Eur. Phys. J. Spec. Top. (2021) 230:2875–2886
https://doi.org/10.1140/epjs/s11734-021-00151-8

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Bi-layer voter model: modeling intolerant/tolerant positions and bots in opinion dynamics

Didier A. Vega-Oliveros[1,2,a], Helder L. C. Grande[3,b], Flavio Iannelli[4,c], and Federico Vazquez[5,d]

[1] Institute of Computing, University of Campinas (Unicamp), Campinas, SP, Brazil
[2] Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA
[3] National Institute for Space Research (INPE), Sao Jose dos Campos, SP, Brazil
[4] URPP Social Networks, Universität Zürich, Andreasstrasse 15, 8050 Zurich, Switzerland
[5] Instituto de Cálculo, FCEN, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina

Received 5 December 2020 / Accepted 23 April 2021 / Published online 6 June 2021

Abstract The diffusion of opinions in social networks is a relevant process for adopting positions and attracting potential voters in political campaigns. Opinion polarization, bias, targeted diffusion, and the radicalization of postures are key elements for understanding the voting dynamics' challenges. In particular, social bots are currently a new element that can have a pronounced effect on the formation of opinions during electoral processes by, for instance, creating fake accounts in social networks to manipulate elections. Here, we propose a voter model incorporating bots and radical or intolerant individuals in the decision-making process. The dynamics of the system occur in a multiplex network of interacting agents composed of two layers, one for the dynamics of opinions where agents choose between two possible alternatives, and the other for the tolerance dynamics, in which agents adopt one of the two tolerance levels. The tolerance accounts for the likelihood to change opinion in an interaction, with tolerant (intolerant) agents switching opinion with probability 1.0 ($\gamma \leq 1$). We find that intolerance leads to a consensus of tolerant agents during an initial stage that scales as $\tau^{+} \sim \gamma^{-1} \ln N$, who then reach an opinion consensus during the second stage in a time that scales as $\tau \sim N$, where N is the number of agents. Therefore, very intolerant agents ($\gamma \ll 1$) could considerably slow down dynamics towards the final consensus state. We also find that the inclusion of a fraction $\sigma_{\mathbb{B}}^{-}$ of bots breaks the symmetry between both opinions, driving the system to a consensus of intolerant agents with the bots' opinion. Thus, bots eventually impose their opinion to the entire population, in a time that scales as $\tau_B^{-} \sim \gamma^{-1}$ for $\gamma \ll \sigma_{\mathbb{B}}^{-}$ and $\tau_B^{-} \sim 1/\sigma_{\mathbb{B}}^{-}$ for $\sigma_{\mathbb{B}}^{-} \ll \gamma$.

1 Introduction

The voter model describes a simple process for opinion dynamics and consensus in a population of agents that can hold one of the two different opinions (A and B) [1,2]. In a single step of the dynamics, a voter chosen at random adopts a random neighbor's opinion. This step is repeated until voters' population eventually reaches a state of consensus in a finite system, where all agents share the same opinion. Due to its simplicity and analytical tractability, the voter model has become a paradigmatic model to study basic properties of opinion diffusion, and the dynamics of elections [3]. After its introduction in two independent works, by Clifford in 1973 [1] and soon lately by Liggett in 1975 [2], many extensions of the voter model have been proposed in the scientific literature to mimic more realistic or complex scenarios of social dynamics, such as considering multiple opinions [5–7], heterogeneity in transition rates [8,9], and complex interaction topologies that are static [10–15] or evolve in time [16,17], where clusters of opposite opinions coexist. Other works have studied how the presence of agents that never change opinion (stubborn individuals) affects the dynamics and consensus properties of the system [18–20]. Moreover, the introduction of personalized information [21], reinforcing the political orientation of an agent when its opinion changes, has shown to prevent global consensus for strong captured information change, showing the phenomena of strengthening political positions observed in many countries. This polarization behavior has also recently been explored through multistate voter models that include a mechanism of opinion reinforcement, which is a consequence of exchanging persuasive arguments [22–25]. Another implementation of the voter model has investigated the role of confidence in indi-

[a] e-mail: davo@unicamp.br
[b] e-mail: heldercgrande@gmail.com (corresponding author)
[c] e-mail: lavio.iannelli@business.uzh.ch
[d] e-mail: fede.vazmin@gmail.com

viduals by introducing two states per agent, its opinion, and its level of commitment to the opinion: unsure or tolerant and confident or intolerant [26]. After interacting with an agent of the opposite opinion, a tolerant agent can change its opinion, while an intolerant agent becomes tolerant but keeps its opinion. It is found that consensus is achieved very quickly in a mean-field setup (all-to-all interactions). At the same time, in square lattices of finite dimensions, the system reaches a metastable state where clusters of opposite opinions coexist for very long times until consensus is eventually reached.

Given the propensity of polarization in societies and the emergence of echo chambers within political conversations in online social networks (OSN) [27], social bots can be used to interfere in the political dialogue as a biased attack vector for opinion manipulation. For instance, some works showed evidence of the prevalence of bots in the 2016 US presidential elections [28], the UK-EU Brexit referendum [29], the 2018 Italian general election [30], and the 2019 Spanish general election [31]. Social bots can be defined as automatic agents designed to mimic or impersonate humans' behavior. They are prevalent as social actors in OSN platforms, amplifying misinformation effects in several magnitudes [27,32]. Due to their artificial nature, bots have specific aims, and they do not change their opinion, neither their posture about some parties, candidates, or topics. Therefore, it is natural to wonder how the inclusion of a minimum fraction of bots could modify the behavior of tolerant and intolerant individuals and what could be the impact on a given electoral process. How are the results of a simple model with bots compared to those obtained from "human" stubbornness in the voter model?

In this article, we introduce and study an extension of the voter model that incorporates bots and the tolerance level of agents. Each agent is endowed with an opinion (A, B) and a tolerance $(+, -)$ that is updated according to the voter dynamics. The opinion and tolerance processes are coupled to each other and take place on two different networks, forming a multiplex network topology. The dynamic on the opinion layer is affected by that of the tolerance layer by a mechanism that makes intolerant agents more resilient to switch opinion. This framework also allows the introduction of bots, modeled as agents that try to change other agents' opinions but are not influenced by them. Thus, these bots can be seen as stubborn agents that try to model the presence of opinion makers or the use of a false profile by political actors on a social network to influence electoral results.

We need to mention that some previous related works have also implemented voter-like dynamics on multiplex networks [33–37]. However, the models in these works explore how the propagation of an opinion, rumor, or information affects the spreading of a disease in a population. Therefore, they couple the voter dynamics in one layer with that of the SI, SIS or SIR dynamics in the other layer (S, I and R stands for susceptible, infected and recovered individuals), unlike in our model where both layer support a voter dynamics.

The rest of the article is organized as follows. In Sect. 2, we define the model and its dynamics on a bi-layer network. In Sect. 3, we develop a mean-field approach to study the version of the model without bots. We perform a stability analysis of the steady states and estimate the consensus times. Section 4 is dedicated to the study of the model with bots. Results from Monte Carlo simulations are presented in Sect. 5. Finally, in Sect. 6, we summarize and give the conclusions.

2 Multilayer voter model

We consider a population of interacting agents in which each agent can adopt one of the two possible opinions $\mathcal{O} = A$ or B. Besides, agents are endowed with a tolerance value $\mathcal{T} = +$ or $-$ that indicates the willingness of an agent to change its opinion, where the positive posture $(+)$ means that the agent is more tolerant and open to switching between both opinions, and the negative posture $(-)$ indicates that the agent is more radical or convinced about its own opinion, and thus less likely to change. The system of agents and their interactions are represented by a multiplex network composed of two layers of networks with an equal number of nodes (see Fig. 1), where nodes in layers 1 (tolerance layer $\pm$) and 2 (opinion layer AB) describe the tolerance and opinions of agents, respectively. The multiplex topology means that each node in the $\pm$-layer is connected to a node in AB-layer by an inter-layer link (dashed vertical arrow), representing an agent's opinion and its tolerance, but the configuration of links within

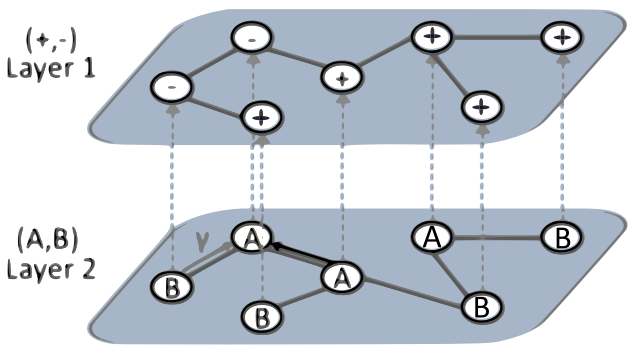

Fig. 1 Schematic representation of the bi-layer voter model. The network of interactions between agents' opinions (A and B) is represented in layer 2, while layer 1 represents the interactions between tolerance levels ($+$ and $-$) of agents. Opinion and tolerance states are updated according to the voter dynamics, i.e., by copying the state of a random neighbor in the corresponding layer. A tolerant agent (tolerance $+$) copies a neighbor's opinion with probability 1, while an intolerant agent (tolerance $-$) adopts the opinion of a neighbor with a smaller probability $\gamma \leq 1$, and becomes tolerant

each network layer (intra-layer connections) could be different. Besides, we consider that the networks have no degree correlations, i.e., nodes are randomly connected.

To simplify notation, we denote by $\left[\begin{smallmatrix}\mathcal{T}\\ \mathcal{O}\end{smallmatrix}\right]$ the state of a node in the bi-layer system, and thus there are four possible node states:

$$\left[\begin{smallmatrix}\mathcal{T}\\ \mathcal{O}\end{smallmatrix}\right] = \left\{\left[\begin{smallmatrix}+\\ A\end{smallmatrix}\right], \left[\begin{smallmatrix}-\\ A\end{smallmatrix}\right], \left[\begin{smallmatrix}+\\ B\end{smallmatrix}\right], \left[\begin{smallmatrix}-\\ B\end{smallmatrix}\right]\right\}. \tag{1}$$

In a single time step $\Delta t = 1/N$ of the dynamics, a node i with state $\left[\begin{smallmatrix}\mathcal{T}_i\\ \mathcal{O}_i\end{smallmatrix}\right]$ is chosen at random, and its tolerance $\mathcal{T}_i$ and opinion $\mathcal{O}_i$ are updated according to the voter dynamics. That is, a random neighbor j with state $\left[\begin{smallmatrix}\mathcal{T}_j\\ \mathcal{O}_j\end{smallmatrix}\right]$ is chosen from the $\pm$-layer, and a random neighbor k with state $\left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right]$ is chosen from the AB-layer. Then, node i copies the tolerance of node j in layer $\pm$ ($\mathcal{T}_i \to \mathcal{T}_i = \mathcal{T}_j$):

$$\left[\begin{smallmatrix}\mathcal{T}_i\\ \mathcal{O}_i\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_j\\ \mathcal{O}_j\end{smallmatrix}\right] \xrightarrow{1} \left[\begin{smallmatrix}\mathcal{T}_j\\ \mathcal{O}_i\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_j\\ \mathcal{O}_j\end{smallmatrix}\right]. \tag{2}$$

Also, node i copies the opinion of node k in layer AB ($\mathcal{O}_i \to \mathcal{O}_i = \mathcal{O}_k$) with probability 1 if its tolerance is $\mathcal{T}_i = +$:

$$\left[\begin{smallmatrix}+\\ \mathcal{O}_i\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right] \xrightarrow{1} \left[\begin{smallmatrix}+\\ \mathcal{O}_k\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right], \tag{3}$$

and with probability γ if $\mathcal{T}_i = -$:

$$\left[\begin{smallmatrix}-\\ \mathcal{O}_i\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right] \xrightarrow{\gamma} \left[\begin{smallmatrix}+\\ \mathcal{O}_k\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right], \quad \text{when } \mathcal{O}_i \neq \mathcal{O}_k \quad \text{and} \tag{4}$$

$$\left[\begin{smallmatrix}-\\ \mathcal{O}_i\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right] \xrightarrow{1} \left[\begin{smallmatrix}-\\ \mathcal{O}_k\end{smallmatrix}\right] \left[\begin{smallmatrix}\mathcal{T}_k\\ \mathcal{O}_k\end{smallmatrix}\right], \quad \text{when } \mathcal{O}_i = \mathcal{O}_k. \tag{5}$$

Table 1 in Appendix A shows explicitly all possible transitions when a pair of nodes interact.

In other words, agents adopt the tolerance of a random neighbor in the tolerance $\pm$-layer, following a known mechanism called social influence by which a tolerant individual tends to become intolerant or radical when most of their acquaintances are intolerant, and vice versa. In the opinion AB-layer, each agent copies a random neighbor's opinion with probability 1 if it is tolerant, but with probability γ if it is intolerant. This tries to capture the fact that intolerant or radical individuals are less likely to change opinion than tolerant or moderate individuals, which is modeled by assuming that intolerant agents change their minds with a reduced probability $\gamma \leq 1$. Additionally, if an intolerant agent does change opinion, we assume that it also becomes tolerant, as it is expected that a radical individual that changes its mind is prone to become more tolerant or, similarly, it is rarely expected that radical individuals suddenly adopt a radical position of the opposite view. As we can see, the dynamics of the two layers affect each other. On the one hand, the $\pm$-layer influences the dynamics on the AB-layer by reducing the rate at which intolerant agents switch opinion. On the other hand, the AB-layer influences the tolerance states in the $\pm$-layer by turning intolerant agents to tolerant when they change opinion.

3 Mean-field approach

The state of the system at the macroscopic level is well characterized by the global densities of nodes in each of the four tolerance–opinion states, $\sigma^{\mathcal{T}}_{\mathcal{O}} = \sigma^+_A, \sigma^-_A, \sigma^+_B$ and σ^-_B. Given that the number of nodes is conserved in each layer, the conditions $\sigma^+_A(t) + \sigma^-_A(t) + \sigma^+_B(t) + \sigma^-_B(t) = 1$, $\sigma_A(t) + \sigma_B(t) = 1$ and $\sigma^+(t) + \sigma^-(t) = 1$ must be fulfilled for all time $t \geq 0$, where $\sigma_{\mathcal{O}} = \sigma^+_{\mathcal{O}} + \sigma^-_{\mathcal{O}}$ and $\sigma^{\mathcal{T}} = \sigma^{\mathcal{T}}_A + \sigma^{\mathcal{T}}_B$ are the density of agents with opinion $\mathcal{O}$ and tolerance $\mathcal{T}$, respectively. Within a mean-field (MF) approach, the time evolution of the densities is given by the following set of rate equations:

$$\frac{d\sigma^+_A}{dt} = \sigma^-_A\sigma^+ + \gamma\sigma^-_B\sigma_A + \sigma^+_B\sigma_A - \sigma^+_A\sigma^- - \sigma^+_A\sigma_B, \tag{6a}$$

$$\frac{d\sigma^-_A}{dt} = \sigma^+_A\sigma^- - \sigma^-_A\sigma^+ - \gamma\sigma^-_A\sigma_B, \tag{6b}$$

$$\frac{d\sigma^+_B}{dt} = \sigma^-_B\sigma^+ + \gamma\sigma^-_A\sigma_B + \sigma^+_A\sigma_B - \sigma^+_B\sigma^- - \sigma^+_B\sigma_A, \tag{6c}$$

$$\frac{d\sigma^-_B}{dt} = \sigma^+_B\sigma^- - \sigma^-_B\sigma^+ - \gamma\sigma^-_B\sigma_A. \tag{6d}$$

This approach neglects state correlations between neighboring nodes in the networks. It, thus, should work reasonably well for random networks with homogeneous degree distributions and without degree correlations, such as the Erdös–Rényi networks. The gain and loss terms in Eqs. (6a, 6b, 6c, 6d) correspond to the different transitions between node states. For instance, the gain term $\sigma^-_A\sigma^+$ in Eq. (6a) corresponds to the transition of a node to state $\left[\begin{smallmatrix}-\\ A\end{smallmatrix}\right]$ to state $\left[\begin{smallmatrix}+\\ A\end{smallmatrix}\right]$ in a time step, when its tolerance switches from $-$ to $+$: a $\left[\begin{smallmatrix}-\\ A\end{smallmatrix}\right]$-node i is chosen with probability σ^-_A and copies the tolerance of a random neighbor j, which has tolerance $\mathcal{T}_j = +$ with probability σ^+. Within an MF approximation, we are assuming here that the fraction of neighbors of node i with tolerance $\mathcal{T} = +$ is approximately equal to σ^+.

Expanding the expressions for $\sigma^+, \sigma^-, \sigma_A$ and σ_B in Eqs. (6a, 6b, 6c, 6d) in terms of the four densities $\sigma^+_A, \sigma^-_A, \sigma^+_B$ and σ^-_A we obtain, after rearranging terms,

the following closed system of rate equations:

$$\frac{d\sigma_A^+}{dt} = 2\sigma_A^-\sigma_B^+ + \gamma\sigma_B^-(\sigma_A^+ + \sigma_A^-) - 2\sigma_A^+\sigma_B^-, \quad (7a)$$

$$\frac{d\sigma_A^-}{dt} = \sigma_A^+\sigma_B^- - \sigma_A^-\sigma_B^+ - \gamma\sigma_A^-(\sigma_B^+ + \sigma_B^-), \quad (7b)$$

$$\frac{d\sigma_B^+}{dt} = 2\sigma_B^-\sigma_A^+ + \gamma\sigma_A^-(\sigma_B^+ + \sigma_B^-) - 2\sigma_B^+\sigma_A^-, \quad (7c)$$

$$\frac{d\sigma_B^-}{dt} = \sigma_B^+\sigma_A^- - \sigma_B^-\sigma_A^+ - \gamma\sigma_B^-(\sigma_A^+ + \sigma_A^-). \quad (7d)$$

To study the behavior of the multilayer system, we numerically integrated Eqs. (7a, 7b, 7c, 7d) subject to the symmetric initial condition in opinion $\sigma_A(0) = \sigma_B(0) = 0.5$ and tolerance $\sigma^+(0) = \sigma^-(0) = 0.5$, and for six different values of $\sigma_B^-(0) = 0.25, 0.3, 0.35, 0.4, 0.45$ and 0.5, so that the other three node densities are $\sigma_A^+(0) = 0.5 - \sigma_B^+(0) = \sigma_B^-(0)$ and $\sigma_A^-(0) = \sigma_B^+(0) = 0.5 - \sigma_B^-(0)$. To explore how radical agents of a given opinion affect the final outcome of the model, we are considering an initial state that favors intolerant agents with opinion B ($\sigma_B^-(0) \geq 0.25$), compared to the perfectly symmetric condition $\sigma_A^+(0) = \sigma_A^-(0) = \sigma_B^+(0) = \sigma_B^-(0) = 0.25$.

3.1 Steady states

The system of Eqs. (7a, 7b, 7c, 7d) has four trivial fixed points $(\sigma_A^+, \sigma_A^-, \sigma_B^+, \sigma_B^-)$ =(1,0,0,0), (0, 1, 0, 0), (0, 0, 1, 0) and (0, 0, 0, 1) corresponding to a consensus in states A^+, A^-, B^+ and B^-, respectively. These are absorbing (inactive) states where there are no more possible updates, as all agents have the same opinion and tolerance. Besides, Eqs. (7a, 7b, 7c, 7d) have infinitely many non-trivial fixed points $\boldsymbol{\sigma}^* = (1 - \sigma_B^*, 0, \sigma_B^*, 0)$ that correspond to a consensus of tolerant agents ($\sigma^+ = 1, \sigma^- = 0$), where $\sigma_B^* = \sigma_B(t = \infty) = \sigma_B^+(t = \infty)$ ($\sigma_B^* \in [0, 1]$) is the stationary density of agents with opinion B. As there are only agents with + tolerance at the steady state, we have $\sigma_A(t = \infty) = \sigma_A^+(t = \infty) = 1 - \sigma_B^*$. This can be considered as a steady state of coexistence between A and B tolerant agents, with constant densities over time. This happens because the system is reduced to a simple 2-state symmetric voter model where the fraction of voters that make a transition from state A^+ to state B^+ per unit time, $\sigma_A^+\sigma_B^+$, is equal to the fraction of voters making the reverse transition (from B^+ to A^+); thus, the net flow is zero and the densities are conserved.

We have checked that the density of tolerant agents with opinion B at the stationary state σ_B^* depends on the initial condition, controlled by the initial density of opinion B intolerant agents $\sigma_B^-(0)$. This can be seen in Fig. 2a, where we plot σ_B^* vs the likelihood γ of intolerant agents to change opinion, for various values of $\sigma_B^-(0)$. We observe that, for a fixed value of γ, σ_B^* increases with $\sigma_B^-(0)$, meaning that a larger initial number of B-agents leads to a larger final number of B-agents. We also see a more intriguing effect, that σ_B^* increases as γ decreases. We can obtain an insight into these results from a closer inspection of Eqs. (7a, 7b, 7c, 7d). Adding Eqs. (7c) and (7d), we obtain that the density of opinion B agents evolves according to

$$\frac{d\sigma_B}{dt} = (1 - \gamma)(\sigma_A^+\sigma_B^- - \sigma_A^-\sigma_B^+), \quad (8)$$

while adding Eqs. (7a) and (7d) leads to the following evolution of the density of intolerant (−) agents:

$$\frac{d\sigma^-}{dt} = -\gamma\left(\sigma_A^+\sigma_B^- + 2\sigma_A^-\sigma_B^- + \sigma_A^-\sigma_B^+\right). \quad (9)$$

Given that all four initial densities can be written in terms of $\sigma_B^-(0)$, we arrive from Eq. (8) that at $t = 0$ is

$$\frac{d\sigma_B(0)}{dt} = (1 - \gamma)\left[\sigma_B^-(0) - 0.25\right], \quad (10)$$

which is larger than zero for all initial conditions $\sigma_B^-(0) > 0.25$ of Fig. 2a. Therefore, it is expected that, for any $\gamma > 0$, σ_B increases from 0.5 at $t = 0$ to a stationary value σ_B^* larger than 0.5 as $t \to \infty$, explaining why all curves of Fig. 2a are above 0.5, except the initially symmetric case $\sigma_B^-(0) = 0.25$ for which the densities are conserved. Another exception is the $\gamma = 1$ case, where opinion densities are conserved [see Eq. (8)], and so $\sigma_B(t) = \sigma_B(0) = 0.5$ and $\sigma_A(t) = \sigma_A(0) = 0.5$ for all $t \geq 0$.

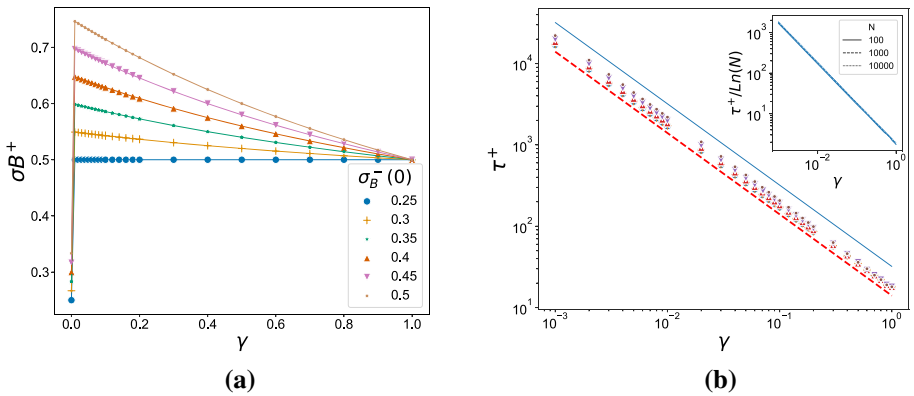

Fig. 2 Effects of the initial conditions, tolerance level γ and the initial density of bias radical individuals σ_B^-, in the outcome of the bi-layer voter model without bots: **a** the final density of tolerant agents with opinion B (σ_B^+); **b** the consensus time of tolerant agents with respect of γ. The red dotted line below is the $1/\gamma$ curve

As we described above, the initial asymmetric state that favors B^- agents leads to a stationary state with a majority of B-agents ($\sigma_B^* > 0.5$). This behavior is more pronounced as γ decreases [Eq. (10)], and it seems to be the reason why σ_B^* increases as γ approaches zero, as we see in Fig. 2a, showing a maximum (peak) in the $\gamma \to 0$ limit.

The $\gamma = 0$ case is special because the tolerance densities are conserved [see Eq. (20)], and so $\sigma^+(t) = \sigma^+(0) = 0.5$ and $\sigma^-(t) = \sigma^-(0) = 0.5$ for all $t \geq 0$. As a consequence, $\sigma_A^+ = 0.5 - \sigma_B^+$ and $\sigma_A^- = 0.5 - \sigma_B^-$ for $t \geq 0$. Replacing these expressions for σ_A^+ and σ_A^- in Eq. (8) we obtain

$$\frac{d\sigma_B}{dt} = 0.5\left(\sigma_B^- - \sigma_B^+\right), \tag{11}$$

and thus

$$\frac{d\sigma_B(0)}{dt} = 0.5\left[2\sigma_B^-(0) - 0.5\right], \tag{12}$$

at $t = 0$. Then, for $\sigma_B^-(0) \geq 0.25$ we expect that σ_B increases from 0.5 at $t = 0$ and reaches a value $\sigma_B^* \geq 0.5$. Finally, given that $\sigma_B^+ = \sigma_B^-$ at the stationary state [Eq. (11)], we have that $2\sigma_B^+ = \sigma_B^* \geq 0.5$ and thus $\sigma_B^+ \geq 0.25$, as we can check in Fig. 2a for $\gamma = 0$.

For $\gamma > 0$, the right-hand side of Eq. (20) is always negative, thus σ^- decreases and eventually approaches zero in the $t \to \infty$ limit, corresponding to a consensus in the tolerant state (+) at the steady-state ($\sigma^- = 0$, $\sigma^+ = 1$) as we mentioned before. A magnitude of interest is the time to reach the tolerant consensus τ^+. Given that the rate Eqs. (7a, 7b, 7c, 7d) describe an infinitely large system where finite-size fluctuations are neglected, we estimated τ^+ as the time for which the density of tolerant (+) agents becomes larger than $1 - 1/N$, that is, when there is less than one agent with state −. Results are shown in Fig. 2b, where we plot τ^+ as a function of γ for different initial conditions. We can see that τ^+ diverges as $\tau^+ \sim 1/\gamma$ when γ approaches zero. The intuition behind this result is that for $\gamma \ll 1$ the consensus time is determined by the slowest time scale of the system, associated with the transition of all intolerant agents − to the tolerant state + at rate γ, which takes a time of order $1/\gamma$. We also see that τ^+ is not strongly affected by the initial state that favors B^- agents ($\sigma_B^-(0) > 0.25$).

In summary, these results show that the system eventually reaches a tolerant consensus in the long run. Still, the convergence could be extremely slow when radical agents are unlikely to change their opinion, and that it becomes infinitely large (there is never consensus) for the extreme case of stubborn or intolerant agents ($\gamma = 0$).

3.2 Stability analysis and consensus times

A better estimation of the tolerant consensus time τ^+ can be obtained from a linear stability analysis of the tolerant fixed point $\boldsymbol{\sigma}^* = (1 - \sigma_B^*, 0, \sigma_B^*, 0)$. For that, we consider small perturbations ϵ_i ($i = 1, 2, 3, 4$) of the components of $\boldsymbol{\sigma}^*$ and write $\sigma_A^+ = 1 - \sigma_B^* + \epsilon_1$, $\sigma_A^- = \epsilon_2$, $\sigma_B^+ = \sigma_B^* + \epsilon_3$ and $\sigma_B^- = \epsilon_4$, where $\sum_{i=0}^4 \epsilon_i = 0$. Inserting these expressions for the densities into Eqs. (7a, 7b, 7c, 7d) and neglecting terms of order 2 we obtain, to first order in ϵ_i, the following system of linear equations in matrix representation:

$$\frac{d\boldsymbol{\epsilon}}{dt} = \mathbf{A}\,\boldsymbol{\epsilon},$$

where

$$\mathbf{A} \equiv \begin{pmatrix} 0 & 2\sigma_B^* & 0 & (1-\sigma_B^*)(\gamma-2)) \\ 0 & -\sigma_B^*(1+\gamma) & 0 & 1-\sigma_B^* \\ 0 & \sigma_B^*(\gamma-2) & 0 & 2(1-\sigma_B^*) \\ 0 & \sigma_B^* & 0 & -(1-\sigma_B^*)(1+\gamma) \end{pmatrix},$$

and $\boldsymbol{\epsilon} \equiv (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$. Matrix $\mathbf{A}$ has two negative eigenvalues

$$\lambda_\pm = \frac{-1-\gamma \pm \sqrt{(1+\gamma)^2 - 4\sigma_B^*(1-\sigma_B^*)\gamma(2+\gamma)}}{2}, \tag{13}$$

associated with a perturbation in the total densities of + and − agents 1.0 and 0, respectively, but that keeps the densities of A and B-agents $1 - \sigma_B^*$ and σ_B^*, respectively, unchanged. This means that the tolerance consensus state is stable. The other two eigenvalues are zero. One corresponds to the conservation of the total density of agents 1.0, and the other describes the instability of $\boldsymbol{\sigma}^*$ after a perturbation that changes the densities of A and B-agents. Then, the perturbations evolve according to $\epsilon_i = a_i + b_i\, e^{\lambda_+ t} + c_i\, e^{\lambda_- t}$, where a_i, b_i and c_i are constants given by the initial condition, and thus the density of tolerant agents $\sigma^+ = \sigma_A^+ + \sigma_B^+$ evolves after a smaller perturbation as

$$\begin{aligned}\sigma^+(t) \simeq 1 + \epsilon_1(t) + \epsilon_3(t) = 1 + (a_1 + a_3) \\ +(b_1 + b_3)e^{\lambda_+ t} + (c_1 + c_3)e^{\lambda_- t}.\end{aligned} \tag{14}$$

As we know that σ^+ approaches 1 as $t \to \infty$ [see Eq. (20)] and that $\lambda_- < \lambda_+ < 0$ for $\gamma > 0$, the coefficient corresponding to the 0 eigenvalue $a_1 + a_3$ must be zero. Besides, at long times, only the term corresponding to the largest eigenvalue λ_+ survives (smallest absolute value), and thus Eq. (14) becomes

$$\sigma^+(t) \simeq 1 + (b_1 + b_3)e^{\lambda_+ t}. \tag{15}$$

The time to reach consensus can be estimated from Eq. (15) as the time τ^+ for which the density of tolerant + agents reaches the value $1 - 1/N$, that is, $\sigma^+(\tau) = 1 + (b_1 + b_3)e^{\lambda_+ \tau} = 1 - 1/N$, from where we arrive at the approximate expression

$$\tau^+ \simeq \frac{\ln\left[-(b_1 + b_3)N\right]}{-\lambda^+}. \tag{16}$$

We notice that, as $b_1 + b_3 < 0$ [Eq. (15)] and $\lambda_+ < 0$, expression Eq. (16) gives a physical time $\tau^+ > 0$. In Fig. 2b, we see that the approximate expression Eq. (16) (solid lines) captures quite well the behavior of τ^+ with γ obtained from the integration of Eqs. (7a, 7b, 7c, 7d) (symbols).

4 Inclusion of bots

We now include in the model a fraction of Bots $\sigma_{\mathbb{B}}^-$ that remains constant over time. Bots are artificial entities that diffuse opinions related to a specific position. Due to their artificial nature, bots do not change opinion neither the posture. In this section, we analyze the effects of including bots that have a fixed opinion B, and so they can be considered as extremist intolerant agents in the state B^-. The total density of agents is now decomposed in five terms,

$$\sigma_A^+ + \sigma_A^- + \sigma_B^+ + \sigma_B^- + \sigma_{\mathbb{B}}^- = 1, \tag{17}$$

where $\sigma_{\mathbb{B}}^-(t) = \sigma_{\mathbb{B}}^-(0)$ for all $t \geq 0$. We also consider the same initial conditions as that without bots, determined by $\sigma_B^-(0)$, i.e., $\sigma_A^+(0) = \sigma_B^-(0)$ and $\sigma_A^-(0) = \sigma_B^+(0) = 0.5 - \sigma_B^-(0)$, leading to $\sigma^+(0) = 0.5$ and $\sigma_A(0) = 0.5$. The rates equations for the evolution of the densities can be derived following the same procedure as that for the model with no bots at the beginning of Sect. 3, considering an extra compartment $\mathbb{B}^-$ that behaves as an intolerant state B^-, but with the important distinction that transitions from state $\mathbb{B}^-$ to states B^+ and A^+ are not allowed (see Table 2 in Appendix A for a detailed description of all possible transitions). We make clear that agents "see" a bot as another B^- agent. Still, they make transitions only between the four states A^+, A^-, B^+ and B^+ (never to the bots' state $\mathbb{B}^-$), so that the total density of agents $1 - \sigma_{\mathbb{B}}^-$ as well as the density of bots $\sigma_{\mathbb{B}}^-$ are conserved quantities. The resulting set of MF equations reads

$$\frac{d\sigma_A^+}{dt} = 2\sigma_A^-\sigma_B^+ + \gamma\sigma_B^-(\sigma_A^+ + \sigma_A^-) - 2\sigma_A^+(\sigma_B^- + \sigma_{\mathbb{B}}^-), \tag{18a}$$

$$\frac{d\sigma_A^-}{dt} = \sigma_A^+(\sigma_B^- + \sigma_{\mathbb{B}}^-) - \sigma_A^-\sigma_B^+ - \gamma\sigma_A^-(\sigma_B^+ + \sigma_B^- + \sigma_{\mathbb{B}}^-), \tag{18b}$$

$$\frac{d\sigma_B^+}{dt} = \sigma_A^+(2\sigma_B^- + \sigma_{\mathbb{B}}^-) + \gamma\sigma_A^-(\sigma_B^+ + \sigma_B^- + \sigma_{\mathbb{B}}^-) - \sigma_B^+(2\sigma_A^- + \sigma_{\mathbb{B}}^-), \tag{18c}$$

$$\frac{d\sigma_B^-}{dt} = \sigma_B^+(\sigma_A^- + \sigma_{\mathbb{B}}^-) - \sigma_B^-\sigma_A^+ - \gamma\sigma_B^-(\sigma_A^+ + \sigma_A^-). \tag{18d}$$

4.1 Steady states

We integrated the rate Eqs. (18a, 18b, 18c, 18d) for different fractions of bots $\sigma_{\mathbb{B}}^-$ and different initial conditions that favor σ_B^-, to explore how different proportions of bots, combined with tolerant agents and asymmetric initial conditions, affects the outcome of the model. Results are shown in Fig. 3. In Fig. 3a and b, we observe that the stationary density of B^- agents for different initial conditions and $\gamma > 0$ is $\sigma_B^- = 1 - \sigma_{\mathbb{B}}^-$, that is, there is always a consensus of intolerant B-agents, except for $\gamma = 0$. It seems that bots break the symmetry of A and B opinions observed in the baseline model without bots of Sect. 3.1, introducing a bias towards B^- agents that prevents the tolerant (+) consensus found for the case with no bots. Indeed, as we see in Fig. 3b, for the no bots case $\sigma_{\mathbb{B}}^- = 0$ is $\sigma_B^- = 0$, while adding a small fraction of bots is enough to remove the + consensus and drive the system to the B^- consensus. For $\gamma = 0$, agents that become intolerant of the A opinion never escape from that state, and thus, a consensus in B^- is never reached.

4.2 Consensus times and stability analysis

In Fig. 3c, we plot the time to reach the B^- consensus τ_B^- as a function of $\sigma_{\mathbb{B}}^-$ for various values of γ. We see that τ_B^- decays with $\sigma_{\mathbb{B}}^-$ as $\tau_B^- \sim 1/\sigma_{\mathbb{B}}^-$ for $\sigma_{\mathbb{B}}^- \ll 1$, independent of γ (solid line). In Fig. 3d, we plot τ_B^- as a function of γ for various values of $\sigma_{\mathbb{B}}^-$, where the y-axis was rescaled by $\sigma_{\mathbb{B}}^-$ to collapse the data for values of γ close to 1.0. We can see that $\tau_B^- \sim C/\gamma$ for $\gamma \ll 1$, with an amplitude $C(\sigma_{\mathbb{B}}^-)$ that depends on $\sigma_{\mathbb{B}}^-$. To gain a better understanding of these results, below we derive equations for the evolution of the density of A-agents and +-agents. For that, we add Eqs. (18a) and (18b) to obtain

$$\frac{d\sigma_A}{dt} = -(1-\gamma)(\sigma_A^+\sigma_B^- - \sigma_A^-\sigma_B^+) - \sigma_{\mathbb{B}}^-(\sigma_A^+ + \gamma\sigma_A^-), \tag{19}$$

and Eqs. (18a) and (18c) to arrive at

$$\frac{d\sigma^+}{dt} = -\sigma_{\mathbb{B}}^-\sigma^+ + \gamma(\sigma_A^-\sigma_B^+ + 2\sigma_A^-\sigma_B^- + \sigma_A^+\sigma_B^- + \sigma_{\mathbb{B}}^-\sigma_A^-). \tag{20}$$

Although these equations cannot be solved exactly, it proves instructive to analyze the $\gamma = 1$ case, for which Eq. (19) adopts the simple form

$$\frac{d\sigma_A}{dt} = -\sigma_{\mathbb{B}}^-\sigma_A, \tag{21}$$

with solution $\sigma_A(t) = \sigma_A(0)\, e^{-\sigma_{\mathbb{B}}^- t}$. Then, σ_A decays exponentially fast to zero in a time that scales as $1/\sigma_{\mathbb{B}}^-$. Once the fraction of A-agents is less than $1/N$ (negligible small for $N \gg 1$), the second term of Eq. (20) can be neglected assuming that all terms inside the brackets are of order $1/N$ (they depend on $\sigma_A^\pm$), and thus we have

$$\frac{d\sigma^+}{dt} = -\sigma_{\mathbb{B}}^-\sigma^+, \tag{22}$$

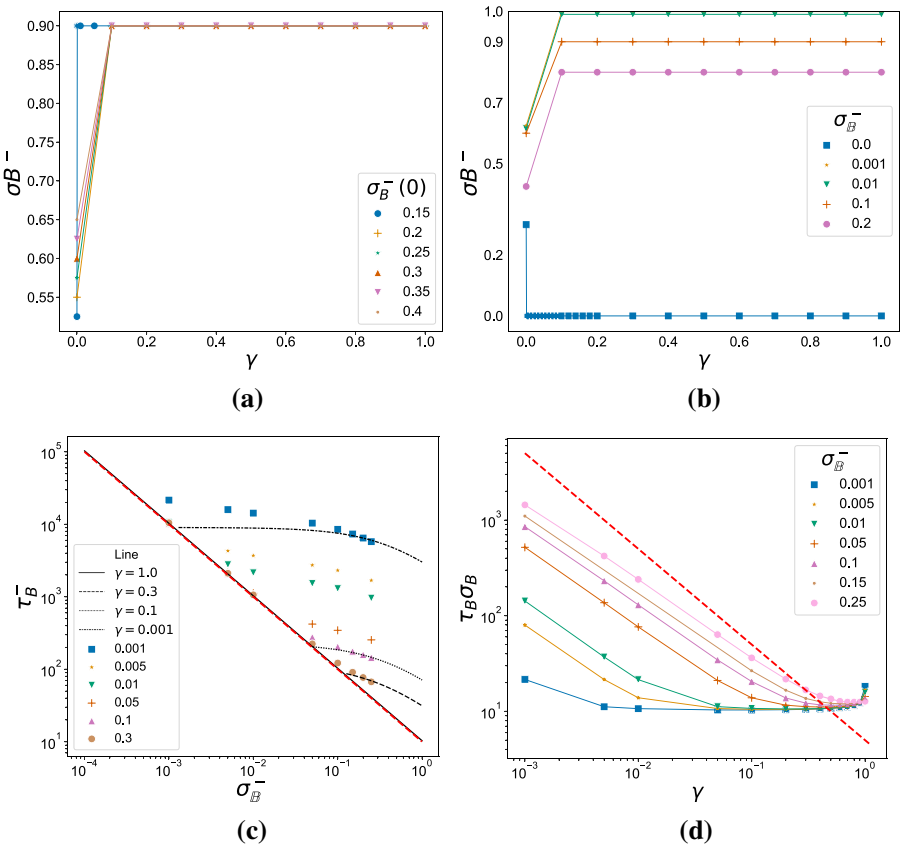

Fig. 3 Effects of the initial conditions in the outcome of the bi-layer voter model with the inclusion of bots. **a** The final density of intolerant agents B when the density of bots is $\sigma_{\mathbb{B}}^- = 0.1$ and different bias densities of intolerant B. **b** The effect of the density of bots in the final density of intolerant B agents, with no initial intolerance bias, i.e., $\sigma_B^+ = 0.25$. **c** The numerical and approximated consensus time to reach the final density of intolerant B. The dotted line represents $1/\sigma_{\mathbb{B}}^-$. **d** The numerical consensus time to reach the final density of B agents, in terms of the density of bots $\sigma_{\mathbb{B}}^-$. The dotted line is γ^{-1} curve

from where we obtain a consensus to the $-$ state that also scales as $1/\sigma_{\mathbb{B}}^-$. Therefore, as both the initial B-consensus and the subsequent $-$ consensus scale as $1/\sigma_{\mathbb{B}}^-$, we find that $\tau_B^- \sim 1/\sigma_{\mathbb{B}}^-$. This explains the pure power law behavior of τ_B^- with $\sigma_{\mathbb{B}}^-$ for $\gamma = 1$ (solid line in Fig. 3a). For $\gamma < 1$ the arguments above are not valid any more, because both time scales $1/\sigma_{\mathbb{B}}^-$ and $1/\gamma$ are at play. A more precise approach to the general case of any $\sigma_{\mathbb{B}}^-$ and γ is given by a linear stability analysis similar to that of Sect. 3.2 for the case without bots, as we describe below.

all densities fall in the $[0, 1]$ interval, but the analysis is also valid for $\epsilon_i < 0$. Inserting these expressions for the densities into Eqs. (18a, 18b, 18c, 18d) and expanding to first order in $|\epsilon_i| \ll 1$, we obtain $d\boldsymbol{\epsilon}/dt = \mathbf{A}\,\boldsymbol{\epsilon}$, where

$$\mathbf{A} \equiv \begin{pmatrix} \gamma - 2(1+\sigma_{\mathbb{B}}^-) & \gamma & 0 & 0 \\ 1+\sigma_{\mathbb{B}}^- & -\gamma(1+\sigma_{\mathbb{B}}^-) & 0 & 0 \\ 2+\sigma_{\mathbb{B}}^- & \gamma(1+\sigma_{\mathbb{B}}^-) & -\sigma_{\mathbb{B}}^- & 0 \\ 1+\gamma & \gamma & -\sigma_{\mathbb{B}}^- & 0 \end{pmatrix}.$$

The eigenvalues of matrix $\mathbf{A}$ are

$$\lambda_1 = 0, \tag{23}$$

$$\lambda_2 = -\sigma_{\mathbb{B}}^-, \tag{24}$$

$$\lambda_{3,4} = \frac{-2-(2+\gamma)\sigma_{\mathbb{B}}^- \pm \sqrt{\left[2+(2+\gamma)\sigma_{\mathbb{B}}^-\right]^2 - 4\gamma(1+\sigma_{\mathbb{B}}^-)(1-\gamma+2\sigma_{\mathbb{B}}^-)}}{2}. \tag{25}$$

The only fixed point in the system of Eqs. (18a, 18b, 18c, 18d) is $(0, 0, 0, 1)$, corresponding to a B^- consensus, as we mentioned above. We consider a small generic perturbation of this absorbing state of the form $\sigma_A^+ = \epsilon_1$, $\sigma_A^- = \epsilon_2$, $\sigma_B^+ = \epsilon_3$ and $\sigma_B^- = 1 - \epsilon_4$, such that $\sum_{i=1}^3 \epsilon_i - \epsilon_4 = 0$. The reason why we chose the $-\epsilon_4$ perturbation is to give a physical meaning to all perturbations, considering that $\epsilon_i > 0$ $(i = 1, \ldots, 4)$, thus

The eigenvalue $\lambda_1 = 0$ expresses the conservation of the total density of agents excluding bots $1 - \sigma_{\mathbb{B}}^-$. Given that λ_2, λ_3 and λ_4 are negative, the consensus fixed point $(0, 0, 0, 1)$ is stable, as expected. As we explained in Secti. 3.2, the consensus time is estimated by the exponential decay of the slowest mode $e^{\lambda_{\max} t}$ $(\lambda_{\max} < 0)$ to the fixed point after a perturbation, $\tau_B^- \sim -\ln N/\lambda_{\max}$, which corresponds to the mode with the largest negative eigenvalue $\lambda_{\max}$. Then, given that

$\lambda_4 < \lambda_3$, the consensus time is given by the largest of the two eigenvalues λ_2 and λ_3, which depends non-trivially on the relation between $\sigma_{\mathbb{B}}^-$ and γ. That is, for a fixed value of $\gamma > 0$ and decreasing $\sigma_{\mathbb{B}}^-$, we have that λ_3 approaches the value $-1+\sqrt{1-\gamma(1-\gamma)} < 0$, while $\lambda_2 = -\sigma_{\mathbb{B}}^-$ approaches zero from bellow. Therefore, λ_2 becomes larger than λ_3 for $\sigma_{\mathbb{B}}^-$ small enough, and thus

$$\tau_B^- \sim \frac{\ln N}{\sigma_{\mathbb{B}}^-} \quad \text{for } \sigma_{\mathbb{B}}^- \to 0. \tag{26}$$

This is the behavior observed in Fig. 3c, where τ_B^- decays as power law of $\sigma_{\mathbb{B}}^-$ with an amplitude that is γ independent for $\gamma \geq 0.1$ (solid line). For $\gamma = 0.001$, it seems that the values of $\sigma_{\mathbb{B}}^-$ plotted are not small enough, so we expect that $\lambda_3 > \lambda_2$, and thus $\tau_B^- \sim -\ln N/\lambda_3$. In general, for a fixed $\gamma > 0$ there is a "crossover" value $\hat{\sigma}_{\mathbb{B}}^-$ for which $\lambda_2 = \lambda_3$, so that τ_B^- is determined by λ_2 for $\sigma_{\mathbb{B}}^- < \hat{\sigma}_{\mathbb{B}}^-$ and by λ_3 for $\sigma_{\mathbb{B}}^- > \hat{\sigma}_{\mathbb{B}}^-$. This is equivalent to setting τ_B^- to the largest of the two functions λ_2^{-1} and λ_3^{-1} vs $\sigma_{\mathbb{B}}^-$, as plotted in Fig. 3c by solid and dashed lines, respectively. We can see that the behavior $\tau_B^- \sim 1/\sigma_{\mathbb{B}}^-$ (solid line) fits the data very well for small values of $\sigma_{\mathbb{B}}^-$, while for larger values of $\sigma_{\mathbb{B}}^-$ the behavior of τ_B^- is dominated by λ_3 (dashed lines). Discrepancies around the crossover point $\hat{\sigma}_{\mathbb{B}}^-$ are due to fact that both time scales are similar close to this point, and thus τ_B^- is determined by both time scales.

A similar analysis can be done for the τ_B^- vs γ plot (Fig. 3d). A Taylor series expansion of expression Eq. (25) for λ_3 to first order in γ leads to $\lambda_3 \simeq -(1/2 + \sigma_{\mathbb{B}}^-)\gamma$. Therefore, the consensus time can be approximated as

$$\tau_B^-(\gamma) \simeq \begin{cases} \frac{\ln N}{(1/2+\sigma_{\mathbb{B}}^-)\gamma} & \text{for } \gamma \lesssim \hat{\gamma}, \\ \frac{\ln N}{\sigma_{\mathbb{B}}^-} \quad \gamma \gtrsim \hat{\gamma}, \end{cases} \tag{27}$$

where $\hat{\gamma} = 2\sigma_{\mathbb{B}}^-/(1+2\sigma_{\mathbb{B}}^-)$. In Fig. 3d, we can see that the approximation from Eq. (27) works well for γ small (dashed line), while approximating τ_B^- as a constant of γ for larger values of γ is not a good estimation. However, it seems to give the right scaling $\tau_B^- \simeq \ln N/\sigma_{\mathbb{B}}^-$ for $\gamma \lesssim 1$, as curves for different $\sigma_{\mathbb{B}}^-$ collapse into one curve when the y-axis is rescaled by $\sigma_{\mathbb{B}}^-$. Indeed, at $\gamma = 1$ we have $\lambda_3(\gamma = 1) = -\sigma_{\mathbb{B}}^-$.

5 Monte Carlo simulation results

We performed extensive Monte Carlo (MC) simulations of the dynamics of the bi-layer voter model described in Sect. 2 without bots and in Sect. 4 with bots, to check the results obtained with the MF approach (Sects. 3, 4). We run the simulations on a multiplex network composed of two networks of $N = 10^4$ nodes and mean degree $\langle k \rangle = 20$ each, which are strongly coupled to each other, i.e., every node in one network is connected to one node in the other network. In the first set of simulations, we used two Erdös–Renyi (ER) networks (Poisson degree distribution), while in the second set, we used two Barabasi–Albert (BA) or scale-free networks.

We notice that the only possible final state in the simulations is the fully ordered or consensus state, in which all agents have the same opinion and tolerance level, unlike in the MF analysis of the model without bots, where a stationary coexistence of both opinions is possible. This is because fluctuations in finite-size networks make the system ultimately fall in an absorbing state of complete order, where the system is trapped and can no longer evolve, while MF equations are for infinite large systems and neglect fluctuations. The results we shall present below correspond to average values over 500 independent realizations of the dynamics for each initial condition.

In Fig. 4, we show simulation results of the model without bots. Top panels (a) and (b) correspond to ER networks, while bottom panels (c) and (d) correspond to BA networks. In panels (a) and (c) we plot the average value of the final density of tolerant agents with opinion B, σ_B^*, as a function of γ, where we observe for both ER and BA networks a behavior that is similar to that found with the MF approach (see Fig. 2a), that is, the smaller the γ, the larger the σ_B^*. Panels (b) and (d) show the mean consensus time to the tolerant state (τ^+) as a function of γ, where we can see the decay of τ^+ with γ that approximately follows a power law with an exponent close to -1 (dashed line), in close agreement with the MF approach (Fig. 2b). This confirms that the system reaches a tolerant consensus which takes a time that increases as the intolerant agents become more resilient, i.e., as γ decreases. However, as we mentioned before, the system ultimately reaches consensus by fluctuations, something not captured by the MF equations. In the insets of Fig. 4b and (d), we plot the mean opinion consensus time τ_{AB}, where we see that τ_{AB} is independent of γ and of order $N = 10^4$. This is because the dynamic that leads to the final opinion consensus is that of the voter model between two symmetric states A^+ and B^+, which scales as $\tau_{AB} \sim N$, and does not depend on γ because there are no intolerant agents.

In Fig. 5, we show simulation results of the model with bots. Panels (a) and (b) show the final density of opinion B intolerant agents (σ_B^-) as a function of γ for different initial conditions. In agreement with MF results (Fig. 3a, b), the system always reaches a consensus of B^- agents for $\gamma > 0$ and $\sigma_{\mathbb{B}}^- > 0$, independent of the initial condition, while for $\gamma = 0$ the final state consists of an absorbing configuration with a coexistence of A^- and B^- agents. Panels (c) and (d) show the mean consensus time to the B^- state (τ_B^-) for ER and BA networks, respectively. We observe that τ_B^- decays as power law with exponent close to -1 (dashed line), as predicted by the MF theory (Fig. 3c). That is, the consensus time increases as the fraction of bots decreases. In the insets of panels (c) and (d) we see

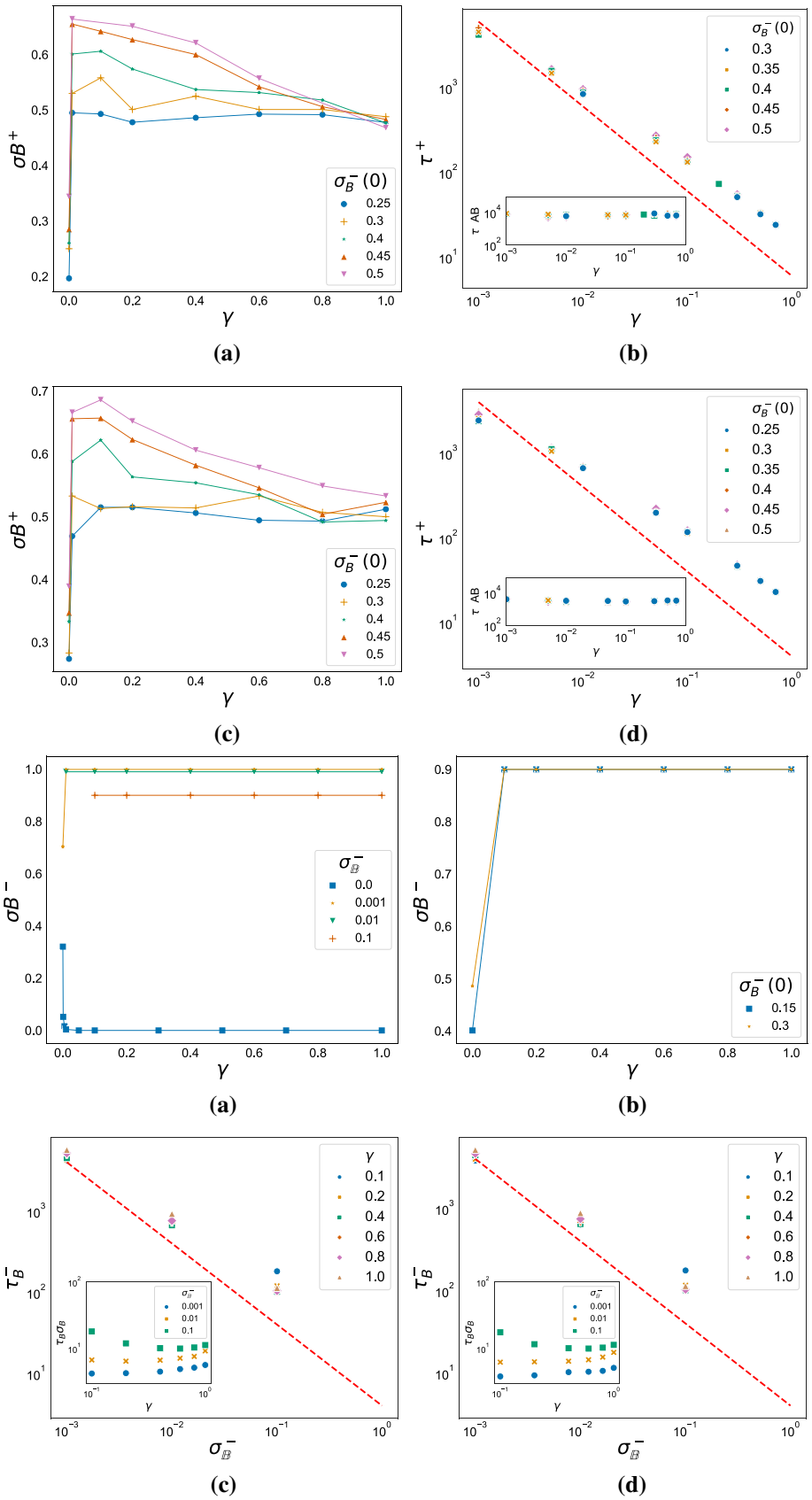

Fig. 4 Monte Carlo results of the simulated bi-layer voter model without the inclusion of bots. In **a** and **b**, we have the average value of the final densities of tolerant individuals with opinion B (σ_B^+) and the consensus time, both as a function of γ. The dotted red line represents the γ^{-1} curve. In the inset figure, we have the consensus time for the Layer AB. Top panels **a** and **b** are for the synthetic Erdös–Renyi (ER) network, and the bottom panels **c** and **d** are the analysis for the Barabási–Albert (BA) network. The legends indicate the initial density of radical individuals with opinion B

Fig. 5 Monte Carlo simulation results of the bi-layer voter model with the inclusion of bots. Figures **a** and **b** show the final density of opinion B intolerant agents (σ_B^-) as a function of γ for different initial conditions; in (b) we fixed to 10% the number of bots. Figures **c** and **d** show the average consensus time for the ER and BA networks, respectively; in the inset figures, we have the consensus time to reach the final density of B agents, in terms of the density of bots $\sigma_{\mathbb{B}}^-$. The dotted lines are the $1/\sigma_{\mathbb{B}}^-$ curve

that τ_B^- does not change much with γ. This is probably due to the fact that the values of γ used in simulations were not small enough (simulation are computationally very costly for $\gamma < 0.1$), possible hindering the power law behavior $\tau_B^- \sim \gamma^{-1}$ found in the MF approximation (Fig. 3d).

6 Summary and conclusions

We proposed a voter model on a multiplex network to study the interplay between the dynamics of opinions (A and B) and the tolerance (tolerant/intolerant) of individuals to accept others' opinions. Intolerant agents are less likely to change opinion, and they become tolerant when they do. We have also explored the effects of introducing a fraction of agents that play the role of bots, which are entities that never change opinion but can intentionally align other agents' opinions in a given direction. We performed simulations on Erdös–Renyi and Barabási–Albert networks and studied the system using an MF approach. When there are no bots in the population, both opinion states are symmetric. The system is initially driven towards a state where all agents are tolerant, with fractions of A and B-opinion agents that depend on the initial condition. This consensus of tolerant agents happens because there is a bias of agents from the intolerant to the tolerant state. After this first stage of tolerant consensus, there is a second stage where both opinions of tolerant agents evolve under the voter dynamics. As this dynamics in a finite system is only driven by finite-size fluctuations, the fractions of voters of each opinion perform a symmetric random walk until a consensus in one opinion is eventually reached. This final state of consensus is absorbing, as opinion and tolerance states can no longer evolve, unlike the initial tolerant consensus that is an active state where both opinions coexist. The time to reach the initial tolerant consensus scales as $\tau^+ \sim \gamma^{-1}$, given that it is controlled by the rate γ at which intolerant agents become tolerant. Consequently, radical agents can slow down the dynamics towards consensus by a factor that diverges as they become more persistent in their opinions ($\gamma \to 0$). The time to reach the final opinion consensus scales as in the voter model, $\tau_{AB} \sim N$, where N is the number of agents. Thus, the overall consensus time of the system is determined by γ in the case of very intolerant agents ($\gamma \ll 1/N$) and by N for very large systems ($N \gg \gamma^{-1}$).

Adding in the population bots that hold opinion B breaks the system's symmetry in both opinion and tolerance states, introducing a bias towards the intolerant opinion B state. This broken symmetry dramatically changes the model's outcome, where bots eventually impose their opinion to the rest of the system. As bots behave as intolerant agents, the final (absorbing) consensus state consists of all intolerant agents with opinion B. The consensus time has a non-trivial dependence on γ and the fraction of bots $\sigma_{\mathbb{B}}^-$, where the first controls the time scale associated with the persistence of intolerant agents and the second controls the bias towards intolerant opinion B. In the limiting case scenarios the consensus time is determined by the slowest of these two time scales, that is, $\tau_B^- \sim \gamma^{-1}$ for $\gamma \ll \sigma_{\mathbb{B}}^-$ and $\tau_B^- \sim 1/\sigma_{\mathbb{B}}^-$ for $\sigma_{\mathbb{B}}^- \ll \gamma$.

The results described above mean that radical individuals who are resilient to change their minds can significantly impact the consensus of opinions, slowing down the overall opinion consensus process. However, a striking consequence of the existence of radical or extremist individuals is that the entire population eventually becomes tolerant, in a state having only moderate individuals of both opinions, which are more prone to change and reach consensus. Therefore, the consensus of opinions in the model is a two-step process characterized by an initial extinction of extremists—who hinder opinion consensus—and a later debate between moderate individuals that facilitate consensus. Contrary to this result, bots can have the negative effect of preventing the state of tolerant consensus and leading the population to a state where every individual is an extremist of the opinion imposed by bots, which can be risky in democratic societies.

It might be interesting to study an extension of the model where intolerant agents switch opinion with a probability that depends on its opinion A or B, i.e., γ_A and γ_B, respectively. This could model a society where the level of individuals' tolerance depends on their opinion orientations, for instance, rightist or leftist. Also, it would be worthwhile to explore a version of the model with a quote of free will by adding the possibility that agents switch opinion spontaneously, modeled as external noise. These are variants of the model for future investigation.

Acknowledgements This research is supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant No.: 2015/50122-0 and the German Research Council (DFG-GRTK) Grant No.: 1740/2. D.A.V.O acknowledges the computational resources from the Center for Mathematical Sciences Applied to Industry (CeMEAI) under Grant 2013/07375-0, and FAPESP Grants 2016/23698-1, 2018/24260-5, and 2019/26283-5. F.V. acknowledges financial support from Agencia Nacional de Promoción Cienítfica y Tecnológica (Grant No. PICT 2016 Nro 201-0215). H.L.C.G. was funded by the research scholarship PCI-INPE, process 301113/2020-3. We thank Prof. Dr. Alessandro Vespignani, Dr. Dario Mazzilli, and PhD(c) Daniele Notarmuzi for useful comments and intellectual discussions.

Appendix A: complement of the explicit transitions rules

In this section, we explicitly write all transitions between opinion and tolerance states of agents in the model without bots (Table 1) and with bots (Table 2). The notations A^+, A^-, B^+ and B^- correspond to states of agents with

Table 1 Explicit transitions rules of the baseline bi-layer voter model without bots

Bi-layer voter transitions without bots						
Layer ±	$A^+\ A^-$	$\longrightarrow$	$A^-\ A^-$	$B^+\ B^-$	$\longrightarrow$	$B^-\ B^-$
	$A^-\ A^+$	$\longrightarrow$	$A^+\ A^+$	$B^-\ B^+$	$\longrightarrow$	$B^+\ B^+$
	$A^+\ B^-$	$\longrightarrow$	$A^-\ B^-$	$B^+\ A^-$	$\longrightarrow$	$B^-\ A^-$
	$A^-\ B^+$	$\longrightarrow$	$A^+\ B^+$	$B^-\ A^+$	$\longrightarrow$	$B^+\ A^+$
Layer AB	$A^+\ B^+$	$\longrightarrow$	$B^+\ B^+$	$B^+\ A^+$	$\longrightarrow$	$A^+\ A^+$
	$A^+\ B^-$	$\longrightarrow$	$B^+\ B^-$	$B^+\ A^-$	$\longrightarrow$	$A^+\ A^-$
	$A^-\ B^+$	$\xrightarrow{\gamma}$	$B^+\ B^+$	$B^-\ A^+$	$\xrightarrow{\gamma}$	$A^+\ A^+$
	$A^-\ B^-$	$\xrightarrow{\gamma}$	$B^+\ B^-$	$B^-\ A^-$	$\xrightarrow{\gamma}$	$A^+\ A^-$

Table 2 Explicit transition rules of the bi-layer voter model including bots

Bi-layer voter transitions including bots						
Layer ±	$A^+\ A^-$	$\longrightarrow$	$A^-\ A^-$	$B^+\ (B^- + \mathbb{B}^-)$	$\longrightarrow$	$B^-\ (B^- + \mathbb{B}^-)$
	$A^-\ A^+$	$\longrightarrow$	$A^+\ A^+$	$(B^- + \mathbb{B}^-)\ B^+$	$\longrightarrow$	$(B^+ + \mathbb{B}^-)\ B^+$
	$A^+\ (B^- + \mathbb{B}^-)$	$\longrightarrow$	$A^-\ (B^- + \mathbb{B}^-)$	$B^+\ A^-$	$\longrightarrow$	$B^-\ A^-$
	$A^-\ B^+$	$\longrightarrow$	$A^+\ B^+$	$(B^- + \mathbb{B}^-)\ A^+$	$\longrightarrow$	$(B^+ + \mathbb{B}^-)\ A^+$
Layer AB	$A^+\ B^+$	$\longrightarrow$	$B^+\ B^+$	$B^+\ A^+$	$\longrightarrow$	$A^+\ A^+$
	$A^+\ (B^- + \mathbb{B}^-)$	$\longrightarrow$	$B^+\ (B^- + \mathbb{B}^-)$	$B^+\ A^-$	$\longrightarrow$	$A^+\ A^-$
	$A^-\ B^+$	$\xrightarrow{\gamma}$	$B^+\ B^+$	$(B^- + \mathbb{B}^-)\ A^+$	$\xrightarrow{\gamma}$	$(A^+ + \mathbb{B}^-)\ A^+$
	$A^-\ (B^- + \mathbb{B}^-)$	$\xrightarrow{\gamma}$	$B^+\ (B^- + \mathbb{B}^-)$	$(B^- + \mathbb{B}^-)\ A^-$	$\xrightarrow{\gamma}$	$(A^- + \mathbb{B}^-)\ A^-$

opinion and tolerance A and $+$, A and $-$, B and $+$, and B and $-$, respectively. In a single time step $\Delta t = 1/N$ of the dynamics, one node is chosen at random. Then, this node copies the tolerance of a random neighbor in the $\pm$-layer, and the opinion of a random neighbor in the AB-layer. In Tables 1 and 2, the states on the left and right of a given pair correspond, respectively, to the focal agent—who changes state—and the random neighbor on the corresponding layer. Only situations that lead to a state change are included in the tables.

References

1. P. Clifford, A. Sudbury, Biometrika **60**, 581 (1973)
2. T.M. Liggett, *Interacting particle systems* (Springer, New York, 1975)
3. C. Castellano, S. Fortunato, V. Loreto, Rev. Mod. Phys. **81**, 591 (2009)
4. F. Vazquez, P.L. Krapivsky, S. Redner, J. Phys. A Math. Gen. **36**, L61–L68 (2003)
5. F. Vazquez, S. Redner, J. Phys. A Math. Gen. **37**, 8479 (2004)
6. D. Volovik, M. Mobilia, S. Redner, Europhys. Lett. **85**, 48003 (2009)
7. F. Vazquez, E.S. Loscar, G. Baglietto, Phys. Rev. E **100**, 042301 (2019)
8. N. Masuda, N. Gibert, S. Redner, Phys. Rev. E **82**, 010103 (2010)
9. D.A. Vega-Oliveros, L. Da F. Costa, F.A. Rodrigues, J. Stat. Mech. Theory Exp. **v.2**, 023401 (2017)
10. K. Suchecki, V.M. Eguíluz, M. San Miguel, EPL **69**(2), 228 (2004)
11. V. Sood, T. Antal, S. Redner, Phys. Rev. E **77**, 041121 (2008)
12. K. Suchecki, V.M. Eguíluz, M. San Miguel, Phys. Rev. E **72**, 036132 (2005)
13. V. Sood, S. Redner, Phys. Rev. Lett. **94**, 178701 (2005)
14. D.A. Vega-Oliveros, L. da Fontoura Costa, F.A. Rodrigues, Commun. Nonlinear Sci. Numer. Simul. **83**, 105094 (2020)
15. F. Vazquez, V.M. Eguíluz, N. J. Phys. **10**, 063011 (2008)
16. F. Vazquez, V.M. Eguíluz, M. San Miguel, Phys. Rev. Lett. **100**, 108702 (2008)
17. G. Demirel, F. Vazquez, G.A. Böhme, T. Gross, Phys. D **267**, 68–80 (2014)
18. S. Galam, Phys. A Stat. Mech. Appl. **333**, 49–55 (2004)
19. G. Serge, F. Jacobs, Phys. A. Stat. Mech. Appl. **381**, 366–376 (2007)
20. M. Mobilia, A. Petersen, S. Redner, J. Stat. Mech. Theory Exp. **2007**, P08029 (2007)
21. G. De Marzo, A. Zaccaria, C. Castellano, Phys. Rev. Res. **2**, 043117 (2020)
22. C.E. La Rocca, L.A. Braunstein, F. Vazquez, Europhys. Lett. **106**, 40004 (2014)
23. F. Velásquez-Rojas, F. Vazquez, J. Stat. Mech. Theor. Exp. **2018**, 043403 (2018)
24. F. Vazquez, N. Saintier, J.P. Pinasco, Phys. Rev. E **101**, 012101 (2020)
25. N. Saintier, J. Pablo Pinasco, F. Vazquez, Chaos **30**, 063146 (2020)
26. D. Volovik, S. Redner, J. Stat. Mech. Theory Exp. **2012**, P04003 (2012)
27. E. Colleoni, A. Rozza, A. Arvidsson, Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. J. Commun. **64**(2), 317–332 (2014)
28. O. Boichak, S. Jackson, J. Hemsley, S. Tanupabrungsun, Automated diffusion? Bots and their influence during the 2016 US presidential election. In: International conference on information (Springer, 2018), pp. 17–26

29. A. Duh, M. Slak Rupnik, D. Korošak, Collective behavior of social bots is encoded in their temporal twitter activity. Big Data **6**(2), 113–123 (2018)
30. M. Stella, M. Cristoforetti, M. De Domenico, Influence of augmented humans in online interactions during voting events. PLoS One **14**(5), e0214210 (2019)
31. J. Pastor-Galindo, M. Zago, P. Nespoli, S.L. Bernal, A.H. Celdrán, M.G. Pérez, J.A. Ruipérez-Valiente, G.M. Pérez, F.G. Mármol, Spotting political social bots in twitter: a use case of the 2019 Spanish general election. arXiv preprint arXiv:2004.00931 (2020)
32. D.M.J. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild et al., The science of fake news. Science **359**(6380), 1094–1096 (2018)
33. F. Velásquez-Rojas, F. Vazquez, Phys. Rev. E **95**, 052315 (2017)
34. P.C.V. da Silva, F. Velásquez-Rojas, C. Connaughton, F. Vazquez, Y. Moreno, F.A. Rodrigues, Phys. Rev. E **100**, 032313 (2019)
35. F. Velásquez-Rojas, P. Cesar Ventura, C. Connaughton, Y. Moreno, F.A. Rodrigues, F. Vazquez, Phys. Rev. E **102**, 022312 (2020)
36. W. Wang, M. Tang, H. Yang, Y. Do, Y.-C. Lai, G. Lee, Sci. Rep. **4**, 5097 (2014)
37. C. Granell, S. Gómez, A. Arenas, Phys. Rev. E **90**, 012808 (2014)