using NLP techniques, we presented a data-driven model based on vector stochastic differential equations. Then, in order to analyze the model and compute analytically the stationary solutions for its spatial marginals, we

considered its mean-field approximation in the form of a system of Fokker-Planck equations, where the dissemination of fake news was carried on through a compartmental approach. Finally, we compared the evolution of the model computed numerically with the one of the dataset time series extract using sentiment analysis. Our results show a good agreement between them, allowing us to observe the formation of bimodal distributions indicating the polarization of opinions toward very negative sentiments as manifested in the real data. We emphasize that the present model, due to its generality, naturally lends itself to many other areas of application in relation to the analysis of fake-news dissemination using NLP techniques in different contexts.

Supporting information

S1 Data.
(CSV)

Acknowledgments

This work has been written within the activities of GNFM and GNCS groups of INdAM (National Institute of High Mathematics).

Author Contributions

Conceptualization: Lorenzo Pareschi.

Data curation: Elena Bellodi, Marco Gavanelli.

Formal analysis: Lorenzo Pareschi.

Funding acquisition: Marco Bresadola.

Investigation: Jonathan Franceschi, Lorenzo Pareschi.

Methodology: Lorenzo Pareschi, Marco Bresadola.

Project administration: Lorenzo Pareschi, Marco Bresadola.

Software: Jonathan Franceschi, Elena Bellodi, Marco Gavanelli.

Supervision: Lorenzo Pareschi, Marco Bresadola.

Validation: Lorenzo Pareschi, Elena Bellodi, Marco Gavanelli, Marco Bresadola.

Visualization: Jonathan Franceschi.

Writing – original draft: Jonathan Franceschi.

Writing – review & editing: Lorenzo Pareschi, Elena Bellodi, Marco Gavanelli, Marco Bresadola.

References

1. World Health Organization. Reported cases of vaccine-preventable diseases (VPDs) globally; 2022. https://immunizationdata.who.int/.

2. Centers for Disease Control and Prevention. Measles Cases and Outbreaks; 2022. https://www.cdc.gov/measles/cases-outbreaks.html.

3. Ren J, Wagner AL, Zheng A, Sun X, Boulton ML, Huang Z, et al. The demographics of vaccine hesitancy in Shanghai, China. PLOS ONE. 2018; 13(12). https://doi.org/10.1371/journal.pone.0209117 PMID: 30543712

4. Roy DN, Biswas M, Islam E, Azam MS. Potential factors influencing COVID-19 vaccine acceptance and hesitancy: A systematic review. PLOS ONE. 2022. https://doi.org/10.1371/journal.pone.0265496 PMID: 35320309

5. Müller J, Tellier A, Kurschilgen M. Echo chambers and opinion dynamics explain the occurrence of vaccination hesitancy. Royal Society Open Science. 2022; 9(10):220367. https://doi.org/10.1098/rsos.220367 PMID: 36312563

6. Salathé M, Bonhoeffer S. The effect of opinion clustering on disease outbreaks. Journal of The Royal Society Interface. 2008; 5(29):1505–1508. https://doi.org/10.1098/rsif.2008.0271 PMID: 18713723

7. Dror AA, Eisenbach N, Taiber Sea. Vaccine hesitancy: the next challenge in the fight against COVID-19. European Journal of Epidemiology. 2020; 35:775–779. https://doi.org/10.1007/s10654-020-00671-y PMID: 32785815

8. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine. 2018; 36(25):3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040 PMID: 29773322

9. Cascini F, Pantovic A, Al-Ajlouni YA, Failla G, Puleo V, Melnyk A, et al. Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature. eClinicalMedicine. 2022; 48:101454. https://doi.org/10.1016/j.eclinm.2022.101454 PMID: 35611343

10. Alvarez-Zuzek LG, La Rocca CE, Iglesias JR, Braunstein LA. Epidemic spreading in multiplex networks influenced by opinion exchanges on vaccination. PloS one. 2017; 12(11). https://doi.org/10.1371/journal.pone.0186492 PMID: 29121056

11. Funk S, Salathé M, Jansen VA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface. 2010; 7(50):1247–1256. https://doi.org/10.1098/rsif.2010.0142 PMID: 20504800

12. Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. Social science & medicine. 2014; 112:1–11. https://doi.org/10.1016/j.socscimed.2014.04.018 PMID: 24788111

13. Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. Public Health. 2021; 194:245–251. https://doi.org/10.1016/j.puhe.2021.02.025 PMID: 33965796

14. Törnberg P, Andersson C, Lindgren K, Banisch S. Modeling the emergence of affective polarization in the social media society. PLOS ONE. 2021; 16(10). https://doi.org/10.1371/journal.pone.0258259 PMID: 34634056

15. Bessi A, Zollo F, Del Vicario M, Puliga M, Scala A, Caldarelli G, et al. Users Polarization on Facebook and Youtube. PLOS ONE. 2016; 11(8). https://doi.org/10.1371/journal.pone.0159641 PMID: 27551783

16. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proceedings of the National Academy of Sciences. 2021; 118(9). https://doi.org/10.1073/pnas.2023301118 PMID: 33622786

17. Fu F, Rosenbloom DI, Wang L, Nowak MA. Imitation dynamics of vaccination behaviour on social networks. Proceedings of the Royal Society B: Biological Sciences. 2011; 278(1702):42–49. https://doi.org/10.1098/rspb.2010.1107 PMID: 20667876

18. FINE PEM, Clarkson JA. Individual versus public priorities in the determination of optimal vaccination policies. American journal of epidemiology. 1986; 124(6):1012–1020. https://doi.org/10.1093/oxfordjournals.aje.a114471 PMID: 3096132

19. Galvani AP, Reluga TC, Chapman GB. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. Proceedings of the National Academy of Sciences. 2007; 104(13):5692–5697. https://doi.org/10.1073/pnas.0606774104

20. Bauch CT, Galvani AP, Earn DJ. Group interest versus self-interest in smallpox vaccination policy. Proceedings of the National Academy of Sciences. 2003; 100(18):10564–10567. https://doi.org/10.1073/pnas.1731324100 PMID: 12920181

21. Pires MA, Oestereich AL, Crokidakis N. Sudden transitions in coupled opinion and epidemic dynamics with vaccination. Journal of Statistical Mechanics: Theory and Experiment. 2018; 2018(5):053407. https://doi.org/10.1088/1742-5468/aabfc6

22. Sznajd-Weron K. Sznajd model and its applications. Acta Physica Polonica B. 2005; 36:2537.

23. Axelrod R. The dissemination of culture: A model with local convergence and global polarization. Journal of conflict resolution. 1997; 41(2):203–226. https://doi.org/10.1177/0022002797041002001

24. Garnier J, Papanicolaou G, Yang TW. Consensus convergence with stochastic effects. Vietnam Journal of Mathematics. 2017; 45(1):51–75. https://doi.org/10.1007/s10013-016-0190-2

25. Noorazar H, Vixie KR, Talebanpour A, Hu Y. From classical to modern opinion dynamics. International Journal of Modern Physics C. 2020; 31(07):2050101. https://doi.org/10.1142/S0129183120501016

26. Albi G, Herty M, Pareschi L. Kinetic description of optimal control problems and applications to opinion consensus. Communications in Mathematical Sciences. 2015; 13(6):1407–1429. https://doi.org/10.4310/CMS.2015.v13.n6.a3

27. Pareschi L, Vellucci P, Zanella M. Kinetic models of collective decision-making in the presence of equality bias. Physica A: Statistical Mechanics and its Applications. 2017; 467:201–217. https://doi.org/10.1016/j.physa.2016.10.003

28. Mastroeni L, Vellucci P, Naldi M. Agent-based models for opinion formation: A bibliographic survey. IEEE Access. 2019; 7:58836–58848. https://doi.org/10.1109/ACCESS.2019.2913787

29. Goddard BD, Gooding B, Short H, Pavliotis G. Noisy bounded confidence models for opinion dynamics: the effect of boundary conditions on phase transitions. IMA Journal of Applied Mathematics. 2022; 87 (1):80–110. https://doi.org/10.1093/imamat/hxab044

30. Toscani G. Kinetic models of opinion formation. Communications in mathematical sciences. 2006; 4 (3):481–496. https://doi.org/10.4310/CMS.2006.v4.n3.a1

31. Pareschi L, Toscani G. Interacting multiagent systems: kinetic equations and Monte Carlo methods. OUP Oxford; 2013.

32. Iacomini E, Vellucci P. Contrarian effect in opinion forming: Insights from Greta Thunberg phenomenon. The Journal of Mathematical Sociology. 2023; 47(2):123–169. https://doi.org/10.1080/0022250X.2021.1981310

33. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. 2000; 42(4):599–653. https://doi.org/10.1137/S0036144500371907

34. Daley DJ, Kendall DG. Epidemics and rumours. Nature. 1964; 204(4963):1118–1118. https://doi.org/10.1038/2041118a0 PMID: 14243408

35. Dimarco G, Pareschi L, Toscani G, Zanella M. Wealth distribution under the spread of infectious diseases. Physical Review E. 2020; 102(2):022303. https://doi.org/10.1103/PhysRevE.102.022303 PMID: 32942503

36. Franceschi J, Pareschi L. Spreading of fake news, competence and learning: kinetic modelling and numerical approximation. Philosophical Transactions of the Royal Society A. 2022; 380 (2224). PMID: 35400178

37. Franceschi J, Pareschi L, Zanella M. From agent-based models to the macroscopic description of fake-news spread: the role of competence in data-driven applications. Partial Differential Equations and Applications. 2022; 3(68).

38. Zanella M. Kinetic models for epidemic dynamics in the presence of opinion polarization. preprint ArXiv:221209130. 2022;.

39. Bernardi E, Pareschi L, Toscani G, Zanella M. Effects of vaccination efficacy on wealth distribution in kinetic epidemic models. Entropy. 2022; 24(2):216. https://doi.org/10.3390/e24020216 PMID: 35205511

40. Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowledge and Information Systems. 2019; 60(2):617–663. https://doi.org/10.1007/s10115-018-1236-4

41. Nemes L, Kiss A. Social media sentiment analysis based on COVID-19. Journal of Information and Telecommunication. 2021; 5(1):1–15. https://doi.org/10.1080/24751839.2020.1790793

42. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned Sentiment Analysis of COVID-19 Vaccine–Related Social Media Data: Comparative Study. Journal of Medical Internet Research. 2022; 24(10):e40408. https://doi.org/10.2196/40408 PMID: 36174192

43. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, et al. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. Vaccines. 2021; 9(1):28. https://doi.org/10.3390/vaccines9010028 PMID: 33430428

44. Nicola G. Bidirectional Attentional LSTM for Aspect Based Sentiment Analysis on Italian. In: EVALITA@CLiC-it; 2018.

45. Albi G, Pareschi L, Toscani G, Zanella M. Recent advances in opinion modeling: control and social influence. In: Active particles. Vol. 1. Advances in theory, models, and applications. Model. Simul. Sci. Eng. Technol. Birkhäuser/Springer, Cham; 2017. p. 49–98.

46. Slanina F, Lavicka H. Analytical results for the Sznajd model of opinion formation. The European Physical Journal B-Condensed Matter and Complex Systems. 2003; 35:279–288. https://doi.org/10.1140/epjb/e2003-00278-0

47. Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents. Advances in Complex Systems. 2000; 3(01n04):87–98. https://doi.org/10.1142/S0219525900000078

48. Sznajd-Weron K, Sznajd J. Opinion evolution in closed community. International Journal of Modern Physics C. 2000; 11(06):1157–1165. https://doi.org/10.1142/S0129183100000936

49. Yin X, Wang H, Yin P, Zhu H. Agent-based opinion formation modeling in social network: A perspective of social psychology. Physica A: Statistical Mechanics and its Applications. 2019; 532:121786. https://doi.org/10.1016/j.physa.2019.121786

50. Aletti G, Naldi G, Toscani G. First-order continuous models of opinion formation. SIAM Journal on Applied Mathematics. 2007; 67(3):837–853. https://doi.org/10.1137/060658679

51. Hegselmann R, Krause U, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. Journal of artificial societies and social simulation. 2002; 5(3).

52. Golse F. The mean-field limit for the dynamics of large particle systems. Journées équations aux dérivées partielles. 2003; p. 1–47.

53. Jabin PE, Wang Z. In: Bellomo N, Degond P, Tadmor E, editors. Mean Field Limit for Stochastic Particle Systems. Cham: Springer International Publishing; 2017. p. 379–402.

54. Kolokoltsov VN. Nonlinear Markov Processes and Kinetic Equations. Cambridge Tracts in Mathematics. Cambridge University Press; 2010.

55. Risken H. Fokker-planck equation. In: The Fokker-Planck Equation. Springer; 1996. p. 63–95.

56. Chiou L, Tucker C. Fake news and advertising on social media: A study of the anti-vaccination movement. National Bureau of Economic Research; 2018.

57. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. Health economics. 2019; 28(11):1377–1382. https://doi.org/10.1002/hec.3937 PMID: 31429153

58. Catalan-Matamoros D, Elías C. Vaccine hesitancy in the age of coronavirus and fake news: analysis of journalistic sources in the Spanish quality press. International Journal of Environmental Research and Public Health. 2020; 17(21):8136. https://doi.org/10.3390/ijerph17218136 PMID: 33158059

59. Waszak PM, Kasprzycka-Waszak W, Kubanek A. The spread of medical fake news in social media–the pilot quantitative study. Health policy and technology. 2018; 7(2):115–118. https://doi.org/10.1016/j.hlpt.2018.03.002

60. Lyu H, Zheng Z, Luo J. Both rates of fake news and fact-based news on twitter negatively Correlate with the state-level COVID-19 vaccine uptake. arXiv preprint arXiv:210607435. 2021;.

61. Montagni I, Ouazzani-Touhami K, Mebarki A, Texier N, Schück S, Tzourio C, et al. Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy. Journal of Public Health. 2021; 43(4):695–702. https://doi.org/10.1093/pubmed/fdab028 PMID: 33693905

62. Waisbord S. Truth is What Happens to News. Journalism Studies. 2018; 19(13):1866–1878. https://doi.org/10.1080/1461670X.2018.1492881

63. Rubin A, Pedroni ML, Fabbri M, Bellodi E, Gavanelli M, Franceschi J, et al. Le argomentazioni degli esitanti italiani sulla dose di richiamo del vaccino anti Covid. Mediascapes Journal. 2022; 20.

64. Barbieri F, Basile V, Croce D, Nissim M, Novielli N, Patti V. Overview of the evalita 2016 sentiment polarity classification task. In: Proceedings of third Italian conference on computational linguistics (CLiC-it 2016) & fifth evaluation campaign of natural language processing and speech tools for Italian. Final Workshop (EVALITA 2016); 2016.

65. Basile P, Croce D, Basile V, Polignano M, et al. Overview of the EVALITA 2018 Aspect-based Sentiment Analysis task (ABSITA). In: EVALITA Evaluation of NLP and Speech Tools for Italian. CEUR; 2018. p. 1–10.

66. Klimiuk K, Czoska A, Biernacka K, Balwicki Ł. Vaccine misinformation on social media–topic-based content and sentiment analysis of Polish vaccine-deniers' comments on Facebook. Human Vaccines & Immunotherapeutics. 2021; 17(7):2026–2035. https://doi.org/10.1080/21645515.2020.1850072 PMID: 33517844

67. Agustiningsih KK, Utami E, Al Fatta H. Sentiment Analysis of COVID-19 Vaccine on Twitter Social Media: Systematic Literature Review. In: 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE; 2021. p. 121–126.

68. Albi G, Pareschi L, Zanella M. Control with uncertain data of socially structured compartmental epidemic models. J Math Bio. 2021; 82(7):63. https://doi.org/10.1007/s00285-021-01617-y PMID: 34023964

69. Pareschi L, Zanella M. Structure preserving schemes for nonlinear Fokker–Planck equations and applications. Journal of Scientific Computing. 2018; 74(3):1575–1600. https://doi.org/10.1007/s10915-017-0510-z