

Modeling communication asymmetry and content personalization in online social networks

Franco Galante^{a,*}, Luca Vassio^a, Michele Garetto^b, Emilio Leonardi^a

^a*Politecnico di Torino, Corso Duca Degli Abruzzi, 24, 10129 Torino*

^b*Università degli Studi di Torino, Corso Svizzera 185, 10149 Torino*

Abstract

The increasing popularity of online social networks (OSNs) attracted growing interest in modeling social interactions. On online social platforms, a few individuals, commonly referred to as *influencers*, produce the majority of content consumed by users and hegemonize the landscape of the social debate. However, classical opinion models do not capture this communication asymmetry. We develop an opinion model inspired by observations on social media platforms with two main objectives: first, to describe this inherent communication asymmetry in OSNs, and second, to model the effects of content personalization. We derive a Fokker-Planck equation for the temporal evolution of users' opinion distribution and analytically characterize the stationary system behavior. Analytical results, confirmed by Monte-Carlo simulations, show how strict forms of content personalization tend to radicalize user opinion, leading to the emergence of *echo chambers*, and favor *structurally advantaged* influencers. As an example application, we apply our model to Facebook data during the Italian government crisis in the summer of 2019. Our work provides a flexible framework to evaluate the impact of content personalization on the opinion formation process, focusing on the interaction between *influential* individuals and *regular* users. This framework is interesting in the context of marketing and advertising, misinformation spreading, politics and activism.

Keywords: opinion dynamics, online social networks, content personalization, Fokker-Planck equation

*Corresponding author

Email addresses: franco.galante@polito.it (Franco Galante), luca.vassio@polito.it (Luca Vassio), michele.garetto@unito.it (Michele Garetto), emilio.leonardi@polito.it (Emilio Leonardi)

1. Introduction

In recent years, the way people communicate has changed dramatically. With the advent of the Internet, new communication channels have emerged that allow people to transcend geographical and language barriers thanks to a global communication network and alternatives to text-only interaction like images and videos. Online social networks (OSNs) are probably the most notable example of such new interaction mechanisms and have greatly influenced our society by fostering discussions and disseminating information. The amount of content produced on such social media platforms is immense. Therefore, to keep users engaged within the social network, the platform performs filtering to select the posts to offer. This filtering mechanism may reinforce the natural tendency of users to interact with like-minded individuals, one of the main drivers of social network building [1], known as *homophily*. And, in turn, it can lead to the formation of *echo chambers* [2], where people who share a similar point of view interact with each other while being isolated from the rest of the users. The reach of certain social media users can be extraordinary. Posts by very influential people, often referred to as *influencers* or *opinion leaders*, can reach a large audience in virtually no time. Another aspect worth mentioning is the diversity of topics discussed, ranging from commentaries on the latest sporting events to debates on sensitive issues such as vaccinations.

From a user's perspective, there are two types of online interactions: those with acquaintances and those with influencers. Our model focuses primarily on the latter interaction, as influencers' sway has not yet received sufficient attention in the literature. Toward the end of the manuscript, we present an extension of the model to include regular users' interactions.

The main objective of this work is to develop an analytical framework tailored to online interactions, incorporating the following aspects:

- The asymmetry of OSNs: a relatively small portion of *well-known* users, i.e., the *influencers*, can reach a vast number of far less known individuals.

- The closed-loop between influencers’ and regular users’ dynamics, triggered by the *content personalization* mechanism applied by the social media platform in one direction and user *feedback* (e.g., likes) in the other.

The concept of *reference direction*, which is the individual’s main topic of interest and expertise, represents a novel aspect of our approach. To our knowledge, the existing literature has not yet considered a reference topic for each influencer on the opinion formation process in multi-dimensional spaces. Following recent work, we support the model’s hypotheses with social network data [3][4] from a large ensemble of Italian influencers, and compare the outcomes with emerging phenomena on two popular OSNs, i.e., Facebook and Instagram [5][2]. The proposed model does not aim to be quantitatively predictive but rather to be a tool to study qualitatively the emerging behavior in online social networks and the impact of content personalization, with a focus on opinion leaders.

For the single-topic case, we derive a Fokker-Planck equation as a second-order approximation of the opinion formation process and prove the existence and uniqueness of the stationary solution. In addition, since this approach is non-constructive we develop a less accurate first-order approximation, i.e., the fluid limit, which provides a closed-form formula expressions for the stationary solution. Then, through a Monte-Carlo approach, we analyze the effects of the influencers’ structural parameters and the interplay with content personalization in a multi-topic environment. The results highlight some of the threats associated with content filtering. We show that it favors *structurally advantaged* influencers and can radicalize users’ opinions, leading to the formation of *echo chambers*.

The paper is organized as follows. Section 2 briefly discusses the relevant work in the literature. The Communication Asymmetry model is presented in Section 3, along with the notation used throughout the article. Section 4 presents some observations from real social networks supporting our modeling assumptions. Section 5 is devoted to the mean-field analysis of the model as the number of users grows large. The theoretical results on the steady-state behavior of the model are proved in the Appendix, together with the experimental

analysis. Section 6 then investigates the impact of the model parameters on a reference scenario with two influencers. Section 7 further validates our model with real-world data collected on Instagram and Facebook. Finally, Section 8 discusses limitations of the work and possible extensions and future directions.

2. Related work

The first steps in the study of opinion dynamics were taken in the late 1950s by a number of social psychologists [6][7][8]. Ash introduced the concept of *social pressure* [6], a conformist tendency in individuals, while French used directed graphs to model interpersonal relationships [7]. A landmark in the field is Festinger’s work on *social comparison*[8]. Individuals tend to evaluate their position by comparing it to others, and it is inversely proportional to the distance between viewpoints. Opinion models are continuous or discrete, according to the description of the opinion variable. As for most models [9], the seminal work in the field is continuous. For example, the DeGroot model [10] considers a networked social system in which individuals interact with their neighbors. Individuals average their current opinion with the opinion of their neighbors. Subsequently, Friedkin and Johnsen [11] developed a linear model which encompasses both the processes of *social conformity* and *social conflict* leading to behavior that goes beyond simple consensus. In the early 2000s, Hegselmann and Krause [12] and Deffuant and Weisbuch [13] proposed two similar models, introducing the idea of bounded confidence: individuals interact only with peers whose beliefs are not too different. The proposed models are nonlinear and challenging to study [14]. A great deal of attention has also been paid to discrete models. A prominent example is the voter model [15][16] and its extensions accounting for evolving networks [17][18], individuals with multiple opinions [19] or spontaneous changes of opinion [20]. A consistent bulk of research on opinion dynamics comes from the physics literature, among which early contributions are Ben-Naim [21] and Toscani [22]. Ben-Naim and Toscani consider two mechanisms of opinion formation: *compromise*, and *introspection* (in other

models, e.g., [20], modeled as noise), which the authors believe represents the impact of external sources of information (e.g., media). The Sznajd model [23] is a generalization of the Ising model, which implements *social validation* and for which [24] derives analytical results. For a comprehensive review of classic opinion models, we refer to the survey by Castellano et al. [25].

Most of the seminal literature on opinion dynamics is suited to describe the decision-making process in small groups of individuals, e.g., a board of directors, or to capture relatively *regular* patterns determined by the daily personal interactions of individuals. Models such as the voter model have been studied extensively on regular lattices [26] [27]. The structure of interactions, especially those online, is far from homogeneous. As mentioned earlier, an inherent asymmetry in communication exists in OSNs where a limited number of individuals (*influencers*) monopolize the discussion. The voter model has been studied over heterogeneous networks (e.g., [28] [29]) to account for this diversity. On such networks, there can exist *hubs* (strongly connected nodes) playing a role similar to *influencers* in our framework. However, the authors did not explicitly make such a distinction. Other works have divided the population into classes, e.g., [30] introduced *stubborn* agents, and if such individuals have opposing opinions, they hinder the possibility of the population converging to consensus. Recent work further draws attention to online platforms by adapting classical frameworks to the specificity of online interactions. [31, 32, 33] have developed an opinion model that embodies *algorithmic personalization*. [31] also compares model predictions and dynamics observed on Facebook and Twitter. Our work differs from [31, 32, 33] as we consider distinct classes of users, precisely characterizing *influencers* and closing the interaction loop between users and the platform by a *feedback* function.

Attempts to validate opinion models are scarce due to several reasons: i) the mapping of opinions into values, ii) an adequate definition of links between agents, and iii) the change of opinion after an interaction is hardly measurable. A recent survey [34] examines the latest research concerning the use of data in opinion dynamics. Usually the approach is either through observational data

[5][2][31] or controlled sociological experiments[35][36]. The first type allows to scale to large number of users and is more related to our work. The political environment has classically been a florid field for opinion dynamics due to the possibility of attributing a person’s opinion to the political orientation of the chosen candidate [5]. Also, a noisy voter model was used to fit data from US elections [37]. The authors of [38] estimate the political ideology of users on Twitter on a single axis (left-right) using the ground truth reconstructed by roll-call votes of members of parliament and their network of followees-followers. In [39], opinions are estimated through a sentiment analysis tool applied to the text of posts on Twitter. Posts are first classified into topics according to their keywords (as we do, see Section 4.1), and then a continuous sentiment score is given between negative and positive. The authors of [40] consider users’ opinions in a multidimensional ideological space. Through doc2vec clustering, they identify the axes of 4 dimensions and then map users to these 4-topics opinions. Validation is made through well-known positions of famous Reddit groups. Other recent approaches [41] have used shared *news* on Facebook to assess the extent to which individuals are exposed to opposing views through their (online) friendship relationships, using users’ self-reported ideological affiliations to infer opinion. More recent work [2] [31] have directly employed data from online social networks, such as Gab, Facebook, Reddit, and Twitter, to observe the emergence of *echo chambers* [2] and to validate a model encompassing algorithmic personalization in the process of opinion formation [31].

3. The Communication Asymmetry opinion model

In this section, we first establish the notation used throughout the paper and then present the Communication Asymmetry model in its most general formulation, along with the limitations of the model.

3.1. Notation

In this work, we adopt the following vectorial notation. We denote vectors by bold characters, whereas we denote their components with normal-font char-

acters whose subscript is the index in the vector, e.g., $\mathbf{a} = \{a_k\}_k$. Lowercase letters denote parameters and dynamical variables associated with an individual. In general, index i runs over the set of *influencers* while the index u runs over that of regular users. For those parameters/variables that can be associated with individuals of both classes (either influencers or regular users), the above indices are indicated between superscript parentheses, e.g., $a^{(i)}, a^{(u)}$, to identify the class to which the individual belongs immediately. If necessary, the dependence of variables on other system parameters is made explicit by specifying the independent variables between parentheses, e.g., $\alpha(\cdot, \cdot)$. Italic capital letters denote sets, e.g., \mathcal{I} is the set of all influencers in the population, while $|\mathcal{I}|$ is its cardinality. Capital letters represent outcomes of stochastic experiments whose characteristic parameters are lowercase letters: e.g., $\Omega(\omega(\cdot, \cdot))$. The operator $\mathbb{E}[\cdot]$ represents an expected value, and a bar over a variable, e.g., \bar{a} , represents its average value. Whenever we need to express the probability of an event, we use the notation $Pr[\cdot]$. We employ $\mathbb{1}_{\{\cdot\}}$ for the indicator function. Lastly, time is denoted by t if considered continuous and by n if discrete.

3.2. Description of the model

We propose a continuous opinion model with two interacting classes of agents. Specifically, the population consists of $N_u = |\mathcal{U}|$ *regular users* and $N_i = |\mathcal{I}|$ *influencers*. This division mimics what happens in real social networks, where a small portion of the population, the *influencers*, has a much larger number of people following their posts on the online social network. The opinion space is $\mathcal{X} \subset \mathbb{R}^d$, where each dimension represents an uncorrelated topic on which users have a belief. Hence, an opinion is a d -dimensional vector $\mathbf{x}^{(u)}(n) \in \mathbb{R}^d$, which evolves as a result of the interaction between a regular user and the influencers on every possible topic. Moreover, we model the tendency toward an a priori opinion $\mathbf{z}^{(u)}$, and we refer to it as the *prejudice* of a user. Unless otherwise specified, we will assume that the user's initial opinion corresponds to the prejudice: $\mathbf{x}^{(u)}(0) = \mathbf{z}^{(u)}$. We assume that the generation of new posts, i.e., messages in the OSN, is a Poisson Point Process (PPP) with inten-

sity λ , where each event of the PPP corresponds to the creation of a new post from an influencer $i \in \mathcal{I}$. The corresponding embedded discrete time will be denoted by the integer $n \in \mathbb{N}_+$, $n = 1, 2, \dots$, where n is the n -th post. Figure 1 illustrates the social media platform’s role in *filtering* the posts sent¹ to regular users and receiving *feedback* from those users. These two aspects implement, respectively, a *selective exposure* effect: namely the tendency of both the platform and the users to suggest/access similar content, and a *confirmation bias*, namely the tendency of users to value content that is close to one’s point of view (see [2] and resources therein).

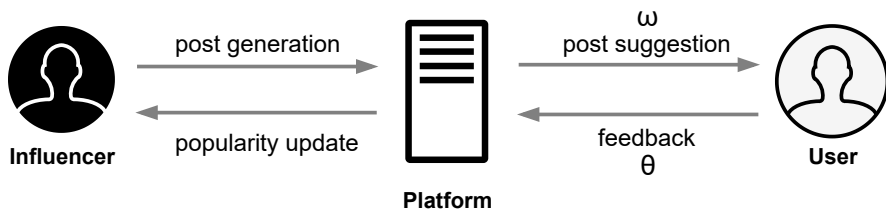


Figure 1: Illustration of the interaction between users, influencers, and social media platforms highlighting the closed loop between user feedback θ and content filtering ω , which will be further characterized below.

Influencers are considered stubborn, i.e., $\mathbf{x}^{(i)}(n) = \mathbf{x}^{(i)}(0) = \mathbf{x}^{(i)} = \mathbf{z}^{(i)} \in \mathbb{R}^d$, $\forall n > 0$ and $i \in \mathcal{I}$. As we will show in Section 4, each influencer has a main topic of interest on which it publishes most of its posts and typically coincides with the topic it is mainly known for on the OSN. We call it the *reference direction* $r^{(i)} \in \{0, \dots, d-1\}$ of the influencer. Another parameter characterizing influencer i is its *consistency* $c^{(i)}(n)$, which denotes the (possibly time-varying) probability of posting on the reference direction. High-consistency individuals preferably post on their reference topic. We denote by $f^{(i)}$ the probability that a post is generated by influencer i at any time instant n , with $\sum_{i \in \mathcal{I}} f^{(i)} = 1$. At last, we introduce the *popularity* vector $\mathbf{p}(n) \triangleq \{p_i(n)\}_{i \in \mathcal{I}}$, containing the current popularity of all influencers at time n , before the emission of the post

¹We use the terms “send”, “suggest” and “reach” interchangeably, referring to a post shown to a user by the platform. We will refer to *regular users* simply as *users*. Moreover, the terms *agent* or *individual* indicate a social network user of either class.

at time n . We also introduce the normalized version of this vector $\boldsymbol{\pi}(n) = \{\pi_i(n)\}_{i \in \mathcal{I}}$ where the components are the normalized popularities $\pi_i = \frac{p_i}{\sum_{j \in \mathcal{I}} p_j}$.

The dynamic variables of users (i.e., their opinion $\boldsymbol{x}^{(u)}$) and influencers (i.e., their popularity p_i) are updated upon every post generation according to Algorithm 1. Figure 2 gives a schematic representation of the model.

Algorithm 1 The Communication Asymmetry opinion model

Input: N_i influencers, N_u users, filtering function ω , feedback function θ
Output: opinion of each regular user $\boldsymbol{x}^{(u)}(n)$, $\forall u$
Output: popularity of each influencer $p_i(n)$, $\forall i$

- 1: **loop**
- 2: select influencer i according to $f^{(i)}$
- 3: select a posting direction j , i.e., $j = r^{(i)}$ with probability $c^{(i)}$, otherwise j is selected uniformly on $j \in \{0, \dots, d-1\} \setminus \{r^{(i)}\}$
- 4: $p_i(n+1) = p_i(n)$
- 5: **for all** regular user u in the population **do**
- 6: $x_j^{(u)}(n+1) = x_j^{(u)}(n)$
- 7: **if** $\Omega \left(\omega(|x_{r^i}^{(i)} - x_{r^i}^{(u)}|, \pi_i(n)) \right) = 1$ **then** {post suggestion}
- 8: get feedback $\Theta \left(\theta(|x_j^{(u)} - x_j^{(i)}|) \right)$
- 9: **if** $\Theta = 1$ **then** {positive feedback}
- 10: $x_j^{(u)}(n+1) = \alpha z_j^{(u)} + \beta x_j^{(u)}(n) + (1 - \alpha - \beta)x_j^{(i)}$
- 11: update popularity of i : $p_i(n+1) += 1/N_u$
- 12: **end if**
- 13: **end if**
- 14: **end for**
- 15: **end loop**

At any time instant n , an influencer i , selected according to distribution $f^{(i)}$, generates a post. The influencer i posts on its reference direction $r^{(i)} = j$ with a probability equal to its consistency $c^{(i)}$. Otherwise, it posts on one of the *secondary* directions $j \in \{0, 1, \dots, d-1\} \setminus \{r^{(i)}\} = \mathcal{N}_r$ according to a given distribution, $Pr[j = k]$ for k in \mathcal{N}_r . In the following, we assume this distribution to be uniform over the set of non-reference directions. The emitted post carries the j -th component of the influencer's opinion vector $\boldsymbol{x}^{(i)}$. We suppose posts accurately reflect the influencer's true belief and no noise affects user perception.

To decide whether the post reaches a given user (independently from other users), we extract a Bernoulli random variable Ω with parameter ω . The user

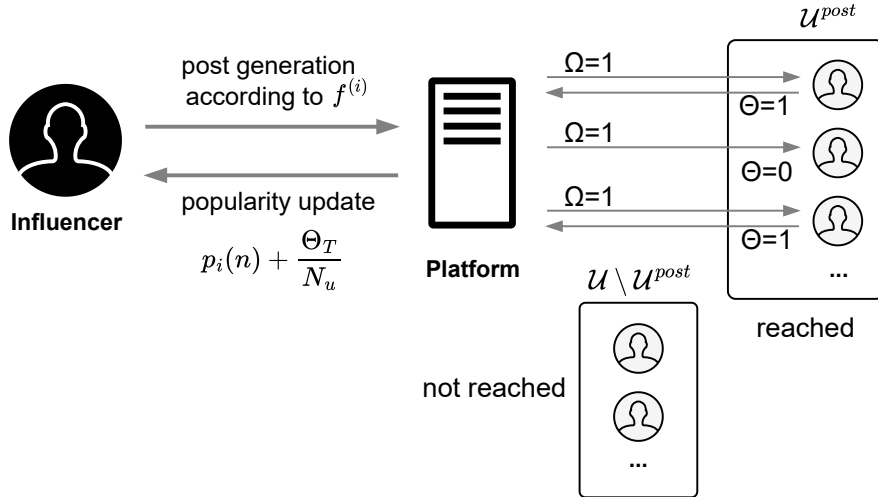


Figure 2: Schematic representation of the model dynamics. The figure highlights the proportions of users who view a particular post \mathcal{U}^{post} , i.e., those for which the random variable Ω equals 1. They react with their feedback Θ (e.g., *likes*), which depends on the opinion distance between them and the influencer i . Then the platform updates the influencer's i popularity.

receives the message when $\Omega(\omega) = 1$. The parameter ω can be interpreted as a *visibility* function from the influencer's perspective, as it affects the subset of users reached by its posts. In principle, a post can reach any user as the interaction network is a dynamic complete bipartite network between the set of nodes \mathcal{U} and \mathcal{I} whose links are defined by $\Omega(\cdot)$ (see Figure 3). The *visibility* function ω is a decreasing function of the opinion distance on the reference direction $d_r(n) = |x_r^{(u)}(n) - x_r^{(i)}(n)|$. This dependence embeds the concept of *homophily*, one of the main drivers of interaction on social networks[1]: individuals with strongly divergent opinions interact less frequently than like-minded individuals. Moreover, ω is increasing in the popularity ratio π_i . The higher the relative popularity of an influencer, the more users it can reach. This posts-users matching process constitutes the content *personalization* we consider in this paper (see Remark 2). Note that the *filtering* process for selecting the subset of users who receive the post is based on the opinion distance along the reference direction between each user and the influencer who made the post. This because we expect that influencers mainly attract users whose opinions are similar to

their main topic. For instance, politicians primarily attract users interested in the political landscape and with similar orientations. Adopting this distance to perform the user selection couples the dynamics in different directions, which would otherwise evolve independently of each other (see Remark 1).

Users express their *feedback* to a post on the platform through a Bernoulli random variable $\Theta\left(\theta(|x_j^{(u)} - x_j^{(i)}|)\right) \in \{0, 1\}$ whose parameter θ depends on the difference in opinion on the *actual* direction j of the contribution. Only posts that receive positive feedback, i.e., $\Theta = 1$, can influence the user's opinion, reflecting the tendency to ignore unappreciated content. The social media platform collects feedback from all reached users to update the popularity p_i of the posting influencer.

The update rule for the popularity of the posting influencer i reads as follows:

$$p_i(n+1) = p_i(n) + \frac{\Theta_T(\theta, \mathcal{U}^{post})}{N_u} \quad (1)$$

$$\Theta_T(\theta, \mathcal{U}^{post}) = \sum_{u \in \mathcal{U}^{post}} \Theta\left(\theta(|x_j^{(u)}(n) - x_j^{(i)}(n)|)\right) \quad (2)$$

where \mathcal{U}^{post} is the subset of users who were made aware of the post by the platform, i.e., those for whom $\Omega(\omega)$ takes the value one. The summation in the formula gives the aggregate feedback of all users who saw the post, which is normalized by the size of the population of regular users $|\mathcal{U}| = N_u$ to update the popularity. This normalization is introduced only to avoid excessive popularity growth of the influencers when the number of users becomes large. It does not affect the system dynamics, which depends only on the normalized popularity π_i , which is not affected by the scaling factor $1/N_u$.

The core of the dynamics is the opinion update rule, which prescribes how the user's opinion changes on the direction j of the post:

$$x_j^{(u)}(n+1) = \begin{cases} \alpha z_j^{(u)} + \beta x_j^{(u)}(n) + \gamma x_j^{(i)} & \text{if } \Omega(\omega(d_r, \pi_i)) = 1, \Theta(\theta(d_j)) = 1 \\ x_j^{(u)}(n) & \text{otherwise} \end{cases} \quad (3)$$

When the post reaches the user ($\Omega = 1$) who likes it ($\Theta = 1$), then the updated opinion is a convex combination, i.e., $\alpha + \beta + \gamma = 1$, of the current opinion $x_j^{(u)}(n)$, the prejudice $z_j^{(u)}$, and the opinion $x_j^{(i)}$ conveyed by the influencer through the post. The opinion is not updated if the user does not receive the post ($\Omega=0$) or does not like it even if it reaches them ($\Theta = 0$). While it is common in the literature to express the opinion update as the convex combination in Eq. (3), we present an equivalent formulation that sheds more light on the meaning of the update parameters, which are practically two:

$$x_j^{(u)}(n+1) = (1 - \beta) \left[\delta z_j^{(u)} + (1 - \delta)x_j^{(i)} \right] + \beta x_j^{(u)}(n), \quad \delta, \beta \in [0, 1] \quad (4)$$

(4) can be easily derived from (3) by noting that $\alpha + \gamma = 1 - \beta$, and defining $\delta \triangleq \frac{\alpha}{\alpha + \gamma}$. Thus, β represents the *inertia* of users, i.e., how slowly they change, and δ their *degree of stubbornness*, i.e., the relative impact of external opinions w.r.t their prejudice.

Remark 1. *The distance on the reference direction drives filtering action because we assume the platform is unaware of the specific topic associated with the post just created. At the same time, homophilic connections between users and influencers primarily depend on opinion similarity on the main topic of discussion. Note the joint effect in the model of the distance between the user’s opinion and the influencer’s opinion on the reference direction and the distance along the direction defined by the post’s topic. Both contribute to determining the likelihood for the user to provide positive feedback to the message.*

Remark 2. *Most OSNs have explicit subscriptions to influencers, i.e., the follow mechanism. Our approach does not explicitly represent such “long-term” relationships. However, by applying the function $\Omega(\omega(\cdot))$, we dynamically determine the set of users reached by each influencer. Essentially, followers are regenerated at each post-emission. The resulting network is a dynamic bipartite graph, see Figure 3, whose structure reflects a given degree of homophily of users’ connections. Typically homophily is one of the elements that mainly*

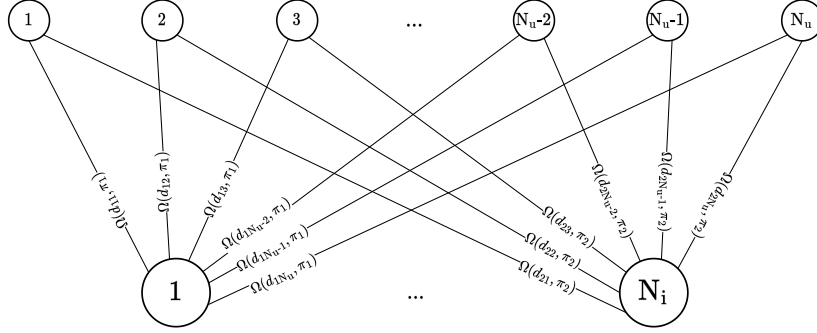


Figure 3: The bipartite dynamic structure of the underlying graph between regular users (top) and influencers (bottom) defined by the visibility function $\Omega \left(d_{iu}(n) = |x_r^{(u)}(n) - x_r^{(i)}(n)|, \pi_i(n) = \frac{p_i(n)}{\sum_i p_i(n)} \right)$ which depends on both the user-influencer opinion distance and the relative popularity of the influencer at a given time instant n .

influences users' choices when they select individuals to connect to [1]. In particular, for some domains (e.g., product adoption, which is also a good fit for our “competing” scenario in Section 6) homophily has emerged as the key driver governing the structure of the network [42].

At last, observe that, nowadays, most social media platforms (e.g., Facebook, Instagram, Twitter) do not only offer their users content they explicitly subscribe to, but also what they may like. The selection of such users is based on their previous activity on the platform. This mechanism reinforces the homophilic structure of the network and resembles what we are modeling.

Remark 3. In our framework, regular users are passive, as they merely consume content produced by influencers: this constitutes a rather simplistic assumption. First, users can share the posts they receive, which increases their reach. Secondly, users themselves write posts that reflect their opinion, influencing other users. The impact of active users is briefly discussed in Appendix G.

4. Observations from Online Social Networks

This section motivate some of our modeling choices by analyzing real-world social networks. We monitored on Facebook and Instagram the posts of 649 influencers for over 5 years. For a detailed description of the dataset used, see

Appendix A. One of the most important features introduced in this paper is the concept of *reference direction*, i.e., the main topic an influencer is interested in and on which they publish most of their posts, which is validated here. Moreover, we examine the post-generation process to justify the choice of a Poisson Point Process to describe it.

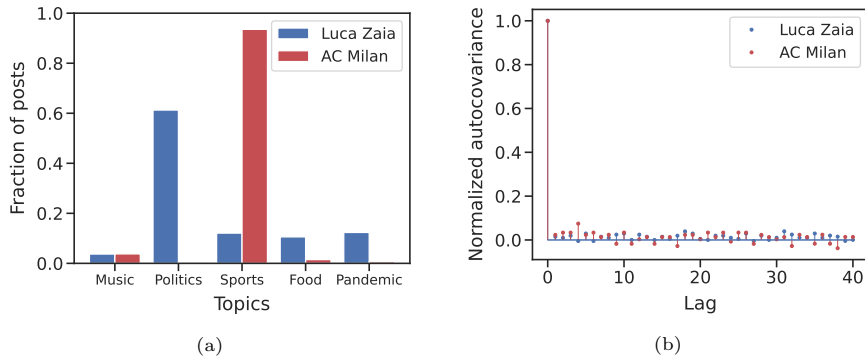


Figure 4: (4a) Percentage of labeled posts on each of the considered topics for Luca Zaia, an Italian politician, and AC Milan, an Italian football club. (4b) Normalized autocovariance function for both influencers on a *secondary* topic, i.e., music.

4.1. The reference direction

This section shows that influencers prefer to post about a specific topic. We have developed a post classifier that flags posts based on their topic. See Appendix B for details on the classification and filtering process on the data. We should point out that classifying posts on OSNs into topics is not straightforward, and interpreting the results should be done cautiously. First, the range of possible subjects discussed in a social network is practically countless. For practical reasons, we will only focus on a subset of five topics: Sports, Politics, Food and Cooking, Music, and Pandemics. These can be considered popular and general enough to cover a substantial fraction of the influencers' posts taken into consideration in our dataset.

After classification, we examined the distribution of posts on the topics for each influencer. In Figure 4a, we show two example influencers. In these two cases, the influencers have one topic on which they write most of their posts. Luca Zaia, an Italian politician, posts mainly about politics, and AC Milan, a

soccer club, discuss sports predominantly. This behavior supports the existence of a reference direction for influencers. Figure 5a shows the distribution of the proportion of posts dealing with the main topic of each influencer. Recall that this proportion was called *consistency* in the jargon of our model.

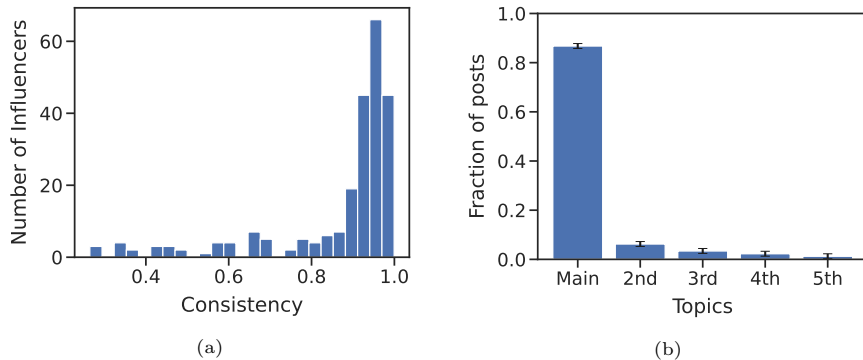


Figure 5: (5a) Distribution of the fraction of posts published on the main topic of interest by the subset of influencers considered in this experiment, i.e., their consistency. (5b) The average percentage of labeled posts on each topic in decreasing order for all the influencers considered. The 95% confidence interval for each average value is reported in the figure.

Most influencers have a clear *reference topic* on which they write more than half of their posts, i.e., with high consistency. Figure 5b shows the average per-topic percentage of all influencers in the dataset in descending order, regardless of the specific topic. On average, almost 90% of the posts are in the reference direction. We discovered that influencers with low consistency values are affected by the presence of news outlets in the considered profiles, for which the lack of a sharp main topic is sensible.

4.2. Independence of posts' generation process on secondary directions

Users interact in an OSN by posting content (i.e., text, images, videos) and receiving suggestions about what other users of the OSN posted, according to the filtering process set up by the social media platform. We examine the normalized autocovariance² between posts on each topic by looking at the chronological

²Given a wide-sense stationary process $\{X_n\}_n \in \mathbb{Z}$ with average μ and variance σ^2 , the normalized autocovariance is given by: $\rho_{XX}(i) = \frac{\mathbb{E}[(X_n - \mu)(X_{n+i} - \mu)]}{\sigma^2}$

sequence of the messages of the individual influencers. We perform this analysis only on *secondary* topics, i.e., those that differ from the influencer’s *reference*. We do it since influencers post less frequently on these topics, and it would be easier to detect a bursty behavior pattern (which would not be well captured by the Poisson process). Regarding the main topic, since the consistency of the influencers is generally relatively high, we expect the covariance to be rather small, (see Figure 5a). Indeed, the covariance on the main topic tends to zero by construction as $c^{(i)}$ approaches one.

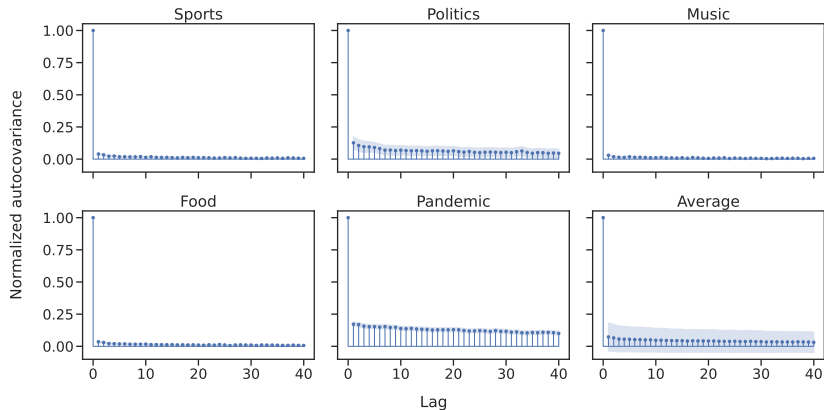


Figure 6: Mean normalized autocovariance values of the post-generation process for each secondary topic of all the influencers. The last plot represents the average value over all topics. The 95% confidence interval is shown in each plot.

In the previous section, we were able to assign a *reference direction* $r^{(i)}$ to each influencer. Here we look at the time series of the Influencers’ labeled posts. For each secondary direction $s_j^{(i)}$, we define an indicator function $\mathbb{1}_{\{post_{label}=s_j^{(i)}\}}$ that takes the value 1 if the post was labelled as $s_j^{(i)}$ and 0 otherwise. For each influencer, we thus obtain four sequences (recall that we consider five topics in total) of Bernoulli random variables indicating whether a post belongs to that particular direction. We calculated the normalized autocovariance function $a(t)$ for these sequences. Figure 4b shows two examples of such functions, limited to 40-time lags, for the profiles of Luca Zaia and AC Milan. The time is discretized, i.e., the actual time between postings is not considered: only the posting events matter. An autocovariance that equals zero everywhere except at $\tau = 0$ would

represent uncorrelated samples. In our case, the autocovariance takes moderate values in most cases ($\ll 1$). Therefore, it is reasonable to assume that the post-generation is independent, and a Poisson Point Process is an appropriate choice. Lastly, note that the autocovariance function for the *pandemic* topic takes larger values than for the other topics (see Figure 6), suggesting that the samples are weakly correlated. This fact is due to the exceptional public interest in the topic and because the outbreak of the epidemic only interested the last part of the considered time horizon.

5. Asymptotic Analysis of the Model

This section is devoted to the analytical study of the model. In particular, results are derived through a mean-field approach obtained by letting the number of users $N_u \rightarrow \infty$. In this situation, we will show that, under mild assumptions, the system converges to a unique steady state, independently from the initial condition. Moreover, in some cases, it is possible to analytically characterize the equilibrium value for the influencers' mean-popularity ratios $\bar{\pi}_i$ as well as users' mean opinion value $\bar{x}(z)$. Furthermore, transient analysis of the system can be carried out by describing the dynamics of the users through a Fokker-Plank equation. For simplicity, we restrict our investigation to the situation where the opinion space is one-dimensional. However, we remark that it is possible to extend the analysis to the more general case by following the same approach.

5.1. Mean field approach

When the number of users grows large, it is convenient to characterize the system state by the users' opinion *distribution* over the space. Moreover, from now on, we will refer to system dynamics over continuous time t .

Let $(X(t), Z(t)) = (X(t), Z)$ be the current position (opinion) and prejudice of a randomly selected user. We introduce the cumulative distribution function $F(x, z, t) = Pr[X(t) < x, Z < z]$. The corresponding probability density function is $f(x, z, t) = \frac{\partial^2}{\partial x \partial z} F(x, z, t)$. Note that, by hypothesis, there are no

dynamics along the z -axes, thus $h(z) = \int_x f(x, z, t) dx$ does not depend on t and corresponds to the initial distribution of users' prejudice. In Section 5.2, we will derive a Fokker-Plank equation for the evolution of the opinion distribution over time and space.

For what concerns the evolution of the popularity of a generic influencer i , recall that we distinguish between its absolute popularity value $p_i(t)$ and the normalized value $\pi_i = \frac{p_i(t)}{\sum_j p_j(t)}$. Influencer's popularities concentrate around their average $\bar{p}_i(t)$ as N_u grows large, as it can be easily shown. We can write down the equation for the evolution of the mean popularity $\bar{p}_i(t)$:

$$\frac{d\bar{p}_i(t)}{dt} = \frac{1}{N_u} \lambda f^{(i)} \int_x \int_z f(x, z, t) \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i, |x - x^{(i)}|) dz dx \quad (5)$$

Indeed, the rate at which the popularity of influencer i grows is proportional to its posting rate (term $\lambda f^{(i)}$) times the probability that a generic user at (x, z) provides positive feedback to the post generated at time t (integral term). Moreover, recall from Algorithm 1 that each positive feedback increases the absolute popularity of the influencer by $1/N_u$.

5.2. Fokker-Planck equation for the opinion distribution

The Fokker-Planck (FP) equation [43] is a standard tool that describes the evolution of an asymptotically large population of particles moving over a given domain according to a Brownian motion, which is locally characterized by the instantaneous average velocity and the relative variance. In general, both average and variance may depend on the point and time at which they are evaluated.

Furthermore, the FP approach has been successfully applied in the literature to approximate the evolution of a large population of particles moving according to more general laws. The stochastic process describing the movement of particles is approximated by a Brownian motion which fits the first two moments (average and variance of the velocity). For these reasons, FP description is referred to in the literature as a second-order approximation. Here, we essentially identify each user opinion with a particle moving over the space \mathcal{X}^2 (i.e., the set (x, z) with $x, y \in [a, b]$), and characterize by an average velocity $v_x(x, z, t)$

along the x -axes, and a variance $\sigma_x^2(x, z, t)$, in Section 5.3. As a consequence, the Fokker-Plank equation for the probability density function $f(x, z, t)$ (where $x, z \in [a, b]$) is given by:

$$\frac{\partial f(x, z, t)}{\partial t} = -\frac{\partial v_x(x, z, t)f(x, z, t)}{\partial x} + \frac{1}{2} \frac{\partial^2 \sigma_x^2(x, z, t)f(x, z, t)}{\partial x^2} \quad (6)$$

5.3. Identification of the parameters $v_x(x, z, t)$ and $\sigma_x^2(x, z, t)$

To compute $v_x(x, z, t)$, in the continuous-time FP approximation, we assume that for the effect of a post, “users/particles” reach their new position by moving at a constant speed during the interval ΔT equal to the average time $1/\lambda$ that elapses between the generation of two successive posts. Therefore, assuming that at time t a post is generated by user i , the following equation describes how the opinion of a user with prejudice z evolves from t to $t + \Delta T$:

$$x(t + \Delta T) = \alpha z + \beta x(t) + \gamma x^{(i)}(t)$$

Thus, the increment is:

$$\Delta x(i) = x(t + \Delta T) - x(t) = \alpha(z - x^{(i)}(t)) + (1 - \beta)(x^{(i)}(t) - x(t)) \quad (7)$$

where we remark that $\Delta x(i)$ here represents the change in position of a user in position x , providing positive feedback to a post of influencer i . We can then compute its average velocity as:

$$\begin{aligned} \mathbb{E}[v_x(x, z, t) \mid X(t) = x, Z = z] &= \mathbb{E} \left[\frac{[X(t + \Delta T) - X(t) \mid X(t) = x, Z = z]}{\Delta T} \right] \\ &= \sum_i \lambda f^{(i)} \Delta T \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i(t), |x - x^{(i)}|) \frac{\Delta x(i)}{\Delta T} \\ &= \sum_i \lambda f^{(i)} \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i(t), |x^{(i)} - x|) \Delta x(i) \quad (8) \end{aligned}$$

where $\theta(|x - x^{(i)}|)$ is the probability of providing positive feedback (users move only in this case), while $\omega(\bar{\pi}_i(t), |x - x^{(i)}|)$ is the probability with which a user in x is exposed to a post created by influencer i at time t . Indeed, users only move if they are exposed to the post and provide positive feedback. Note

that, to avoid a cumbersome notation, we have omitted the dependency on the time of the distance term $|x - x^{(i)}|$.

The variance of the velocity is given by the relation:

$$\begin{aligned}\sigma_x^2(x, z, t) &= \sum_i \lambda f^{(i)} \Delta T \theta \left(|x - x^{(i)}| \right) \omega \left(\bar{\pi}_i(t), |x - x^{(i)}| \right) \frac{(\Delta x^{(i)} - \mathbb{E}[v_x(x, z, t)] \Delta T)^2}{\Delta T^2} \\ &= \frac{1}{(\Delta T)^2} \sum_i f^{(i)} \theta \left(|x - x^{(i)}| \right) \omega \left(\bar{\pi}_i(t), |x - x^{(i)}| \right) (\Delta x^{(i)} - \mathbb{E}[v_x(x, z, t)] \Delta T)^2\end{aligned}$$

5.4. Steady state analysis

Now we direct our attention to the existence of stationary solutions for the system. Stationary solutions of (6) necessarily satisfy:

$$\frac{\partial}{\partial x} \left(-v_x(x, z) f(x, z) + \frac{1}{2} \frac{\partial \sigma_x^2(x, z) f(x, z)}{\partial x} \right) = 0$$

where $v_x(x, z)$ and $\sigma_x^2(x, z)$ must be constant over time. This requires the normalized popularities to be static (i.e. $\omega(\cdot)$ to be constant over time). From the previous equation, integrating both sides with respect to x , we get:

$$\left(-v_x(x, z) f(x, z) + \frac{1}{2} \frac{\partial \sigma_x^2(x, z) f(x, z)}{\partial x} \right) = c_0(z) \quad (9)$$

where $c_0(z)$ is a uni-dimensional arbitrary in z . Now, observe that, for every z , previous equation is a first order linear ODE in x , and therefore an explicitly solution for $f(x, z)$ can be obtained:

$$\begin{aligned}f(x, z) &= \left(c_1(z) \exp(A(x, z) - A(a, z)) + \right. \\ &\quad \left. c_0(z) \exp(-A(x, z)) \int_a^x \exp(A(\theta, z)) d\theta \right) h(z) \quad (10)\end{aligned}$$

where

$$A(x, z) = \int_a^x \eta(u, z) du \quad \eta(x, z) = -2 \frac{v_x(x, z) - \frac{1}{2} \frac{\partial \sigma_x^2(x, z)}{\partial x}}{\sigma_x^2(x, z)}.$$

Function $c_0(z)$ can be obtained by imposing boundary conditions:

$$\left(-v_x(x, z) f(x, z) + \frac{1}{2} \frac{\partial}{\partial x} \sigma_x^2(x, z) f(x, z) \right) \Big|_{x=a, b} = 0. \quad \forall z$$

which leads to $c_0(z) = 0$, while function $c_1(z)$ is determined by imposing the normalization condition:

$$\int f(x, z) dx = h(z).$$

Observe that when $\sigma_x^2(x, z) \rightarrow 0$ and $\frac{\partial \sigma_x^2(x, z)}{\partial x} \rightarrow 0$, from (9), with $x_0(z) = 0$, we obtain that necessarily the mass concentrates around the points for which $v_x(x, z) = 0$. Such points, improperly referred to in the following as *equilibrium points*, will be characterized analytically later on.

Turning our attention to popularity dynamics, recall that stationary conditions necessarily imply normalized popularities to be constant over time:

$$\bar{\pi}_i(t) = \bar{\pi}_i \quad \forall i$$

On the other hand, absolute popularities naturally grow over time. However, the ratio between any two of them (say i, j) must converge to a constant value c_{ij} equal to the ratio of their corresponding normalized popularities:

$$\frac{\bar{p}_i(t)}{\bar{p}_j(t)} = c_{ij} = \frac{\bar{\pi}_i}{\bar{\pi}_j} \quad \forall i, j \in \mathcal{I}, i \neq j \quad (11)$$

Now observe that in stationary conditions the right-hand side. of (5) does not depend on time. Therefore (5) admits the following trivial solution:

$$\bar{p}_i(t) = \left(\lambda f^{(i)} \int_x \int_z \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i, |x - x^{(i)}|) dF(x, z) \right) \frac{t}{N_u} + \bar{p}_i(0) \quad (12)$$

Therefore, we meet conditions (11) for any $t \geq 0$, iff normalized popularities of influencers $\{\bar{\pi}_i\}_i$ satisfy the following system of equations:

$$\begin{aligned} \lambda f^{(i)} \int_x \int_z \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i, |x - x^{(i)}|) dF(x, z) &= c \bar{\pi}_i \quad \forall i, \text{ for some } c \in \mathbb{R}^+ \\ \text{s.t. } \bar{\pi}_i &\geq 0 \text{ and } \sum_i \bar{\pi}_i = 1. \end{aligned} \quad (13)$$

and the initial condition $\{p_i(0)\}_i$ satisfies (11), i.e., $p_i(0) = k \bar{\pi}_i$ for some $k > 0$.

Let

$$k_i(\bar{\pi}_i) \triangleq \lambda f^{(i)} \int_x \int_z \theta(|x - x^{(i)}|) \omega(\bar{\pi}_i, |x - x^{(i)}|) dF(x, z) \quad \bar{\pi}_i \in [0, 1] \quad (14)$$

We can show that:

Theorem 5.1. *Solutions of (13) always exist whenever $k_i(\cdot) \in C_1[0, 1]$, $k_i(\cdot)$ is increasing, continuous and strictly concave.*

The proof is reported in Appendix C.

We remark that when $k_i(0) > 0 \forall i$, the solution is always unique with $\bar{\pi}_i \in (0, 1)$. Instead when $k_i(0) = 0$ for some i , the solution is not guaranteed to be unique.

Now, the problem is how to jointly solve for stationary solutions of $\{\bar{\pi}_i\}_i$ and $F(x, z)$. In a schematic way, on the one hand, we have shown that given $\bar{\pi} = \{\bar{\pi}_i\}_i$, and $h(z)$, we can uniquely determine a $F_{\bar{\pi}}(x, z) = \mathcal{H}(\bar{\pi})$, where $F_{\bar{\pi}}(x, z) = \int_{-\infty}^x \int_{-\infty}^z f_{\bar{\pi}}(y, w) dy dw$ is the opinion distribution of users resulting from fixed influencers' popularities $\bar{\pi}$ (by (10)).

On the other hand, under the conditions: $k_i(\cdot) \in C_1[0, 1]$, $k_i(\cdot)$ is increasing and strictly concave, $k_i(0) > 0 \forall i$, given $F(x, z)$, we can obtain a $\bar{\pi}_F = \mathcal{G}(F(x, z))$ that uniquely corresponds to i (Theorem 5.1). The existence of a unique fixed point for the joint system of (stationary) users' opinions and influencers' popularities is guaranteed under the condition that the operator $\mathcal{H} \circ \mathcal{G}(\cdot)$ is a contraction over a complete space.

Theorem 5.2. *Under the assumption that both $\omega(\cdot, \cdot)$ and $\theta(\cdot)$ exhibit a sufficiently weak dependence on their variables, the operator $\mathcal{H} \circ \mathcal{G}(\cdot)$ is a contraction over a complete space, and therefore a unique stationary solution exists.*

The proof is reported in Appendix C.

5.5. Asymptotic analysis of the fluid limit

Previous theoretical analysis is, unfortunately, non-constructive, meaning that it does not allow for direct computation of stationary solutions of our dynamical system. To complement the previous analysis, in this section, we propose a methodology to numerically compute stationary solutions, even in multi-dimensional scenarios, under stricter assumptions. In particular, if the

FP approach is a second-order approximation (matching the first to moments of the instantaneous velocity), the approach proposed in this section, and referred to as *fluid limit*, is a first-order approach (matching only the first moment of the velocity and assuming the variance, as well as its spatial derivative, to be negligible). Therefore we expect that *fluid limit* provides reasonable good predictions when $N_u \rightarrow \infty$, $\lambda \rightarrow \infty$, $1 - \beta \rightarrow 0$ (keeping $\lambda(1 - \beta)$ constant). Previous assumptions, indeed, imply that $\sigma_x^2(x, z) \rightarrow 0$ and $\frac{\partial \sigma_x^2(x, z)}{\partial x} \rightarrow 0$.

5.5.1. Mean opinion assuming that normalized popularities converge

As already observed in Section 5.4, recall that, given $\bar{\pi} = \{\bar{\pi}_i\}_i$, the distribution of users with a given prejudice z concentrates around *equilibrium points*, i.e., points $\bar{x}(z)$ at which $v(x, z)$, as given in (8), is null (i.e. $v(\bar{x}(z), z) = 0$). Therefore, points $\bar{x}(z)$ must satisfy equation:

$$0 = \sum_i f^{(i)} \omega(\bar{\pi}_i, |\bar{x} - x^{(i)}|) \theta(|\bar{x} - x^{(i)}|) \left(\alpha(z - x^{(i)}) + (1 - \beta)(x^{(i)} - \bar{x}) \right) \quad (15)$$

Defining for compactness $d^{i, \bar{x}} = |\bar{x} - x^{(i)}|$ and recalling $\gamma = 1 - \alpha - \beta$, from (15) we get:

$$\bar{x}(z) = \frac{\alpha}{1 - \beta} z + \frac{\gamma}{1 - \beta} \frac{\sum_{i \in \mathcal{I}} f^{(i)} \omega(\bar{\pi}_i, d^{i, \bar{x}}) \theta(d^{i, \bar{x}}) x^{(i)}}{\sum_{i \in \mathcal{I}} f^{(i)} \omega(\bar{\pi}_i, d^{i, \bar{x}}) \theta(d^{i, \bar{x}})} \quad (16)$$

We can rewrite this relation in terms of the *degree of stubbornness* of Eq. (4) as $\bar{x} = \delta z + (1 - \delta) \frac{\sum_{i \in \mathcal{I}} f^{(i)} \omega(\bar{\pi}_i, d^{i, \bar{x}}) \theta(d^{i, \bar{x}}) x^{(i)}}{\sum_{i \in \mathcal{I}} f^{(i)} \omega(\bar{\pi}_i, d^{i, \bar{x}}) \theta(d^{i, \bar{x}})}$, $\delta \in [0, 1]$. By so doing it should be clear that the assumption $\beta \rightarrow 1$ is well-founded, reinforcing the idea that β is associated with the *inertia* of the system (see the end of Section 3.2) but does not affect the equilibrium points. This assumption is required to avoid too large oscillations of users' opinions in response to a single post generated by an influencer, which may reduce the accuracy of our mean-field approximation, validated in Appendix D.

The hypothesis is not restrictive: since β represents the weight individuals give to their current opinion, we can reasonably assume that users do not dramatically change their opinion in response to single post events.

5.5.2. Normalized popularities assuming opinion convergence

Here we assume that users with prejudice z are concentrated in opinion point $\bar{x}(z)$, and we look for the stationary popularity ratios $\bar{\pi}_i$. To simplify the expressions, we introduce the quantity $F_i(\bar{\pi}_i) \triangleq \int_z f^{(i)} \omega(\bar{\pi}_i, d^{i, \bar{x}(z)}) \theta(d^{i, \bar{x}(z)}) h(z) dz$.

Observe that solutions of (13) are necessarily of the form:

$$\bar{\pi}_i = \frac{F_i(\bar{\pi}_i)}{\sum_{j \in \mathcal{I}} F_j(\bar{\pi}_j)} \quad (17)$$

where c appearing in (13) is given by $c = \frac{1}{\sum_{j \in \mathcal{I}} F_j(\bar{\pi}_j)}$. Under the assumption that $\omega(\cdot, \cdot)$ is concave in its first argument (for any choice of the second), Theorem 5.1 guarantees the existence of such solutions for every choice of function $\bar{x}(z)$. Moreover, even in the more general case, i.e., when $\omega(\cdot, \cdot)$ is non-concave in its first argument, solutions of (17) can be found numerically in many cases, through a fixed point iteration method.

To conclude, observe that a pair $(\bar{x}(z), \{\bar{\pi}_i\}_i)$ represents a stationary solution if it jointly satisfies (16) and (17). The existence of such a solution can be, again, only verified *numerically* through a fixed point approach.

At last, note that, in the special case in which all users have the same prejudice z we can rewrite (16) as:

$$\bar{x} = \frac{\alpha}{1 - \beta} z + \frac{\gamma}{1 - \beta} \frac{\sum_{i \in \mathcal{I}} F_i(\bar{\pi}_i, \bar{x}) x^{(i)}}{\sum_{i \in \mathcal{I}} F_i(\bar{\pi}_i, \bar{x})} = \frac{\alpha}{1 - \beta} z + \frac{\gamma}{1 - \beta} \sum_{i \in \mathcal{I}} \bar{\pi}_i x^{(i)} \quad (18)$$

which provides a direct formula for the mean opinion \bar{x} in terms of the normalized popularities $\bar{\pi}_i$ and the influencers' opinions $x^{(i)}$.

6. Monte-Carlo approach

This section presents a selection of results obtained through a Monte-Carlo approach. We vary model parameters to explore the impact of content personalization and influencers' characteristics. We focus on the two main dynamic variables of the system: the average opinion \bar{x} of regular users and the normalized popularities $\{\bar{\pi}_i\}_i$ for influencers. Note that the quantities shown in this

section, i.e., the pair $(\bar{x}, \bar{\pi})$, are empirical averages over multiple runs, and over all the regular users, as far as \bar{x} is concerned. Hence, they can be regarded as empirical, finite system approximations of the quantities defined in the previous section, which refer to the limiting case of an infinite population of users with the same prejudice z , and where $\lambda \rightarrow \infty$ and β approaches 1. Moreover, in some cases, we omit the results on average user opinion to save space as it is tightly coupled with the normalized popularities, as observed in the previous section.

Lastly, to facilitate the interpretation of results, we restrict ourselves to the case of two “*competing*” influencers. We are interested in determining the conditions under which an individual attains higher $\bar{\pi}_i$ than the other. We say that influencer i “wins” over the opponent, which is not anymore visible over the platform, when $\bar{\pi}_i \rightarrow 1 \implies \omega(\cdot, \bar{\pi}_j) \rightarrow 0 \forall j \in \mathcal{I}$. Note that this scenario is by no means trivial and is relevant in various applications, for example, in marketing (e.g., two brands promoting the same product) or in election campaigns (e.g., two candidates of different parties). We provide further details on the scenario in Section 6.1. In Section 6.2 we show final opinion distributions of the regular users in a few paradigmatic cases. Then, in Section 6.3, we present the behavior as function of publication frequency $f^{(i)}$, and in Section 6.4 as function of consistency $c^{(i)}$. We provide additional details in Appendix E.

6.1. Description of the scenario

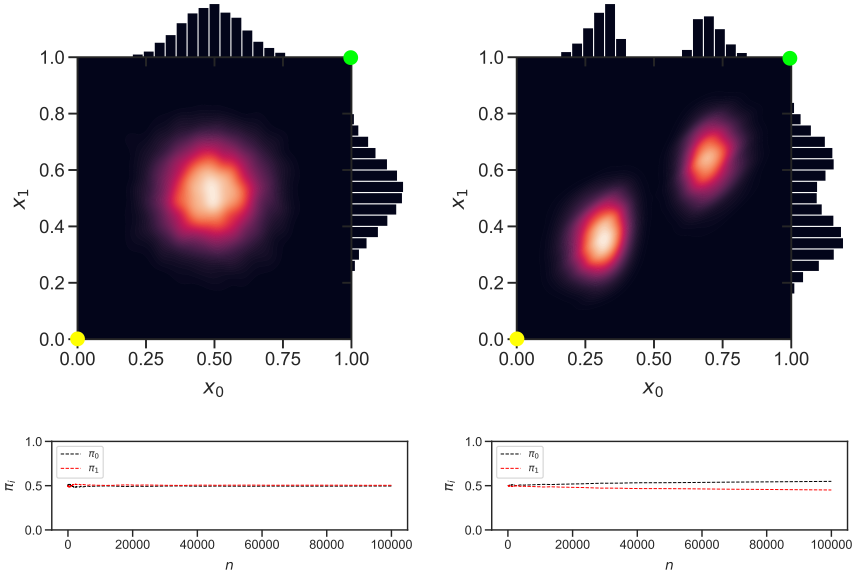
The default parameters of our reference scenario are reported in the Appendix (Table E.2), unless otherwise explicitly stated. As mentioned earlier, we consider the case of two “competing” influencers, i.e., $N_i = 2$. The opinion space is supposed to be bi-dimensional. We assume that $x_j^{(0)} = 0$ and $x_j^{(1)} = 1 \forall j \in \{1, 2\}$ hence the influencers are placed on two antipodal vertices of the square $[0, 1]^2$. We consider influencers with different reference directions $r^{(0)} \neq r^{(1)}$. Furthermore, most regular users initially take a *moderate* position on both topics. More precisely, initial opinions, which coincide with prejudices, are distributed according to a Beta distribution, independently on each axis, with shape parameters $a = b = 10$ (Fig. E.17 in the Appendix).

We take as $\omega(\cdot)$ a Gaussian function similar to the *trust* function in [44], but modulated by $\bar{\pi}_i$: $\omega(d_r^{i,u}, \bar{\pi}_i) = e^{-\rho \frac{(x_r^{(u)} - x_r^{(i)})^2}{\bar{\pi}_i}}$. Here, the coefficient ρ is a parameter that controls the extent to which the social media platform filters content, i.e., which expresses the homophilic degree over the network in a synthetic way. Small values of ρ correspond to *smooth* personalization, i.e., influencers can reach users whose opinion strongly differs from theirs. Conversely, high values of ρ correspond to *sharp* personalization: only close users (in the opinion space) are reachable with non-negligible probability. The function $\theta(\cdot)$ is assumed to be a decreasing, linear function of the opinion difference: $\theta(d_j^{i,u}) = 1 - |x_j^{(i)} - x_j^{(u)}|$.

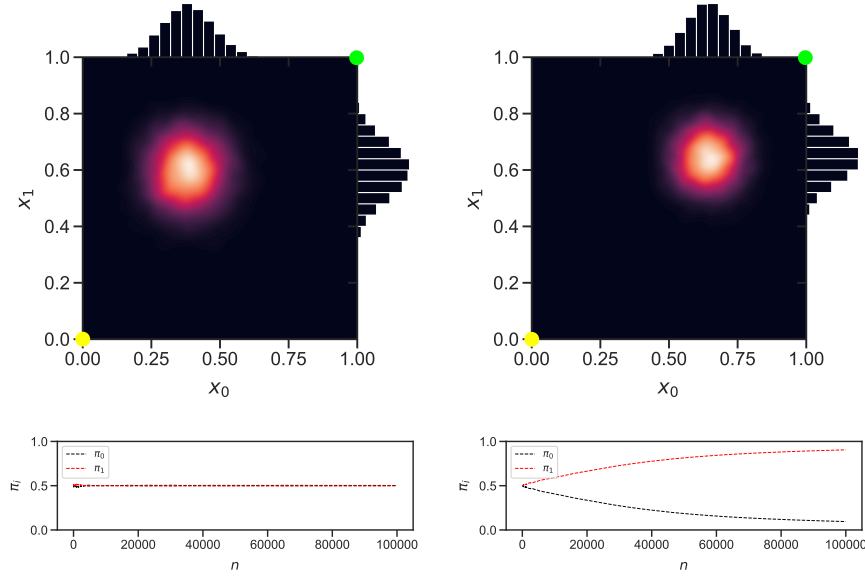
6.2. Opinion configuration considering combinations of reference directions

In the following sections, we will focus primarily on the influencer perspective by observing $\bar{\pi}_i$ as a function of their parameters. Here, we present possible final opinion configurations of the users' population, in a symmetric scenario in terms of influencers' characteristics, i.e., frequency of publication $f^{(0)} = f^{(1)}$ and consistency $c^{(0)} = c^{(1)}$. As before, they hold opinions $\mathbf{x}^{(0)} = (0, 0)$ and $\mathbf{x}^{(1)} = (1, 1)$. We consider both the case of same (Fig. 7a and 7b) and different (Fig. 7c and 7d) reference directions, assessing the impact of smooth (Fig. 7a and Fig. 7c) and strict content personalization (Fig. 7b and Fig. 7d).

In Figure 7a, we observe only a negligible perturbation with respect to the initial distribution (see Figure E.17). In this case, the platform practically does not filter the content, so every post reaches all users. From a regular user's perspective, individuals are exposed to nearly identical forces, i.e., "opposite" stimuli from the two influencers, which almost entirely cancel each other out. In Figure 7b, the impact of *sharp* personalization is clear: the filtering effect introduced by the platform leads to the emergence of two *echo chambers*, whose membership is determined mainly by the user's prejudice. Each user reaches an equilibrium point at which the resultant attraction induced by the two influencers is balanced by the attraction exerted by its prejudice. Interestingly, users also tend to cluster in the non-reference direction (x_1 in Fig. 7b) and align



(a) Same reference dir. $r^{(0)} = r^{(1)} = 0$
Smooth personalization $\rho = 5.0$ (b) Same reference dir. $r^{(0)} = r^{(1)} = 0$
Strict personalization $\rho = 5.0$



(c) Different reference dir. $r^{(0)} = 0, r^{(1)} = 1$
Smooth personalization $\rho = 5.0$ (d) Different reference dir. $r^{(0)} = 0, r^{(1)} = 1$
Strict personalization $\rho = 5.0$

Figure 7: Users' opinion distribution and normalized popularities. The two influencers have same consistency $c^{(0)} = c^{(1)} = 0.8$ and same frequency $f^{(0)} = f^{(1)} = 0.5$. The distributions are obtained as the time average of the opinion distribution in one realization of the process.

their opinion with the influencer associated with the echo chamber they end up in. We remark that this is a metastable condition, i.e., the influencers have not yet reached a stable equilibrium, as the gap in the π_i in Fig. 7b hints. One of the two influencers will eventually “win” (similar to Fig. 7d) but in a much longer time horizon, which may be unreasonable.

Figures 7c and 7d refer to the case of different reference directions: the two influencers do not primarily compete on the same topic. In Figure 7c, it is clear that there is no competition on their reference directions as the two influencers are able to attract users to their *reference opinion*, i.e., $x_{r(0)} = 0$ for $i = 0$ and $x_{r(1)} = 1$ for $i = 1$. This is a particularly relevant case, whose occurrence is linked indissolubly to the newly introduced concept of *reference direction*. In the last scenario, Figure 7d, the influencer $i = 1$ “wins”, i.e., $\bar{\pi}_1 \rightarrow 1$, which brings public opinion closer to its belief $\mathbf{x}^{(1)}$ on both issues. The final users’ opinion does not coincide with $\mathbf{x}^{(1)}$ because users are anchored by their prejudice. Note that here *sharp* personalization leads to a situation where only one individual monopolizes the public scene. To better understand the dynamics, we simulated this scenario with 10 different simulator seed selections: 5 times influencer $i = 1$ won, 3 times influencer $i = 0$ won, and in 2 cases, the system did not reach full convergence after $N_i = 100000$ iterations. The nature of the equilibrium point ($\pi_0 = \frac{1}{2}, \pi_1 = \frac{1}{2}$) appears to be *unstable*. Stochastic fluctuations of the system state bring it to one of the two *asymptotically stable* configurations: ($\pi_0 = 1, \pi_1 = 0$), where $i = 0$ wins, or ($\pi_0 = 0, \pi_1 = 1$), where $i = 1$ wins.

6.3. Behavior as a function of the frequency of publication

The frequency of publication $f^{(i)}$ is one of the basic parameters that characterize influencers. The higher $f^{(i)}$, the higher the *structural advantage* of the influencer because it more frequently reaches users through posts, attracting them to its own opinion. In this section, we examine the value of mean normalized popularity $\bar{\pi}_0$ as a function of $f^{(0)}$. Note that in the case of two influencers, $f^{(1)} = 1 - f^{(0)}$. We performed this experiment by fixing the consistency of the two influencers: $c^{(0)} = c^{(1)} = 0.8$, which is approximately the

average consistency observed on real-world data (Figure 5a).

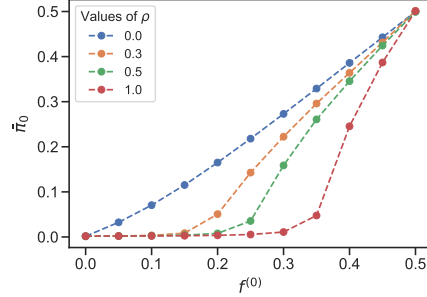


Figure 8: Popularity ratio $\bar{\pi}_0$ of influencer 0 as function of the publication rate $f^{(0)}$. Each point is obtained by averaging over 100 time samples and 10 different process realizations. We consider different levels of personalization by varying the parameter ρ . The two influencers have the same consistency $c^{(0)} = c^{(1)} = 0.8$. Recall that $f^{(0)} = 1 - f^{(1)}$ and the identity of the influencer is arbitrary. This “symmetry” allows us to readily infer the behavior of the normalized popularity $\bar{\pi}_0$ for $f^{(0)}$ in the range of $[0.5, 1.0]$.

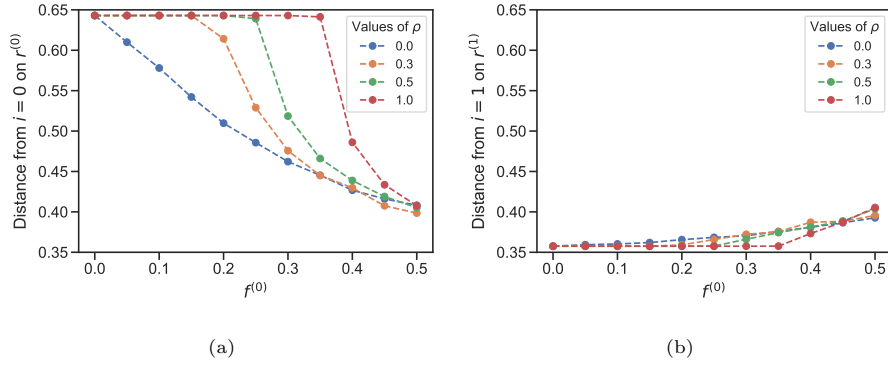


Figure 9: Opinion distance of the influencers’ opinion on their reference direction $x_{r^{(i)}}^{(i)}$ and the average opinion of the regular users’ population on the same direction $\bar{x}_{r^{(i)}}$. Various degrees of personalization are considered, tuning the parameter ρ . The setting is the same as that of Figure 8.

In Figure 8, we consider different levels of personalization by varying the parameter ρ in the exponent of the visibility function ω . We see that the higher the degree of personalization (i.e., the higher the value of ρ), the lower the normalized popularity of influencer $i = 0$, for any given $f^{(0)}$. This result suggests that algorithmic personalization favors the *structurally advantaged* individual, i.e., the one with higher $f^{(i)}$. This mechanism, in turn, leads to more radical po-

sitions in the population of regular users, as the platform preferentially exposes them to the belief of the *advantaged* influencer. Figure 9 clearly shows this behavior. Note that for high values of ρ , the average user opinion exhibits a significant bias toward the structurally advantaged influencer. Such bias persists up to a critical value of posting frequency. For example, when the personalization parameter is $\rho = 1.0$, the critical posting frequency value is roughly 0.35; when the personalization parameter is $\rho = 0.5$, the critical posting frequency value is approximately 0.25. Below this critical threshold the *advantaged* influencers “wins”, i.e., its normalized popularity $\bar{\pi}_i$ approaches 1, completely shadowing the opposing influencer. In Figure 9b, the opinion variation is limited since $f^{(1)} = 1 - f^{(0)} > 0.5$ implies that $i = 1$ exerts a strong influence over $r^{(1)}$ and suffers little competition from $i = 0$ (as $c^{(0)} = c^{(1)} = 0.8$ and $r^{(0)} \neq r^{(1)}$). For instance, for $f^{(0)} = 0$ the opinion values are those admitted by users’ prejudice and a single *winning* influencer (similar to what we discuss Appendix D.2.1). Fig. 9a and 9b provide complementary information. Moreover due to the aforementioned symmetry, the reader can easily understand how the system would evolve for $f_0 \in [0.5, 1]$.

We argue that content filtering in OSN potentially threatens opinion diversity. This premise is inextricably linked to the goal of usage maximization [33] pursued by the social media platform. Indeed, many platforms indeed prefer to suggest just *similar* content rather than exposing individuals to radically different opinions, hence often avoiding the so-called *serendipity*.

6.4. Behavior as a function of the consistency

In Section 4.1, we showed the existence of a reference direction for real influencers. Here, we investigate the impact on dynamics of the extent to which an influencer publishes on its reference direction, i.e., its consistency $c^{(i)}$. In this experiment, we consider two influencers with the same posting frequency $f^{(0)} = f^{(1)} = 0.5$, different reference directions $r^{(0)} = 0$, $r^{(1)} = 1$, and we let $c^{(0)}$ vary while keeping $c^{(1)}$ fixed. We report plots for a few choices of $c^{(1)}$ since the popularity pattern as a function of $c^{(0)}$ depends on the characterization of

the competing influencer $i = 1$. Figure 10 shows that consistency does not significantly affect the normalized popularities when personalization is *smooth* ($\rho = 0.0001$). In contrast, it becomes relevant when the platform applies *sharp* personalization to the content ($\rho = 1$). We consider two cases. First, that of *high* $c^{(1)}$ (Figs 10a and 10b), which is in line with the empirical evidence of Section 4.1. Second, the *low-consistency* scenario (Figs 10c and 10d) which has an interesting interpretation that goes beyond the scope of this article and is briefly discussed in Appendix G.2.

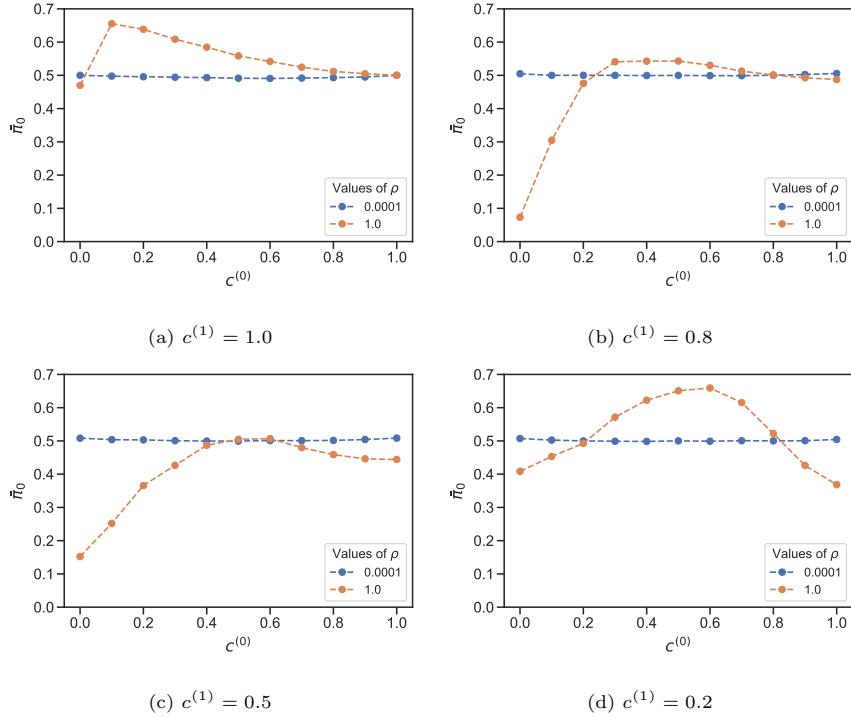


Figure 10: Popularity ratio $\bar{\pi}_0$ of i_0 as function of its consistency $c^{(0)}$, while considering $f^{(0)} = f^{(1)} = 0.5$ and keeping fixed the consistency of the second influencer at (10a) $c^{(1)} = 1.0$, (10b) $c^{(1)} = 0.8$, (10c) $c^{(1)} = 0.5$, and (10d) $c^{(1)} = 0.2$. The two colors represent two different levels of personalization (i.e., smooth and sharp). Each point is obtained by averaging over 100 time samples and 10 different realizations of the process.

In Figure 10a, $c^{(1)} = 1$ means that influencer $i = 1$ posts exclusively in the reference direction $r^{(1)}$. The corner cases, in which both influencers post all their posts in one direction, are i) $c^{(0)} = 1 - c^{(1)}$, where both post on $r^{(1)}$ but $i = 1$

has a slight advantage as it is posting on its reference direction where filtering occurs, and ii) $c^{(0)} = c^{(1)}$, which is a symmetric scenario. This dichotomy is also found in Fig. 10b, 10c and 10d. Whenever the orange curve approaches $\bar{\pi}_i = 0.5$ and $c^{(0)} = c^{(1)}$ or $c^{(0)} = 1 - c^{(1)}$, the influencer with the largest share of posts in the reference direction has a slight advantage. Again in Fig. 10a, if $0 < c^{(0)} < 1$, the influencer $i = 0$ posts in both directions and does not face competition over $r^{(0)}$. Therefore, $i = 0$ attracts users towards its “reference opinion” $x_0^{(0)}$, which in turn increases the chance of reaching users while competing with the other influencer in the non-reference direction (content filtering is performed with respect to the distance on the reference direction). In this rather extreme case, the lower the consistency $c^{(0)}$, the higher the proportion of posts on $r^{(1)}$, and the higher the final value of $\bar{\pi}_0$ as more posts “compete” for users’ attention with $i = 1$. The other scenarios are not as easy to interpret. However, all 10a-10d are consistent in pointing out that influencer $i = 0$ has a *structural advantage* roughly when its consistency is $\min(c^{(1)}, 1 - c^{(1)}) < c^{(0)} < \max(c^{(1)}, 1 - c^{(1)})$.

Figure 10 suggests that a value of consistency around 0.5 is nearly optimal for any value of the opponent’s $c^{(1)}$. This observation reflects the natural tendency of people to seek varied content. We evaluated $\operatorname{argmax}_{c^{(0)}} \bar{\pi}_0(c^{(1)})$, and this is indeed true for $c^{(1)} < 0.8$. However, for high values of $c^{(1)}$, $i = 0$ is better off reducing its consistency, i.e., post less on its reference direction $r^{(0)}$ and more on the reference topic $r^{(1)}$ of the opponent (recall $r^{(0)} \neq r^{(1)}$). This behavior points towards another potential hazard of *content personalization*. If influencer $i = 0$ is well-known on a platform as it deals with topic $r^{(0)}$ and starts posting massively on the other topic $r^{(1)}$ ($c^{(0)}$ drops low), it will gain an advantage over the opponent. This is because the platform filters posts considering the reference direction $r^{(0)}$ where $i = 0$ faces little competition from $i = 1$ (since $c^{(1)}$ is high) and therefore can attract users more easily over that topic. This results in an increase in popularity, allowing it to reach more users. This is an indication of how an influencer can leverage its *importance* in the reference direction to attract users in a non-reference topic favored by content personalization.

7. Online social network data

This section examines data collected from Facebook and Instagram social networks (see Appendix A) and compares the observed behavior with some of the findings of our Communication Asymmetry model.

7.1. Correlation between frequency of publication and popularity

In previous sections, especially in Section 6, we discussed *structural advantage* from the influencer’s point of view. A key advantage parameters is the publication frequency $f^{(i)}$: the higher $f^{(i)}$, the greater the advantage (see Figure 8). In this section, we attempt to validate this finding by correlating the frequency of publication of influencers with their popularity growth, using the total number of *followers*, i.e., the number of people subscribed to the *profile*, as a proxy for popularity. We consider temporal sequences from Instagram on a sample set of 110 influencers.

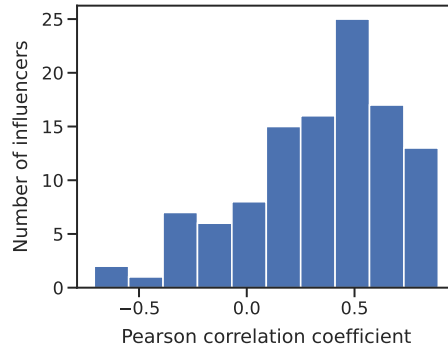


Figure 11: Distribution of the Pearson correlation coefficient between the monthly number of posts and popularity growth (in terms of *followers*).

For each influencer, we considered a temporal granularity of one month, determined the number of posts during this period, and calculated the relative change in the number of *followers* considering the values at the beginning and end of the interval. Then for each user, we calculated the Pearson correlation coefficient between the number of posts and the relative variation of followers

in the month. In Figure 11, we show the distribution of these correlation coefficients. Results suggest that there exists, in general, a positive correlation between the two quantities, i.e., influencers with aggressive posting habits tend (but not always) to get more followers, which likely favors them when in competition with other influencers on social media platforms. This is consistent with the model predictions shown in Section 6.3.

7.2. Case Study: Italian government crisis in August 2019

In June 2018, a few months after the general elections, Giuseppe Conte was appointed Italian Prime Minister. Two parties formed his supporting coalition: Movimento 5 Stelle (his party, holding the relative majority of the Italian Parliament) and Lega, whose leader was Matteo Salvini. In August 2019, Salvini decided to withdraw Lega’s support to the government, starting a crisis aimed at driving Italians to new elections and gaining more votes. However, Movimento 5 Stelle reached an agreement with other parties to form a new government, and on September 5, 2019, Giuseppe Conte became Prime Minister for the second time, excluding Lega from the new administration.

In this section, we apply the proposed model to reproduce the sudden rise of Giuseppe Conte’s popularity in social networks during this government crisis. We exploit the multidimensional capability of the model considering two directions: *Politics*, reference topic for Salvini and Conte, and attitude toward government fall, *End government* (see Figure 12).

In the opinion space, we assume Salvini has a more radical political viewpoint ($x_{Politics} = 0$), while Conte has an opposing and more moderate position ($x_{Politics} = 0.76$). We set these values in a somewhat discretionary manner. However, we provide a sensitivity analysis in Appendix F, proving that results are robust. Conversely, it is safe to assume that the two politicians take opinions at the extreme of the spectrum on the attitude toward government fall, i.e., Salvini has $x_{Endgovernment} = 1$ and Conte has $x_{Endgovernment} = 0$. Moreover, we consider a population with a *moderate* initial opinion on *Politics* (centered at $x_{Politics} = 0.5$, see Figure 12a). On *End government*, we sought a distribution

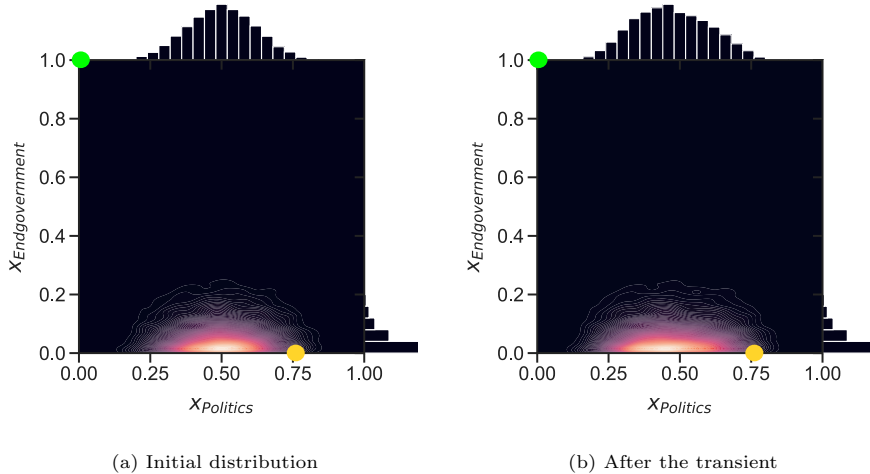


Figure 12: (12a) Initial population’s distribution along the *Politics* and *End government* direction and (12b) after transient, at the start of the observed period. The opinion position of the two leaders in the space is depicted with a green (Salvini) and a yellow (Conte) point.

that could explain the sudden popularity leap of Conte. We found that the user population must be strongly biased towards Conte’s opinion (Figure 12).

Some further simplifying assumptions are necessary to apply the model. We assume that the two politicians have a *consistency* $c^{(i)}$ of exactly one (real values are often close to this value, see Figure 5a). Moreover, Giuseppe Conte and Matteo Salvini are the only influencers. Although this hypothesis is restrictive, in the scenario studied, the two influencers were the main (active and popular) protagonists during the government crisis. Moreover, we consider the simplest scenario in which personalization is not employed: $\rho = 0$ and thus $\omega \equiv 1$. We consider a *feedback* function of the form $\theta = e^{-8.25(x^{(u)} - x^{(i)})^2}$ for both opinion directions. For an exhaustive list of the parameters, we refer to Table F.3.

A period of eleven weeks is considered, from July 7 to September 22, during which data was collected weekly from Facebook. A total of 1162 posts were published, of which 125 were by Conte. The *frequency of publication* $f^{(i)}$ is calculated as the number of posts by an influencer relative to the total number of posts ($f^{(Conte)} = 0.108$, $f^{(Salvini)} = 0.892$).

Figure 13 shows the timeline of the experiment. The two influencers start

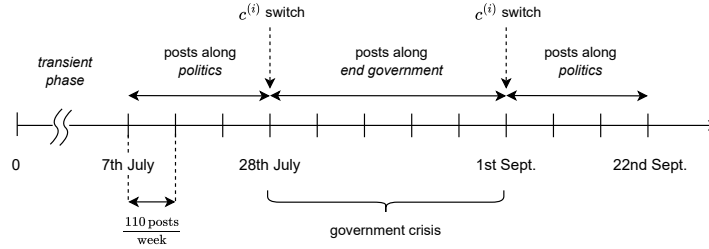


Figure 13: Timeline of modelled scenario from July 7, to September 22. From July 28 to September 1 we have a consistency switch, with posts along *End government* direction.

with the same initial popularity. We consider a transient of $N_t = 10000$ discrete time-units, after which the stationary normalized popularities π_i approximately correspond to the empirical normalized popularities obtained by dividing the number of *followers* of each influencer by the total number of the two. After the transient, we can see in Figure 12b that the distribution of public opinion is skewed towards Salvini, who has a higher popularity ratio due to his much higher publication frequency. After the transient, the crisis starts, and both influencers post in the *End government* direction (i.e., we observe a consistency switch for both influencers) during a time window of five weeks that approximates the duration of the government crisis, after which the two politicians switch back to posting on the *Politics* direction.

Even with these simplifications, it is possible to reproduce the observed social behavior as a whole: it corresponds to a situation where an influencer is in stark contrast to the opinions of its user base and loses ground with respect to the other influencer. In the model, only a very unbalanced population distribution towards Conte’s opinion (against the government fall) can explain the sudden increase in Conte’s popularity, despite the remarkable differences in popularity ratios in favor of Salvini. Figure 14 compares the simulation results of the described setting and Facebook’s measurements.

Clearly, our model does not precisely fit empirical observations but provides qualitative insights into the possible causes of the rather sudden popularity shift that was observed. Many of the model’s parameters are unknown, such as the

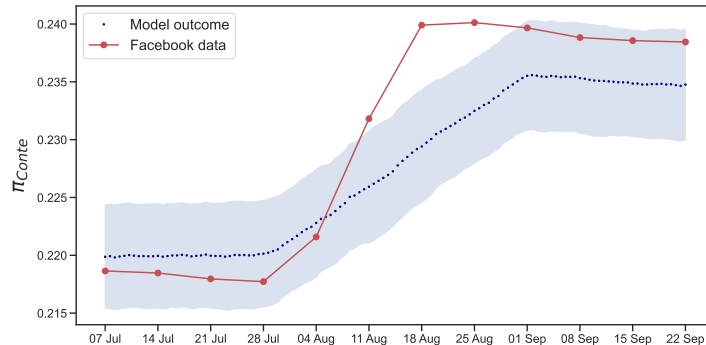


Figure 14: Popularity ratio π_{Conte} for Conte, from Facebook data (red) and from the model (blue) along with its 95% confidence interval, computed over 10 realizations of the process. One can see how the model follows the increase in popularity during August 2019.

opinion distribution, the weights of the updating rule, or the *feedback* function. However, by measuring some parameters, e.g., $f^{(0)}$, the phases' duration, $\pi_{Conte}(7 \text{ July})$, and making reasonable assumptions about the others, i.e., α , β , θ , $\mathbf{x}^{(i)}$, the emerging behavior is consistent for several choices of the parameters as demonstrated in Appendix F. We conclude that the observed popularity trends can be explained mainly by considering the fear of political instability in the user base.

8. Conclusion

Online social interactions have recently played an increasingly important role in opinion formation. Understanding the mechanisms underlying this modern communication paradigm requires the development of new, flexible frameworks suitable for describing interactions on social media platforms. In this work, we developed an opinion model tailored to online interactions, with particular attention to the interplay between *regular users* and *influencers*. In addition, we characterize influential individuals on online platforms by grounding our design decisions in data from real-world online social networks. Similar to other work in the recent literature, we integrated *content personalization* in a flexible and tunable manner. We have shown how content personalization reinforces

inequality by favoring the *structurally advantaged* influencer and, in most cases, preventing the “competing” influencer from remaining visible to the population, potentially hindering diversity of opinion among the population. Furthermore, even under structurally balanced conditions, personalization can lead to the emergence of *echo chambers* in which users’ opinions become radicalized by the influencer’s point of view, even on topics that are not the influencer’s *reference* one. Or to unstable situations in which a single individual hegemonizes the online scene.

The proposed model is one of the first attempts to describe the complexity of online interactions faithfully and comes with some limitations. In our model, users are passive entities, and influencers are stubborn agents. Moreover, *homophily* is the primary driver of user interaction, as no other relationship structure was considered. Nonetheless, despite the simplifying assumptions, the emergent behavior of the model proved rich enough to show the impact of content personalization and shed light on the dynamics of influencer popularity.

Our work points to several research directions, such as considering users as active agents capable of publishing their own posts and forwarding (i.e., sharing) posts from influencers, as we discuss in Appendix G.1. This may pose significant challenges in terms of analytical tractability. Another promising direction might be to consider influencers as “strategic” actors who aim to maximize their popularity on the platform by exploiting the internal mechanisms of the platform itself, e.g., content filtering, as we briefly discuss in Appendix G.2.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Description of the dataset

We collected data from real online social networks to support the hypotheses of our model and compare emergent behaviors. We focus on two popular social

networks: Facebook (FB) and Instagram (IG). Facebook has long been the most popular social media application, while Instagram has undergone a surge in popularity in recent years.

In Facebook and Instagram, a *profile*, i.e., a social network user, can be followed by other profiles, i.e., its *followers*. A profile with a large number of followers is also called an *influencer* - we consider profiles with more than ten thousand followers as influencers. Influencers post content (i.e., *posts*) that profile’s followers and anyone registered on the platform can see, *like* and comment. Note that when we use the term influencer, we do not only mean individuals but also groups, soccer teams, newspapers, or companies.

To get the list of such popular profiles, we exploited the online analytics platform hypeauditor.com for IG, and www.socialbakers.com and www.publicodelirio.it for FB. We restricted the analysis to influencers with at least 10,000 followers on June 1, 2021. The obtained 649 influencers are the same of our previous paper [45] and are publicly available.³ In this work, we are interested in the posts of influencers and their temporal sequence. For each monitored influencer, we downloaded all the data related to the posts published between January 1, 2016, and June 1, 2021, using the CrowdTangle tool and its API⁴. CrowdTangle is a content discovery and social analytics tool owned by Meta and available to researchers and analysts worldwide to support research, subject to a partnership agreement. Finally, we stored the data, which takes around 110 GB of disk space, on a Hadoop-based cluster, and we used PySpark for scalable processing.

Appendix B. Details on post classification

We developed a classifier that can categorize posts according to a particular set of subjects, similarly to what we have done in our previous work [46]. First, we arbitrarily identified a subset of topics that sufficiently characterize

³<https://mplanestore.polito.it:5001/sharing/P4WnRClQn>

⁴<https://github.com/CrowdTangle/API>

the discussions on the monitored profiles. Specifically, these topics are *sports*, *politics*, *food and cooking*, *music*, and *pandemics*, which are intentionally loose and relatively uncorrelated to each other. We developed a keyword classifier to classify the posts. For each topic, we manually defined a list of representative keywords. For example, if we consider *pandemic*, we search for words like COVID, pandemic, and coronavirus in Italian (and commonly used terms in other languages). We search for the topic-specific terms in the text corpus of the post, and if we find a match, we mark the post as belonging to the topic. Notice that since keywords of various topics may be present in the same corpus, we can flag a message as discussing *multiple* topics. In this work, we discard posts marked as multiple and only consider posts associated with a single topic.

We are not interested in classifying all posts by an influencer, first because our list of topics does not cover all possible ones, and second because we only need a large enough subsample of posts to make some statistical considerations. Conversely, it is of utmost importance that the accuracy of the classifier is high since misclassified posts could lead to wrong conclusions about the distribution among the available topics. Therefore, we manually validate the accuracy of our methodology for topic detection, as described in the following paragraph.

Appendix B.1. Classifier Precision Evaluation

We empirically evaluated the accuracy of the classifier by taking a random subsample of the labeled posts, i.e., 100 posts for each topic for a total of 500 messages, and manually classifying them. To this end, we defined a lower and upper bound for accuracy. Indeed, even for a human being, it is challenging to univocally classify posts based on their content. Therefore, we defined three possible states for each classification decision: “t” correct classification, “f” incorrect classification, and “ncc” standing for not completely correct (indicating that the assigned topic is related to the post but may not be the main topic of the post or the classification of the post is difficult). Given this states subdivision, the precision bounds are as follows:

$$P_L = \frac{N_t}{N_t + N_f + N_{ncc}} \quad (\text{B.1})$$

$$P_U = \frac{N_t + N_{ncc}}{N_t + N_f + N_{ncc}} \quad (\text{B.2})$$

We refined our term selection for each topic to improve precision based on this analysis.⁵ The classifier’s precision is subject-dependent but was consistently above 80% considering the upper bound defined in (B.1). The classification is particularly effective in the case of *politics* and *pandemic*, where the precision goes above 90%. Table B.1 summarises the bounds on precision achieved by the procedure described above. These results are sufficient to use the classification to support our modelling assumptions.

Table B.1: Per-topic Precision

Topic	Precision l.b.	Precision u.b.
Sports	76.9	83.2
Politics	87.0	94.4
Music	53.4	84.5
Food	65.5	82.4
Pandemic	76.6	93.1

The average percentage of messages classified is 27.8% for all influencers in the dataset. Considering the final classifier and the analysed dataset, we automatically flagged about one million posts⁶ with at least one topic. Of these, only 6.7% of the posts were flagged with multiple labels, indicating the message dealt with more than one topic. We decided to consider in the rest of the work only influencers for whom it was possible to classify more than a thousand posts in the observed period. At the end of this filtering process, we could keep 237 influencers for whom the average posts’ classification percentage is 53.2%.

The dataset used contains a subset of Italian politicians. To check the cor-

⁵We make the final list of terms available at <https://mplanestore.polito.it:5001/sharing/0wD5oU6xr>.

⁶1167963 posts were tagged with at least one label.

rectness of the labelling procedure, we checked whether the derived reference topic for all politicians was *politics*. It turned out that two politicians did not have *politics* as reference: Vincenzo De Luca had *pandemic*, and Renata Briano had *food*. However, this is entirely understandable as the latter runs a food blog and the former was known for his firm and frequent statements on the pandemic situation during the COVID -19 pandemic.

Appendix C. Proofs of Theorems (5.1) and (5.2)

Appendix C.1. Proof of Theorem (5.1)

Let us start assuming $k_i(0) > 0 \forall i$. In such a case we denote with $i_0 = \arg \max_i k_i(1)$ and with $K := k_{i_0}(1)$. First we show that the problem:

$$k_i(y_i) - cy_i = 0, \quad \text{with } y_i \in [0, 1] \quad \forall i \quad (\text{C.1})$$

admits a solution for any $c \geq K$. Indeed by choosing $c \geq K$ we have that necessarily $k_i(1) \leq k_{i_0}(1) \leq c \cdot 1 \forall i$ while $k_i(0) > c \cdot 0 = 0$; therefore a zero $z_i(c)$ must exist for every i . This zero is unique as a consequence of the concavity of $k_i(\cdot)$. The set of zeros $z_i(c)_i$ provides a solution of (C.1). Now to get a solution of the original problem (13) we need to show that there exist a c such that $\{z_i(c)\}_i$ are normalized. Observe that for $c = K$ by construction $z_{i_0}(K) = 1$ while $0 < z_i(K) \leq 1$ for $i \neq i_0$, therefore $\sum_i z_i(K) > 1$. Now, due to the monotonicity and concavity of $k_i(\cdot)$, $z_i(c)$ is by construction decreasing with respect to c , moreover $z_i(c) \rightarrow 0$ as $c \rightarrow \infty \forall i$, therefore since $\sum_i z_i(\cdot)$ is a continuous function of its argument, there will necessarily be a c_0 in correspondence of which $\sum_i z_i(c_0) = 1$. In the case in which $k_i(0) = 0$, observe that 0 is a solution of (C.1) for any c , i.e. $z_i(c) = 0$. Moreover for any $c \geq K$ a second zero may exist. For example, by construction, $z_{i_0}(K) = \{0, 1\}$. Therefore for $c = K$, as before, we can always choose as set of zeros $\{z_i(K)\}_i$, such that $z_i(K) = 0$ if $k_i(0) = 0$, and $i \neq i_0$, $z_{i_0}(K) = 1$. By construction $\sum_i z_i(K) \geq 1$. In particular $\sum_i z_i(K) > 1$ is there exists a i such that $k_i(0) > 0$. In this latter case, by increasing c all the non null zeros decrease, therefore, as before, there

will necessarily be a c_0 in correspondence of which $\sum_i z_i(c_0) = 1$. \square

Appendix C.2. Proof of Theorem (5.2)

We first show that $\|\bar{\pi}^{(1)} - \bar{\pi}^{(2)}\|_{L_\infty} = \max_i |\bar{\pi}_i^{(1)} - \bar{\pi}_i^{(2)}| = \|\mathcal{G}(F_1(x, z)) - \mathcal{G}(F_2(x, z))\|_{L_\infty} \leq M\|F_1(x) - F_2(x)\|_{L_\infty}$; then we show that we can always enforce: $\|F_1(x, z) - F_2(x, z)\|_{L_\infty} = \|\mathcal{H}(\bar{\pi}^{(1)}) - \mathcal{H}(\bar{\pi}^{(2)})\| \leq 1/(2M)\|\bar{\pi}^{(1)} - \bar{\pi}^{(2)}\|_{L_\infty}$ by properly choosing $\omega(\cdot, \cdot)$ and $\theta(\cdot)$. Therefore, we can conclude that $\|\mathcal{H} \circ \mathcal{G}(F_1(x, z)) - \mathcal{H} \circ \mathcal{G}(F_2(x, z))\| \leq 1/(2M)\|(\mathcal{G}(F_1(x)) - \mathcal{G}(F_2(x)))\| \leq M/(2M)\|F_1(x, z) - F_2(x, z)\| = 1/2\|F_1(x, z) - F_2(x, z)\|$.

First note that $\|F_1(x, z) - F_2(x, z)\|_{L_\infty} = \sup_x |F_1(x, z) - F_2(x, z)|$ coincides with the Kolmogorov distance between the two distributions.

Let us denote with

$$k_i(y, F_1(x, z)) = \lambda f^{(i)} \int \int \theta(|x - x_i|) \rho(\bar{\pi}_i, |x - x_i|) dF_1(x, z),$$

and similarly for $k_i(y, F_2(x, z))$ we assume that:

$$\sup_{y \in [0, 1], i} |k_i(y, F_1(x, z)) - k_i(y, F_2(x, z))| := \Delta K(F_1, F_2) \leq a \|F_1(x, z) - F_2(x, z)\|_{L_\infty} \quad a \in \mathbb{R}^+$$

and

$$\frac{dk_i(y, F_1(x, z))}{dy} \Big|_{y=0} < \max_i k_i(1, F_1(x, z)) \quad \frac{dk_i(y, F_2(x, z))}{dy} \Big|_{y=0} < \max_i k_i(1, F_2(x, z)) \quad \forall i.$$

Without lack of generality we assume $\max_i k_i(1, F_1(x)) \geq \max_i k_i(1, F_2(x))$. Let

the pair $(\bar{\pi}^{(1)} = \{\bar{\pi}_i^{(1)}\}_i, c_1)$ be the solution of

$$k_i(y_i, F_1(x, z)) - c_1 y_i = 0 \quad \text{s.t.} \sum_i y_i = 1, y_i \geq 0, \forall i$$

now let $(\{\hat{p}_i^{(2)}\}_i)$ the non necessarily normalized solution of

$$k_i(y_i, F_2(x, z)) - c_1 y_i = 0 \quad \text{s.t.} y_i \geq 0, \forall i.$$

by means of elementary geometric considerations we can bound:

$$|\bar{\pi}_i^{(1)} - \hat{p}_i^{(2)}| \leq \frac{\Delta K(F_1, F_2)}{c_1 - h_1}$$

where $h_1 = \frac{dk_i(y, F_1(x))}{dy} \Big|_{y=\min(\bar{\pi}_i^{(2)}, \hat{p}_i^{(2)})} \leq \frac{dk_i(y, F_1(x))}{dy} \Big|_{y=0}$. We recall that by construction (see proof of Theorem 5.1) we have $c_1 > \max_i k_i(1, F_1(x))$.

Denoting with $|\hat{p}^{(2)}| = \sum_i \hat{p}_i^{(2)}$, we have

$$1 - \sum_i |\hat{p}_i^{(2)} - \bar{\pi}_i^{(1)}| \leq |\hat{p}^{(2)}| \leq 1 + \sum_i |\hat{p}_i^{(2)} - \bar{\pi}_i^{(1)}|$$

Now denoted with $(\{\bar{\pi}_i^{(2)}\}_i, c_2)$ the solution of

$$k_i(y_i, F_2(x, z)) - cy_i = 0 \quad \text{s.t.} \quad \sum_i y_i = 1, y_i \geq 0, \forall i$$

we have, by construction, that:

$$\frac{1}{\max(1, |\hat{p}^{(2)}|)} < \frac{c_1}{c_2} < \frac{1}{\min(1, |\hat{p}^{(2)}|)}$$

and therefore, exploiting again elementary geometrical arguments, we can bound:

$$|\hat{p}_i^{(2)} - \bar{\pi}_i^{(2)}| \leq \left| \left(\frac{c_1 - h_2}{c_2 - h_2} - 1 \right) \hat{p}_i^{(2)} \right|$$

where $h_2 = \frac{dk_i(y, F_2(x, z))}{dy} \Big|_{y=\min(\hat{p}_i^{(2)}, \bar{\pi}_i^{(2)})} = \frac{dk_i(y, F_2(x, z))}{dy} \Big|_{y=0}$. Putting everything together, we have proved that:

$$\max_i \|\bar{\pi}_i^{(1)} - \bar{\pi}_i^{(2)}\| = \|\mathcal{G}(F_1(x, z)) - \mathcal{G}(F_2(x, z))\|_{L_\infty} \leq M \|F_1(x, z) - F_2(x, z)\|_{L_\infty}$$

To conclude the proof, first note that by properly choosing $\rho(\cdot, \cdot)$ and $\theta(\cdot)$ we can assume $v_x(x, z)$ and $\sigma_x^2(x, z)$ to depend sufficiently smoothly on $\bar{\pi}$, i.e. $\forall \varepsilon > 0$ we can assume:

$$\sup_x \left\| v_x^{(1)}(x, z) - v_x^{(2)}(x, z) \right\|_{L_\infty} \leq \varepsilon \|\bar{\pi}^{(1)} - \bar{\pi}^{(2)}\|_{L_\infty} \quad \forall z,$$

$$\sup_x \left\| \sigma_x^{2,(1)}(x, z) - \sigma_x^{2,(2)}(x, z) \right\|_{L_\infty} \leq \varepsilon \|\bar{\pi}^{(1)} - \bar{\pi}^{(2)}\|_{L_\infty} \quad \forall z,$$

and

$$\sup_x \left\| \frac{\partial \sigma_x^{2,(1)}(x, z)}{\partial x} - \frac{\partial \sigma_x^{2,(2)}(x, z)}{\partial x} \right\|_{L_\infty} \leq \varepsilon \|\bar{\pi}^{(1)} - \bar{\pi}^{(2)}\|_{L_\infty} \quad \forall z.$$

Then observe that the solution of the Fokker-Planck equation given in (10) on a compact interval (and so also its primitive) depends smoothly on function $v_x(x, z)$, function $\sigma_x^2(x, z)$ and its first derivative, as long as $\inf_{x,z} \sigma_x^2(x, z)$ is bounded away from zero. As final remark note that the set of weakly-increasing functions $F(x)$, such that $F(a) = 0$ and $F(b) = 1$ equipped with the L_∞ -norm forms a closed set in a complete metric space. \square

Appendix D. Validation of the fluid limit approximation

In this section, we compare predictions of the simplified *fluid limit* against simulation results of the full stochastic model described by algorithm 1 (obtained through a Monte-Carlo approach). We restrict ourselves to a one-dimensional opinion space, as in Section 5, and assume that all users share the same prejudice z . Again, we consider two “competing” influencers. A similar analysis could be performed in scenarios with any number of influencers at any point in the opinion space, but this would be computationally more challenging since multiple stationary points may exist, each with its own attraction basin.

First, we derive in Appendix D.1 some preliminary analytical results for the case of two influencers, using the results of the *fluid limit* introduced in Section 5.5. Then in Appendix D.2 two extreme instances of the model are solved in closed form. Appendix D.3 is devoted to comparing the analytical results of the *fluid* model with simulations. Finally, we discuss the impact of content personalization.

Appendix D.1. Two competing influencers

Let us specialize the equations presented in Section 5.5 for the mean opinion $\bar{x}(z)$ (Eq. (18)) and the normalized popularities $\bar{\pi}_i$ (Eq. (17)). Note that for

$N_i = 2$, $\bar{\pi}_0 = 1 - \bar{\pi}_1$, so it is sufficient to study $\bar{\pi}_1$.

As for the mean user opinion $\bar{x}(z)$, equation (18) allows us to write the asymptotic mean directly as a function of $\bar{\pi}_1$ and the opinions of the two influencers $x^{(0)}, x^{(1)}$:

$$\bar{x}(z) = \frac{\alpha}{1-\beta}z + \frac{\gamma}{1-\beta} \left[(1 - \bar{\pi}_1) x^{(0)} + \bar{\pi}_1 x^{(1)} \right] \quad (\text{D.1})$$

Substituting the functional forms of the *visibility* ω and *feedback* θ (as given in Table E.2) into equation (17), we obtain the following expression for the normalized popularity $\bar{\pi}_1$:

$$\bar{\pi}_1 = \frac{f^{(1)} e^{-\rho \frac{(x^{(1)} - \bar{x})^2}{\bar{\pi}_1}} (1 - |x^{(1)} - \bar{x}|)}{\sum_{i \in \{0,1\}} f^{(i)} e^{-\rho \frac{(x^{(i)} - \bar{x})^2}{\bar{\pi}_1}} (1 - |x^{(i)} - \bar{x}|)} = f(\bar{\pi}_1, \bar{x}) \quad (\text{D.2})$$

Moreover, if we combine the above expression with equation (D.1) for \bar{x} , we get $\bar{\pi}_1 = f(\bar{\pi}_1)$, which can be solved numerically through a fixed-point approximation (FPA) (a graphical representation is shown on Fig. D.16). The outcome of this FPA and the corresponding simulation results are compared in Figure D.15.

Appendix D.2. Closed form computations in extremal cases

The combination of equations (D.1) and (D.2) cannot be solved in closed form in the general case. However, there are at least two scenarios in which this is possible, separately considered in the following subsections.

Appendix D.2.1. When an influencer “wins”

We consider an influencer a “winner” if its normalized popularity $\bar{\pi}_i$ approaches 1. Suppose that the influencer whose opinion is $x^{(1)} = 1$ *wins*, then $\bar{\pi}_1 \rightarrow 1$. This implies $\bar{\pi}_0 \rightarrow 0$ and thus $\omega \rightarrow 0^+$: the influencer with $x^{(0)} = 0$ is seen by a negligible fraction of users and in practice, only influencer $i = 1$ remains visible. Note that in the extreme case in which influencer 1 wins, users see only $x^{(1)}$, and asymptotically all users move towards it. In this case, the final opinion $\bar{x}(z)$ can be easily calculated with a recursion of the update rule (3):

$$x^{(u)}(n) = \sum_{i=0}^n \beta^i (\alpha z + \gamma x^{(1)}) + \beta^n x(0)$$

For $n \rightarrow \infty$ and considering $\beta < 1$ (the case $\beta = 1$ coincides with the trivial case where users remain fixed at their initial opinion) we get:

$$x^{(u)} = \frac{\alpha}{1-\beta} z + \frac{\gamma}{1-\beta} x^{(1)}, \quad (\text{D.3})$$

which is in agreement with (D.1) if one sets $\bar{\pi}_1 = 1$. This corresponds to one of the extreme cases that we will use later to examine the model behavior as a function of the personalization parameter ρ . It should be noted that this construction relies on the knowledge of the *winning* influencer, which is unknown in advance. However in the fluid limit, we expect that the winning influencer, if any, is the one that has a structural advantage over the others at the beginning (e.g., a higher posting rate $f^{(i)}$, see Figure 8).

Appendix D.2.2. Constant personalization function

The other extreme case we consider is the one in which $\rho = 0$. In this case, the personalization function ω no longer depends on $\bar{\pi}_i$, and it is easy to see from Table E.2 that it returns $\omega \equiv 1$. Moreover, we consider $x^{(1)} = 1, x^{(0)} = 0$, which further simplifies (D.1). The above formulas (Eq. D.2 and Eq. D.1) can then be solved in closed form. In particular, equation (D.2) for the normalized popularity $\bar{\pi}_1$ becomes:

$$\bar{\pi}_1 = \frac{f^{(1)}(q + m \bar{\pi}_1)}{f^{(0)}(1 - (q + m \bar{\pi}_1)) + f^{(1)}(q + m \bar{\pi}_1)}$$

where $m \triangleq \frac{\gamma}{1-\beta}$ and $q \triangleq \frac{\alpha}{1-\beta} z$ for compactness. This leads to a second order equation which can be easily solved for $\bar{\pi}_1$:

$$\bar{\pi}_1^2 m (f^{(1)} - f^{(0)}) + \bar{\pi}_1 \left[f^{(0)}(1 - q) + f^{(1)}(q - m) \right] - f^{(1)} q = 0 \quad (\text{D.4})$$

Appendix D.3. Comparison of analytical results and Monte-Carlo simulations

This section is devoted to comparing the analytical results derived in Section 5 with simulations of the model. Numerical and graphical solutions of equation (D.2) are also provided, shedding light on the impact of the algorithmic personalization performed by the platform.

Appendix D.3.1. Description of the scenario

The scenario setting is analogous to that described in Section 6.1 and Table E.2. However, here, we consider a one-dimensional opinion space $[0, 1]$ and we assume all users to have the same prejudice, i.e., $z^{(u)} = z = 0.4, \forall u \in \mathcal{U}$ matching their initial opinion $x^{(u)}(0)$. The “competing ” influencers have opinions at the extremes of the domain, and their posting frequencies are $f^{(1)} = 0.7$ and $f^{(0)} = 0.3$, i.e., influencer $i = 1$ has a *structural advantage* over influencer $i = 0$. Note that in a one-dimensional space, the reference direction $r^{(i)}$, and hence the consistency $c^{(i)}$, lose their significance. To avoid obtaining trivial results in which influencer 1 obviously wins, regular users are initially placed closer to the disadvantaged influencer $i = 0$.

Appendix D.3.2. Simulation, fluid limit and fixed-point approximation

Comprehensive validation and comparison of the approaches used to obtain the system equilibria are shown in Figure D.15. First, the stochastic model described by Algorithm 1 is “simulated” by obtaining 100 different sample whose length is 500000 elementary steps. The variables of interest $\bar{x}(z)$ and $\bar{\pi}_1$ are obtained by averaging the process over both discrete times steps n and sample paths and are represented by circle marks. Second, equation (D.1), which is a specialization of (18) obtained from the fluid limit, indicates that the state of the system lies on a line in the plane $\bar{\pi}_1, \bar{x}$ (dashed line in Figure D.15). Third, the extreme cases of the model analyzed in Appendix D.2, for which we derived a closed-form solution, are represented by star-like marks. Lastly, diamonds are solutions of (D.2) employing the fixed-point approximation.

We observe that, for given ρ , simulation marks match well with analytical marks. The only exception is for $\rho = 0.5$, for which simulations provide $\bar{\pi}_1 \approx 0.79$, whereas the analysis provides $\bar{\pi}_1 \approx 1$ (see also the table on Fig. D.16). This mismatch is due to the fact that $\rho = 0.5$ is close to a ‘phase transition’, at which the system switches from a regime in which two stable solutions exist (in particular, one in which both influencers survive) to a regime in which influencer $i = 1$ wins. In such a situation, the population is exposed to the opinions of a single individual, hindering diversity on the social platform. This behavior is better illustrated in Fig. D.16, where the curve corresponding to $\rho = 0.5$ is almost tangent to the bisector. It should be noted that the “empty” region in Figure D.15 is directly related to this behavior since no stable solutions can exist for that values of $\bar{\pi}_i$. In fact, there is no *stable* intersection with the bisector in Figure D.16 in the corresponding interval.

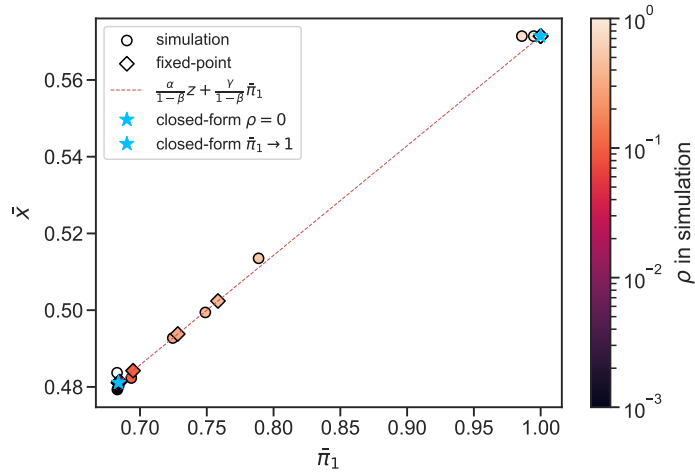


Figure D.15: Comparison between analytical results, including the exact extreme points calculated in Appendix D.2.2 and Appendix D.2.1, and the linear relationship between \bar{x} and $\bar{\pi}_1$ according to Equation (D.1). Diamonds represent the fixed-point approximation for the solution of equation (D.2). Simulation results of the stochastic dynamics, represented by circles, were obtained by averaging 100 realizations of the process as described by Algorithm 1. We consider a scenario in which $\alpha = 0.05, \beta = 0.93$, with two influencers at the extremes of the domain, with $f^{(0)} = 0.3, f^{(1)} = 0.7$ and the same initial absolute popularity $p_0 = p_1 = 100$. Numerical values from simulation and fixed-point approximation are reported in the table alongside the plot in Fig. D.16.

ρ	$\bar{\pi}_1$	
	<i>SIM</i>	<i>FPA</i>
0.0	0.682	0.684
0.001	0.683	0.684
0.01	0.684	0.685
0.1	0.693	0.695
0.3	0.725	0.728
0.4	0.749	0.758
0.5	0.789	0.999
0.8	0.986	0.999
1.0	0.995	1.0

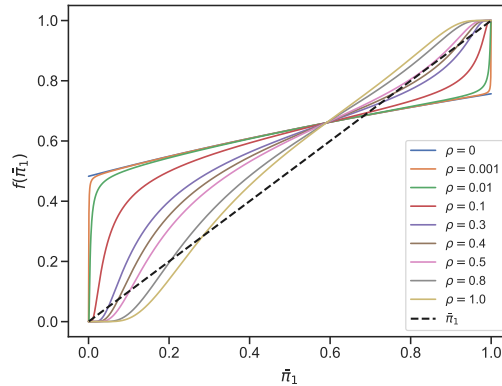


Figure D.16: Graphical solution of $\bar{\pi}_i = f(\bar{\pi}_i)$, (D.2). Stable solutions corresponds to intercepts between $f(\bar{\pi}_i)$ and the bisector, such that $f'(\bar{\pi}_i) < 1$. We observe that non-trivial solutions (i.e., solutions in which both influencers survive) exist, roughly in the interval $[0.7, 0.8]$, provided that ρ is not too large (i.e., $\rho < 0.5$). For $\rho > 0.5$, the only stable solution is $\bar{\pi} = 1$. This explains the results in Figure D.15. Simulation results reported on the alongside table confirm the validity of the analytical predictions.

Appendix E. Additional simulation results

Here we report all the *default* parameters used in the simulations of Section 6 (Table E.2) and the behavior of the model as a function of the *degree of stubbornness* which is interesting but not the focus of our investigation.

Appendix E.1. Details of the simulation scenario

In this section, we provide further details on the simulation setting of Section 6. In Figure E.17 we show the initial distribution of *regular* users' opinions and, since $\mathbf{x}^{(u)}(0) = z^{(u)}$ also of their prejudice. Table E.2 contains all the *default* choices of the model's parameters.

Appendix E.2. Behavior as function of the updating weights

The behavior of the system depends not only on the characteristics of the influencers and the composition of public opinion, but also on the parameters controlling the opinion update rule in equation (3). The update is a convex combination of the prejudice, the current opinion, and the opinion conveyed by the post. We chose to hold fixed the weight β (inertia) and then vary the *degree of stubbornness*, $\frac{\alpha}{\alpha+\gamma}$.

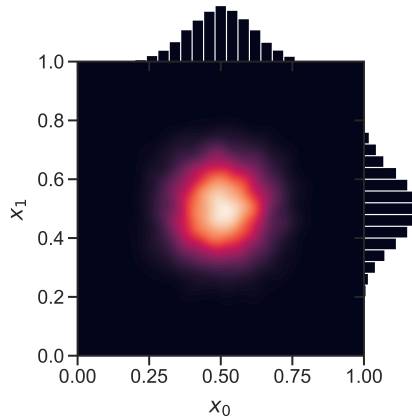


Figure E.17: Initial opinion (prejudice) distribution of the regular users, recall $\mathbf{x}^{(u)}(0) = \mathbf{z}^{(u)}$.

Table E.2: Parameters and functions shared across experiments

Symbol	Value - Form	Description
N_i	2	Number of influencers
$x_j^{(0)}$	0	Opinion of influencer 0 on direction j
$x_j^{(1)}$	1	Opinion of influencer 1 on direction j
$r_j^{(0)}$	0	Reference direction of influencer 0
$r_j^{(1)}$	1	Reference direction of influencer 1
$p_{0,1}(0)$	100	Initial absolute popularity of both influencers
N_u	10000	Number of regular users
N_{iter}	100000	Number of iterations for each simulation
α	0.05	First weight in the updating rule in Eq. 3
β	0.93	Second weight in the updating rule in Eq. 3
$\theta(\cdot)$	$1 - \frac{ x_j^{(i)} - x_j^{(u)} }{(x_r^{(u)} - x_r^{(i)})^2}$	Functional form of the <i>feedback</i> function
$\omega(\cdot)$	$e^{-\rho \frac{\cdot}{\bar{\pi}_i}}$	Functional form of the <i>visibility</i> function
a	10	First parameter of the initial Beta distribution
b	10	Second parameter of the initial Beta distribution
$z_j^{(u)}$	$x_j^{(u)}(0)$	Prejudice coincides with initial opinion

We considered an unbalanced scenario in which influencer $i = 1$ has a structural advantage, i.e., $f^{(1)} = 0.7 > f^{(0)}$. Figure E.18 again shows that personalization favors the structurally advantaged individual (consistent with Section 6.3). Note that the x -scale is logarithmic to highlight the sudden drop of $\bar{\pi}_0$ for $\frac{\alpha}{\alpha+\gamma} \approx 10^{-3}$ (corresponding to small values of α) when *sharp* personalization is applied. The shape of the two curves is quite similar, only the decrease is

observed at different values of $\frac{\alpha}{\alpha+\gamma}$. *Smooth* personalization allows the coexistence of influencers on the whole domain, while with *sharp* personalization, for a wide range of parameters, influencer $i = 1$ “wins”.

In both cases, there is an initial phase (for low values of $\frac{\alpha}{\alpha+\gamma}$) in which the two influencers coexist, and this is followed by a drop of the normalized popularity of the disadvantaged influencer. This can be explained by the fact that small values of $\frac{\alpha}{\alpha+\gamma}$ imply that a negligible weight is given to the prejudice, and therefore regular users concentrate around the two influencers’ opinions on their reference direction. This can be easily confirmed by looking at the final opinion configuration of users, who concentrate in the upper corners of the opinion space (around $[0, 1]$ and $[1, 1]$). This is because the influencer $i = 1$, whose opinion is $x^{(1)} = [1, 1]$ is stronger than the other in terms of popularity and is able to pull users along its non-reference direction as well. We remark that when users are very close in opinion to a particular influencer, it is difficult for the other to persuade them, as the probability of this happening is proportional to the product $\omega \cdot \theta$, both of which are a function of opinion distance. In these scenarios, the distance from the “further” influencer is $d_j \approx 1$, which drastically reduces the probability of reaching the users. Thus, as long as $\frac{\alpha}{\alpha+\gamma}$ is small enough, both influencers can build their user base. These situations represent rather degenerate cases where the population almost disregards their prejudice in favor of the opinion conveyed by the post. It might be interesting to consider users with varying degrees of “volatility” who are able to pull along the opinion of their neighborhood.

As the degree of stubbornness increases, users are more entrenched in their prejudice and therefore no longer concentrate in a small neighborhood of the influencer’s opinions. This favors the structurally advantaged influencer, as the other (i.e., $i = 0$) is unable to build its user base because users do not get close enough to it (see Figure E.18 for $i = 1$ and $\rho = 0.0001$ we have $\bar{\pi}_0 \rightarrow 0$). The subsequent rise in $\bar{\pi}_0$ depends on the fact that when the degree of stubbornness approaches one, users give importance only to their prejudice, and therefore they do not deviate too much from their initial position. As a consequence, it can

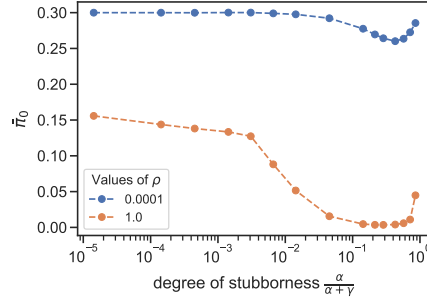


Figure E.18: Normalized popularity $\bar{\pi}_0$ as a function of the degree of stubbornness $\frac{\alpha}{\alpha+\gamma}$; the points are obtained considering 50 realizations of the process and averaging over 100 discrete time instants. Again, smooth ($\rho = 0.0001$) and strict ($\rho = 1$) personalization are considered.

not be triggered the positive feedback between users’ opinion and influencers’ popularity that leads to the complete victory of one influencer.

Appendix F. Sensitivity analysis of the case study

For the *Case Study* in Section 7.2, we were able to measure or define some characteristic parameters (e.g., $f^{(i)}$ and $x_{Endgovernment}$). However, others were unknown, and even if there are methods to measure, for example, user’s opinions (see Section 2), this is far beyond the scope of the present work. We could not measure the updating weights α , β , the *feedback* function θ , and the users’ opinion on *Politics*. Nevertheless, it is sufficient to assume that a) Conte and Salvini have “opposing views” over *Politics*, i.e., $x^{(Salvini)} < 0.5$ and $x^{(Conte)} > 0.5$ and that b) Salvini has a more extreme viewpoint, i.e., $|x^{(Salvini)} - 0.5| > |x^{(Conte)} - 0.5|$ c) the population is not too *volatile* in their viewpoints, i.e., the *degree of stubbornness* is not too low c) the *feedback* function θ decreases quickly with the opinion distance (exponential), motivated by the fact that about *politics* people do not easily like viewpoints too far from theirs, we were able to obtain the results in Figure 14. These are obtained with the parameters specified in Table F.3.

To justify our choices, we run a *sensitivity analysis* on the parameters we did not measure, defining the scenarios in Table F.4. Figure F.19 presents the

Table F.3: Parameters and functions for the Case Study

Symbol	Value - Form	Description
N_i	2	Number of influencers
$x_0^{(Conte)}$	0.76	Opinion of Giuseppe Conte on direction j
$x_0^{(Salvini)}$	0.0	Opinion of influencer 1 on direction j
$f^{(Conte)}$	0.108	Opinion of Giuseppe Conte on direction j
$f^{(Salvini)}$	0.892	Opinion of influencer 1 on direction j
$r^{(Conte),(Salvini)}$	0	Reference direction of both influencers
$p_{Conte,Salvini}(0)$	20	Initial absolute popularity of both influencers
N_u	10000	Number of regular users
N_{iter}	15000	Number of iterations for each simulation
N_t	10000	Duration of the transient phase
w	550	Length of the government crisis
α	0.3	First weight in the updating rule in Eq. 3
β	0.65	Second weight in the updating rule in Eq. 3
$\theta(\cdot)$	$e^{-8.25(x^{(u)} - x^{(i)})^2}$	Functional form of the <i>feedback</i> function
$\omega(\cdot)$	$\rho = 0 \implies \omega \equiv 1$	Functional form of the <i>visibility</i> function

results. It is evident that the increase in the normalized popularity π_{Conte} does not depend on these parameters and is due to the timing (see Fig. 13) and the skewness in the population distribution, see Figure 12. Note that in Figure F.19 there is a negative *offset* that tends to underestimate Conte’s normalized popularity. This is because we use the measured rate $f^{(Conte)} = 0.108$, which puts him at quite a disadvantage compared to Salvini, who has a much higher publication frequency.

Appendix G. Possible model extensions

Appendix G.1. Modeling interactions between users

In our model we have purposely ignored interactions among “regular” users. This choice can be justified by the need to keep the model sufficiently simple to be analyzed using the tools introduced so far, and by the hypothesis that opinion dynamics are primarily driven by the interaction between influencers and regular users, at least for some topics.

Nevertheless, our model could be extended to incorporate also interactions among regular users, and here we outline one possible way to proceed along this direction. Firstly, to preserve the scalability of the model, we do not represent individual interactions among regular users, as done by other authors, but we

Table F.4: Definition of the scenarios for the sensitivity analysis

Scenario	α	β	θ param.	$x_0^{(Conte)}$	$x_0^{(Salvini)}$
1	0.25	0.708	8.0	0.74	0.01
2	0.25	0.708	8.0	0.74	0.05
3	0.25	0.708	8.0	0.78	0.01
4	0.25	0.708	8.0	0.78	0.05
5	0.45	0.475	8.0	0.74	0.01
6	0.45	0.475	8.0	0.74	0.05
7	0.45	0.475	8.0	0.78	0.01
8	0.45	0.475	8.0	0.78	0.05
9	0.25	0.708	8.25	0.74	0.01
10	0.25	0.708	8.25	0.74	0.05
11	0.25	0.708	8.25	0.78	0.01
12	0.25	0.708	8.25	0.78	0.05
13	0.45	0.475	8.25	0.74	0.01
14	0.45	0.475	8.25	0.74	0.05
15	0.45	0.475	8.25	0.78	0.01
16	0.45	0.475	8.25	0.78	0.05
17	0.25	0.708	8.5	0.74	0.01
18	0.25	0.708	8.5	0.74	0.05
19	0.25	0.708	8.5	0.78	0.01
20	0.25	0.708	8.5	0.78	0.05
21	0.45	0.475	8.5	0.74	0.01
22	0.45	0.475	8.5	0.74	0.05
23	0.45	0.475	8.5	0.78	0.01
24	0.45	0.475	8.5	0.78	0.05

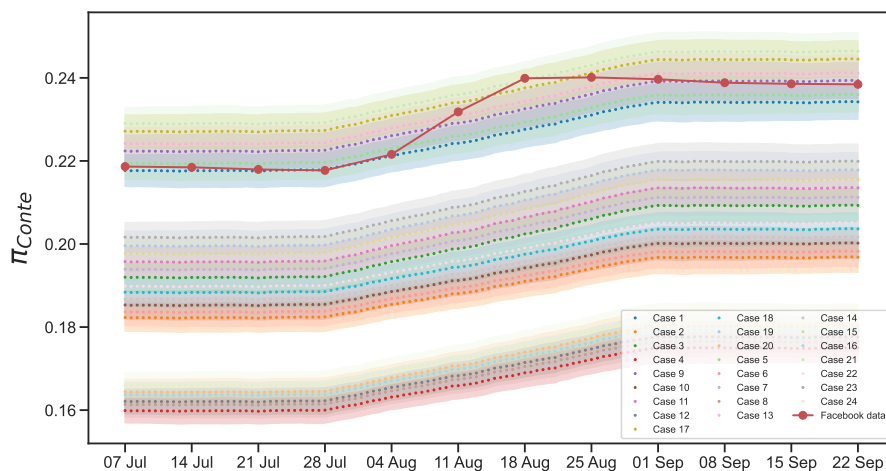


Figure F.19: Sensitivity analysis for the Figure 14 considering the 24 different scenarios (in terms of parameters set) defined in Table F.4.

rather represent the average effect of such interactions on the generic regular user u through a mean-field approach. Secondly, we distinguish between two classes of additional interactions originated by the activity of regular users: *direct* (pairwise) interactions between regular users and *influencer-triggered* users interactions.

For what concerns the first class, we introduce a monotonic non-increasing function $\zeta(d)$, representing the probability that two “regular” users whose current opinion distance is d influence each other. Note that $\zeta(d)$ measures the degree of homophily of regular users as function of their separation in the opinions space: more skewed $\zeta(d)$ shapes correspond to larger degrees of homophily. Moreover, we again assume that user u provides a positive feedback to a message generated by user u' at time t on topic j with probability $\theta(|x_j^{(u)(t)} - x_j^{(u')(t)}|)$. It follows that, to account for the overall effect on u of pairwise interactions, we need to compute the ‘center of mass’ of neighbors receiving positive feedback from u along direction j :

$$\begin{aligned}\widehat{x}_j^{(u)}(n) &= \sum_{u' \neq u} \mathbf{x}^{(u')}(n) \theta(|x_j^{(u)}(n) - x_j^{(u')}(n)|) \zeta(\|\mathbf{x}^{(u)}(n) - \mathbf{x}^{(u')}(n)\|) \\ &\rightarrow \int f(\mathbf{x}, t) \theta(|x_j^{(u)} - x_j|) \zeta(\|\mathbf{x}^{(u)} - \mathbf{x}\|) d\mathbf{x}\end{aligned}$$

Then we extend the user opinion update rule (3) as follows:

$$x_j^{(u)}(n+1) = \begin{cases} \alpha z_j^{(u)} + \beta x_j^{(u)}(n) + \gamma x_j^{(i)} + \delta \widehat{x}_j^{(u)} & \text{if } \Omega(\omega(d_r, \pi_i)) = 1, \Theta(\theta(d_j)) = 1 \\ \frac{\alpha}{1-\gamma} z_j^{(u)} + \frac{\beta}{1-\gamma} x_j^{(u)}(n) + \frac{\delta}{1-\gamma} \widehat{x}_j^{(u)} & \text{otherwise} \end{cases} \quad (\text{G.1})$$

with $\alpha + \beta + \gamma + \delta = 1$. Different ratios δ/γ correspond to different relative strengths between *direct* interactions between regular users and *influencer-triggered* users interactions.

Influencer-triggered users interactions represent cascading effect of reactions by a user on a post generated by an influencer, appearing on the timeline of its friends. This is another form of (indirect) interaction among regular users which

might play a significant role in the overall opinion dynamics. It is reasonable to assume that all follow-up reactions to posts of influencer i mainly circulate among the *followers* of i with the resulting effect of amplifying the strength of the original post on the users providing positive feedback to it. As a consequence, one can take into account the impact the *influencer-triggered* users interactions by playing on weights α , β , γ and δ , i.e., by increasing the relative weight of γ .

Appendix G.2. Strategic influencers

The influence exerted on non-reference directions depends heavily on the size of the user audience reached by the influencer’s post, which in turn depends on how popular an influencer is in its main field of expertise. For example, famous public figures (e.g., athletes, models) may express their views on potentially sensitive topics and resonate more than experts due to their popularity in their own field. This mechanism can be deliberately exploited by an influencer to gain popularity and/or strategically exert influence on a topic other than its own reference topic.

For example, we have seen in Section 6.4 how *low-consistency* values allow an influencer to gain a structural advantage whenever the consistency of the opposing influencer (who has different reference topic) tends to one. We must here remark that a consistency $c^{(i)} < 0.5$ in a bidimensional setting is slightly in contrast with the definition of reference direction itself, if the parameters are static. However, if we consider time-varying consistency, an influencer that is well-known on topic j , i.e., $r^{(i)}(n_1) = j$ may modify its consistency at time n_2 if it learns to have little competition on its reference direction.

Our model can be rather easily adapted to study such scenarios, by representing influencers as *strategic agents*, whose selected topics and relative opinions expressed in their generated posts can be dynamically adapted so to maximize a given utility function.

References

- [1] M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks, *Annual Review of Sociology* 27 (1) (2001) 415–444. doi:10.1146/annurev.soc.27.1.415.
URL <https://doi.org/10.1146/annurev.soc.27.1.415>
- [2] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media, *Proceedings of the National Academy of Sciences* 118 (9) (Feb. 2021). doi:10.1073/pnas.2023301118.
URL <https://doi.org/10.1073/pnas.2023301118>
- [3] A. Das, S. Gollapudi, K. Munagala, Modeling opinion dynamics in social networks, in: *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM '14*, Association for Computing Machinery, New York, NY, USA, 2014, p. 403–412. doi:10.1145/2556195.2559896.
URL <https://doi.org/10.1145/2556195.2559896>
- [4] F. Xiong, Y. Liu, Opinion formation on social media: An empirical approach, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 24 (1) (2014) 013130. doi:10.1063/1.4866011.
- [5] S. Fortunato, C. Castellano, Scaling and universality in proportional elections (2006). doi:10.48550/ARXIV.PHYSICS/0612140.
URL <https://arxiv.org/abs/physics/0612140>
- [6] S. E. Asch, Opinions and social pressure, *Scientific American* 193 (5) (1955) 31–35.
URL <http://www.jstor.org/stable/24943779>
- [7] J. R. P. French, A formal theory of social power., *Psychological Review* 63 (3) (1956) 181–194. doi:10.1037/h0046123.
URL <https://doi.org/10.1037/h0046123>

- [8] L. Festinger, A theory of social comparison processes, *Human Relations* 7 (2) (1954) 117–140. doi:10.1177/001872675400700202.
URL <https://doi.org/10.1177/001872675400700202>
- [9] L. Mastroeni, P. Vellucci, M. Naldi, Agent-based models for opinion formation: A bibliographic survey, *IEEE Access* 7 (2019) 58836–58848. doi:10.1109/access.2019.2913787.
URL <https://doi.org/10.1109/access.2019.2913787>
- [10] M. H. Degroot, Reaching a consensus, *Journal of the American Statistical Association* 69 (345) (1974) 118–121. doi:10.1080/01621459.1974.10480137.
URL <https://doi.org/10.1080/01621459.1974.10480137>
- [11] N. E. Friedkin, E. C. Johnsen, Social influence and opinions, *The Journal of Mathematical Sociology* 15 (3–4) (1990) 193–206. doi:10.1080/0022250X.1990.9990069.
- [12] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence: Models, analysis and simulation, *Journal of Artificial Societies and Social Simulation* 5 (2002) 1–24.
- [13] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, *Advances in Complex Systems* 03 (01n04) (2000) 87–98. doi:10.1142/s0219525900000078.
URL <https://doi.org/10.1142/s0219525900000078>
- [14] J. Lorenz, Continuous opinion dynamics under bounded confidence: A survey, *International Journal of Modern Physics C* 18 (12) (2007) 1819–1838, arXiv:0707.1762 [physics]. doi:10.1142/S0129183107011789.
- [15] P. CLIFFORD, A. SUDBURY, A model for spatial conflict, *Biometrika* 60 (3) (1973) 581–588. doi:10.1093/biomet/60.3.581.
URL <https://doi.org/10.1093/biomet/60.3.581>

- [16] R. A. Holley, T. M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, *The Annals of Probability* 3 (4) (Aug. 1975). doi:10.1214/aop/1176996306.
URL <https://doi.org/10.1214/aop/1176996306>
- [17] P. Holme, M. E. J. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions, *Physical Review E* 74 (5) (2006) 056108. doi:10.1103/PhysRevE.74.056108.
- [18] R. Durrett, J. P. Gleeson, A. L. Lloyd, P. J. Mucha, F. Shi, D. Sivakoff, J. E. S. Socolar, C. Varghese, Graph fission in an evolving voter model, *Proceedings of the National Academy of Sciences* 109 (10) (2012) 3682–3687. doi:10.1073/pnas.1200709109.
URL <https://doi.org/10.1073/pnas.1200709109>
- [19] C. Nardini, B. Kozma, A. Barrat, Who’s talking first? consensus or lack thereof in coevolving opinion formation models, *Physical Review Letters* 100 (15) (apr 2008). doi:10.1103/physrevlett.100.158701.
URL <https://doi.org/10.1103/physrevlett.100.158701>
- [20] B. L. Granovsky, N. Madras, The noisy voter model, *Stochastic Processes and their Applications* 55 (1) (1995) 23–43. doi:10.1016/0304-4149(94)00035-R.
- [21] E. Ben-Naim, Opinion dynamics: Rise and fall of political parties, *Europhysics Letters* 69 (55) (2005) 671–677. doi:10.1209/epl/i2004-10421-1.
- [22] G. Toscani, Kinetic models of opinion formation (2006). doi:10.48550/ARXIV.MATH-PH/0605052.
URL <https://arxiv.org/abs/math-ph/0605052>
- [23] K. SZNAJD-WERON, J. SZNAJD, OPINION EVOLUTION IN CLOSED COMMUNITY, *International Journal of Modern Physics C* 11 (06) (2000)

1157–1165. doi:10.1142/s0129183100000936.

URL <https://doi.org/10.1142/2Fs0129183100000936>

- [24] F. Slanina, H. Lavicka, Analytical results for the sznajd model of opinion formation, *The European Physical Journal B - Condensed Matter* 35 (2) (2003) 279–288. doi:10.1140/epjb/e2003-00278-0.
URL <https://doi.org/10.1140/epjb/e2003-00278-0>
- [25] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, *Reviews of Modern Physics* 81 (2) (2009) 591–646, arXiv:0710.3256 [cond-mat, physics:physics]. doi:10.1103/RevModPhys.81.591.
- [26] J. T. Cox, Coalescing random walks and voter model consensus times on the torus in z_d , *The Annals of Probability* 17 (4) (1989) 1333–1366.
URL <http://www.jstor.org/stable/2244439>
- [27] L. Frachebourg, P. L. Krapivsky, Exact results for kinetics of catalytic reactions, *Physical Review E* 53 (4) (1996) R3009–R3012. doi:10.1103/physreve.53.r3009.
URL <https://doi.org/10.1103/physreve.53.r3009>
- [28] K. Suchecki, V. M. Eguíluz, M. S. Miguel, Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution, *Physical Review E* 72 (3) (Sep. 2005). doi:10.1103/physreve.72.036132.
URL <https://doi.org/10.1103/physreve.72.036132>
- [29] V. Sood, S. Redner, Voter model on heterogeneous graphs, *Physical Review Letters* 94 (17) (May 2005). doi:10.1103/physrevlett.94.178701.
URL <https://doi.org/10.1103/physrevlett.94.178701>
- [30] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, A. Scaglione, Binary opinion dynamics with stubborn agents, *ACM Transactions on Economics and Computation* 1 (4) (2013) 1–30. doi:10.1145/2538508.
URL <https://doi.org/10.1145/2538508>

- [31] C. M. Valensise, M. Cinelli, W. Quattrociocchi, The dynamics of online polarization (2022). doi:10.48550/ARXIV.2205.15958.
URL <https://arxiv.org/abs/2205.15958>
- [32] A. F. Peralta, M. Neri, J. Kertész, G. Iñiguez, Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions, *Physical Review E* 104 (4) (oct 2021). doi:10.1103/physreve.104.044312.
URL <https://doi.org/10.1103/physreve.104.044312>
- [33] N. Perra, L. E. C. Rocha, Modelling opinion dynamics in the age of algorithmic personalisation, *Scientific Reports* 9 (1) (May 2019). doi:10.1038/s41598-019-43830-2.
URL <https://doi.org/10.1038/s41598-019-43830-2>
- [34] A. F. Peralta, J. Kertész, G. Iñiguez, Opinion dynamics in social networks: From models to data (2022). doi:10.48550/ARXIV.2201.01322.
URL <https://arxiv.org/abs/2201.01322>
- [35] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, A. Volfovsky, Exposure to opposing views on social media can increase political polarization, *Proceedings of the National Academy of Sciences* 115 (37) (2018) 9216–9221. arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.1804840115>, doi:10.1073/pnas.1804840115.
URL <https://www.pnas.org/doi/abs/10.1073/pnas.1804840115>
- [36] T. J. H. Morgan, L. E. Rendell, M. Ehn, W. Hoppitt, K. N. Laland, The evolutionary basis of human social learning, *Proceedings of the Royal Society B: Biological Sciences* 279 (1729) (2011) 653–662. doi:10.1098/rspb.2011.1172.
URL <https://doi.org/10.1098/rspb.2011.1172>
- [37] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel, V. M. Eguíluz, Is the voter model a model for voters?, *Phys. Rev. Lett.* 112 (2014)

158701. doi:10.1103/PhysRevLett.112.158701.

URL <https://link.aps.org/doi/10.1103/PhysRevLett.112.158701>

- [38] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, Tweeting from left to right: Is online political communication more than an echo chamber?, *Psychological Science* 26 (10) (2015) 1531–1542, PMID: 26297377. arXiv:<https://doi.org/10.1177/0956797615594620>, doi: 10.1177/0956797615594620.

URL <https://doi.org/10.1177/0956797615594620>

- [39] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, S. Chakrabarti, Learning linear influence models in social networks from transient opinion dynamics, *ACM Trans. Web* 13 (3) (nov 2019). doi:10.1145/3343483.

URL <https://doi.org/10.1145/3343483>

- [40] C. Monti, G. Manco, C. Aslay, F. Bonchi, Learning ideological embeddings from information cascades, in: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21*, Association for Computing Machinery, New York, NY, USA, 2021, p. 1325–1334. doi:10.1145/3459637.3482444.

URL <https://doi.org/10.1145/3459637.3482444>

- [41] E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on facebook, *Science* 348 (6239) (2015) 1130–1132. doi: 10.1126/science.aaa1160.

URL <https://doi.org/10.1126/science.aaa1160>

- [42] S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, *Proceedings of the National Academy of Sciences* 106 (51) (2009) 21544–21549. doi:10.1073/pnas.0908800106.

URL <https://doi.org/10.1073/pnas.0908800106>

- [43] H. Risken, Fokker-planck equation, in: *The Fokker-Planck Equation*, Springer, 1996, pp. 63–95.

- [44] R. Cohen, A. Tsang, K. Vaidyanathan, H. Zhang, Analyzing opinion dynamics in online social networks, *Big Data & Information Analytics* 1 (4) (2016) 279–298.
- [45] L. Vassio, M. Garetto, C. Chiasserini, E. Leonardi, Temporal dynamics of posts and user engagement of influencers on facebook and instagram, in: *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '21*, Association for Computing Machinery, New York, NY, USA, 2022, p. 129–133. doi:10.1145/3487351.3488340.
URL <https://doi.org/10.1145/3487351.3488340>
- [46] M. Trevisan, L. Vassio, D. Giordano, Debate on online social networks at the time of covid-19: An italian case study, *Online Social Networks and Media* 23 (2021) 100136. doi:<https://doi.org/10.1016/j.osnem.2021.100136>.
URL <https://www.sciencedirect.com/science/article/pii/S2468696421000203>