Online Supplementary Materials for: “Optimal control in opinion dynamics models: diversity of influence mechanisms and complex influence hierarchies”[1]

Ivan V. Kozitsin

Laboratory of Active Systems, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznaya street, Moscow, 117997, Russian Federation

Email: kozitsin.ivan@mail.ru

[1] Hereafter – Optimal control in opinion dynamics models: diversity of influence mechanisms and complex influence hierarchies.

Table of contents

A. Code *3*

B. Social influence mechanisms and their mathematical operationalization via transition matrices *4*

B.1. Assimilative influence **4**

B.2. Bounded confidence **6**

B.3. Empirically calibrated transition matrices (coexistence of assimilative & dissimilative influence mechanisms) **8**

B.4. Agents that are immune to social influence **9**

C. Solving the Cauchy problem (5), (6) from the Main Manuscript numerically *11*

D. Agreement between the mean-field predictions and simulations with the Advanced Model on sparse complex networks *14*

E. Auxiliary formulas and computations *22*

E.1. Supporting computations for the derivation of the mean-field approximation **22**

E.2. First Integrals of system (3) (Main Manuscript) **23**

E.3. The right side of the Euler-Lagrange equations (see (16) in the Main Manuscript) **23**

E.4. Singular control **24**

F. Technical details of the Direct and FBS methods *27*

F.1. Direct method **27**

F.2. FBS method **27**

F.3. Drawbacks **27**

G. Experiments *29*

G.1. Assimilative influence **29**

G.2. Bounded confidence **31**

G.3. Systems with several types of ordinary agents **34**

G.4. Systems with empirically calibrated transition matrices **39**

Literature *41*

A. Code

All experiments, visualizations, and analyses were performed in JupiterHub using the Python 3 language (version 3.10.8). For better structuring, the following manually created Python modules were developed (they can be found in the root folder at https://doi.org/10.7910/DVN/6D6OGG):

- *InitProblem_library* (for solving the initial value problem (6), (7) from the Main Manuscript)
- *DM_library* (for solving the control problem via the Direct method).
- *FBSM_library* (for solving the control problem via the FBS method).
- *ABM_library* (for performing simulations with the Advanced Model).
- *Support_library* (for visualization).
- *Transition_Matrices_Bank* (for storing transition matrices).

The following files (also located in the root folder) perform various numerical experiments and replicate the results presented in the Main Manuscript and Online Supplementary Materials:

Code (Part 0 - Computational Routine).ipynb	This file investigates the issue of solving the Cauchy problem (6), (7) from the Main Manuscript numerically and using simulations with the Advanced model.
Code (Part 1 - Case studies - Assimilative).ipynb	This file performs numerical experiments on solving the control problem for systems that follow the assimilative influence mechanism.
Code (Part 2 - Case studies - Bounded Confidence).ipynb	This file performs numerical experiments on solving the control problem for systems that follow the bounded confidence mechanism.
Code (Part 3 - Case studies - Several Types).ipynb	This file performs numerical experiments on solving the control problem when there are several types of ordinary agents.
Code (Part 4 - Case studies - Empirics).ipynb	This file performs numerical experiments on solving the control problem for systems in which influence structures are described by empirically calibrated transition matrices.

B. Social influence mechanisms and their mathematical operationalization via transition matrices

This section presents transition matrices that were involved in numerical experiments. These matrices can be grouped into three major classes based on the micro-level influence mechanisms they are drawing upon. To apply these mechanisms, it is necessarily to assume that the opinion alphabet is ordered: $x_1 \prec x_2 \prec \cdots \prec x_m$ and can approximate a one-dimensional opinion scale. Besides, we introduce a distance in the opinion space, which is measured as the absolute difference in opinion indices, as follows: $|x_i - x_j| \triangleq |i - j|$.

All matrices presented in this section are stored in the module *Transition_Matrices_Bank*. Each transition matrix is organized as a list of 2D NumPy arrays[2]. These arrays stand for matrix slices over the first index[3].

B.1. Assimilative influence

The assimilative influence mechanism postulates that interactions between agents reduce differences in the agents' opinions, and the greater the initial opinion discrepancy, the more distant an opinion shift should be expected (in other words, more distant opinions are more attractive). The classic example of this assumption can be found in the famous DeGroot model (DeGroot, 1974), in which agents follow the *linear* assimilative social influence mechanism whereby the magnitude of the opinion shift grows linearly with the distance in opinions between communicating agents.

Assimilative influence can be operationalized in different ways. Figure B1 presents three alternative formalizations in the case of a 3-element opinion space ($m = 3$). Whereas the matrices in panels A and B signify what is sometimes referred to as the *strict* assimilative influence mechanism, the transition matrix depicted on panel C realizes a *smooth* version of assimilative influence in which intermediate opinion shifts[4] are also allowed. Motivated by empirical studies (Carpentras et al., 2022; Kozitsin, 2022b, 2023; Moussaïd et al., 2013), the transition matrices presented on panels B and C showcase high resistance to social influence: the highest probability of opinion change outlined by them is only 0.2 (this happens only if the most distant opinions x_1 and x_3 interact).

[2] https://numpy.org/doc/stable/reference/generated/numpy.array.html

[3] A Transition matrix $\mathcal{P} = [p_{s,l,k}]_{s,l,k\in[m]}$ can be naturally presented as a list of 2D slices over the first index: $\mathcal{P}_1$, …, $\mathcal{P}_m$, where $\mathcal{P}_s = [p_{s,l,k}]_{l,k\in[m]}$.

[4] Let assume that agents i and j with opinions $o_i(t) = x_s$ and $o_j(t) = x_l$ interact. Agent j influences agent i. An opinion shift $x_s \to x_k$ undertaken by i is said to be intermediate if $x_s \prec x_k \prec x_l$ or $x_l \prec x_k \prec x_s$.

A

a		
1	0	0
0	1	0
0	0	1

b		
1	0	0
0	1	0
0	0	1

c		
1	0	0
0	1	0
0	0	1

B

a		
1	0	0
0.9	0.1	0
0.9	0	0.1

b		
0.1	0.9	0
0	1	0
0	0.9	0.1

c		
0.1	0	0.9
0	0.1	0.9
0	0	1

C

a		
1	0	0
0.9	0.1	0
0.8	0.15	0.05

b		
0.2	0.8	0
0	1	0
0	0.8	0.2

c		
0.05	0.15	0.8
0	0.1	0.9
0	0	1

Figure B1. In this figure, and in ongoing ones, the transition matrices are presented by their consecutive slices over the first index (see footnote 2 on the previous page). On each panel, tables a, b, and c stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ respectively. The matrix depicted on panel A indicates that a focal agent always adopts the influencing opinion. Panel B considers a different situation: the act of adoption occurs with probability of 0.1. Finally, according to panel C, the probability that the opinion of the focal agent will change grows with the absolute difference in opinions, and the more distant is the opinion of the influence source, the larger is the expected magnitude of opinion change: if the distance is one, then we get the magnitude of 0.1. If the polar opinions x_1 and x_3 interact, then the expected magnitude is 0.15 + 2*0.05 = 0.25.

Further, Figure B2 defines a transition matrix that outlines a smooth version of assimilative influence in the case of a 5-element opinion space ($m = 5$).

a

0.99	0.01	0	0	0
0.97	0.03	0	0	0
0.94	0.04	0.02	0	0
0.91	0.05	0.03	0.01	0
0.87	0.06	0.04	0.02	0.01

b

0.03	0.97	0	0	0
0.01	0.98	0.01	0	0
0	0.97	0.03	0	0
0	0.94	0.04	0.02	0
0	0.91	0.05	0.03	0.01

c

0.02	0.04	0.94	0	0
0	0.03	0.97	0	0
0	0.01	0.98	0.01	0
0	0	0.97	0.03	0
0	0	0.94	0.04	0.02

d

0.01	0.03	0.05	0.91	0
0	0.02	0.04	0.94	0
0	0	0.03	0.97	0
0	0	0.01	0.98	0.01
0	0		0.97	0.03

e

0.01	0.02	0.04	0.06	0.87
0	0.01	0.03	0.05	0.91
0	0	0.02	0.04	0.94
0	0	0	0.03	0.97
0	0	0	0.01	0.99

Figure B2. Tables a, b, c, d, and e stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$, $\mathcal{P}_5$ respectively. As it can be seen, both the probability and the expected magnitude of opinion change grow if the distance between interacting opinions goes up. In compliance with empirical observations (Carpentras et al., 2022; Kozitsin, 2022b, 2023; Moussaïd et al., 2013), even if the interacting opinions are equal, opinion changes are allowed with a small probability. This phenomenon is sometimes referred to as *anticonformity* (Krueger et al., 2017). Further, such a situation can be understood as an act of *reflection* in the sense that an individual rethinks their current belief after being exposed to it by someone else.

B.2. Bounded confidence

Within the bounded-confidence (BC) mechanism, only assimilative opinion shifts can happen. But they occur only if the interacting opinions are not too distant. Similar to the assimilative influence mechanism, the BC mechanism can also be operationalized in several ways. The literature highlights strict and smooth formalizations (Kurahashi-Nakamura et al., 2016). Examples of the first one can be found in Figure B3. In turn, the transition matrix depicted in Figure B4 presents a smooth version of the BC mechanism. All these matrices are designed in a 5-element opinion space because it provides more room for the formalization of various forms of influence-response functions (that define mathematically how opinions react to influence).

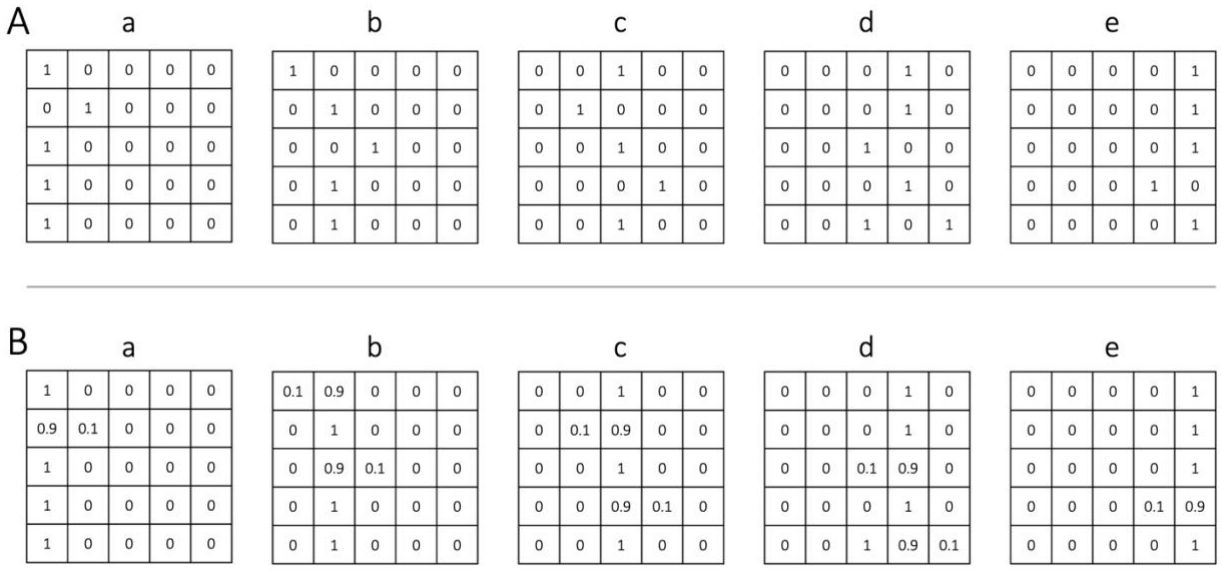


Figure B3. On each panel, tables a, b, c, d, and e stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$, $\mathcal{P}_5$ respectively. The transition matrix from panel A indicates that opinion changes may occur only if neighboring opinions interact, and the probability of adoption equals one on this occasion (the recipient with opinion x_s always adopts the opinion x_l of the influence source if $|x_s - x_l| = 1$). On panel B, we account for the high resistance to social influence observed in real social systems. Thus, communications between neighboring opinions lead to opinion changes in only one of every ten cases.

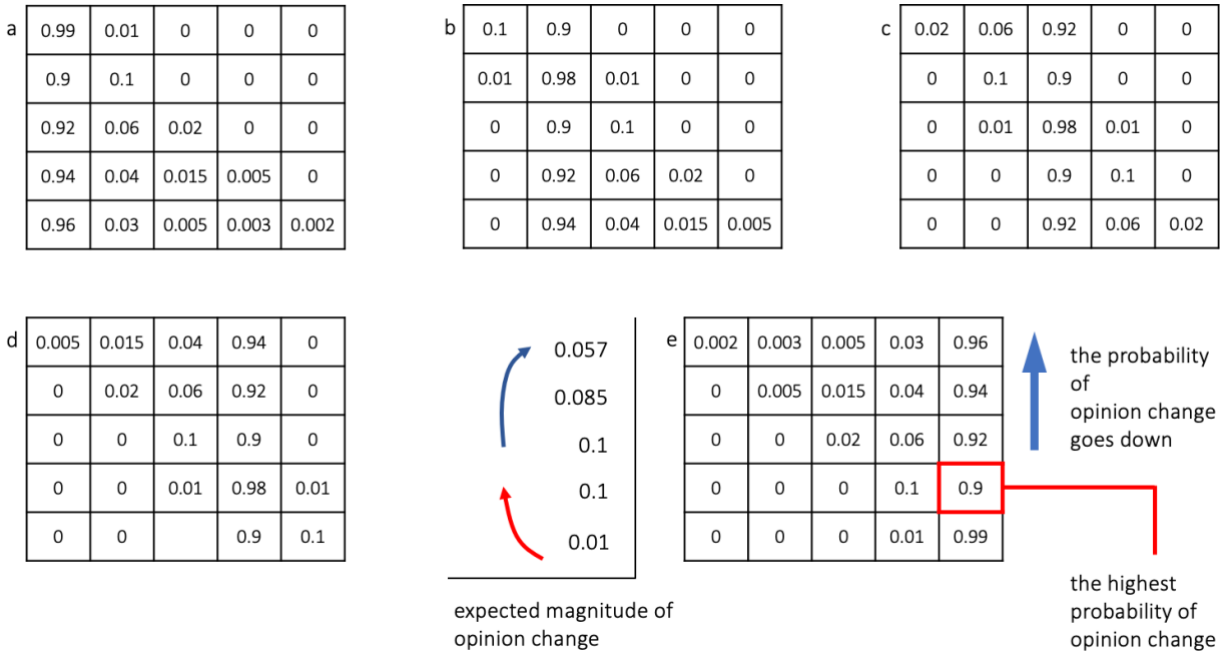


Figure B4. Tables a, b, c, d, and e stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$, $\mathcal{P}_5$ respectively. According to this transition matrix, the probability that an agent's opinion will change depends on the opinion distance between interacting agents via an inverted U-shape form (first grows, reaching the maximum of 0.9 at $|x_s - x_l| = 1$ (neighboring opinions), and then gradually decreases). Further, using the last slice $\mathcal{P}_5$, we demonstrate how the expected magnitude of opinion change varies with the distance between the opinions of communicating agents.

B.3. Empirically calibrated transition matrices (coexistence of assimilative & dissimilative influence mechanisms)

This subsection presents transition matrices calibrated on empirical data from Refs. (Kozitsin, 2022b, 2023). The data represent the opinion dynamics of a large-scale sample (~1.6 M) of online social network users around a political topic. As was reported in Ref. (Kozitsin, 2023), within these data, one can observe a quite complex asymmetric[5] map of influence processes, with both assimilative and dissimilative (directed opposite to the influence source) opinion changes. Figures B5 and B6 define the transition matrices calibrated on the data in 3- and 5-element opinion spaces (to calibrate these matrices, the methodology presented in Ref. (Kozitsin, 2022a) was implemented). Both the matrices are demonstrated to three decimal places.

a

0.96	0.04	0
0.942	0.057	0.001
0.907	0.091	0.002

b

0.04	0.952	0.008
0.021	0.969	0.01
0.02	0.944	0.036

c

0.001	0.082	0.917
0.001	0.07	0.929
0.001	0.054	0.945

Figure B5. The transition matrix calibrated on the data from Ref. (Kozitsin, 2022a) in a 3-element opinion space. Tables a, b, and c stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ respectively. All components are presented subject to 3 decimal places ($p_{1,1,3} > 0$)

[5] Users' opinions were estimated on a one-dimensional scale [0,1]. The aforementioned asymmetry was manifested in the fact that the left and right opinions display different levels of influence power.

a

0.947	0.044	0.008	0.001	0
0.954	0.04	0.006	0	0
0.938	0.051	0.01	0.001	0
0.911	0.07	0.016	0.003	0
0.919	0.063	0.009	0.009	0

b

0.055	0.872	0.071	0.002	0
0.046	0.898	0.055	0.001	0
0.032	0.893	0.073	0.002	0
0.04	0.859	0.097	0.004	0
0.047	0.858	0.092	0.003	0

c

0.002	0.057	0.919	0.022	0
0.002	0.052	0.935	0.011	0
0.001	0.038	0.943	0.018	0
0.001	0.039	0.913	0.046	0.001
0.001	0.043	0.896	0.059	0.001

d

0.001	0.002	0.102	0.876	0.019
0.001	0.005	0.093	0.886	0.015
0	0.002	0.078	0.905	0.015
0.001	0.002	0.068	0.898	0.031
0.001	0.003	0.069	0.884	0.043

e

0	0	0.011	0.097	0.892
0	0.003	0.009	0.081	0.907
0.001	0.001	0.006	0.072	0.92
0	0.001	0.006	0.069	0.924
0.001	0	0.005	0.071	0.923

Figure B6. The transition matrix calibrated on the data from Ref. (Kozitsin, 2022a) in a 5-element opinion space. Tables a, b, c, d, and e stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$, $\mathcal{P}_5$ respectively. All components are presented subject to 3 decimal places (those components who were equal to zero before rounding are colored gray).

B.4. Agents that are immune to social influence

To model the situation when ordinary agents are not subject to social influence (never change their opinions), in the case of a 3-element opinion space, we use the transition matrix visualized in Figure B7.

a

1	0	0
1	0	0
1	0	0

b

0	1	0
0	1	0
0	1	0

c

0	0	1
0	0	1
0	0	1

Figure B7. Tables a, b, and c stand for slices $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ respectively. According to this transition matrix, agents never change their opinions, regardless of what opinions they are exposed to.

Table B1. Glossary of transition matrices

Figure	Notes	Symbol

Figure B1, panel A	Assimilative influence, strict form ($m = 3$)	(B1, A)
Figure B1, panel B	Assimilative influence, strict form, high confidence of agents ($m = 3$)	(B1, B)
Figure B1, panel C	Assimilative influence, smooth form, high confidence of agents ($m = 3$)	(B1, C)
Figure B2	Assimilative influence, smooth form, high confidence of agents ($m = 5$)	(B2)
Figure B3, panel A	Bounded confidence, strict form ($m = 5$)	(B3, A)
Figure B3, panel B	Bounded confidence, strict form, high confidence of agents ($m = 5$)	(B3, B)
Figure B4	Bounded confidence, smooth form, high confidence of agents ($m = 5$)	(B4)
Figure B5	Empirically calibrated ($m = 3$)	(B5)
Figure B6	Empirically calibrated ($m = 5$)	(B6)
Figure B7	Agents are immune to social influence ($m = 3$)	(B7)

C. Solving the Cauchy problem (6), (7) from the Main Manuscript numerically

Both the Direct and FBS methods need to solve the initial value problem (6), (7) from the Main Manuscript for a given control function $u(\tau_k)$ on the grid of points $\tau_0, \tau_1, \dots, \tau_T$ that cover the interval $[\tau_0, \tau_T]$. The function *RKM_Dyn_System* (solver) that is realized in the module *InitProblem_library* makes this possible. The solver obtains the state function $y(\tau_k)$ drawing upon the Runge-Kutta 4th order direct method. We first tried to apply already existing Python procedures, such as *odeint*, *solve_ivp*, or *RK45* (see the Python subpackage *scipy.integrate*), but they appeared to be hardly adaptable to our needs.

Preliminary experiments revealed that for long integration intervals, the solver tends to violate the phase constraints $y_{1,f} + \dots + y_{m,f} = n_f, f \in [M]$. Typically, it reflects in the fact that state variables tend to zero or, inversely, grow unexpectedly (see Figure C1). To exclude these situations, we slightly modified the Runge-Kutta method: each time when the Runge-Kutta step is implemented, the newly approximated value $y(\tau_{k+1})$ of the state function is checked for whether it meets the phase restrictions. If not, then an *adjustment procedure* is launched: the nearest (in L^2 norm) to $y(\tau_{k+1})$ $m \times M$ matrix that meets the phase constraints is searched (via solving an auxiliary optimization problem) and then replaces $y(\tau_{k+1})$ as a new Runge-Kutta approximation $y^*(\tau_{k+1})$. A possible way to check if this ad-hoc modification gives reliable outputs is to launch simulations with the Advanced Model under the mean-field settings—trajectories of phase variables obtained in such simulations should mark or be close to ground-truth. This option is provided by the function *ABM_Dyn_System* that is a part of the module *ABM_library*. To make such a comparison possible, one should recall that the times in the mean-field differential equations (τ) and in the underlying agent-based model (t) are connected by $\tau = t/N$. In other words, one time unit of the scaled time equals N iterations of the model. It means that the time span $[\tau_0, \tau_T]$, where, for example, $\tau_0 = 0, \tau_T = 100$ corresponds to $100 * N$ model iterations. By the way, *ABM_Dyn_System* performs simulations not only in mean-field settings, but also in the cases where the structure of the system is described by complex sparse graphs that are much closer to real-world social networks (see Section D).

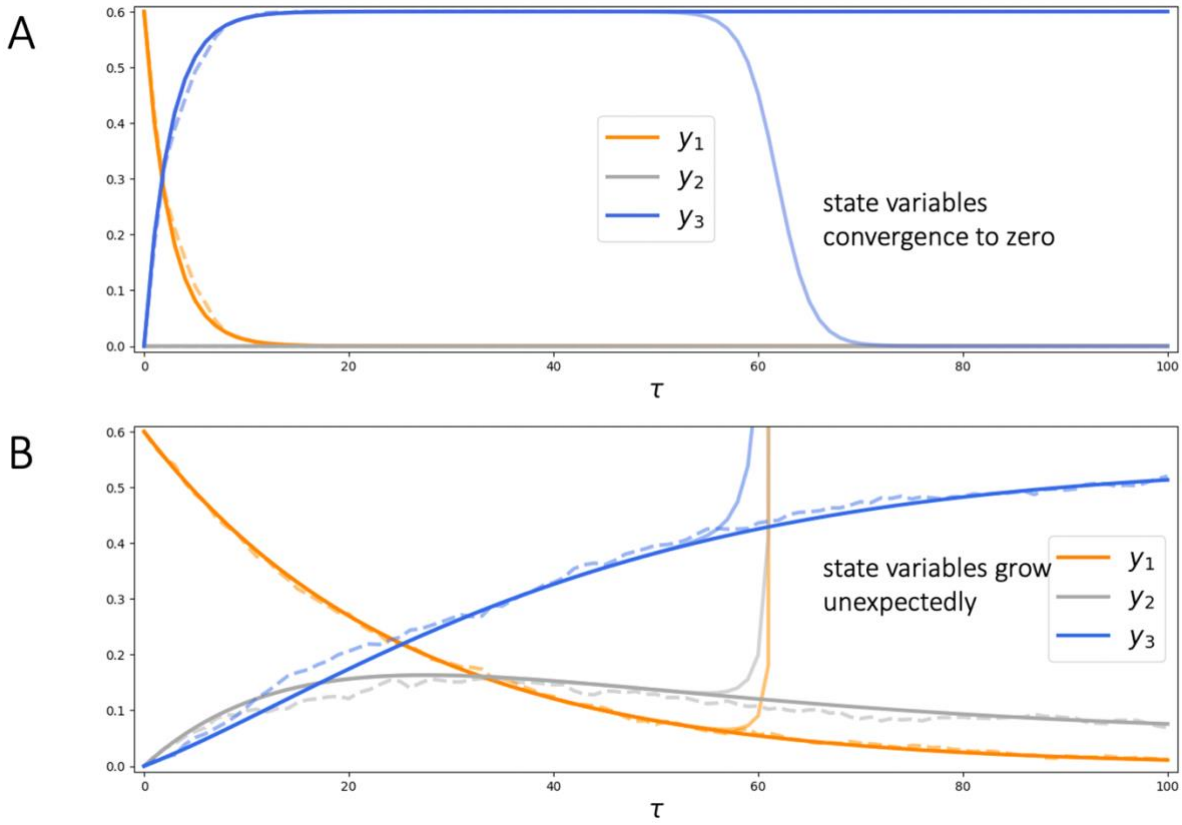


Figure C1. This figure compares the outputs of the function *RKM_Dyn_System* in the cases when the adjustment procedure is applied (solid lines) and not (transparent lines). As ground-truth (dashed lines), we use the results of a simulation run with the Advanced Model performed by the function *RKM_Dyn_System* under mean-field settings ($N = 1{,}000$). The system starts from the point $y_0 = [0.6 \quad 0 \quad 0]^T$ $(m = 3, M = 1)$ and is controlled by $u(\tau) \equiv [0 \quad 0 \quad 0.4]^T$. On this occasion, the phase constraints are given by $y_{1,1} + y_{2,1} + y_{3,1} = 0.6, y(\tau) \geq 0$. On the upper panel, we apply the transition matrix (B1, A) to model both in-group (ordinary agent – ordinary agent) and inter-group (ordinary agent – stubborn agent) interactions ($\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B1, A)). On the bottom panel, the transition matrix (B1, B) defines the in-group interactions ($\mathcal{P}^{1,1}$ = (B1, B)), whereas (B1, C) describes how stubborn agents influence ordinary ones ($\mathcal{P}^{1,2}$ = (B1, C)).

We report that after implementing the adjustment procedure, solutions derived by the function *RKM_Dyn_System* yield agreement with model simulations, a result that indicates reliability of our approach (see Figures C1 and C2). However, experiments revealed that more "complex" systems (in which, for example, several types of ordinary agents coexist) need more agents to make model simulations coincide with numerical solutions—in particular, results presented in Figure C3 were obtained for a system of $N = 1{,}000$ agents, whereas Figure C2 presents a simulation run with $N = 5{,}000$ agents.

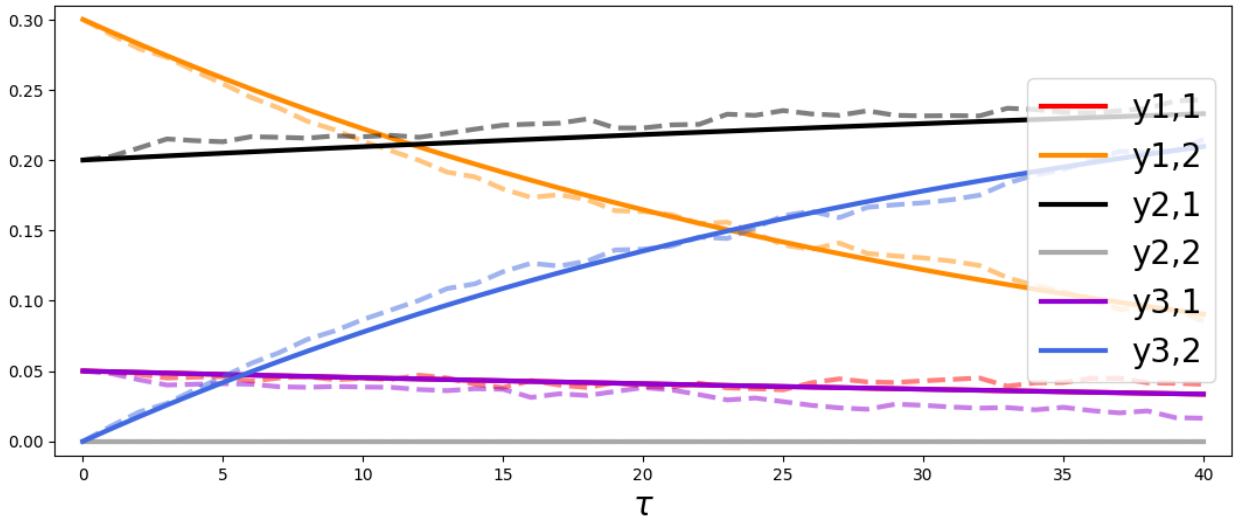


Figure C2. The solution of the function *RKM_Dyn_System* (solid lines) nearly coincides with the model simulations (dashed lines) under the mean-field settings ($N = 5000$). We consider a system in which two types of ordinary agents are presented ($M = 2$), with three possible opinions ($m = 3$). The starting point of the system is $y_0 = \begin{bmatrix} 0.05 & 0.3 \\ 0.2 & 0 \\ 0.05 & 0 \end{bmatrix}$, the control $u(\tau) \equiv \begin{bmatrix} 0 & 0 \\ 0.1 & 0 \\ 0 & 0.3 \end{bmatrix}$ is applied. The phase constraints are given by $y_{1,1} + y_{2,1} + y_{3,1} = n_1, y_{1,2} + y_{2,2} + y_{3,2} = n_2, y(\tau) \geq 0$ ($n_1 = 0.3, n_2 = 0.3$). The structure of influence processes is given by: $\mathcal{P}^{1,1} = \mathcal{P}^{2,2}$ = (B1, A) (ordinary agents are open-minded to in-group influence), $\mathcal{P}^{2,1} = \mathcal{P}^{1,2}$ = (B7) (ordinary agents reject inter-group influence from ordinary agents), and $\mathcal{P}^{2,3} = \mathcal{P}^{1,3}$ = (B1, B) (ordinary agents perceive influence from stubborn agents with a high confidence level).

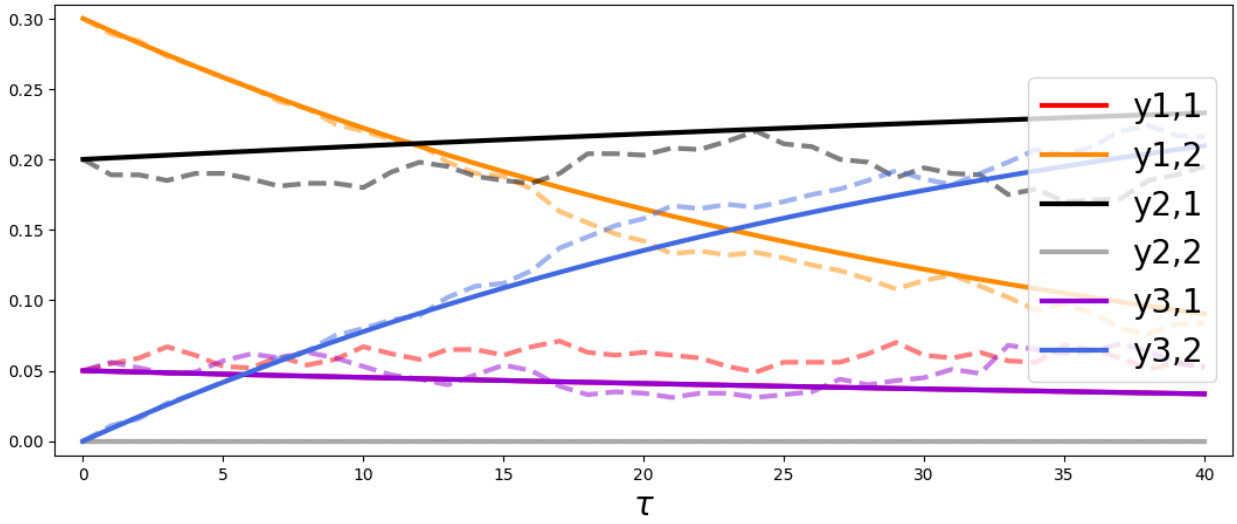


Figure C3. This figure replicates the results presented in Figure C2, but with a lower number of agents in the simulation ($N = 1000$).

D. Agreement between the mean-field predictions and simulations with the Advanced Model on sparse complex networks

The Ref. (Kozitsin, 2022a) reported that the mean-field predictions derived for the Basic Model remain valid even if considering sparse complex topologies, which, in fact, violate the underlying mean-field assumptions (in that paper, the author supposed that the network is a complete graph and the number of agents is huge). This phenomenon was observed *empirically* and was especially prominent for transition matrices that do not include components equal to one (see, for example, matrices (B2), (B4), (B5), or (B6)). Such behavior of the Basic Model gives us hope that the mean-field predictions (see (6), (7) in the Main Manuscript) and, correspondingly, our framework for finding the optimal control (obtained, however, for the Advanced Model), can be reliable even in the case of network topologies that appear in real life.

We test this hypothesis by relaxing the mean-field assumptions on network structure (see Subsection 3.4 in the Main Manuscript for details), just as prescribed in Subsection 5.3 of the Main Manuscript. This algorithm of network creation is realized in the module *ABM_library*.

Using the network creation algorithm described above, we compared our theoretical predictions based on the mean-field equations against simulations with the Advanced Model. Numerical experiments generally indicate agreement between these two alternative approaches (see Figures D1–D4). Some distinctions appear in the cases of large-dimension phase spaces (when m or M are huge) or non-smooth transition matrices (see Figures D2–D4). These effects, however, can be effectively mitigated by increasing the number of agents in simulations (subject to the density remains virtually the same).

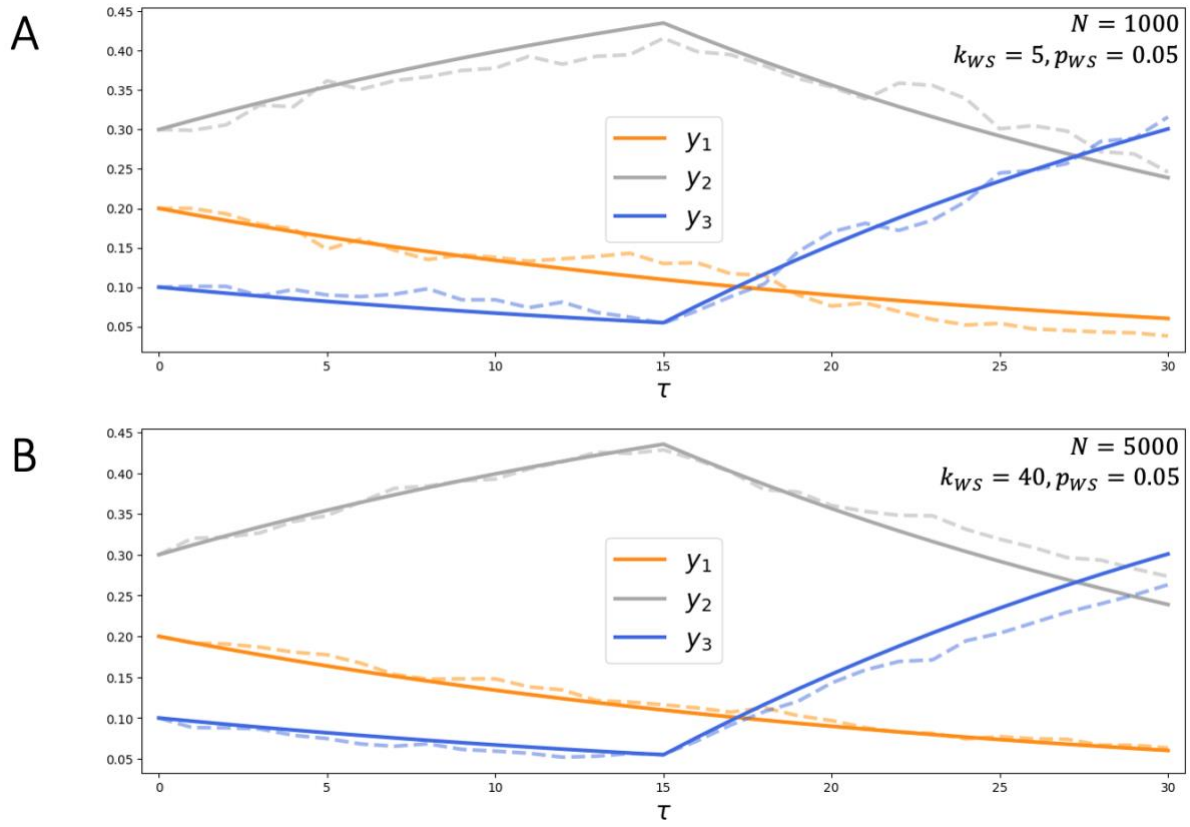


Figure D1. We check if the mean-field equations solved numerically (solid lines) can approximate the results of simulations with the Advanced Model performed on sparse complex topologies (dashed lines). In this example, the system starts from the point $y_0 = [0.2 \quad 0.3 \quad 0.1]^T$ $(m = 3, M = 1)$ and is controlled by:

$$u(\tau_k) = \begin{cases} [0 \quad 0.4 \quad 0]^T, & k \leq 15, \\ [0 \quad 0 \quad 0.4]^T, & k > 15. \end{cases}$$

The focal interval $[0,30]$ is covered by a grid of step $h = 1$. The structure of influence processes is as follows: $\mathcal{P}^{1,1}$ = (B1, A) (ordinary agents are open-minded to in-type influence), $\mathcal{P}^{1,2}$ = (B1, B) (ordinary agents perceive influence from stubborn ones with a high confidence level). The network that connects ordinary agents was developed using the Watts–Strogatz model[6]. On panel A, an experiment with $N = 1000$ agents is visualized (with the network parameters $k_{WS} = 8$ and $p_{WS} = 0.05$), and panel B reports results obtained for $N = 5000$ agents (with the network parameters $k_{WS} = 40$ and $p_{WS} = 0.05$).

[6] https://networkx.org/documentation/networkx-1.9/reference/generated/networkx.generators.random_graphs.watts_strogatz_graph.html

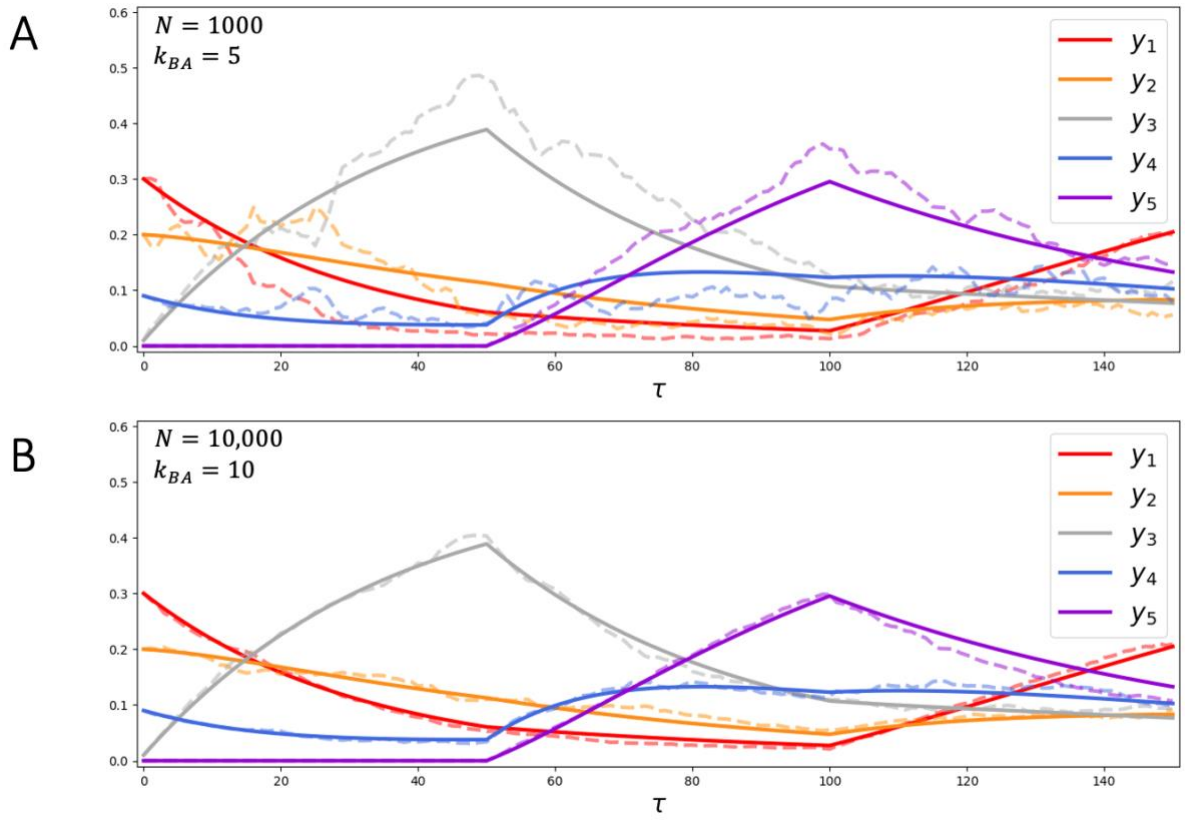


Figure D2. We check if the mean-field equations solved numerically (solid lines) can approximate the results of simulations with the Advanced Model performed on sparse complex topologies (dashed lines). The system starts from the point $y_0 = [0.3 \quad 0.2 \quad 0.01 \quad 0.09 \quad 0]^T$ $(m = 5, M = 1)$ and is controlled by:

$$u(\tau_k) = \begin{cases} [0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T, & k \leq 50, \\ [0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T, & 50 < k \leq 100, \\ [0.4 \quad 0 \quad 0 \quad 0 \quad 0]^T, & k > 100. \end{cases}$$

The focal interval $[0{,}150]$ is covered by a grid of step $h = 1$. The structure of influence processes is as follows: $\mathcal{P}^{1,1}$ = (B3, A), $\mathcal{P}^{1,2}$ = (B4). The network that connects ordinary agents was developed using the Barabasi–Albert model[7]. On panel A, an experiment with $N = 1000$ agents is visualized (network parameter $k_{BA} = 5$), panel B reports results obtained for $N = 10{,}000$ agents (network parameter $k_{BA} = 10$).

[7] https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.barabasi_albert_graph.html

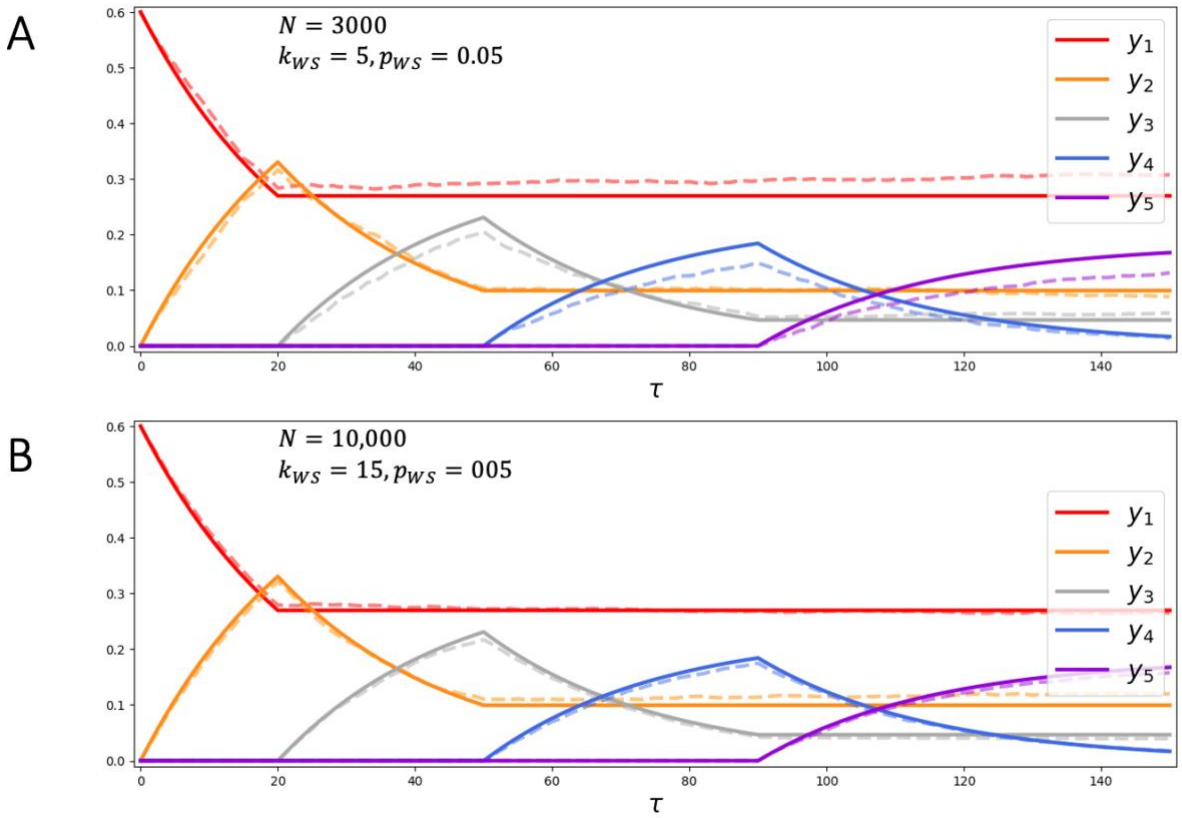


Figure D3. We check if the mean-field equations solved numerically (solid lines) can approximate the results of simulations with the Advanced Model performed on sparse complex topologies (dashed lines). The system starts from the point $y_0 = [0.6 \quad 0 \quad 0 \quad 0 \quad 0]^T$ $(m = 5, M = 1)$ and is controlled by:

$$u(\tau_k) = \begin{cases} [0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T, & k \leq 20, \\ [0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T, & 20 < k \leq 50, \\ [0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T, & 50 < k \leq 90, \\ [0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T, & k > 90. \end{cases}$$

The focal interval $[0{,}150]$ is covered by a grid of step $h = 1$. The structure of influence processes is as follows: $\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B3, B). The network that connects ordinary agents was developed using the Watts–Strogatz model. On panel A, an experiment with $N = 3000$ agents is visualized (network parameters $k_{WS} = 5$ and $p_{WS} = 0.05$), panel B reports results obtained for $N = 10{,}000$ agents (network parameters $k_{WS} = 15$ and $p_{WS} = 0.05$).

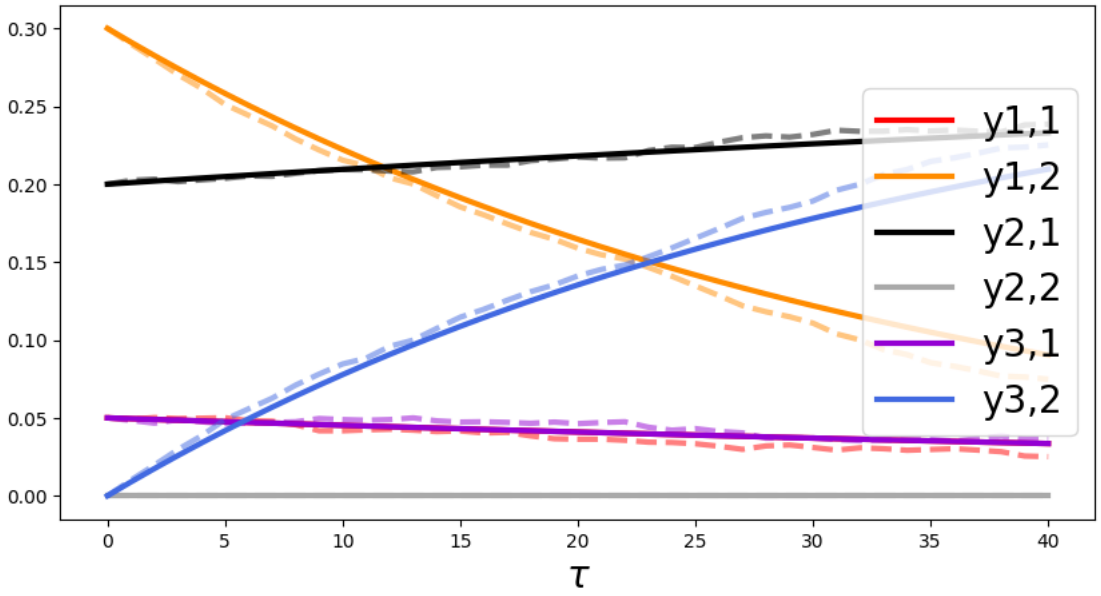


Figure D4. We check if the mean-field equations solved numerically (solid lines) can approximate the results of simulations with the Advanced Model performed on sparse complex topologies (dashed lines). This figure replicates the experiment presented in Figure C2, but now the ordinary agents are connected by a Watts–Strogatz graph with parameters $k_{WS} = 15$ and $p_{WS} = 0.05$. The number of agents in the simulation run was $N = 10{,}000$.

In all experiments presented above, nodal types and initial opinions were nested randomly, without regard to the network topology. It is worth noting that agreement between model simulations and the mean-field predictions can disappear for systems where the network topology correlates with nodal characteristics (see Figure D5). A prominent example of such settings is a situation when the network includes echo-chambers (see Figure D6).

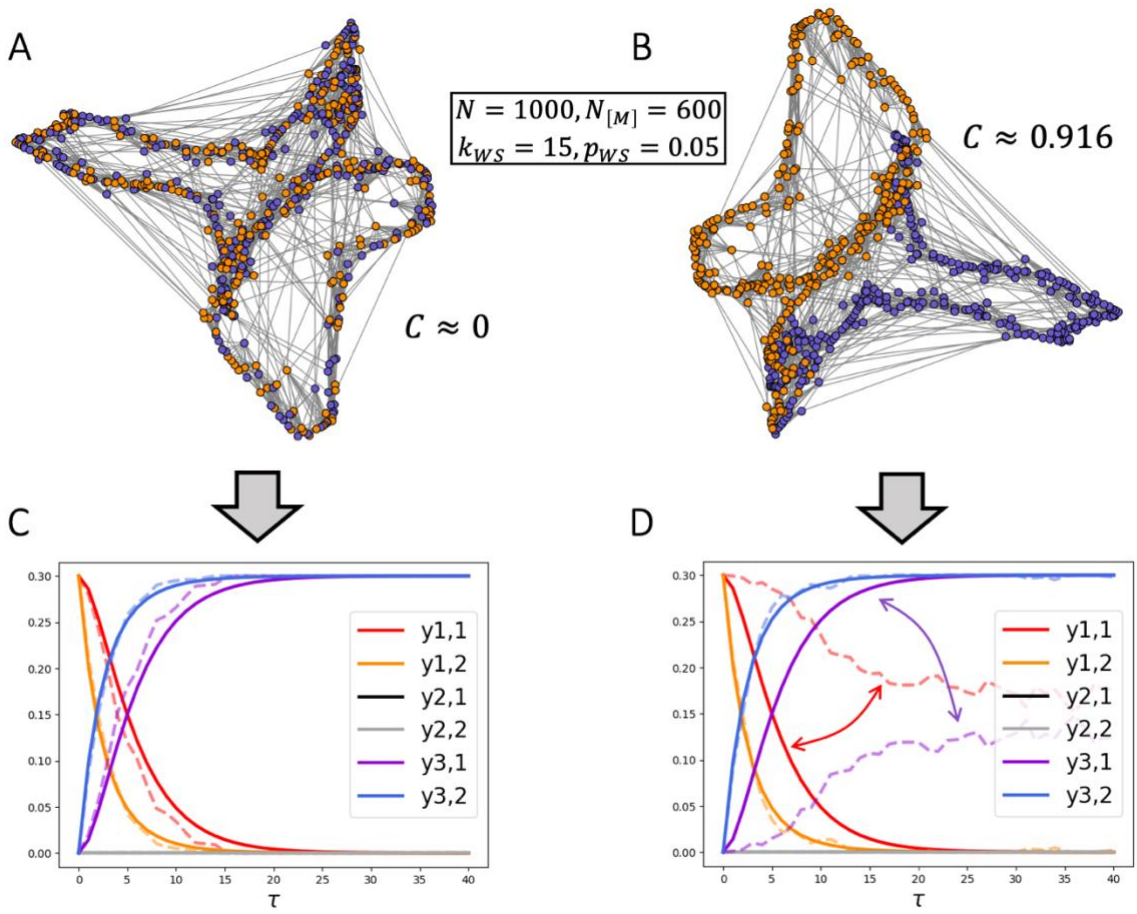


Figure D5. We check if the mean-field equations solved numerically (solid lines) can approximate the results of simulations with the Advanced Model performed on sparse complex topologies (dashed lines) – see panels C and D. This figure demonstrates the behavior of the system defined in the Main Manuscript (see Example 3) in the case the control function is

$$u(\tau) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0.4 \end{bmatrix}.$$

The system starts from the point

$$y_0(\tau) = \begin{bmatrix} 0.3 & 0.3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Panels C and D demonstrates the effect type-based assortativity on opinion dynamics by the example of the Watts–Strogatz network with parameters $N = 1000, k_{WS} = 15, p_{WS} = 0.05$ (note that only $0.6N = 600$ nodes are employed in network development, as the other nodes stands for stubborn agents). If types are nested among the nodes randomly (see Panel A), then the between the mean-field predictions affords a quite precise description of the system evolution (see Panel C). However, in the case of a high level of assortativity (see panel B), which can be achieved if the first 300 nodes will be nested with type Ξ_1 and next 300 nodes will be endowed with type Ξ_2, then a remarkable discord between the mean-field predictions and the behavior of the agent-based model will be observed

(see panel D). On panels A and B, we depict the corresponding values of the assortativity coefficient (calculated with respect to the attribute type) [8].

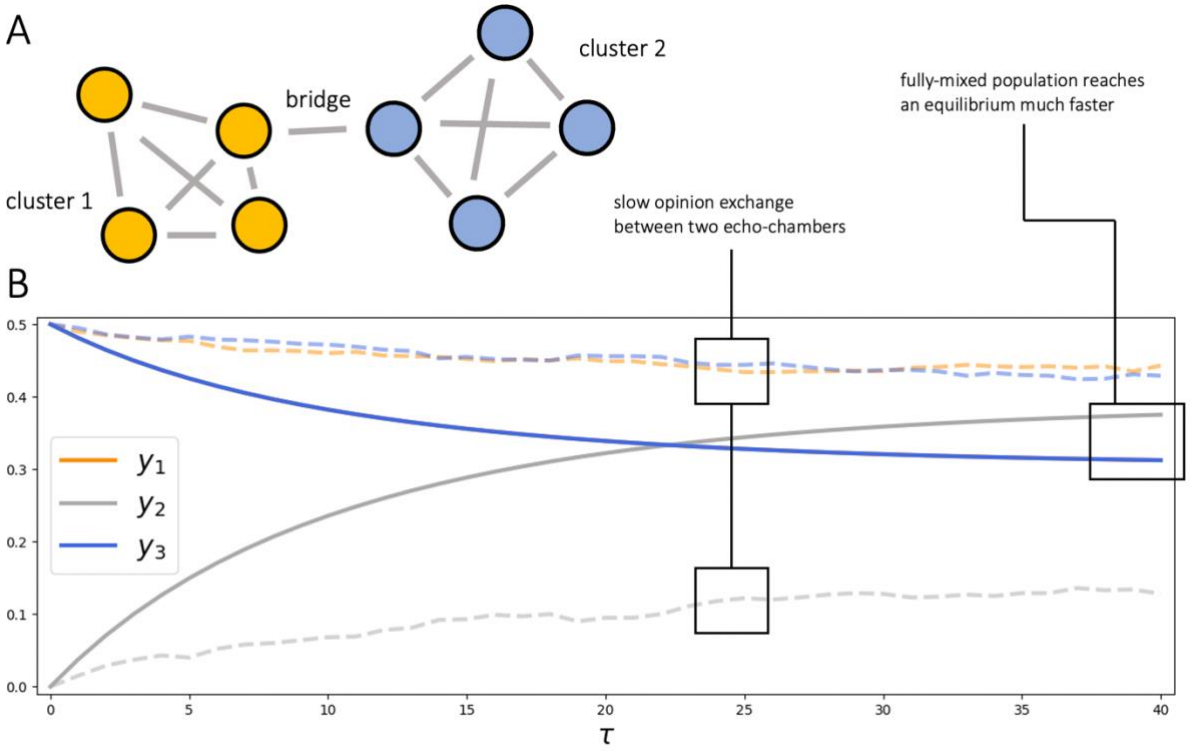


Figure D6. Let us consider a situation when ordinary agents are segregated into two dense clusters (for example, complete subgraphs) connected by one tie (bridge) – see panel A. For simplicity, we presume that there is no external control. Further, suppose that these clusters are opinion-homogeneous and polarized (say, the first cluster includes individuals that only support COVID-19 vaccination (x_1), whereas individuals from the second cluster are purely against (x_3))—$m = 3, M = 1$. If opinions are updated according to the assimilative influence mechanism (for example, as is outlined in the transition matrix (B1, A)), then under mean-field settings, one should expect that agents should quickly reach an equilibrium – see panel B, solid lines. However, if we account for the underlying structure ($N = 1000$), then opinions converge extremely slowly (dashed lines). The assortativity coefficient for this system is approximately equal to one—the highest possible rate.

It is worth noting that not all empirical networks feature such high assortativity rates. For example, networks of online friendship connections are quite moderate from the perspective of

[8] A possible approach to measure to what extent the network at stake is homophilic with respect to the a (numeric) attribute φ is to calculate the *assortativity coefficient*:

$$C(A,\varphi) = \frac{\sum_{i,j}\left(a_{ij} - \frac{\varphi_i\varphi_j}{2q}\right)\varphi_i\varphi_j}{\sum_{i,j}\left(k_i\delta_{ij} - \frac{\varphi_i\varphi_j}{2q}\right)\varphi_i\varphi_j} \in [-1,1].$$

In the above definition, $A = [a_{ij}]$ is the adjacency matrix of the network, k_i represents the degree of node i, and $q = \sum_{i,j} a_{ij}\,/2$ denotes the total number of edges in the network.

opinion assortativity (Balietti et al., 2021). For examples, Kozitsin (2023) reported the value of 0.14 (in this paper, opinions were measured on the scale $[0,1]$). Nonetheless, there could be other (non-opinion) attributes along which social networks could display assortative mixing patterns, such as age, gender, religion, etc. To be more specific, the network reported in Ref. (Kozitsin et al., 2023) was characterized by the age assortativity coefficient of 0.493 and by the opinion assortativity coefficient of 0.226 (opinions were measured on the scale $[0,1]$). On this occasion, one could wonder if such age-level correlations could challenge the quality of the mean-field predictions. To reconcile this question, we split the vertices of this empirical network into two sets ($M = 2$)—with age under (type Ξ_1) and over 35 (type Ξ_2), where 35 is the median value. Next, we discretized the opinion scale into three subintervals $\left[0,\frac{1}{3}\right), \left[\frac{1}{3},\frac{2}{3}\right), \left[\frac{2}{3},1\right]$ that mark opinion values x_1, x_2, and x_3 correspondingly. Finally, we fixed the structure of the network (which included 27,861 users after extraction of the giant component), as well as the joint distribution of opinions and types given by (subject to three decimal places).

$$y_0 = \begin{bmatrix} 0.2 & 0.24 \\ 0.237 & 0.236 \\ 0.071 & 0.016 \end{bmatrix}.$$

For this social system, we performed a simulation run and compared its results against the mean-field predictions (see Figure D7). We specified that in-type communications are governed by the transition matrix (B5), whereas inter-type interactions do not affect agents' opinions (transition matrix (B7)). As it can be seen from Figure D8, the mean-field predictions and the model simulation give roughly the same outputs.

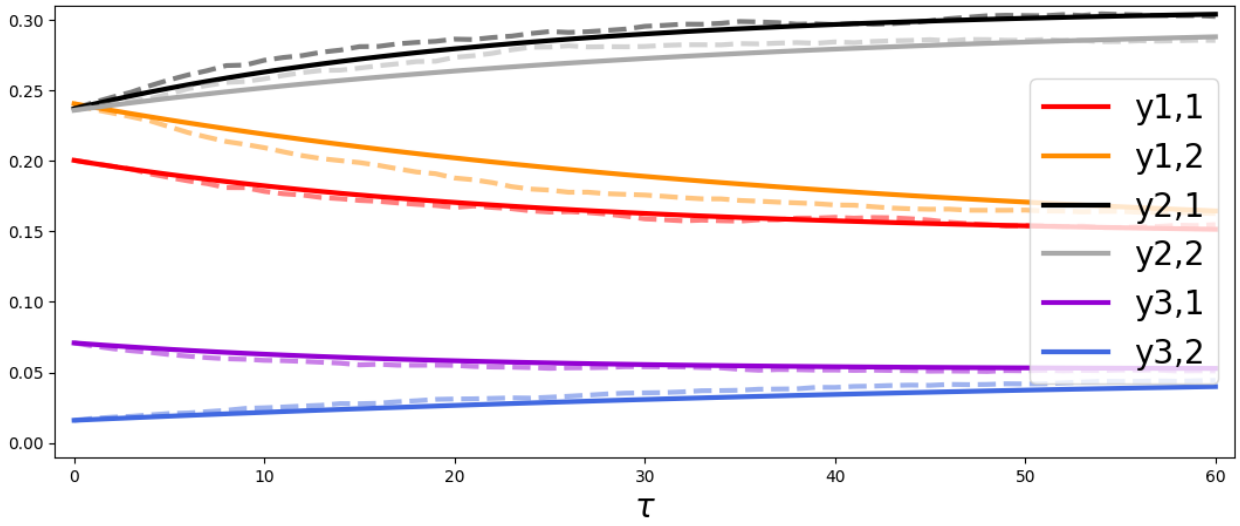


Figure D7. Solid lines mark mean-field predictions, whereas dashed curves represent the results of a simulation run with the Advanced Model

E. Auxiliary formulas and computations

E.1. Supporting computations for the derivation of the mean-field approximation

Let us estimate the probability that for some $s \in [m]$ and $a \in [m]$, the opinion shift $x_s \to x_a$ will occur at time t, subject to the influence object is characterized by type Ξ_f and the influencing message coming from an ordinary agent that has type Ξ_r ($f, r \in [M]$) and opinion x_l. This occurs if (i) an agent with opinion x_s and type Ξ_f is selected as an influence object (probability $\frac{Y_{s,f}(t)}{N}$), (ii) an agent with opinion x_l and type Ξ_r is selected as an influence source (probability $\frac{Y_{l,r}(t)-\delta_{s,l}\delta_{f,r}}{N}$), and, finally, if (iii) the desirable opinion shift happens (probability $p_{s,l,a}^{f,r}$):

$$\frac{Y_{s,f}(t)}{N}\frac{Y_{l,r}(t)-\delta_{s,l}\delta_{f,r}}{N}p_{s,l,a}^{f,r}.$$

If the influence source is a stubborn agent (with opinion x_l), then we should slightly modify the previous expression as follows:

$$\frac{Y_{s,f}(t)}{N}\frac{U_l^f(t)}{N}p_{s,l,a}^{f,M+1}.$$

In ongoing computations, it would be convenient to present quantities $p_{s,l,a}^{f,r}$ and $p_{s,l,a}^{f,M+1}$ in more complicated form:

$$p_{s,l,a}^{f,r}=\sum_{k\in[m]}p_{s,l,k}^{f,r}\,\delta_{k,a},\qquad p_{s,l,a}^{f,M+1}=\sum_{k\in[m]}p_{s,l,k}^{f,M+1}\,\delta_{k,a}.$$

Now we are ready to calculate the probability that the population of ordinary agents that have type Ξ_f and hold opinion x_a at time t will increase by one in the next time moment (we denote it by $P\big(Y_{a,f}(t)\uparrow\big)$). This occurs if a randomly selected agent (i) has type Ξ_f, (ii) holds opinion x_s different from x_a, and (iii) changes it to x_a:

$$P\big(Y_{a,f}(t)\uparrow\big)=\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}\frac{Y_{s,f}(t)}{N}\big(1-\delta_{s,a}\big)\left[\sum_{r\in[M]}\frac{Y_{l,r}(t)-\delta_{s,l}\delta_{f,r}}{N}p_{s,l,k}^{f,r}+\frac{U_l^f(t)}{N}p_{s,l,k}^{f,M+1}\right]\delta_{k,a}.$$

Analogously, we can compute the probability that the population of ordinary agents that have type Ξ_f and hold opinion x_a at time t will decrease by one in the next time moment:

$$P\big(Y_{a,f}(t)\downarrow\big)=\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}\frac{Y_{s,f}(t)}{N}\delta_{s,a}\left[\sum_{r\in[M]}\frac{Y_{l,r}(t)-\delta_{s,l}\delta_{f,r}}{N}p_{s,l,k}^{f,r}+\frac{U_l^f(t)}{N}p_{s,l,k}^{f,M+1}\right]\big(1-\delta_{k,a}\big).$$

As such, we can easily compute the mathematical expectation of the number of agents that have type Ξ_f and espouse opinion x_a at time $t+1$ given the number of such agents at time t:

$$\mathbb{E}\big[Y_{a,f}(t+1)\big]=Y_{a,f}(t)+P\big(Y_{a,f}(t)\uparrow\big)-P\big(Y_{a,f}(t)\downarrow\big).$$

E.2. First Integrals of system (4) (Main Manuscript)

From equation (4) (Main Manuscript), we can write:

$$\begin{aligned}\sum_{a\in[m]}\frac{dy_{a,f}(\tau)}{d\tau}&=\sum_{a\in[m]}\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]\big(\delta_{k,a}-\delta_{s,a}\big)\\&=\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]\sum_{a\in[m]}\big(\delta_{k,a}-\delta_{s,a}\big)\\&=\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]\sum_{a\in[m]}\delta_{k,a}\\&-\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]\sum_{a\in[m]}\delta_{s,a}\\&=\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]\\&-\sum_{s\in[m]}\sum_{l\in[m]}\sum_{k\in[m]}y_{s,f}(\tau)\left[\sum_{r\in[M]}y_{l,r}(\tau)p_{s,l,k}^{f,r}+u_l^f(\tau)p_{s,l,k}^{f,M+1}\right]=0,\end{aligned}$$

as was to be shown.

E.3. The right side of the Euler-Lagrange equations (see (17) in the Main Manuscript)

Derivatives of the Hamiltonian with respect to variables $y_{i,j}$ are given by:

$$\frac{\partial H}{\partial y_{i,j}} = \sum_{a\in[m]} \lambda_{a,j} \sum_{l\in[m]} \left[\sum_{r\in[M]} y_{l,r} p_{i,l,a}^{j,r} + u_l^j p_{i,l,a}^{j,M+1} \right] + \sum_{a\in[m]} \sum_{f\in[M]} \lambda_{a,f} \sum_s y_{s,f} p_{s,l,a}^{f,j}$$

$$- \lambda_{i,j}\left(n_{[M]} + n_{M+1}^j\right) - K v_i.$$

As such, we obtain:

$$\frac{d\lambda_{i,j}(\tau)}{d\tau} = -\frac{\partial H}{\partial y_{i,j}} =$$

$$= K v_i + \lambda_{i,j}(\tau)\left(n_{[M]} + n_{M+1}^j\right) - \sum_{a\in[m]} \lambda_{a,j}(\tau) \sum_{l\in[m]} \left[\sum_{r\in[M]} y_{l,r}(\tau) p_{i,l,a}^{j,r} + u_l^j(\tau) p_{i,l,a}^{j,M+1} \right] -$$

$$- \sum_{a\in[m]} \sum_{f\in[M]} \lambda_{a,f}(\tau) \sum_s y_{s,f}(\tau) p_{s,l,a}^{f,j}.$$

E.4. Singular control

Since

$$\frac{\partial H}{\partial u_i^j} = \sum_{a\in[m]} \lambda_{a,j} \left[\sum_{s\in[m]} y_{s,j} p_{s,i,a}^{j,M+1} - y_{a,j} \right],$$

we can write down the following expression:

$$\frac{d}{d\tau}\frac{\partial H}{\partial u_i^j} = \sum_{a\in[m]} \frac{d\lambda_{a,j}(\tau)}{d\tau} \left[\sum_{s\in[m]} y_{s,j}(\tau) p_{s,i,a}^{j,M+1} - y_{a,j}(\tau) \right] +$$

$$+ \sum_{a\in[m]} \lambda_{a,j}(\tau) \left[\sum_{s\in[m]} \frac{dy_{s,j}(\tau)}{d\tau} p_{s,i,a}^{j,M+1} - \frac{dy_{a,j}(\tau)}{d\tau} \right].$$

Now let us make use of equations (5) and (17) (see the Main Manuscript; detailed presentation of equation (17) is given in Appendix B of the Main Manuscript):

$$\frac{d}{d\tau}\frac{\partial H}{\partial u_i^j} = \sum_{a\in[m]}\Bigg\{ Kv_a + \lambda_{a,j}(\tau)\left(n_{[M]} + n_{M+1}^j(\tau)\right)$$

$$- \sum_{k\in[m]} \lambda_{k,j}(\tau) \sum_{l\in[m]} \left[\sum_{r\in[M]} y_{l,r}(\tau) p_{a,l,k}^{j,r} + u_l^j(\tau) p_{a,l,k}^{j,M+1} \right]$$

$$- \sum_{k\in[m]} \sum_{f\in[M]} \lambda_{k,f}(\tau) \sum_{s} y_{s,f}(\tau) p_{s,l,k}^{f,j} \Bigg\} * \left[\sum_{s\in[m]} y_{s,j}(\tau) p_{s,i,a}^{j,M+1} - y_{a,j}(\tau) \right]$$

$$+ \sum_{a\in[m]} \lambda_{a,j}(\tau) \Bigg[\sum_{q\in[m]} \Bigg\{ \sum_{s\in[m]} \sum_{l\in[m]} y_{s,j}(\tau) \left[\sum_{r\in[M]} y_{l,r}(\tau) p_{s,l,q}^{j,r} + u_l^j(\tau) p_{s,l,q}^{j,M+1} \right]$$

$$- y_{q,j}(\tau)\left(n_{[M]} + n_{M+1}^j(\tau)\right) \Bigg\} p_{q,i,a}^{j,M+1}$$

$$- \sum_{s\in[m]} \sum_{l\in[m]} y_{s,j}(\tau) \left[\sum_{r\in[M]} y_{l,r}(\tau) p_{s,l,a}^{j,r} + u_l^j(\tau) p_{s,l,a}^{j,M+1} \right]$$

$$+ y_{a,j}(\tau)\left(n_{[M]} + n_{M+1}^j(\tau)\right) \Bigg].$$

The above equation can be reorganized as follows:

$$\frac{d}{d\tau}\frac{\partial H}{\partial u_i^j} = \sum_{l\in[m]} c_l^{i,j}(\tau) u_l^j(\tau) + c^{i,j}(\tau), \qquad (E.4.1)$$

where

$$c_l^{i,j}(\tau) = \sum_{a\in[m]} \sum_{q\in[m]} \sum_{s\in[m]} \frac{\left[\sum_{s\in[m]} y_{s,j}(\tau) p_{s,i,a}^{j,M+1} - y_{a,j}(\tau)\right]}{m} \left(\frac{\lambda_{a,j}(\tau)}{m} - \lambda_{q,j}(\tau) p_{a,l,q}^{j,M+1} \right)$$

$$+ \lambda_{a,j}(\tau) \left(y_{s,j}(\tau) p_{s,l,q}^{j,M+1} p_{q,i,a}^{j,M+1} - \frac{y_{s,j}(\tau) p_{s,l,a}^{j,M+1}}{m} - \frac{y_{q,j}(\tau) p_{q,i,a}^{j,M+1}}{m} + \frac{y_{a,j}(\tau)}{m^2} \right)$$

and

$$c^{i,j}(\tau) = \sum_{a\in[m]} \Bigg\{ K v_a + \lambda_{a,j}(\tau) n_{[M]} - \sum_{k\in[m]} \lambda_{k,j}(\tau) \sum_{l\in[M]} \sum_{r\in[M]} y_{l,r}(\tau) p_{a,l,k}^{j,r}$$

$$- \sum_{k\in[m]} \sum_{f\in[M]} \lambda_{k,f}(\tau) \sum_{s} y_{s,f}(\tau) p_{s,l,k}^{f,j} \Bigg\} * \left[\sum_{s\in[m]} y_{s,j}(\tau) p_{s,i,a}^{j,M+1} - y_{a,j}(\tau) \right]$$

$$+ \sum_{a\in[m]} \lambda_{a,j}(\tau) \left[\sum_{q\in[m]} \left\{ \sum_{s\in[m]} \sum_{l\in[m]} y_{s,j}(\tau) \sum_{r\in[M]} y_{l,r}(\tau) p_{s,l,q}^{j,r} - y_{q,j}(\tau) n_{[M]} \right\} p_{q,i,a}^{j,M+1} \right.$$

$$\left. - \sum_{s\in[m]} \sum_{l\in[m]} y_{s,j}(\tau) \sum_{r\in[M]} y_{l,r}(\tau) p_{s,l,a}^{j,r} + y_{a,j}(\tau) n_{[M]} \right].$$

Let us now assume that a pair $y^*(\tau), u^*(\tau)$ is optimal and that $\lambda^*(\tau)$ is the corresponding adjoint function ensured by the Pontryagin Maximum Principle. Let the index pairs $(i_1, j_1), \dots, (i_d, j_d)$ satisfy:

$$(i_1, j_1) = \underset{(i,j)}{\operatorname{argmax}} \frac{\partial H}{\partial u_i^j}, \dots, (i_d, j_d) = \underset{(i,j)}{\operatorname{argmax}} \frac{\partial H}{\partial u_i^j}$$

on some $I \subseteq [\tau_0, \tau_1]$ (note that it could be the case that these pairs have intersections in the first or second indices, but not both). In this case, $(u^*)_i^j(\tau) = 0$ on I if $(i,j) \notin \{(i_1, j_1), \dots, (i_d, j_d)\}$. The other components of the control function are given by:

$$\frac{d}{d\tau} \frac{\partial H}{\partial u_{i_1}^{j_1}} = \frac{d}{d\tau} \frac{\partial H}{\partial u_{i_2}^{j_2}},$$

$$\dots,$$

$$\frac{d}{d\tau} \frac{\partial H}{\partial u_{i_{d-1}}^{j_{d-1}}} = \frac{d}{d\tau} \frac{\partial H}{\partial u_{i_d}^{j_d}},$$

$$(u^*)_{i_1}^{j_1}(\tau) + \dots + (u^*)_{i_d}^{j_d}(\tau) = n_{M+1}.$$

These equations form a linear system with respect to $(u^*)_{i_1}^{j_1}(\tau), \dots, (u^*)_{i_d}^{j_d}(\tau)$. To solve it, one should make use of expression (E.4.1).

F. Technical details of the Direct and FBS methods

As we pointed out in Section C, both the Direct and FBS methods need to solve the initial value problem (6), (7) (Main Manuscript). The function *RKM_Dyn_System* from the module *InitProblem_library* is designed to perform this task. The detailed information about organization and performance of this function can be found in Sections C and D. Now we will discuss other technical details of the Direct and FBS methods.

F.1. Direct method

The Direct method solves a constraint optimization problem where the objective function (22) (Main Manuscript) is minimized. The Python function *DM* from the module *DM_library* realizes the Direct method. The optimization problem is captured by the Python function *minimize* from the subpackage *scipy.optimize*[9].

F.2. FBS method

The FBS method is realized in the module *FBSM_library*, the function *FBSM*.

To solve the Euler-Lagrange system (equation (17) in the Main Manuscript), the FBS method makes use of the Runge-Kutta 4th order method realized as an independent function *RKM_Euler_Lagrange_System* in the module *FBSM_library*.

On **Step 5**, the FBS method checks if the new approximation of the control function differs from the previous one. If the difference in L^2 norm exceeds the value of 0.005, then the two estimations are classified as dissimilar. Otherwise, they are supposed to be the same.

The algorithm works for no longer than 300 iterations. There are three possible convergence statuses: (i) convergence, (ii) cycle (of some length $k \geq 2$—convergence can be understood as a cycle of length 1 on this occasion), and (iii) no convergence. In the second and third cases, the best control function approximation (in terms of the objective functional) is taken as an output. In the first case, the last approximation is the output.

F.3. Drawbacks

Let us discuss the possible drawbacks of the Direct and FBS methods.

First of all, both of these methods rely on numerical algorithms to achieve the general purpose. Whereas the Direct method needs only one auxiliary numerical solver that solves the initial value problem (6), (7) (Main Manuscript), each FBS method iteration solves two auxiliary Cauchy problems (on **Steps 2** and **3**). Each of these problems may be solved with some errors

[9] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html

(especially when integrating long time spans), and when it comes to maximizing the Hamiltonian (on **Step 4**), the resulting error may be quite huge. On the one hand, we managed to fix errors in solutions of the Cauchy problem (6), (7) (Main Manuscript)—see Section C. On the other hand, solutions of the Cauchy problem (17) are beyond our control.

Next, the FBS method builds upon the maximum principle, which is a necessary condition, not a sufficient one. In this regard, even if a pair $u(\tau), y(\tau)$ satisfies (17), (18) (Main Manuscript), it does not necessarily follow that this pair is optimal.

G. Experiments

This section presents the results of our experiments on searching for the optimal control numerically. We tested the Direct and FBS methods in various settings by varying the number of ordinary agent types, the size of the grid, the structure of social influence (outlined by transition matrices), and the structure of the objective functional (parametrized by the relative importance of the terminal term K). In these experiments, we usually considered the situation where the fraction of stubborn agents is $n_{M+1} = 0.4$ as a baseline configuration. This assumption is motivated by the corresponding empirical observation (see Ref. (González-Bailón & De Domenico, 2021). Nonetheless, other cases were also investigated. Further, by default, we assume that the Person's purpose is to push the social system towards the right endpoint of the opinion spectrum. On this occasion, in the case of a 3-element opinion space, we used the opinion-weight vector $v = [2 \quad 1 \quad 0]^T$. For a 5-element opinion space, we applied the opinion-weight vector $v = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T$. The motivation behind this choice can be found in the Main Manuscript (see Remark 3).

Notations that will be used further in this section: $\times$ – the control found is not optimal, $\rightarrow$ – the FBS-method has converged, $\circ$ – the FBS-method is looped (the length of the cycle is indicated in parentheses), $\uparrow$ – the FBS-method has not converge (in order to converge, the FBS method have 300 iterations).

G.1. Assimilative influence

This subsection concerns the optimal control problem when opinion dynamics are governed by the assimilative social influence mechanism (see Subsection B.1). It is straightforward to note that in this case, if the Person focuses on pushing the system towards, say, the right endpoint of the opinion spectrum, then they should make stubborn agents disseminating the most radical right-side opinion x_m. In the case where there is only one type of ordinary agents in the system ($M = 1$), the optimal control should be:

$$u(\tau) \equiv [0 \quad \dots \quad 0 \quad n_2]^T. \qquad (G.1.1)$$

This observation provides us with a perfect opportunity to check the efficacy of our numerical algorithms. Figure G1 tests performance of the Direct and FBS methods on different grids and investigates how the convergence of the FBS method depends on the grid size after controlling for the relative importance of the integral term in the objective functional. Figure G2 discovers the effect of the number of stubborn agents (that is, the Person's control facilities) on convergence properties of the FBS algorithm across different transition matrices.

Experiments revealed that in most situations, both the algorithms are insensitive to the initial guess and the starting point of the system y_0. The main observation is that larger grids impair the performance of the Direct and FBS methods. However, the Direct method displays more sensitivity towards the grid size and becomes too slow on huge grids (it needs ~30 min to derive an output on a grid with 100 points in the case $m = 3, M = 1$). The FBS method is more reliable from this perspective. We obtained that if it converges, then it always reaches the optimal control (G.1.1). Even if the FBS method does not converge, it tends to fluctuate around the optimal solution (see Figures G1, G2). Further, we obtained that the performance of the FBS method can be improved if increasing the value of K, other things being equal. An interesting observation is that if setting $K = 0$, then on sufficiently large grids, the FBS method fails to converge but still yields controls that, despite being different from (G.1.1), zero out the objective functional. This happens because the system is well controlled, and thus, subject to the integral term does not contribute to the objective functional, there are more than one control functions that can bring the system to the desirable point in finite time. Next, we report that the FBS method works better if there are more stubborn agents in the system (that is, more control facilities). Furthermore, the performance of this method and its convergence are affected by the transition matrix configuration (see Figure G2, the second column of the table).

Grid	Direct method	FBS-method
$\tau_0 = 0, \tau_1 = 5, h = 1$	1.379 (0.1 s)	1.379 (0.0 s) →
$\tau_0 = 0, \tau_1 = 20, h = 1$	1.517 (1.9 s)	1.517 (0.1 s) →
$\tau_0 = 0, \tau_1 = 40, h = 1$	3.316 (1.2 min) ×	1.517 (0.3 s) →
$\tau_0 = 0, \tau_1 = 50, h = 1$	6.421 (1.7 min) ×	1.517 (0.2 s) →

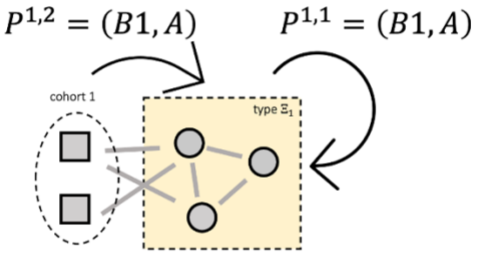
$K = 1$

Transition table (B1, A)

1	0	0
0	1	0
0	0	1

1	0	0
0	1	0
0	0	1

1	0	0
0	1	0
0	0	1



FBS-method

Grid	$K = 0$	$K = 1$	$K = 100$
$\tau_0 = 0, \tau_1 = 30, h = 1$	→	→	→
$\tau_0 = 0, \tau_1 = 40, h = 1$	∘ (10) ×✓	→	→
$\tau_0 = 0, \tau_1 = 50, h = 1$	↑ ×✓	→	→
$\tau_0 = 0, \tau_1 = 65, h = 1$	↑ ×✓	∘ (9)	→
$\tau_0 = 0, \tau_1 = 70, h = 1$	↑ ×✓	↑ ×	∘ (126)
$\tau_0 = 0, \tau_1 = 80, h = 1$	↑ ×✓	↑ ×	↑ ×

Figure G1. (Upper table) We compare the performance (the value of the objective functional and working time) of the Direct and FBS methods for different grids with the fixed step $h = 1$. The underlying social system is characterized by parameters $m = 3, M = 1, y_0 = [0.1 \quad 0.3 \quad 0.2]^T$ and by transition matrix (B1, A) (that governs both in-group and inter-group interactions: $\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B1, A) – see the middle part of the figure). The opinion-weight vector is $v = [2 \quad 1 \quad 0]^T$, $K = 1$. In all experiments presented in this figure, the FBS method successfully and swiftly converged to the same (optimal) output. In turn, the Direct method features relatively slow convergence on large grids, with noticeable deviations from the optimal control (G.1.1). (Lower Table) We analyze how the convergence of the FBS method changes with the size of the grid and the value of K. The underlying social system is characterized by parameters $m = 3, M = 1, y_0 = [0.1 \quad 0.3 \quad 0.2]^T$ and by transition matrix (B1, A) (that governs both in-group and inter-group interactions: $\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B1, A) – see the middle part of the figure). The opinion-weight vector is $v = [2 \quad 1 \quad 0]^T$.

Transition table	The fraction of stubborn agents		
	$n_{M+1} = 0.1$	$n_{M+1} = 0.4$	$n_{M+1} = 0.7$
$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$	↑ ×	∘ (9) ×	→
$\begin{bmatrix}1&0&0\\0.9&0.1&0\\0.9&0&0.1\end{bmatrix}$ $\begin{bmatrix}0.1&0.9&0\\0&1&0\\0&0.9&0.1\end{bmatrix}$ $\begin{bmatrix}0.1&0&0.9\\0&0.1&0.9\\0&0&1\end{bmatrix}$	↑ ×	∘ (3)	→
$\begin{bmatrix}0.99&0.01&0\\0.9&0.1&0\\0.8&0.15&0.05\end{bmatrix}$ $\begin{bmatrix}0.2&0.8&0\\0.01&0.98&0.01\\0&0.8&0.2\end{bmatrix}$ $\begin{bmatrix}0.05&0.15&0.8\\0&0.1&0.9\\0&0.01&0.99\end{bmatrix}$	↑ ×	∘ (4) ×	→

Figure G3. This figure investigates how the convergence of the FBS method depends on the transition matrix configuration (we assume that $\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$) and the fraction of stubborn agents in the system. The underlying social system is characterized by parameters $m = 3, M = 1, y_0 = [n_1 \quad 0 \quad 0]^T$ (all ordinary agents are initially located on the left endpoint of the opinion spectrum). The opinion-weight vector is $v = [2 \quad 1 \quad 0]^T, K = 1$.

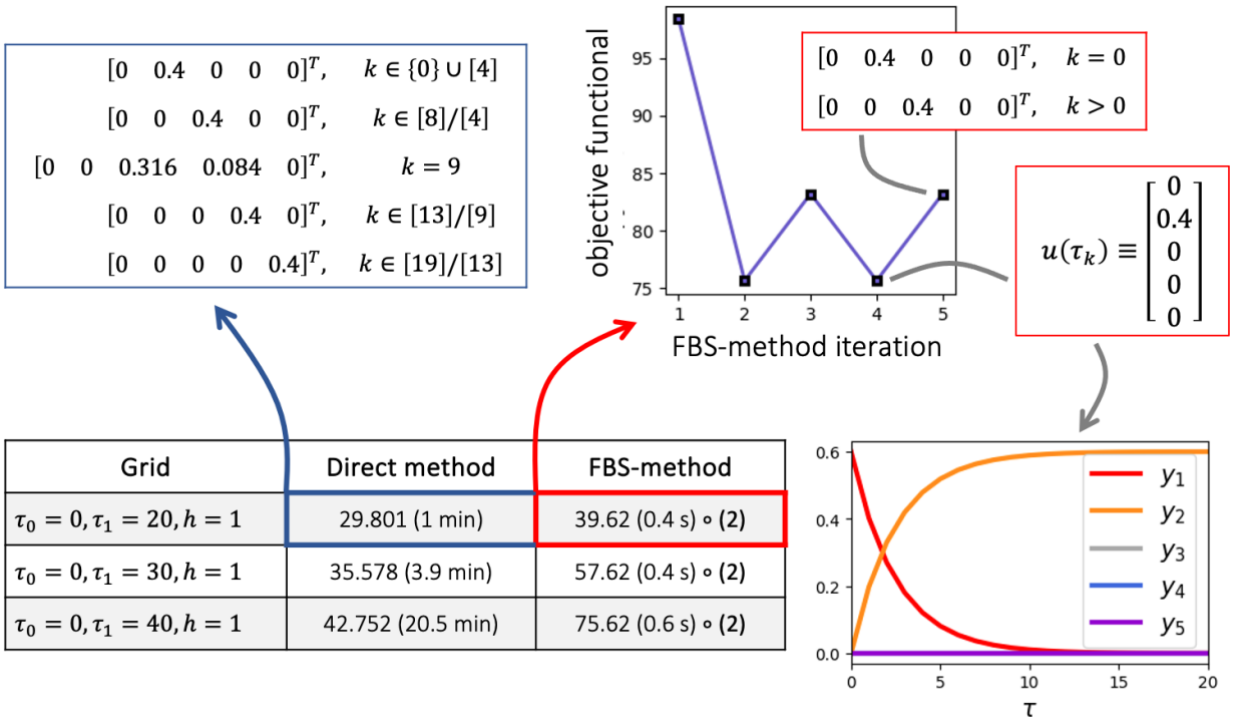
G.2. Bounded confidence

This social influence mechanism is more nuanced from the perspective of solving the control problem. Within the BC mechanism (see Subsection B.2), influencing the system with a radical

opinion without accounting for the current agents' opinions is likely not the best strategy: perhaps, the Person should find some middle ground by toning down their arguments. Consider the following stylized situation. Let assume that the system is located mostly on the left endpoint of the opinions spectrum and that the Person strives to push agents' opinions towards the right endpoint. In this case, the optimal control strategy is quite obvious (at least, from a qualitative perspective). The Person should apply a *nudging* control whereby stubborn agents start with disseminating the closest to x_1 (and thus likely the most profitable) opinion x_2 and then gradually slide their position towards the right (using opinion x_3 instead of x_2 from some point on, then opinion x_4 and so on) as the system moves in the same direction.

From this perspective, the situation when the system is initially located on the left edge of the opinion spectrum can serve as a way to verify the reliability of our numerical algorithms (again, at least from a qualitative perspective).

Experiments with transition matrix (B3, A) revealed that the FBS method falls into a zone of nonoptimal controls (see Figure G4). In contrast, the Direct method gives much more optimal solutions, but, again, it has the disadvantage of being slow on large grids. However, if using the transition matrix (B4) (which is, actually, much closer to empirics, as was reported by Kurahashi-Nakamura et al. (2016), then we see that the FBS method works better in terms of the objective functional and, especially, working time than the Direct one on large grids, whereas their performances on smaller grids are comparable (see Figure G5). Figure G6 also demonstrates that in the case of the BC mechanism, the outputs of the numerical algorithms substantially depend on the initial point of the system.



Grid	Direct method	FBS-method
$\tau_0 = 0, \tau_1 = 20, h = 1$	29.801 (1 min)	39.62 (0.4 s) ∘ **(2)**
$\tau_0 = 0, \tau_1 = 30, h = 1$	35.578 (3.9 min)	57.62 (0.4 s) ∘ **(2)**
$\tau_0 = 0, \tau_1 = 40, h = 1$	42.752 (20.5 min)	75.62 (0.6 s) ∘ **(2)**

Figure G4. We compare the performance (the value of the objective functional and convergence time) of the Direct and FBS methods using the transition matrix (B3, A) for description of influence processes in the system ($\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B3, A)). Ordinary agents' opinions are initially located on the left endpoint of the opinion spectrum ($y_0 = [0.6 \quad 0 \quad 0 \quad 0 \quad 0]^T, m = 5, M = 1$). The opinion-weight vector is $v = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T, K = 1$. The control function derived by the FBS method (highlighted in red) brings up the system to the non-optimal state $y = [0 \quad 0.6 \quad 0 \quad 0 \quad 0]^T$. In turn, the solutions of the Direct method ensure much lower values of the objective functional and generally meet our expectations regarding the structure of optimal control (highlighted by the blue frame).

Grid	Direct method	FBS-method
$\tau_0 = 0, \tau_1 = 10, h = 1$	24.969 (13.1 s)	24.969 (0.1 s) →
$\tau_0 = 0, \tau_1 = 20, h = 1$	45.184 (1.3 min)	45.185 (0.3 s) ∘ **(2)**
$\tau_0 = 0, \tau_1 = 40, h = 1$	81.694 (19.1 min)	79.316 (0.3 s) →

$K = 1$

$[0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T, \quad k = 0$
$[0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T, \quad k \in [13]$
$[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T, \quad k \in [32]/[13]$
$[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T, \quad k \in [39]/[32]$

$[0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T, \quad k = 0$
$[0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T, \quad k \in [15]$
$[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T, \quad k \in [33]/[16]$
$[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T, \quad k \in [39]/[33]$

Grid	Direct method	FBS-method
$\tau_0 = 0, \tau_1 = 10, h = 1$	2.146 (21.9 s)	2.146 (0.1 s) ∘ **(2)**
$\tau_0 = 0, \tau_1 = 20, h = 1$	1.922 (1.7 min)	1.922 (0.3 s) →
$\tau_0 = 0, \tau_1 = 40, h = 1$	1.637 (15.9 min)	1.518 (0.3 s) →

$K = 0$

Figure G5. We compare the performance (the value of the objective functional and convergence time) of the Direct and FBS methods for $K = 0$ and $K = 1$ using the transition matrix (B4) for description of influence processes in the system ($\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B4)). Ordinary agents' opinions are initially located on the left endpoint of the opinion spectrum ($y_0 = [0.6 \quad 0 \quad 0 \quad 0 \quad 0]^T, m = 5, M = 1$). The opinion-weight vector is $v = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T$. We notice that the output of the FBS method changes with K (highlighted by the red and black frames). As an additional check, we sequentially tested each of these two control functions by considering settings $K = 0$ and $K = 1$. We obtained that each of two control functions outperformed the other under the conditions from which it was derived.

Initial point of the system	$K = 0$	$K = 1$
$[0.6 \quad 0 \quad 0 \quad 0 \quad 0]^T$	$[0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T$, $k = 0$ $[0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T$, $k \in [15]$ $[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$, $k \in [33]/[16]$ $[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$, $k \in [39]/[33]$ →	$[0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T$, $k = 0$ $[0 \quad 0 \quad 0.4 \quad 0 \quad 0]^T$, $k \in [13]$ $[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$, $k \in [32]/[13]$ $[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$, $k \in [39]/[32]$ →
$[0.3 \quad 0.2 \quad 0.15 \quad 0.05 \quad 0]^T$	$[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$ $k \in [23]$ $[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$, $k \in [39]/[23]$ →	$[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$ $k \in [19]$ $[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$, $k \in [39]/[19]$ ∘ (2)
$[0.05 \quad 0.1 \quad 0.3 \quad 0.1 \quad 0.05]^T$	$[0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$ $k = 0$ $[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$, $k \in [39]$ →	$[0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$ →

Figure G6. We study how the output of the FBS method varies with y_0 and K on the grid $\tau_0 = 0, \tau_T = 40, h = 1$. The structure of influence is given by $\mathcal{P}^{1,1} = \mathcal{P}^{1,2}$ = (B4). The opinion-weight vector is $v = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T$. If the case the system is located on the left endpoint of the opinion spectrum, then the algorithm recommends using nudging-type strategies. Further, the output depends on the structure of the objective functional: if the integral term matters for the Person, then switches between control regimes occur earlier.

G.3. Systems with several types of ordinary agents

Let us now consider a system in which two types of ordinary agents coexist ($M = 2$). We will assume that ordinary agents perceive in-type influence in accordance with the transition matrix (B1, A): $\mathcal{P}^{1,1} = \mathcal{P}^{2,2}$ = (B1, A) (that is, with a low level of resistance—such a tendency is broadly observed in empirics (Bail et al., 2018)). Next, we will assume that first-type agents are susceptible to messages of second-type agents, whereas the latter ones are substantially more resilient to influence coming from agents of the first type: $\mathcal{P}^{1,2}$ = (B1, A), $\mathcal{P}^{2,1}$ = (B1, B). And finally, all ordinary agents, regardless of their type, are open-minded to control influence: $\mathcal{P}^{1,3} = \mathcal{P}^{2,3}$ = (B1, A). This influence structure is depicted in Figure G7.

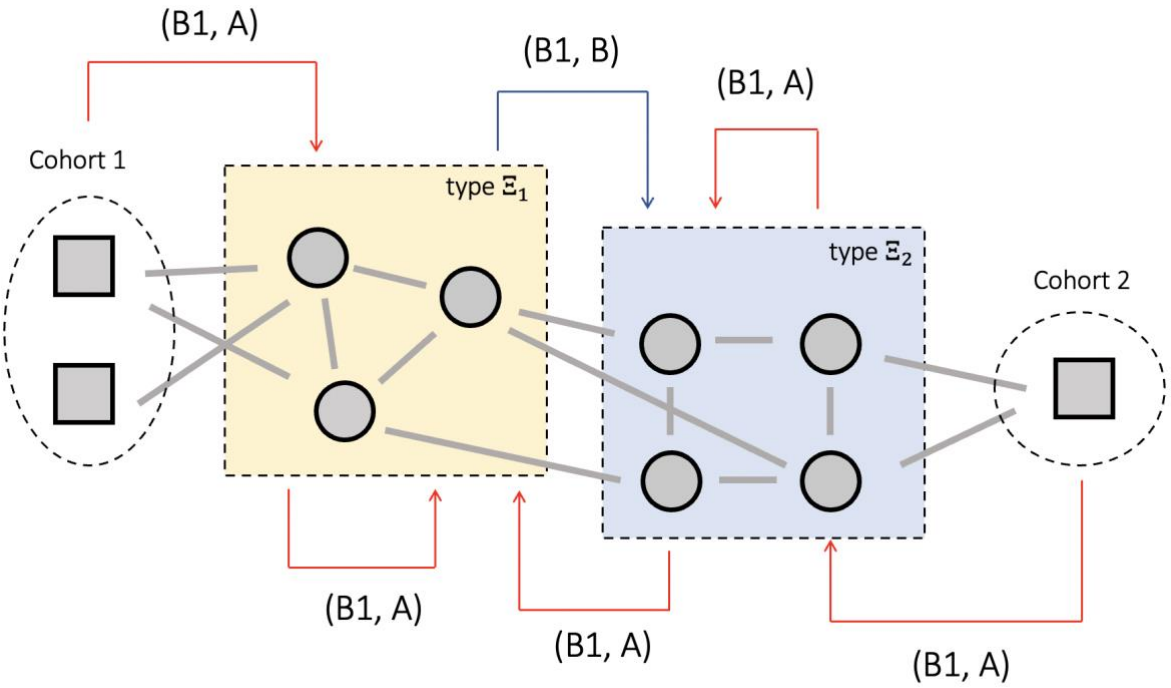


Figure G7. This scheme defines a social system in which two types of ordinary agents ($M = 2$) communicate with each other and are controlled by two cohorts of stubborn agents at the same time. The structure of social influence is demonstrated via arrows. Each arrow outlines how one group affects the other. For example, the arrow directed from the Cohort 2 to type Ξ_2 indicates that $\mathcal{P}^{2,3}$ = (B1, A).

The optimal control in this case is intuitively clear: the Person should primarily focus on influencing agents that have the second type, as they are open-minded to controlling influence and, at the same time, can effectively impact on first-type agents. Of course, it also depends on how types and opinions vary across agents. If, for example, n_1 is substantially greater than n_2 (the number of first-type agents dominates that of second-type ones), then the Person should pay more attention to first-type agents. These conjectures are consistent with the solutions of the FBS method. However, Figure G8 indicates that the FBS method does not necessarily yield the optimal control—the Directed method in some cases demonstrates better performance rates.

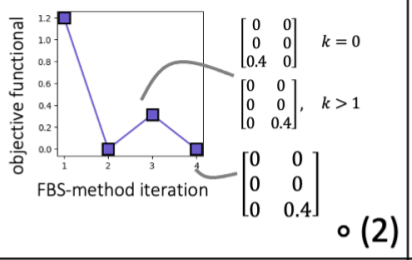
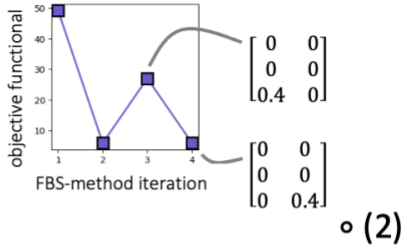
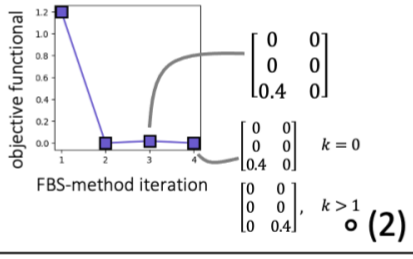
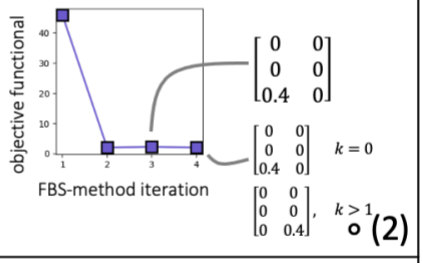
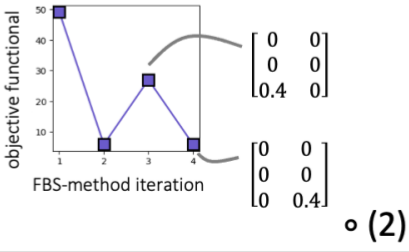
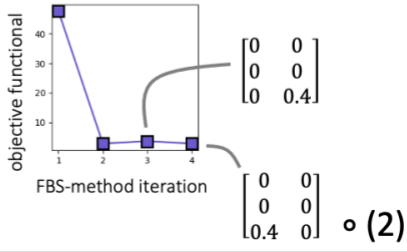
Initial point of the system	$K = 0$	$K = 1$
$\begin{bmatrix} 0.3 & 0.3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$		
$\begin{bmatrix} 0.3 & 0 \\ 0 & 0 \\ 0 & 0.3 \end{bmatrix}$		
$\begin{bmatrix} 0.5 & 0.1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$		

Figure G8. We study how the output of the FBS method depends on the initial point of the system and parameter K ($m = 3, M = 2, v = [2 \quad 1 \quad 0]^T$, the structure of social influence is presented in Figure G7). The underlying grid is defined as $\tau_0 = 0, \tau_T = 40, h = 1$.

K, initial state of the system	Direct method	FBS-method
$K = 0, y_0 = \begin{bmatrix} 0.3 & 0.3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.011 (8 min)	0 (0.5 c) ∘ **(2)**
$K = 0, y_0 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0 \\ 0 & 0.3 \end{bmatrix}$	0.013 (4.1 min)	0 (0.6 c) ∘ **(2)**
$K = 0, y_0 = \begin{bmatrix} 0.5 & 0.1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.036 (5.9 min)	0.007 (0.5 c) ∘ **(2)**
$K = 1, y_0 = \begin{bmatrix} 0.3 & 0.3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	6.847 (5.1 min)	5.92 (0.5 c) ∘ **(2)**
$K = 1, y_0 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0 \\ 0 & 0.3 \end{bmatrix}$	1.548 (5.6 min)	1.979 (0.5 c) ∘ **(2)**
$K = 1, y_0 = \begin{bmatrix} 0.5 & 0.1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	7.487 (9.7 min)	11.272 (0.4 c) ∘ **(2)**

Figure G8. Here, we compare the performance of the Direct and FBS method across different initial points and values of K ($m = 3, M = 2, v = [2 \quad 1 \quad 0]^T$, the structure of social influence is presented in Figure G7). The underlying grid is defined as $\tau_0 = 0, \tau_T = 40, h = 1$.

Consider a different situation. Similar to previous example, there are two types of ordinary agents ($M = 2$), but now we assume that they follow different influence mechanisms: agents of the first type communicate via the transition matrix (B2) (assimilative influence), whereas second-type agents' opinions are moderated by the transition matrix (B4) (BC influence).

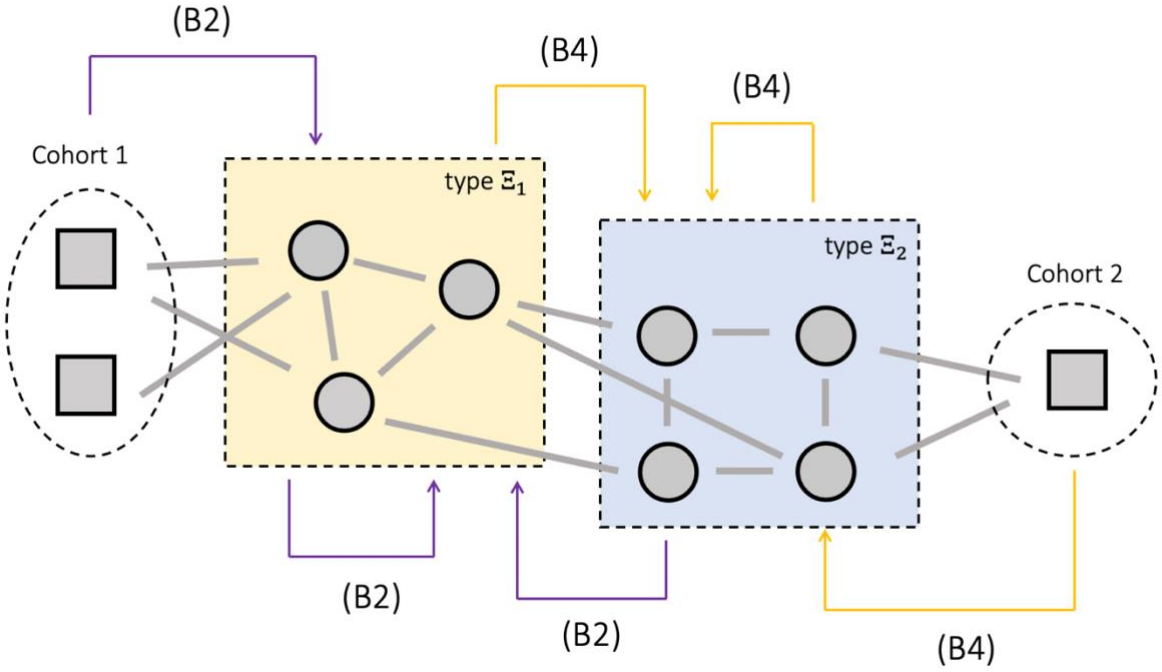


Figure G9. This scheme defines a social system in which two types of ordinary agents ($M = 2$) communicate with each other and are controlled by two cohorts of stubborn agents. The structure of social influence is demonstrated via arrows. Each arrow outlines how one group affects the other. For example, the arrow directed from the Cohort 2 to type Ξ_2 indicates that $\mathcal{P}^{2,3}$ = (B4). The graph notations are borrowed from Figure 2 (Main Manuscript).

In this case, the optimal control strategy substantially depends on the distribution of types among ordinary agents: if first-type agents constitute the majority of ordinary agents, then the Person should apply the following control (subject to the Person wanting to push the system towards the right):

$$u(\tau) \equiv \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ n_3 & 0 \end{bmatrix}. \qquad (G.3.1)$$

Inversely, if n_2 is significantly greater than n_1, then the Person should focus their influence facilities on second-type agents. However, since these agents update their opinions via the BC mechanism, if their opinions are located on, say, the left endpoint of the opinion spectrum and the Person strives to shift them towards the right endpoint, then the following sequence of controls can be a suitable answer:

$$\begin{bmatrix} 0 & 0 \\ 0 & n_3 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & n_3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & n_3 \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & n_3 \end{bmatrix}.$$

These hypotheses are confirmed by experiments with the FBS method (which successfully converges on this occasion). For example, for the initial point

$$y_0 = \begin{bmatrix} 0.5 & 0.1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

the recommended control is exactly (G.3.1). In contrast, if the system is initially located in the point

$$y_0 = \begin{bmatrix} 0.1 & 0.5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

then the FBS method suggest to apply the following control ($\tau_0 = 0, \tau_T = 40, h = 1, \vec{v} = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T, K = 1$):

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0.4 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, 0 \le k \le 12 \rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.4 \\ 0 & 0 \end{bmatrix}, 13 \le k \le 33 \rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.4 \end{bmatrix}, 34 \le k \le 39.$$

Unfortunately, in such settings, the Direct method has a hard time yielding an output due to the high dimensionality of the control space (given by 10 variables).

G.4. Systems with empirically calibrated transition matrices

For the opinion dynamics governed by the transition matrix (B7), compared to previous examples where optimal control strategies were intuitively clear, it is not a trivial task to predict what control the Person should use to achieve their purpose. In this regard, all one can do is rely on the numerical methods and, perhaps, compare their outputs against some benchmark controls.

Let consider a situation where there is only one type of ordinary agents ($M = 1$) and $\mathcal{P}^{1,2}$ = $\mathcal{P}^{1,2}$ = (B6) ($m = 5$). We report that in these settings, the FBS method reveals the influence asymmetry mentioned in Subsection B.3: depending on what direction the Person attempts to push the system, the algorithm obtains sufficiently different controls. For example, on the grid $\tau_0 = 0, \tau_T = 40, h = 1$, for the opinion-weight vector $v = [4 \quad 3 \quad 2 \quad 1 \quad 0]^T$ (the Person attempts to push the agents' opinions towards the right) and $K = 1$, subject to the system is located on the left side of opinion spectrum, the control obtained by the FBS method is:

$$u(\tau_k) = \begin{cases} [0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T, & k \le 29, \\ [0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T, & x \ge 30. \end{cases}$$

In other words, a nudging control should be applied. It is worth mentioning that this control outperforms benchmark controls $u(\tau) \equiv [0 \quad 0 \quad 0 \quad 0.4 \quad 0]^T$ (74.703 vs. 74.705) and $u(\tau) \equiv [0 \quad 0 \quad 0 \quad 0 \quad 0.4]^T$ (74.703 vs. 74.919) in terms of the objective functional.

In contrast, if the initial point of the system is $y_0 = [0 \quad 0 \quad 0 \quad 0 \quad 0.6]^T$ and the opinion-weight vector is $v = [0 \quad 1 \quad 2 \quad 3 \quad 4]^T$, then the FBS method derives:

$$u(\tau_k) = \begin{cases} [0.4 \quad 0 \quad 0 \quad 0 \quad 0]^T, & k \leq 17, \\ [0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T, & k \geq 18. \end{cases}$$

Thus said, the radical opinion x_1 should be employed first, but after that it should be toned down to x_2. This control is somewhat opposite to nudging. We report that it is more effective than monotonic controls $u(\tau) \equiv [0.4 \quad 0 \quad 0 \quad 0 \quad 0]^T$ (67.654 vs. 67.753) and $u(\tau) \equiv [0 \quad 0.4 \quad 0 \quad 0 \quad 0]^T$ (67.654 vs. 68.062). To understand why this counterintuitive control function appears to be so effective, let us take a closer look at the first slice of the transition matrix (B6) (see Figure B6, the first table). For the sake of clarity, we replicate it in Figure G10. We see that the probability that opinion x_1 will remain after an interaction is higher if influence comes not from opinion x_1, but, no matter how strange it may sound, from opinion x_2. In other words, to ensure that agents with opinion x_1 will keep it safe, the Person should influence them with opinion x_2—this is precisely what the FBS method recommends.

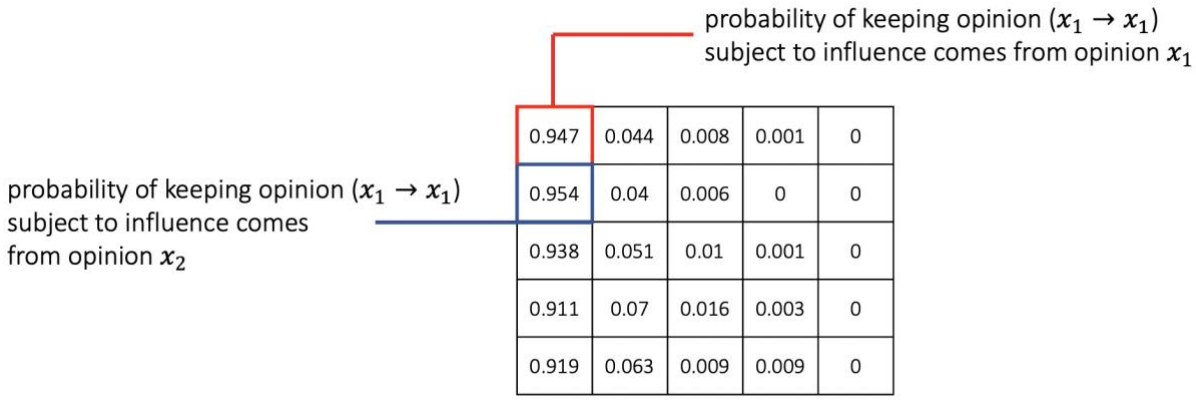


Figure G10. The first slice of the transition matrix (B6). The red frame highlights component $p_{1,1,1}$, the blue frame indicates component $p_{1,2,1}$.

We also report that the Direct method yields less effective controls in these two situations: in the first case, it obtains a control that makes the objective functional equal 77.946 (vs. 74.703 provided by the FBS method); in the second case, the objective functional is 68.663 (vs. 67.654 provided by the FBS method).

Literature

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. *Proceedings of the National Academy of Sciences*, *118*(52). https://doi.org/10.1073/pnas.2112552118

Carpentras, D., Maher, P. J., O'Reilly, C., & Quayle, M. (2022). Deriving an Opinion Dynamics Model from Experimental Data. *Journal of Artificial Societies and Social Simulation*, *25*(4), 4.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121.

González-Bailón, S., & De Domenico, M. (2021). Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences*, *118*(11).

Kozitsin, I. V. (2022a). A general framework to link theory and empirics in opinion formation models. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-09468-3

Kozitsin, I. V. (2022b). Formal models of opinion formation and their application to real data: Evidence from online social networks. *The Journal of Mathematical Sociology*, *46*(2), 120–147. https://doi.org/10.1080/0022250X.2020.1835894

Kozitsin, I. V. (2023). Opinion dynamics of online social network users: A micro-level analysis. *The Journal of Mathematical Sociology*, *47*(1), 1–41. https://doi.org/10.1080/0022250X.2021.1956917

Kozitsin, I. V., Gubanov, A. V., Sayfulin, E. R., & Goiko, V. L. (2023). A nontrivial interplay between triadic closure, preferential, and anti-preferential attachment: New insights from online data. *Online Social Networks and Media*, *34–35*, 100248. https://doi.org/10.1016/j.osnem.2023.100248

Krueger, T., Szwabiński, J., & Weron, T. (2017). Conformity, Anticonformity and Polarization of Opinions: Insights from a Mathematical Model of Opinion Dynamics. *Entropy*, *19*(7), Article 7. https://doi.org/10.3390/e19070371

Kurahashi-Nakamura, T., Mäs, M., & Lorenz, J. (2016). Robust clustering in generalized bounded confidence models. *Journal of Artificial Societies and Social Simulation*, *19*(4).

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social Influence and the

Collective Dynamics of Opinion Formation. *PLOS ONE*, *8*(11), e78433. https://doi.org/10.1371/journal.pone.0078433