scientific reports

OPEN

Modelling opinion dynamics under the impact of influencer and media strategies

Luzie Helfmann[1,2], Nataša Djurdjevac Conrad[1], Philipp Lorenz-Spreen[3] & Christof Schütte[1,4✉]

Digital communication has made the public discourse considerably more complex, and new actors and strategies have emerged as a result of this seismic shift. Aside from the often-studied interactions among individuals during opinion formation, which have been facilitated on a large scale by social media platforms, the changing role of traditional media and the emerging role of "influencers" are not well understood, and the implications of their engagement strategies arising from the incentive structure of the attention economy even less so. Here we propose a novel framework for opinion dynamics that can accommodate various versions of opinion dynamics as well as account for different roles, namely that of individuals, media and influencers, who change their own opinion positions on different time scales. Numerical simulations of instances of this framework show the importance of their relative influence in creating qualitatively different opinion formation dynamics: with influencers, fragmented but short-lived clusters emerge, which are then counteracted by more stable media positions. The framework allows for mean-field approximations by partial differential equations, which reproduce those dynamics and allow for efficient large-scale simulations when the number of individuals is large. Based on the mean-field approximations, we can study how strategies of influencers to gain more followers can influence the overall opinion distribution. We show that moving towards extreme positions can be a beneficial strategy for influencers to gain followers. Finally, our framework allows us to demonstrate that optimal control strategies allow other influencers or media to counteract such attempts and prevent further fragmentation of the opinion landscape. Our modelling framework contributes to a more flexible modelling approach in opinion dynamics and a better understanding of the different roles and strategies in the increasingly complex information ecosystem.

In the era of many-to-many communication on social media, polarization and radicalism can increasingly be understood as self-organized phenomena emerging from opinion dynamics on social networks[1]. Despite the global use of those large communication platforms and the resulting interactions, they have not been thoroughly studied as effects of a dynamical system consisting of multiple interacting components, particularly distinguishing their different roles and goals.

More specifically, social media did not replace traditional media, but became an intermediary entity that allows peer-to-peer communication among individuals but also traditional media outlets to disseminate their content and interact with their audience. In principle, they are treated like individuals on those platforms, but have very different internal dynamics (e.g. editorial processes, reputation and economic incentives). As more and more people around the world access traditional news media through social platforms, this dependency becomes increasingly pronounced[2]. At the same time a new role emerged in this system, namely, the *influencer*, who is a private person with many followers on a social media network[3]. Although mostly following commercial goals, they do turn to politics[4] and can polarize public discussions[5], as well as compete with traditional media and with one another for audience[6]. Potentially this increased competition could contribute to the declining trust in traditional media and polarized debates on social media around the world[7]. We aim to capture the scenario of opinion dynamics under the impact of influencers and media with a generalized modeling framework designed

[1]Department Modelling and Simulation of Complex Processes, Zuse Institute Berlin, 14195 Berlin, Germany. [2]Complexity Science, Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany. [3]Center for Adaptive Rationality, Max Planck Institute for Human Development, 14195 Berlin, Germany. [4]Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany. ✉email: schuette@zib.de

1

to accommodate different opinion dynamic mechanisms while distinguishing these different roles and allowing them to interact with each other.

So far in most cases, opinion dynamics are mathematically modelled by network- or agent-based models (ABMs) that aim to reproduce how individuals change their opinions based on the feedback of their peers in the respective social environment. The largest group of such models describes the change of opinions based on the (dynamical) interaction network between individuals. The well-known voter model[8] in which individuals copy the discrete opinion of a random neighbour may serve as an example. Alternative models use continuous opinion spaces where usually individuals' opinions are drawn to attracting opinions in their social network (in some cases repelled by others). The DeGroot model[9] with individuals being drawn to the weighted average opinion of their neighbours and bounded confidence models[10,11] where attracting interactions only take place with like-minded individuals may be the most prominent. It has been demonstrated that these basic models and their generalizations, e.g.[12–14], allow to describe the emergence of opinion clusters or communities, including phenomena like radicalization, social bubbles and echo chambers[15]. Additionally, biased assimilation models[16,17] have been developed based on the insight that linear weighted averaging of opinions cannot really lead to increasing polarization since people who hold strong opinions are likely to examine information in a biased manner. Further in[18–20], multi-dimensional opinion dynamics models have been proposed where individuals are drawn to the neutral opinion while at the same time reinforcing each other to more extreme opinions. Such models can describe real-world political polarization[21].

The development of these opinion dynamics models has been taken up in the literature with a focus on understanding consensus and community building, or cluster formation. For simple linear models like the DeGroot model this can be studied analytically[9]. For nonlinear opinion models like bounded confidence models this has been studied numerically[10,11,22] or by considering the mean-field limit in terms of partial differential equations (PDEs), providing rigorous theorems on cluster numbers and size, effective timescales and bifurcations based on changing interaction parameters[23–25]. Moreover, it is understood how to extend the theory from agent-based models to mean-field limit models with stochastic partial differential equations[26,27].

A few first steps towards opinion models under the impact of different types of agents have been taken: The influence of external sources on the opinion dynamics of individuals has been studied for example in[12,28,29]. These external sources are often regarded as media or charismatic leaders and have also been termed *zealots* or *stubborn agents* in the literature. More recently their influence has also been analysed in network-based opinion models[30,31] but without explicitly taking coherence or multiple timescales into account. In Ref.[32], the influence of a social group or clan at multiple timescales on the opinion formation on complex networks has been analyzed. However, so far those agents are mostly considered to have constant opinions and relations to their followers that do not change in time.

Here, we present a general framework that generalizes most of the available models (DeGroot, bounded confidence, biased assimilation,...) in a continuous opinion space and complements individual opinion dynamics with traditional media and social media influencers that adapt their opinions on different timescales than the individuals, see Fig. 1. This framework does not introduce a new opinion dynamics mechanisms, but offers a framework to understand various mechanisms and their interplay with different roles in the complex information system of today, while allowing for mean-field approximations that can facilitate more efficient computations.

The question of what role the media and influencers play on the public opinion distribution became increasingly complex. In the attention economy, the main goal of influencers and media is to receive more attention and increase their followership[33], that is, to find strategies that achieve this goal in an *optimal* way. This makes
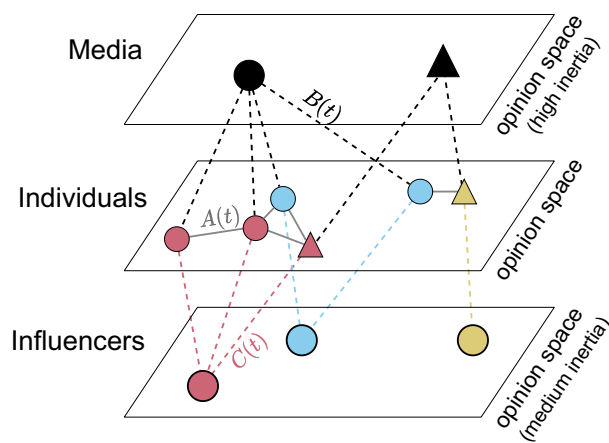
Figure 1. Structure of the agent-based model consisting of individuals, media and influencers. Each agent adapts its opinion in the continuous opinion space by interacting with its network neighbours. The network between individuals is given by $A(t)$ while the relations between individuals and media resp. influencers are defined by $B(t)$ resp. $C(t)$. In the figure, the shape of an individual indicates the medium they read and their color indicates the influencer they follow. Opinion changes happen on different time scales as determined by the inertia parameters.

them themselves dynamic entities that can adjust their positions or agendas in order to achieve those goals[34]. The effects these *optimal strategies* in such an incentive structure could have on the overall opinion dynamics have not yet been fully understood.

Recent attempts to find control strategies of opinion dynamics consist of a variety of seemingly independent approaches. On the one hand, different heuristics have been developed to determine which agents should be targeted by an external controller to maximize their influence[35–37]. On the other hand, in Ref.[38,39], it is studied how *external* controllers have to act to bring the opinion distribution of agents close to a specified target distribution, while, in Ref.[37], it is established how to control opinion distributions by (externally) interfering with the interaction network between individuals, and, in Ref.[40,41], dynamic external controls for mean-field opinion models are considered.

We will approach the design of optimal strategies in the following way: First, we show how to construct the mean-field model related to the proposed opinion dynamics model in the limit of large numbers of individuals and with finitely many influencers and media. This mean-field model takes the form of a system of partial differential equations for the individuals' opinion distribution coupled to stochastic differential equations for the opinions of media and influencers. Then, we will demonstrate how this mean-field model can be used to study the effect of strategies in the attention economy. In particular, we will show how to derive optimal control schemes for counteracting agendas to influence opinion distributions. The advantage of the mean-field model compared to the ABM is that the computational cost is independent of the number of agents, and that the model is deterministic (when assuming that influencers and media are not affected by stochasticity), which, together, means that it is considerably easier to find insight into collective behavior or optimal strategies via the mean-field model.

Opinion model with influencers and media

In the following we start by defining a general opinion model resulting from a large number of *individuals* adapting their opinions through interaction with each other as well as due to the influence of a few specific agents with particular roles, namely traditional *media* and social media *influencers*. See Fig. 1 for the model components. We consider the situation of the early formation of opinions, which is of great importance in the accelerating public discourse[42]. Hence, we focus on the transient model dynamics to study the formation, splitting and merging of opinion clusters, while the asymptotic regime is not of interest here. We assume that in the transient regime, individuals adapt their opinions on a fast time scale, while media and influencer agents change their opinion positions on a significantly slower time scale.

In particular, we consider individuals $i = 1, \ldots, N$ with opinions $x_i(t) \in D$ lying in a continuous d-dimensional opinion space $D \subset \mathbb{R}^d$. We regard the opinion space D to be bounded with no flux boundary conditions. The vector $\boldsymbol{x}(t) = (x_i(t))_{i=1}^N$ summarizes the opinions of all N individuals. In addition, we introduce M highly influential agents that can be considered as media outlets with continuous opinions denoted by the vector $\boldsymbol{y}(t) = (y_m(t))_{m=1}^M \in D^M$, as well as L highly influential influencers with opinions $\boldsymbol{z}(t) = (z_l(t))_{l=1}^L \in D^L$. To describe the real-world situation, we assume that there are much fewer influencers than individuals but still more influencers than media, i.e. $M < L \ll N$. When political opinions are modelled, the opinion space D with $d = 2$ could e.g. span the dimensions economic left $\leftrightarrow$ right and libertarian $\leftrightarrow$ authoritarian.

Relations among individuals such as friendship or connections on social media are defined through a binary adjacency matrix $A(t) \in \{0, 1\}^{N \times N}$ that can depend on time t. The resulting network determines which individuals can interact (when there is an edge between two individuals) and which cannot (when there is no edge). When the edges are additionally weighted, such that $A(t) \in [0, 1]^{N \times N}$, or directed, then the network can also describe the strength of a tie or the direction of social influence as in Reference[9]. Media and influencer agents on the other hand are assumed to only interact with those individuals that are their followers (readers/users). In the binary adjacency matrix $B(t) \in \{0, 1\}^{N \times M}$ of medium-follower relations we store which individual follows which medium at time t while the binary adjacency matrix $C(t) \in \{0, 1\}^{L \times N}$ defines the connections between individuals and influencers at time t. For simplicity, we assume that each individual follows exactly one medium and one influencer, i.e. each row of $B(t)$ and of $C(t)$ contains exactly one entry at all times.

Individuals $i = 1, \ldots, N$ adapt their opinions in time according to the following stochastic differential equation (SDE)

$$dx_i(t) = F_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, t)dt + \sigma \, dW_i(t) \tag{1}$$

where F_i defines the interaction force on individual i, $\sigma > 0$ gives the strength of noise and $W_i(t)$ are i.i.d. $d-$dimensional Wiener processes. The noise can, for example, model unknown external influences, uncertainties in the communication between individuals or free will. The interaction force on individual i is given by a weighted sum of attractive forces from all other connected individuals j scaled by the parameter $a > 0$ as well as attractive forces from the respective media scaled by the parameter $b > 0$ and from the respective influencer scaled by the parameter $c > 0$

$$F_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, t) = \frac{a}{\sum_{j'} w_{ij'}(t)} \sum_{j=1}^N w_{ij}(t)(x_j(t) - x_i(t)) + b \sum_{m=1}^M B_{im}(t) \, (y_m(t) - x_i(t)) + c \sum_{l=1}^L C_{il}(t)(z_l(t) - x_i(t)). \tag{2}$$

The interaction weights between pairs of individuals i, j are given by

$$w_{ij}(t) = A_{ij}(t) \, \phi \big(\| x_j(t) - x_i(t) \| \big), \tag{3}$$

i.e. they depend on the network and on the opinion distance between the pair of individuals. The weights in the force F_i are normalized to ensure a unitary total contribution from all individuals.

Possible choices for the non-negative, pair function $\phi\left(\left\|x_j - x_i\right\|\right)$ are given by:

- $\phi(x) = \exp(-x)$ as in[43] that places exponentially more weight on close-by individuals,
- $\phi(x) = \mathbb{1}_{[0,d]}(x)$ as in bounded confidence models[10–12] that only allows interactions with other individuals that are within a radius d, here $\mathbb{1}_{[0,d]}$ is the indicator function on the set $[0, d]$,
- $\phi(x) = 1$ as in the DeGroot model[9] resulting in interactions irrespective of the opinion distance between individuals.

The first two choices imply that individuals that are already close in opinion space excert higher social influence on each other (*homophily*), while the third choice results in a weight irrespective of the opinions of the interacting individuals.

Remark 1 (Opinion differentiation and higher-order interactions) Apart from the described opinion attraction (assimilation), it is also possible to include opinion repulsion (differentiation) between individuals into the model. In Ref.[44], it is suggested to classify pairs of individuals (i, j) either as friends that attract in their opinions or as enemies that repel in their opinions. These pairs (i, j) will have a positive resp. negative edge weight $A_{ij}(t)$ and thereby can turn the first term of the force (2) from attraction to repulsion. In Ref.[15], another approach that extends bounded confidence models is suggested: pairs of individuals can not only become closer in their opinions when their opinion distance is smaller than the distance d, but they can also repel from each other when they are further apart than the distance $D > d$. By using the pair function $\phi(x) = \mathbb{1}_{[0,d]}(x) - \mathbb{1}_{[D,\infty)}(x)$, this can be incorporated. But when including differentiation, the weights $w_{ij}(t)$ from (3) can become negative and hence can no longer be normalized as before. Instead one can for example normalize the force by the total number of individuals in the system, N, or by the number of neighbours of an individual, thereby ensuring that the interaction strength is independent of the number of individuals in the system.

Additionally, it is possible to describe higher-order interactions among more than two individuals at the same time with this model and thereby to more accurately describe group effects such as peer pressure, see e.g.[27,45].

Not only individuals change their opinions, but also media agents and influencers, however they adapt their opinions on a much slower time scale. The resistance to rapid change is determined by the inertia parameter $\Gamma > 1$ for media agents and by $\gamma > 1$ for influencers. With $\gamma < \Gamma$ media agents are changing their opinions on an even slower time scale than influencers. In the limit when the parameters Γ, γ diverge, the opinions of media and influencers become constant in time. There is a lot of research[12,28,31,46] that studies interactions of individuals with constant agents (also termed *stubborn agents* or *zealots*) but to our knowledge adaptive media and influencers have not been studied so far. In particular here media agents $m = 1, \ldots, M$ slowly adapt their opinions according to the SDE

$$\Gamma dy_m(t) = (\tilde{x}_m(t) - y_m(t))dt + \tilde{\sigma} d\tilde{W}_m(t), \tag{4}$$

such that media agents are drawn in the direction of the average opinion of their followers $\tilde{x}_m(t) = \frac{1}{\sum_k B_{km}(t)} \sum_{i=1}^{N} B_{im}(t)x_i(t)$. In the SDE, $\tilde{\sigma} > 0$ gives the strength of noise on the opinion dynamics and $\tilde{W}_m(t)$ denote i.i.d. d−dimensional Wiener processes. Analogous to media, influencers $l = 1, \ldots, L$ slowly change their opinions in the direction of their average followership according to the SDE

$$\gamma dz_l(t) = (\hat{x}_l(t) - z_l(t))dt + \hat{\sigma} d\hat{W}_l(t), \tag{5}$$

where the average opinion of followers is given by $\hat{x}_l(t) = \frac{1}{\sum_k C_{kl}(t)} \sum_{i=1}^{N} C_{il}(t) x_i(t)$. The noise strength is given by $\hat{\sigma} > 0$, while $\hat{W}_l(t)$ denote i.i.d. d−dimensional Wiener processes.

We have seen that influencers are similar to media agents but are usually more numerous and adapt their opinions on a faster time scale. To reflect that relationships on social media are more dynamic than with traditional media outlets, we further propose an explicit model of how individuals can switch the influencer they are currently following. In particular, each individual i can at any time t switch its current influencer l' to another influencer l with a given rate $\Lambda_m^{\to l}(x, t)$ where m is the medium of individual i and x is the opinion of i. Note that the change rate does not depend on the current influencer of individual i and that an individual can therefore also change to the influencer it is currently following without any effect. The rate could for example take the following form

$$\Lambda_m^{\to l}(x, t) = \eta \, \psi\left(\|z_l - x\|\right) r\left(\frac{n_{m,l}(t)}{\sum_{m'=1}^{M} n_{m',l}(t)}\right) \tag{6}$$

with scaling parameter $\eta > 0$, pair function ψ, the link recommendation function r and $n_{m,l}$ denoting the fraction of individuals that follow medium m and influencer l. By setting the pair function for example to $\psi(x) = \exp(-x)$, an individual has an exponentially higher rate to switch to an influencer that has a similar opinion than to an influencer with a very different opinion, i.e. there is homophily between influencers and individuals when connections are made. On social media platforms, link recommendation algorithms are often used to suggest new

connections to users that have the greatest potential to become established[47,48]. We incorporate link recommendation via the function r by assuming that individuals have a higher chance of switching to an influencer with a structurally similar followership. We measure the structural similarity of the followers of influencer l to individual i by the ratio of followers that are connected to the same influencer (after switching) and medium as i, this proportion is given by $\frac{n_{m,l}}{\sum_{m'} n_{m',l}}$, and we assume that r is an increasing function on $[0, 1]$.

Parameter choice and initial condition

Throughout the paper we will consider one parameter setting. We consider the dynamics for $N = 250$ individuals, $M = 2$ media and $L = 4$ influencers. As an initial condition for $t = 0$, we distribute individuals uniformly in the opinion space and place the media at $y_1(0) = (-1, -1)$ and $y_2(0) = (1, 1)$. To construct the matrix $C(0)$, we assign individuals in each of the 4 quadrants to a different influencer. The initial opinions of influencers are set to the mean opinion of their followers. For simplicity we will choose A and B to be constant on the chosen time scale. We assume an all-to-all connectivity between individuals, i.e. $A_{ij} = 1$ for all (i, j), and choose for each individual i uniformly at random a medium to follow (thereby constructing matrix B). The remaining parameters are chosen as: $a = 1$, $b = 2$, $c = 4$, $\sigma = 0.5$, $\tilde{\sigma} = 0$, $\hat{\sigma} = 0$, $\Gamma = 100$, $\gamma = 10$, $\phi(x) = \psi(x) = \exp(-x)$, $r(x) = \max\{0.1, -1 + 2x\}$ and $\eta = 15$.

Example 1: ABM with media and influencers

In Fig. 2, we show snapshots of one realization of the ABM with the parameter setting and initial conditions just described. This choice of parameters leads to interesting dynamics with clusters being formed, split and merged.

In particular, when we let the model run, the strong attraction force to influencers ($c = 4$) results in individuals quickly being attracted by their respective influencer and forming 4 clusters (compare with $t = 0.1$). After some time, the 4 clusters split further because individuals are also attracted to their medium (compare $t = 0.5$), s.t. individuals now form roughly 8 groups. Some individuals switch the influencer to a more suitable influencer, i.e. one that is closer in opinion space and whose majority of followers are connected to the same medium as the individual. They then get attracted to the new influencer ($t = 1$), until finally ($t = 2$) individuals have formed 2 mixed clusters near the 2 media opinions.

In Fig. 3, we show the evolution of the proportion of individuals that follow a certain influencer and medium. Around $t = 0$, the proportions are roughly the same. But after $t = 0.5$, individuals start switching their influencer (the medium cannot be switched), s.t. quickly the followers of each influencer are dominated by individuals following the same medium. Towards $t = 2$, the proportion of individuals that follow the medium and influencer in the upper right corner (indicated as green triangles) and the proportion of individuals that follow the medium and influencer in the lower left corner (shown by blue circles) dominate.
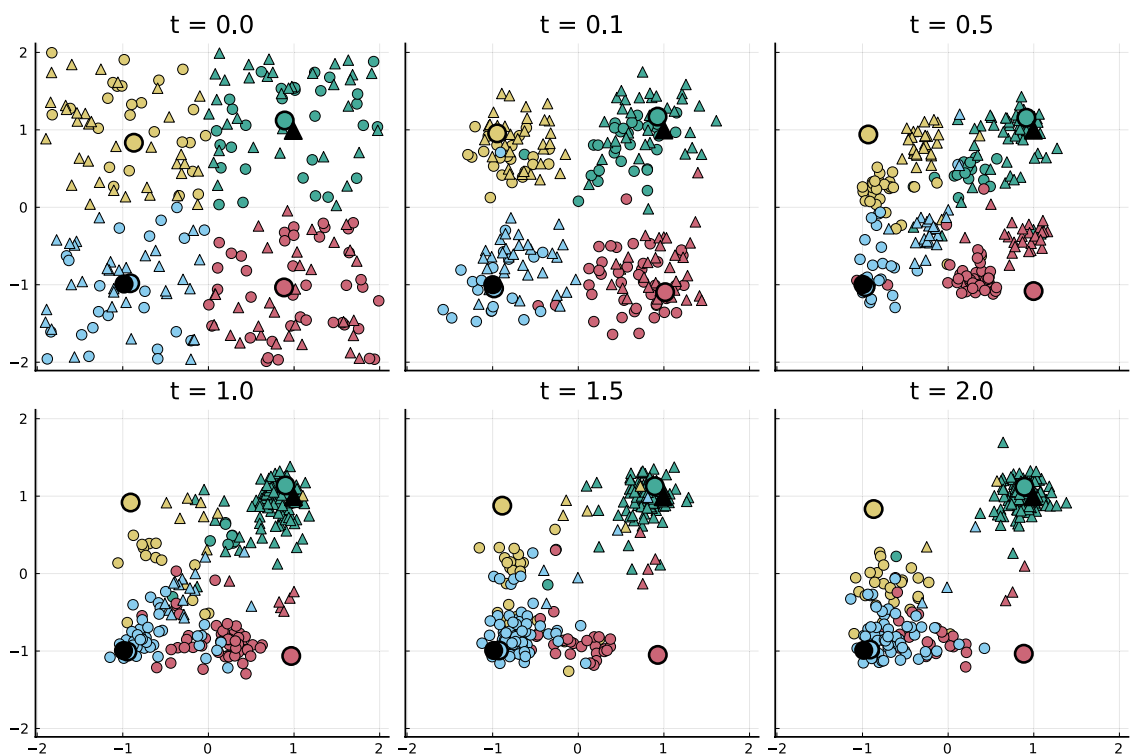
Figure 2. Realization of the ABM with media and influencers (Example 1). The 2 media agents are marked in black, the 4 influencers are indicated by large circles in 4 colours, and the 250 individuals inherit the shape of the medium they read and the color of the influencer they currently follow.
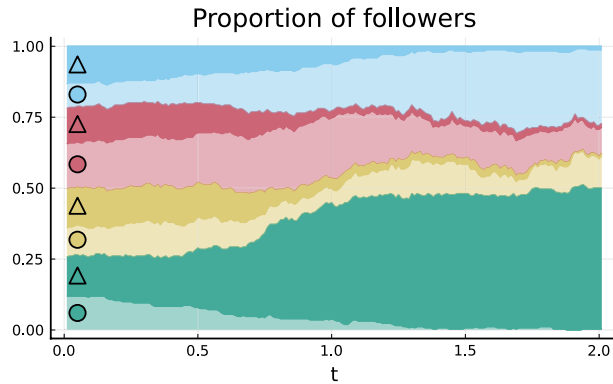
Figure 3. Stack plot showing how the proportion of individuals that follow a certain influencer and medium (marked by the symbols on the left) evolve in time for Example 1.

Also, Fig. 2 shows that at $t = 2$ most of the followers of the medium near $(-1, -1)$ (denoted by circles) follow the same influencer that is also near $(-1, -1)$ (colored in blue) while most followers of the medium near $(1, 1)$ (denoted by triangles) also follow the influencer near $(1, 1)$ (colored in green). The reason for this dominance is that for individuals it is favorable to be close to their medium and their influencer, otherwise they might switch the influencer to a more suitable candidate. The two less suitable influencers in the upper left and lower right corner now very slowly move towards their few remaining followers. On a larger time scale they will reach the clusters. Asymptotically, also the two clusters of individuals, media and influencers will approach another and merge. For different simulation runs, the agents behave qualitatively similar. But the individuals following the influencer initially at $(-1, 1)$ and $(1, -1)$ are sometimes attracted to a different medium than in the shown simulation.

Rich dynamical behaviour

The situation studied in Example 1 is somewhat symmetric and idealistic. In the Supplementary Information, we show realizations for some variations of this setting by changing the strength of influence of individuals, influencers and media (parameters a, b, c). This can qualitatively change the way in which clusters are formed and merged. To better understand these effects, we additionally performed a detailed study on how different choices of influence strength impact (1) the number of clusters and (2) how clustered individuals are around their media resp. influencers, see Section *Parameter studies* in the Supplementary Information. For other initial configurations and other choices of parameters and pair functions ϕ different forms of complex dynamical behavior emerge. This ranges from stable opinion clusters centered around "their" influencers (for $\phi(x) = \psi(x) = \mathbb{1}_{[0,d]}(x)$ and no media) to a complex interplay of cluster formation and reformation for several influencers with smaller and larger inertia γ, and a less symmetric configuration as in Example 1. The dynamics can become even richer, if the interaction network A is already exhibiting clusters. Moreover, different initial conditions lead to different transient dynamics. But we leave a more detailed parameter analysis for future work and concentrate here on deriving and studying the partial mean-field model of our general modeling framework.

Partial mean-field (opinion) model

For situations with many individuals but few influencers and media, one can derive the mean-field limit by a partial differential equation (PDE) that describes the opinion dynamics of individuals in the limit of infinitely many individuals[23,24,49] but is usually already a good approximation to the dynamics for finitely many individuals. Since here the number of influencers and media is still small and finite, their dynamics are still best described by SDEs but now coupled to PDEs for the evolution of the opinion distributions of individuals. We therefore also
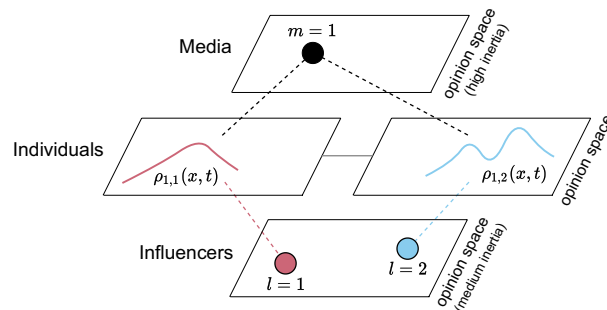
Figure 4. Structure of the partial mean-field model of distinct media agents (indicated by black circles) and influencers (shown as colored circles) in interaction with the opinion distributions of individuals.

call the model a *partial* mean-field model, compare Fig. 4 for the structure of the model. The coupled system of PDEs and SDEs is not only computationally advantageous[26] since the computational effort no longer scales with N^2 (due to the expensive computations of pair-wise distances of individuals in the ABM) but with the number of spatial grid cells and independently of N. Additionally, the model is conceptually easier to study for example to find critical parameters[23–25] or to use the partial mean-field model to control the evolution of the opinion distribution through different influencer and media strategies (see next section).

For the sake of simplicity, we subsequently assume a fully-connected interaction network between individuals, such that $A_{ij} = 1$ for all pairs of individuals (i, j). Without this assumption, one would need to derive mean-field dynamics for interacting agents on realistic complex networks that have a network limit in terms of graphons[50] or graphops[51,52], but here we will make the standard assumption of a fully-connected network between individuals for deriving the mean-field model. Further we assume that each individual follows exactly one medium and one influencer at a time and only influencers can be switched at the rate given in (6).

Then let us define the *empirical distribution* of individuals that follow medium m and influencer l at time t by the sum of Dirac Delta distributions δ placed at the individuals' opinions

$$\rho_{m,l}^{(N)}(x, t) = \frac{1}{N} \sum_{\substack{i \,:\, B_{im} = 1, \\ C_{il}(t) = 1}} \delta(x - x_i(t)). \tag{7}$$

This distribution describes the stochastic opinion instances at a given time t and integrates to $\int_D \rho_{m,l}^{(N)}(x, t)dx =: n_{m,l}^{(N)}(t)$, the proportion of individuals that follow medium m and influencer l.

It can be shown (see Supplementary Information) that as $N \to \infty$, the empirical distribution $\rho_{m,l}^{(N)}(x, t)$ can be replaced by the limiting distribution $\rho_{m,l}(x, t)$ solving the following PDE on the domain D

$$\partial_t \rho_{m,l}(x, t) = \frac{1}{2}\sigma^2 \Delta \rho_{m,l}(x, t) - \nabla \cdot \left(\rho_{m,l}(x, t)\, \mathscr{F}\,(x, y_m, z_l, \rho) \right)$$
$$+ \sum_{l' \neq l} \left(-\Lambda_m^{\to l'}(x, t)\, \rho_{m,l}(x, t) + \Lambda_m^{\to l}(x, t)\, \rho_{m,l'}(x, t) \right) \tag{8}$$

for each $m = 1, \dots, M$ and $l = 1, \dots, L$, where $\nabla\cdot$ denotes the divergence operator and Δ the Laplace operator on opinion space. The PDE is accompanied by boundary conditions ensuring that the number of individuals in the system is conserved.

The changes of $\rho_{m,l}$ are governed by three processes: (i) The first term on the RHS of the PDE is responsible for the stochastic diffusion of opinions. (ii) The divergence term models the interaction of the distribution $\rho_{m,l}$ with the overall distribution of all individuals $\rho = \sum_{m,l} \rho_{m,l}$ as well as with the respective medium y_m and the influencer z_l according to the attraction force at opinion x

$$\mathscr{F}\,(x, y_m, z_l, \rho) = a \frac{\int_D \rho(x', t)\phi(\|x' - x\|)(x' - x)\, dx'}{\int_D \rho(x', t)\phi(\|x' - x\|)\, dx'} + b\,(y_m(t) - x) + c\,(z_l(t) - x). \tag{9}$$

(iii) The last term of the PDE is responsible for the mass exchange between different distributions due to individuals switching the influencer away from l' or towards l. Note that due to the second and third term, the PDE is non-local.

The PDEs are coupled to the SDEs of the media $m = 1, \dots, M$

$$\Gamma dy_m(t) = (\tilde{x}_m(t) - y_m(t))dt + \tilde{\sigma} d\tilde{W}_m(t) \tag{10}$$

where the average opinion of followers is now given by $\tilde{x}_m(t) = \frac{\sum_{l=1}^{L} \int_D x\, \rho_{m,l}(x, t)dx}{\sum_{l=1}^{L} n_{m,l}(t)}$ with $n_{m,l}(t)$ denoting the limit of $n_{m,l}^{(N)}(t)$, as well as to the SDE dynamics of influencers $l = 1, \dots, L$

$$\gamma dz_l(t) = (\hat{x}_l(t) - z_l(t))dt + \hat{\sigma} d\hat{W}_l(t), \tag{11}$$

with $\hat{x}_l(t) = \frac{\sum_{m=1}^{M} \int_D x\, \rho_{m,l}(x, t)dx}{\sum_{m=1}^{M} n_{m,l}(t)}$ denoting the average opinion of individuals that follow influencer l.

Example 2: comparison on ABM and partial mean-field dynamics

In Fig. 5 we show a comparison between the ABM averaged over 1000 simulations and the solution of the partial mean-field model. To compare the ABM configurations against the opinion distributions in the partial mean-field, the opinion distributions resulting from the ABM are visualized via Kernel density estimation with Gaussian kernels. The comparison shows that the average results of the ABM simulation are consistent with the partial mean-field already for $N = 250$ individuals in the system.

Strategies in the partial mean-field model

In the previous section we have stated the partial mean-field model as a reduced model describing the opinion dynamics of infinitely many individuals coupled to the opinion dynamics of finitely many media and influencers. Assuming deterministic opinion changes of influencers and media (i.e. $\hat{\sigma} = \tilde{\sigma} = 0$), the partial mean-field model is deterministic and faster to solve than the ABM, allowing us to study the effect of different (optimal) influencer and media strategies.
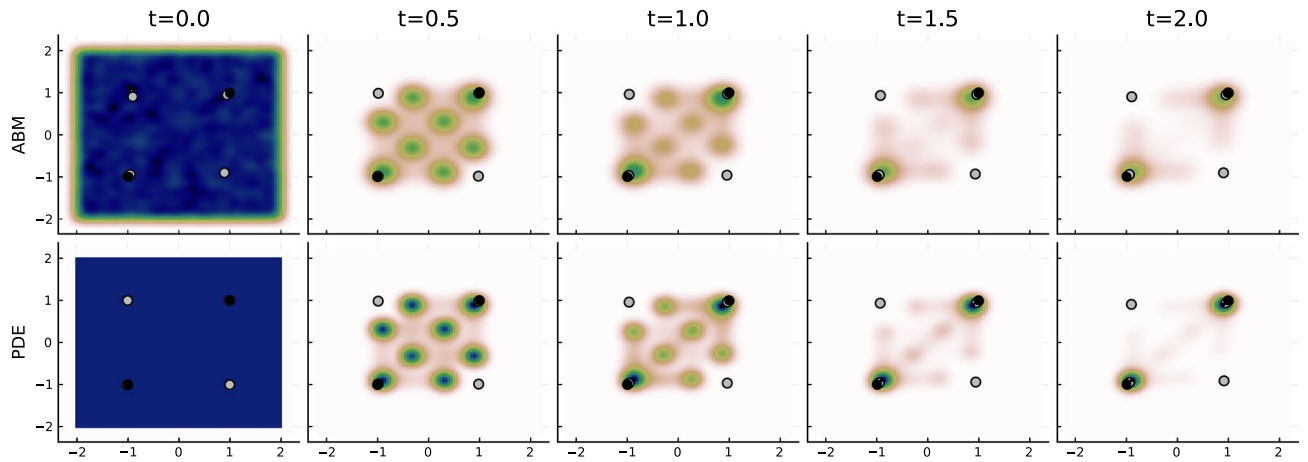
Figure 5. Mean configuration of the ABM (top row) over 1000 simulations and the solution of the partial mean-field PDE. Influencers are now marked by grey circles, and media agents by black circles. The opinion distribution of all individuals (i.e. independent of the media/influencer they follow) is shown with a heat plot.

Increasing followership

In the version of our opinion dynamics model as specified above, influencers adapt their opinions in the direction of the average opinion of their followers. This is a very simple strategy for influencers trying to keep their followers attached to them[33]. Other strategies to keep and even increase the number of followers might be more fruitful. In the following example we discuss a simple strategy for a new influencer with initially zero followers to substantially increase its followership.

Example 3: strategy of an influencer to increase followership

We consider the dynamics as before in Example 1 except with $\eta = 1$. With this choice of parameters, influencers strongly affect and attract individuals (since $c > a, b$) and individuals only slowly change influencers (due to η being small). In this way, once individuals follow a certain influencer, they will remain with that influencer for some time.

At time $t = 5$ we insert an additional influencer into the system with opinion $(0, 0)$ and zero followers. This new influencer then moves at a constant speed to the opinion $(1.5, 1.5)$ during the time interval $[5, 10]$. In Fig. 6, we show snapshots of the realization. The inserted influencer $l = 5$ quickly collects and attracts many followers behind. Starting with initially zero followers, the final followershare is $\sum_{m=1}^{2} n_{m,5}(10.0) \approx 0.18$. Moreover the time-averaged proportion of followers of influencer $l = 5$ over the interval $[5, 10]$ comes out to be approximately $\frac{1}{5}\int_{t=5}^{t=10}\sum_{m=1}^{2} n_{m,5}(t)\,dt \approx 0.10$. Thus the added influencer could substantially increase its followership by moving to an extreme opinion position. Even a new cluster of followers has formed near the influencer.

Optimal counteraction

More generally, the partial mean-field model also allows to apply optimal control techniques in order to derive the best strategy for a single influencer or a medium to achieve a certain objective, such as maximizing the number of their followers or optimally counteracting the goal of another agent.

Let $\varrho(u) = (\rho_{m,l}(t))_{m=1,\ldots,M,l=1,\ldots,L}$, the matrix of all different opinion distributions on the time interval $[0, T]$, satisfy Eqs. (8), (10), and (11), except for the chosen controlled influencer or medium. When we are interested
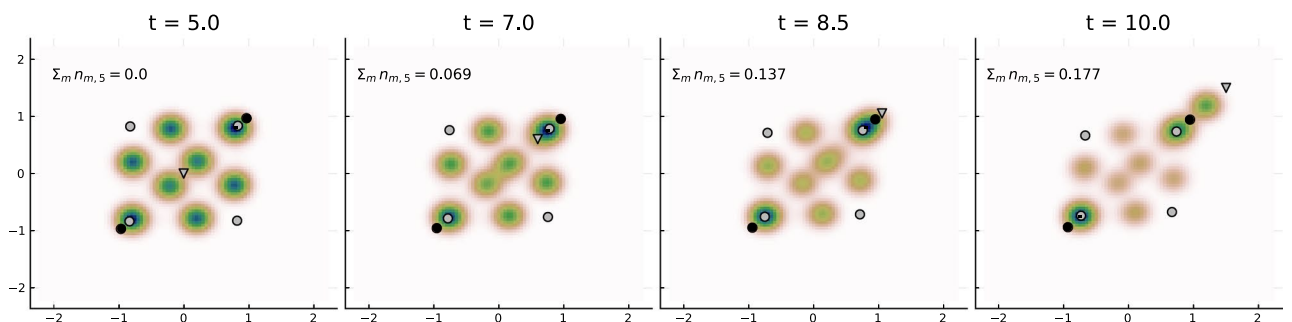
Figure 6. Opinion distribution ρ evolving in time (Example 3) started from a deterministic uniform opinion distribution at $t = 0$. At time $t = 5$, a 5th influencer (marked by a triangle) is inserted into the system and moves at a constant pace in the direction of the right upper corner while collecting new followers. Parameters as before except $\eta = 1$.

in the strategy for an influencer agent $l = l^\star$, we let the control u determine the behaviour of the $l^\star$th influencer, i.e. $z_{l^\star} = u$, when on the other hand we are searching for a media strategy, we control the $m^\star$th medium with $y_{m^\star} = u$. We can then express the aim of influencer $l^\star$ to maximize its temporally aggregated followership with the objective function

$$\mathscr{J}^\rho = \int_0^T \sum_{m=1}^M n_{m,l^\star}(t)\, dt,$$

while maximizing solely the final followership means dropping the integral and taking the integrand at time $t = T$. In contrast, if the goal of influencer $l^\star$ or medium $m^\star$ is to counteract the maximization of followership of another influencer l', this can be achieved by maximizing the objective

$$\mathscr{J}^\rho = -\int_0^T \sum_{m=1}^M n_{m,l'}(t)\, dt.$$

When the control determines the behavior of influencer $l^\star$ or medium $m^\star$, the control u needs to satisfy certain restrictions. That is, the control function u has to come from a set of admissible controls U. On the one hand, U is restricted by the necessary domain constraints $u(t) \in D$ for almost all $t \in [0, T]$. On the other hand, with every control u certain costs are associated. Control discontinuities or even chattering controls could, e.g. increase the risk of the control activities being detected, and countermeasures taken. Such concerns can be included as a penalty term in the objective, e.g. by

$$\mathscr{J}^p = \alpha |\partial_t u|^2_{L^2(]0,T[)},$$

where the parameter α regulates the penalty for large opinion changes. Then, the optimal control problem has the general form

$$\max_{u \in U}\ \mathscr{J}^\rho - \mathscr{J}^p. \tag{12}$$

After an appropriate control discretization with, e.g. piecewise polynomials, the resulting nonlinear programming problem can be solved by derivative-free methods like Nelder-Mead or by more efficient inexact gradient descent[53] or stochastic approximation methods. Sufficiently accurate gradient evaluations can be obtained by finite differencing the PDE/SDE forward equations for $\varrho(u)$ or by Feynman-Kac type gradient sampling.

Example 4: influencer counteraction

We are interested in the optimal counteraction of influencer $l^\star = 6$ when influencer $l' = 5$ moves at a constant speed to the upper right corner (1.5, 1.5) (as in Example 3) and thereby drastically increases its followership. We therefore want influencer $l^\star$ to move in an optimal way to satisfy

$$\max_{u \in U}\ \left(-\int_5^{10} \sum_{m=1}^2 n_{m,5}(t)\, dt - \alpha |\partial_t u|^2_{L^2(]5,10[)} \right) \tag{13}$$

with $\alpha = 0.05$. That is, the goal is to minimize the followershare of influencer l', while avoiding conspicuously drastic opinion changes. The control set U contains all functions with $u(t) \in D$ and with piecewise constant velocities that only change at 3 chosen time points.

The optimal strategy is shown in Fig. 7 and results in an average followershare of influencer l' of $\frac{1}{5} \int_{t=5}^{t=10} \sum_{m=1}^2 n_{m,5}(t)\, dt \approx 0.05$, which is significantly smaller compared to the value 0.1 that was obtained in Example 3 without counteraction. Note also, that the final followershare at time $t = 10.0$ was decreased from 0.18 without counteraction to 0.04 with counteraction. The snapshots in Fig. 7 show that the optimal counterstrategy
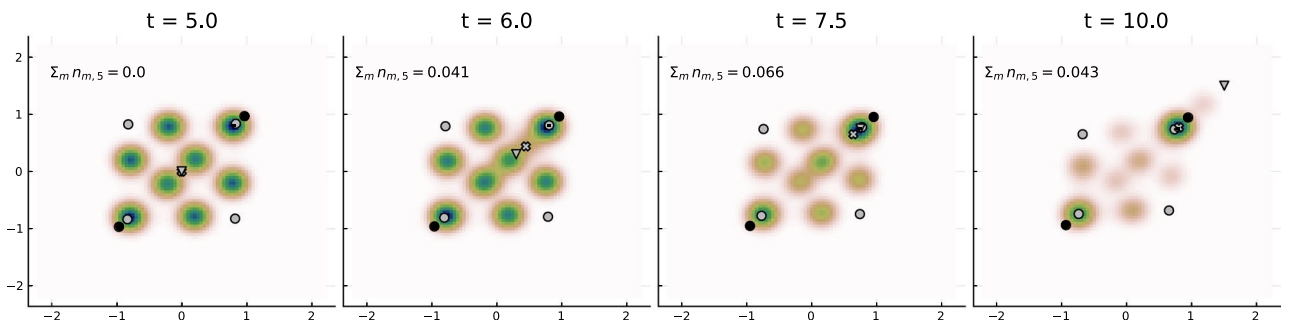
Figure 7. Opinion distribution ρ evolving in time (Example 4) started from a uniform distribution at $t = 0$. At time $t = 5$, a 5th influencer (marked by a grey triangle) is inserted into the system and moves at a constant pace in the direction of the right upper corner. Simultaneously, a 6th influencer (marked by a cross) starts and moves in a optimal way to prevent influencer $l' = 5$ from collecting many followers. Parameters as before except $\eta = 1$.

of influencer $l^\star = 6$ is to steel followers of influencer $l' = 5$ by moving along; it thereby stabilizes the opinion cluster of individuals. Note that according to the objective specified in (13), this counteraction comes out to be more effective than the movement in the opposite direction of influencer $l' = 5$.

Example 5: media counteraction

Next, we study the optimal counterstrategy of the media agent $m^\star = 2$ in the upper right corner when trying to minimize the followershare of influencer $l' = 5$, analogously to the previous example with the objective function given by (13). The resulting optimal strategy is shown in Fig. 8. By changing the opinion drastically from the upper right to the lower right corner, the media agent manages to decrease the average followershare of influencer $l' = 5$ down to $\frac{1}{5} \int_{t=5}^{t=10} \sum_m n_{m,5}(t)\, dt \approx 0.06$. Note that since the system dynamics are symmetric wrt. the diagonal axis $x = y$, an equivalently good strategy would be to move to the upper left corner. The counterstrategy of the media agent is very different from the counteraction of an influencer (given in Example 4). This is because influencers compete for followers and can steal influencers from each other while media agents can only make an opinion topic unattractive to individuals by changing the topic.

Conclusion

In this work we provided new mathematical means for the systematic study of how traditional media and influencers might impact coherent structures of the public opinion distribution as modelled by opinion dynamics models. Our work does neither claim to describe opinion shifts in real-world social networks nor does it state anything about influencing the opinion of an individual human being. It is still an idealized model and, like most contributions to the field, can neither claim to describe human opinion formation processes realistically, nor to be validated with observational data and controlled sociological experiments. However, it may help to describe how shifting individual perspectives and social exchange lead to archetypal states of public opinion distribution like coherent opinion clusters. By providing a strategy for understanding how influencing the public opinion distribution may be done and counteracted in an optimal way, this work may help us to understand how to face current challenges of opinion polarization in complex scenarios.

In particular, our model framework demonstrates the impact influencers and media might have on the opinion distribution by creating stable and coherent opinion clusters that are drawn to the positions of the respective influencer or media. We also see that when influencers move to extreme positions, they fragment the opinion landscape and gain more followers and more attention in the competitive scenario of social media. Hence, these results suggest that those competitive goals could contribute to a polarized and fragmented debate. However, our work also offers avenues towards solutions through optimally counteracting those attempts via other influencers (or media). These counteracting influencers can stabilize the existing opinion landscape against extreme influencers by strengthening existing opinion clusters. Empirical work needs to be done to validate those strategies and interplays that our model is suggesting, potentially through large-scale field experiments on social media. Our theoretical work potentially delivers a good basis for developing such alternative strategies for influencers in the current system to stabilize online discourse.

In this work we did not explore the full generality of the opinion dynamics model for different interaction terms, parameters and different influencer and media strategies. Instead we focused in this work on the feasibility by designing the pipeline from the inclusion of influencers and media via the demonstration of the emergence of temporarily coherent opinion clusters to the options for influencing the public opinion by optimally chosen strategies. The results provided are supported by the analysis of the dynamical patterns that the model exhibits for different parameter combinations, such as numbers of opinion clusters and the study of order parameters. However, a detailed study of further parameter choices is not included, e.g. different time scale choices, pair functions, complex interaction network structures etc.. This was beyond the focus and scope of this study and will be left to future investigations.
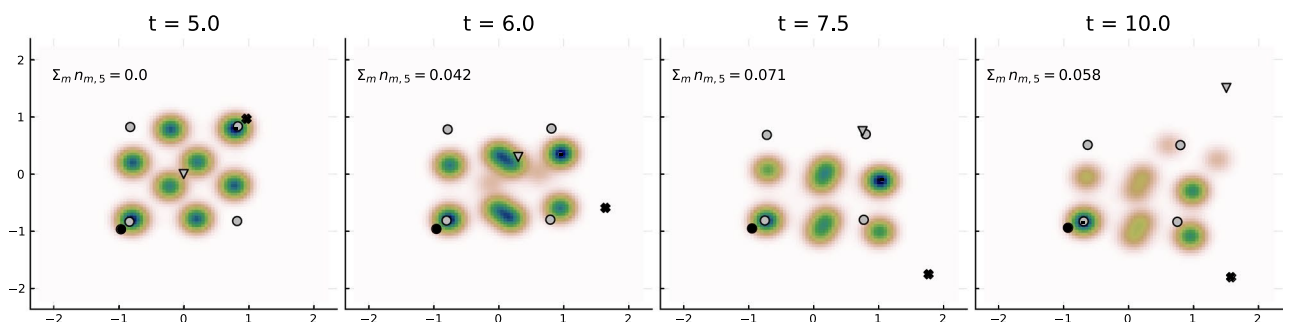
Figure 8. Opinion distribution ρ evolving in time (Example 5) started from a uniform distribution at $t = 0$. At time $t = 5$, a 5th influencer (marked by a triangle) is inserted into the system and moves at a constant pace in the direction of the right upper corner, simultaneously, the media in the upper right corner (marked by a cross) moves in a optimal way to prevent influencer $l' = 5$ from collecting many followers. Parameters as before except $\eta = 1$.

Data availability

References

1. Bak-Coleman, J. B. *et al.* Stewardship of global collective behavior. *Proc. Natl. Acad. Sci.* **118**, e2025764118 (2021).
2. Newman, N. *et al.* Reuters institute digital news report 2021. *Reuters Institute for the study of Journalism* (2021).
3. Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. Everyone's an influencer: Quantifying influence on twitter. In *Proc. of the Fourth ACM International Conference on Web Search and Data mining*, 65–74 (2011).
4. Zimmermann, D. *et al.* Influencers on youtube: A quantitative study on young people's use and perception of videos about political and societal topics. *Curr. Psychol.* https://doi.org/10.1007/s12144-020-0 *(2020)*.
5. Soares, F. B., Recuero, R. & Zago, G. Influencers in polarized political networks on twitter. In *Proc. of the 9th International Conference on Social Media and Society*, 168–177 (2018).
6. Tafesse, W. & Wood, B. P. Social media influencers' community and content strategy and follower engagement behavior in the presence of competition: An instagram-based investigation. *J. Prod. Brand Manag.* **32**, 406–419 (2022).
7. Lorenz-Spreen, P., Oswald, L., Lewandowsky, S. & Hertwig, R. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nat. Hum. Behav.* **7**(74–101), 1–28 (2022).
8. Liggett, T. M. *Interacting Particle Systems* (Springer, 2005).
9. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. assoc.* **69**, 118–121 (1974).
10. Hegselmann, R. & Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simul.* **5**, 1–33 (2002).
11. Weisbuch, G., Deffuant, G., Amblard, F. & Nadal, J.-P. Meet, discuss, and segregate!. *Complexity* **7**, 55–63 (2002).
12. Hegselmann, R. & Krause, U. Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Netw. Heterog. Media* **10**, 477 (2015).
13. Sîrbu, A., Loreto, V., Servedio, V. D. & Tria, F. Opinion dynamics: Models, extensions and external effects. In *Participatory Sensing, Opinions and Collective Awareness* (eds Sîrbu, A. *et al.*) 363–401 (Springer, 2017).
14. Redner, S. Reality-inspired voter models: A mini-review. *Comptes Rendus Phys.* **20**, 275–292. https://doi.org/10.1016/j.crhy.2019.05.004 (2019).
15. Jager, W. & Amblard, F. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Comput. Math. Organ. Theory* **10**, 295–303 (2005).
16. Dandekar, P., Goel, A. & Lee, D. T. Biased assimilation, homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci.* **110**, 5791–5796. https://doi.org/10.1073/pnas.1217220110 (2013).
17. Xia, W., Ye, M., Liu, J., Cao, M. & Sun, X.-M. Analysis of a nonlinear opinion dynamics model with biased assimilation. *Automatica* **120**, 109113. https://doi.org/10.1016/j.automatica.2020.109113 (2020).
18. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**, 048301 (2020).
19. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Emergence of polarized ideological opinions in multidimensional topic spaces. *Phys. Rev. X* **11**, 011012 (2021).
20. Bizyaeva, A., Franci, A. & Leonard, N. E. Nonlinear opinion dynamics with tunable sensitivity. *IEEE Trans. Autom. Control* https://doi.org/10.1109/TAC.2022.3159527 *(2022)*.
21. Leonard, N., Lipsitz, K., Bizyaeva, A., Franci, A. & Lelkes, Y. The nonlinear feedback dynamics of asymmetric political polarization. *Proc. Natl. Acad. Sci.* **118**, e2102149118. https://doi.org/10.1073/pnas.2102149118 (2021).
22. Carro, A., Toral, R. & San Miguel, M. The role of noise and initial conditions in the asymptotic solution of a bounded confidence, continuous-opinion model. *J. Stat. Phys.* **151**, 131–149 (2013).
23. Wang, C., Li, Q., Ee, W. & Chazelle, B. Noisy Hegselmann-krause systems: Phase transition and the 2R-conjecture. *J. Stat. Phys.* **166**(5), 1209–1225 (2017).
24. Garnier, J., Papanicolaou, G. & Yang, T. Consensus convergence with stochastic effects. *Vietnam J. Math.* **45**, 51–75 (2017).
25. Delgadino, M. G., Gvalani, R. S. & Pavliotis, G. A. On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions. *Arch. Ration. Mech. Anal.* **241**, 91–148. https://doi.org/10.1007/s00205-021-01648-1 (2021).
26. Helfmann, L., Djurdjevac Conrad, N., Djurdjevac, A., Winkelmann, S. & Schütte, C. From interacting agents to density-based modeling with stochastic PDES. *Commun. Appl. Math. Comput. Sci.* **16**, 1–32 (2021).
27. Djurdjevac Conrad, N., Köppl, J. & Djurdjevac, A. Feedback loops in opinion dynamics of agent-based models with multiplicative noise. *Entropy* **24**, 1352 (2022).
28. Carletti, T., Fanelli, D., Grolli, S. & Guarino, A. How to make an efficient propaganda. *EPL Europhys. Lett.* **74**, 222 (2006).
29. Mobilia, M., Petersen, A. & Redner, S. On the role of zealotry in the voter model. *J. Stat. Mech. Theory Exp.* **2007**, P08029 (2007).
30. Baumann, F., Sokolov, I. M. & Tyloo, M. A Laplacian approach to stubborn agents and their role in opinion formation on influence networks. *Phys. A Stat. Mech. Appl.* **557**, 124869. https://doi.org/10.1016/j.physa.2020.124869 (2020).
31. Brooks, H. Z. & Porter, M. A. A model for the influence of media on the ideology of content in online social networks. *Phys. Rev. Res.* https://doi.org/10.1103/physrevresearch.2.023041 *(2020)*.
32. Ortiz, E. & Serrano, M. Á. Multiscale voter model on real networks. *Chaos, Solitons & Fractals* **168**, 112847 (2022).
33. Tafesse, W. & Wood, B. P. Followers' engagement with instagram influencers: The role of influencers' content and engagement strategy. *J. Retail. Consum. Serv.* **58**, 102303 (2021).
34. Gilardi, F., Gessler, T., Kubli, M. & Müller, S. Social media and political agenda setting. *Polit. Commun.* **39**, 39–60 (2022).
35. Masuda, N. Opinion control in complex networks. *New J. Phys.* **17**, 033031 (2015).
36. Moreno, G. R., Chakraborty, S. & Brede, M. Shadowing and shielding: Effective heuristics for continuous influence maximisation in the voting dynamics. *PLoS One* **16**(6), e0252515 (2021).
37. Li, Y. & Zhao, H. Evaluation and control of opinion polarization and disagreement: A review. *Int. J. Crowd Sci.* **5**(2), 114–128 (2021).
38. Albi, G., Pareschi, L. & Zanella, M. On the optimal control of opinion dynamics on evolving networks. In *IFIP Conference on System Modeling and Optimization* (eds Albi, G. *et al.*) 58–67 (Springer, 2015).
39. Kozitsin, I. V. Optimal control in opinion dynamics models: towards a unified framework, https://doi.org/10.48550/ARXIV.2207.01300 (2022).
40. Albi, G., Choi, Y.-P., Fornasier, M. & Kalise, D. Mean field control hierarchy. *Appl. Math. Optim.* **76**, 93–135 (2017).
41. Gong, X., Herty, M., Piccoli, B. & Visconti, G. Crowd dynamics: Modeling and control of multiagent systems. *Annu. Rev. Control Robot. Auton. Syst.* **6**, 261–282 (2022).
42. Lorenz-Spreen, P., Mønsted, B. M., Hövel, P. & Lehmann, S. Accelerating dynamics of collective attention. *Nat. Commun.* **10**, 1–9 (2019).

43. Mäs, M., Flache, A. & Helbing, D. Individualization as driving force of clustering phenomena in humans. *PLoS Comput. Biol.* **6**, e1000959 (2010).
44. Martins, T. V., Pineda, M. & Toral, R. Mass media and repulsive interactions in continuous-opinion dynamics. *EPL Europhys. Lett.* **91**, 48003 (2010).
45. Neuhäuser, L., Schaub, M. T., Mellor, A. & Lambiotte, R. Opinion dynamics with multi-body interactions. In *International Conference on Network Games, Control and Optimization* (eds Neuhäuser, L. *et al.*) 261–271 (Springer, 2021).
46. Bhat, D. & Redner, S. Polarization and consensus by opposing external sources. *J. Stat. Mech. Theory Exp.* **2020**, 013402 (2020).
47. Li, Z., Fang, X. & Sheng, O. R. L. A survey of link recommendation for social networks: Methods, theoretical foundations, and future research directions. *ACM Trans. Manag. Inf. Syst. (TMIS)* **9**, 1–26 (2017).
48. Santos, F. P., Lelkes, Y. & Levin, S. A. Link recommendation algorithms and dynamics of polarization in online social networks. *Proc. Natl. Acad. Sci.* **118**, e2102141118 (2021).
49. Goddard, B. D., Gooding, B., Short, H. & Pavliotis, G. Noisy bounded confidence models for opinion dynamics: The effect of boundary conditions on phase transitions. *IMA J. Appl. Math.* **87**, 80–110 (2022).
50. Chiba, H. & Medvedev, G. S. The mean field analysis of the Kuramoto model on graphs. The mean field equation and transition point formulas. *Discrete Contin. Dyn. Syst.* **39**, 131–155. https://doi.org/10.3934/dcds.2019006 (2019).
51. Kuehn, C. Network dynamics on graphops. *New J. Phys.* **22**, 053030 (2020).
52. Gkogkas, M. A. & Kuehn, C. Graphop mean-field limits for Kuramoto-type models. *SIAM J. Appl. Dyn. Syst.* **21**, 248–283 (2022).
53. Nocedal, J. & Wright, S. *Numerical Optimization* (Springer, 1999).

Acknowledgements

Author contributions

L.H., N.D.C., P.L.-S. and C.S. designed research; L.H., N.D.C. and C.S. performed research; L.H., N.D.C., P.L.-S. and C.S. designed simulation experiments; L.H. implemented software; L.H., N.D.C., P.L.-S. and C.S. wrote the paper.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-46187-9.

Correspondence and requests for materials should be addressed to C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.