
Supplementary material**Effects of network topology and trait distribution on collective decision making****Pengyu Liu^{1,2*}and Jie Jian³**¹ Department of Microbiology and Molecular Genetics, University of California, Davis, Davis, CA 95616, USA² Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada³ Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada*** Correspondence:** Email: penliu@ucdavis.edu.

Abstract: Individual-level interactions shape societal or economic processes such as infectious diseases spreading, stock prices fluctuating and public opinion shifting. Understanding how different individuals interacting affect collective outcomes is more important than ever, as the internet and social media develop. Social networks representing individuals' influence relations play a key role in understanding the connections between individual-level interactions and societal or economic outcomes. Recent research has revealed how the topology of a social network affects collective decision making in a community. Furthermore, the traits of individuals that determine how they process received information for making decisions also change a community's collective decisions. In this work, we develop stochastic processes to generate networks of individuals with two simple traits: being a conformist and being an anticonformist. We introduce a novel deterministic voter model for a trait-attributed network, where the individuals make binary choices following simple deterministic rules based on their traits. We show that the simple deterministic rules can drive unpredictable fluctuations of collective decisions which eventually become periodic. We study the effects of network topology and trait distribution on the first passage time for the sequence of collective decisions showing periodicity.

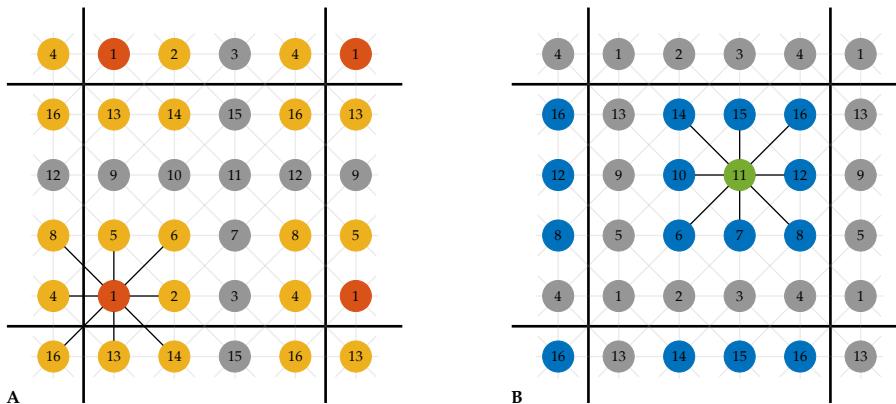
Keywords: interacting system, social network, collective decision making, network topology, trait distribution, first passage time

Mathematics Subject Classification: 60K35, 05C82, 91D30

Supplementary material

Toric lattices and random networks

The toric lattices are boundaryless. A node on a boundary of a lattice is related to some nodes on the opposite boundary. Specifically, for the toric lattice of size m , the node at (x, y) is related to the eight surrounding nodes: the eastern node at $((x+1) \bmod m, y)$, the northern node at $(x, (y+1) \bmod m)$, the western node at $((x-1) \bmod m, y)$, the southern node at $(x, (y-1) \bmod m)$, the northeastern node at $((x+1) \bmod m, (y+1) \bmod m)$, the northwestern node at $((x-1) \bmod m, (y+1) \bmod m)$, the southwestern node at $((x-1) \bmod m, (y-1) \bmod m)$, and the southeastern node at $((x+1) \bmod m, (y-1) \bmod m)$. See Supplementary Figure 1 for an example.



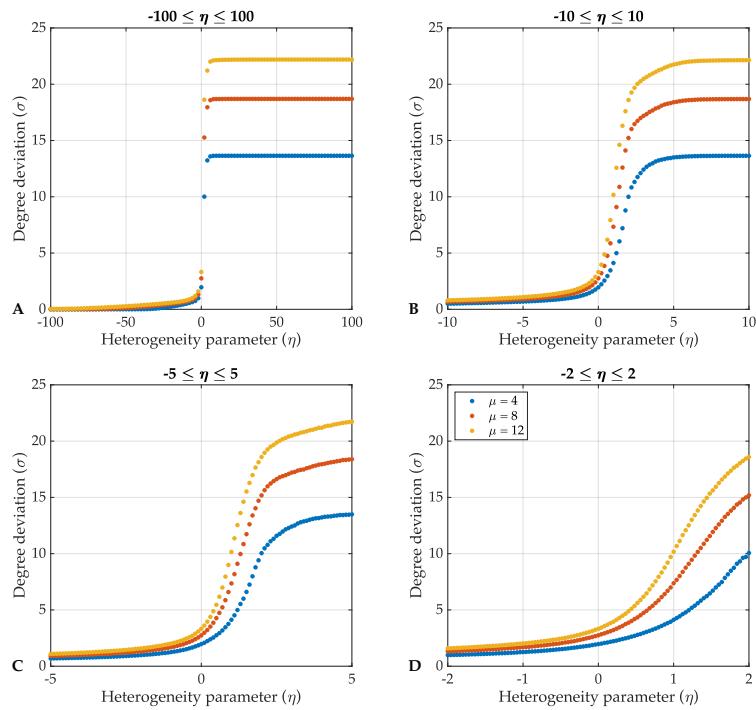
Supplementary Figure 1. The toric lattice of size 4 and the neighbors of two individuals.

The 8 neighbors of individual 1 displayed in red are individuals 2, 4, 5, 6, 8, 13, 14, and 16 displayed in yellow (panel A). The 8 neighbors of individual 11 displayed in green are individuals 6, 7, 8, 10, 12, 14, 15, and 16 displayed in blue (panel B).

The model generating random networks has three parameters: the number of nodes n , the mean degree μ and the heterogeneity parameter η regulating the degree deviation σ . In Supplementary Figure 2, we show the relations between the heterogeneity parameter and the degree deviation. We generate random networks of 100 nodes with mean degree $\mu = 4$, $\mu = 8$ and $\mu = 12$ and η ranging from -100 to 100 . Each data point in Supplementary Figure 2 represents the average degree deviation over 1000 random networks generated with corresponding parameters, and the variance for each data point is smaller than 0.1.

Trait distribution

The process attributing traits to a network topology with the attributing parameter α , which regulates the mixing parameter χ defined to be the average number of conformist neighbors over all anticonformists. Supplementary Figure 3 shows the relations between the attributing parameter and the mixing parameter of random networks. Panel A shows an attributed toric lattice of size 10 generated with half of the nodes being anticonformists ($r = 50$) and the attributing parameter $\alpha = 0.001$. The attributed toric lattice has scattered anticonformists with the mixing parameter $\chi = 5.52$, which means on average a anticonformist has 5.52 conformist neighbors. Panel B shows an attributed toric lattice of



Supplementary Figure 2. Relations between the heterogeneity parameter and degree deviations of generated random networks. The figure shows the relations between average degree deviations of random networks of 100 nodes generated with mean degree $\mu = 4$, $\mu = 8$ and $\mu = 12$, and the heterogeneity parameter η ranging from -100 to 100 in increments of 2 (panel A), from -10 to 10 in increments of 0.2 (panel B), from -5 to 5 in increments of 0.1 (panel C), and from -2 to 2 in increments of 0.04 (panel D). Each data point represents the average degree deviation of 1000 random networks generated with corresponding parameters. The variance for each data point is smaller than 0.1.

size 10 generated with half of the nodes being anticonformists ($r = 50$) and the attributing parameter $\alpha = 10$. The attributed toric lattice has clustered anticonformists with the mixing parameter $\chi = 1.44$. Panel C displays the relations between the attributing parameter and the mixing parameter for random networks generated with different numbers of individuals $n = 50$, $n = 100$ and $n = 150$. Each data point represents an attributed random network generated with other parameters set to $\mu = 8$, $\eta = 0$, $r = 50\%n$ and the initial choices being -1 for all individuals. The relations are linear in general. For $n = 100$, we have a fitted curve $y = -1.13x + 5.25$ with $R^2 = 0.91$; for $n = 50$, we have a fitted curve $y = -1.03x + 5.19$ with $R^2 = 0.84$; for $n = 150$, we have a fitted curve $y = -1.16x + 5.27$ with $R^2 = 0.94$. Panel D displays the relations between the attributing parameter and the mixing parameter for random networks generated with different mean degrees $\mu = 4$, $\mu = 8$ and $\mu = 12$. Each data point represents an attributed random network generated with other parameters set to $n = 100$, $\eta = 0$, $r = 50$ and the initial choices being -1 for all individuals. For $\mu = 8$, we have a fitted curve $y = -1.13x + 5.25$ with $R^2 = 0.91$; for $\mu = 4$, we have a fitted curve $y = -0.67x + 2.75$ with $R^2 = 0.88$; for $\mu = 12$, we have a fitted curve $y = -1.48x + 7.64$ with $R^2 = 0.92$. Panel E displays the relations between the attributing parameter and the mixing parameter for random networks generated with different hetero-

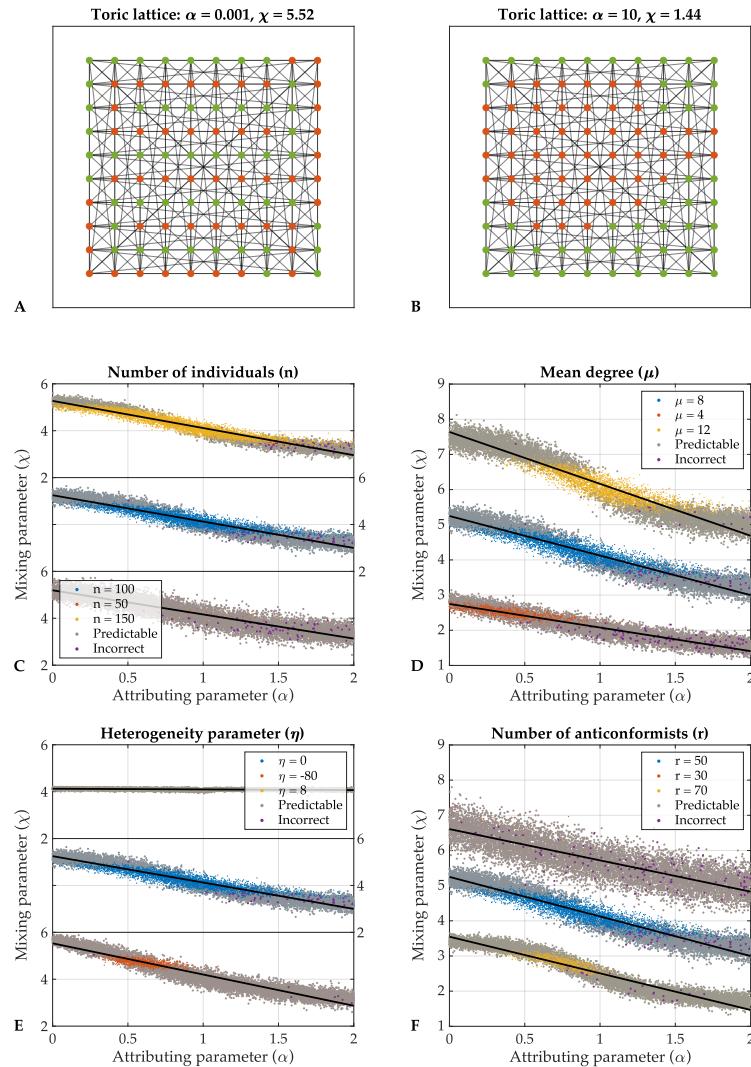
geneity parameters $\eta = -80$, $\eta = 0$ and $\eta = 8$. Each data point represents an attributed random network generated with other parameters set to $n = 100$, $\mu = 8$, $r = 50$ and the initial choices being -1 for all individuals. For $\eta = 0$, we have a fitted curve $y = -1.13x + 5.25$ with $R^2 = 0.91$; for $\eta = -80$, we have a fitted curve $y = -1.33x + 5.53$ with $R^2 = 0.93$; for $\eta = 8$, we have a fitted curve $y = -0.03x + 4.12$ with $R^2 = 0.17$. Panel F displays the relations between the attributing parameter and the mixing parameter for random networks generated with different numbers of anticonformists $r = 30$, $r = 50$ and $r = 70$. Each data point represents an attributed random network generated with other parameters set to $n = 100$, $\mu = 8$, $\eta = 0$ and the initial choices being -1 for all individuals. For $r = 50$, we have a fitted curve $y = -1.13x + 5.25$ with $R^2 = 0.91$; for $r = 30$, we have a fitted curve $y = -0.89x + 6.61$ with $R^2 = 0.72$; for $r = 70$, we have a fitted curve $y = -1.04x + 3.55$ with $R^2 = 0.94$.

First passage time and predictability of cumulative sequences

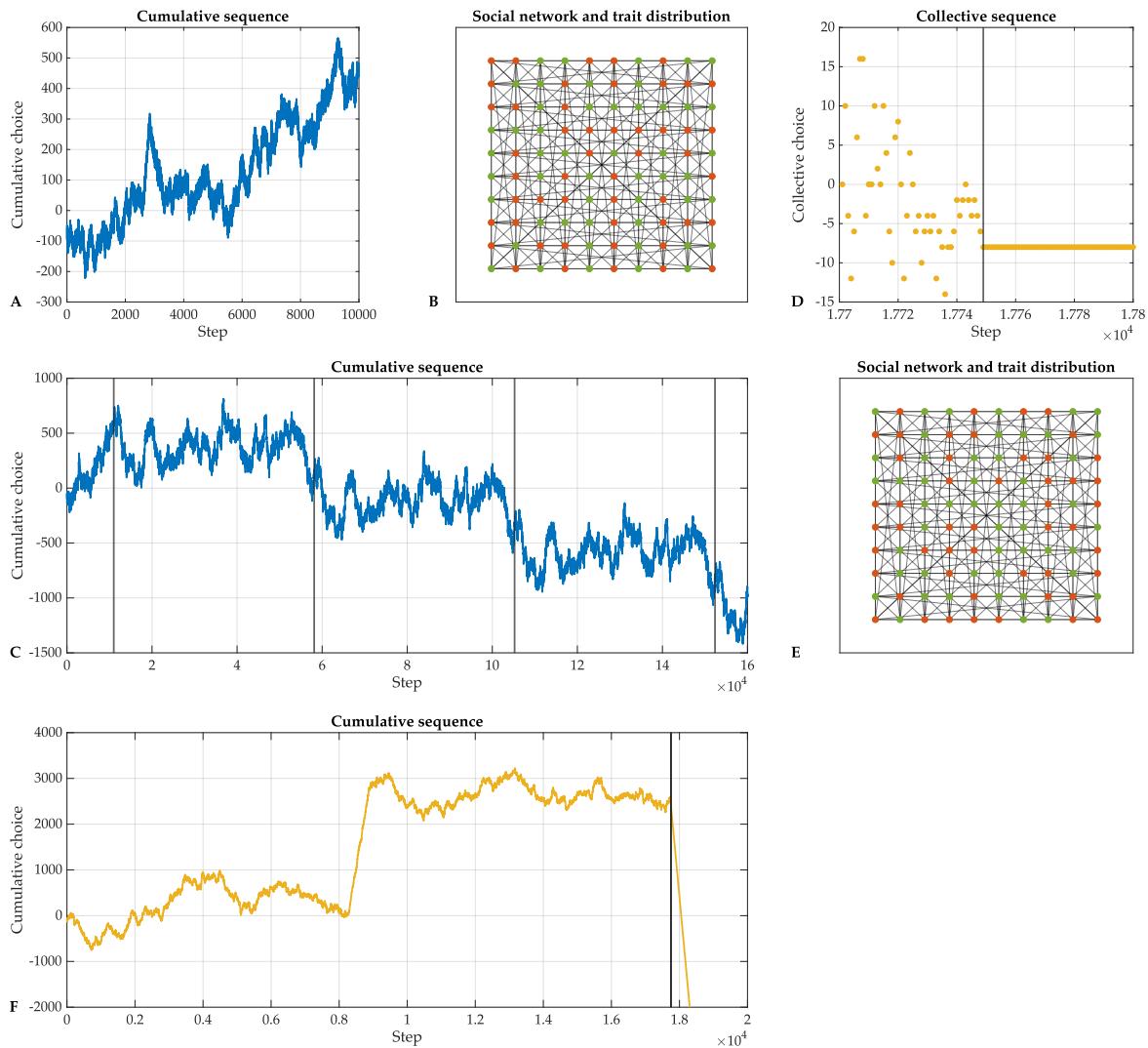
We argue that every collective sequence of choices eventually enters a unique period. For a community of n individuals, there are 2^n unique choice patterns. Note that the deterministic process defined in Section 2.1 is memoryless in the sense that the choice pattern at step k only depends on the choice pattern at step $k - 1$. Moreover, each choice pattern determines a unique succeeding choice pattern. If the deterministic process has more than 2^n steps, then the sequence of choice patterns must have identical elements $C(\cdot, k) = C(\cdot, l)$ due to the pigeon hole principle. Since the process is deterministic, identical subsequences of choice patterns follow $C(\cdot, k)$ and $C(\cdot, l)$, and periodicity appears in the sequence of choice patterns, hence in the collective sequence of the community's choices. Thus, given the network topology, the trait distribution and initial choices for a community of individuals, the collective sequence eventually enters a unique period determined by the three factors.

Recall that the length $L(P)$ of the eventual period P and the length $L(Q)$ of the pre-period subsequence Q of a cumulative sequence are defined to be the numbers of steps that P and Q span respectively. The period gain ΔP of the eventual period is the change in cumulative sequence over the period P . The gradient of the eventual period is defined to be $\nabla P = |\Delta P| / L(P)$. In Supplementary Figure 4, we show the eventual periods and the pre-period subsequences of cumulative sequences of choices of two communities. Panel A shows the first 10000 steps of the cumulative sequence of choices of the community represented by the attributed toric lattice displayed in panel B. The attributed toric lattice has size $m = 10$, and the trait distribution is generated with $r = 50$ and $\alpha = 0.37$. The initial choices of the community are -1 for all individuals. In panel A, the first 10000 steps of the cumulative sequence do not contain a complete eventual period, so the cumulative sequence is unpredictable. Actually, the pre-period subsequence Q displayed in panel C before the first vertical line has length $L(Q) = 11010$. If we extend the length of the process to $t = 160000$, we see three complete eventual periods of the cumulative sequence displayed in panel C. The eventual period P showed in panel C has length $L(P) = 47115$, period gain $\Delta P = -474$ and gradient $\nabla P = 0.01$. Similarly, The attributed toric lattice displayed in panel E has size $m = 10$, and the trait distribution is generated with $r = 50$ and $\alpha = 0.7$. The initial choices of the community are -1 for all individuals. Panel D and panel F show the pre-period subsequence Q of length $L(Q) = 17749$, and the eventual period P has length $L(P) = 1$, period gain $\Delta P = -8$ and gradient $\nabla P = 8$.

To efficiently determine if a collective sequence $s = [s(0), s(1), \dots, s(t)]$ is predictable without recording and comparing choice patterns, we develop the heuristic method as follows. We extract the subsequence $s' = [s(t + 1 - \tau), \dots, s(t - 1), s(t)]$ consisting of the last τ elements in s and search for



Supplementary Figure 3. Relations between the attributing parameter and the mixing parameter of attributed networks. Attributed toric lattices generated with scattered anticonformists and clustered anticonformists are displayed in panel A and B respectively. The rest of the figure shows the relations between the attributing parameter and the mixing parameter of random networks generated with different numbers of individuals (panel C), different mean degrees (panel D), different heterogeneity parameters (panel E) and different number of anticonformists (panel F). Each data point represents a random network generated with corresponding parameters. For each set of parameters, 10000 random networks (data points) are generated. Grey data points represent random networks with predictable cumulative sequences, and colored data points represent random networks with unpredictable cumulative sequences. In particular, the purple data points represent the networks with unpredictable cumulative sequences that are determined to be predictable by the heuristic method.



Supplementary Figure 4. Eventual periods and pre-period subsequences of cumulative sequences. The first three panels show the first 10000 steps of the cumulative sequence (panel A) of the attributed network generated with $m = 10$, $r = 50$ and $\alpha = 0.37$ (panel B) and the first 160000 steps of the cumulative sequence (panel C). The last three panels show the 100 steps of collective sequence (panel D) near the appearance of the first eventual period for the attributed network generated with $m = 10$, $r = 50$ and $\alpha = 0.7$ (panel E) and the first 20000 steps of the cumulative sequence (panel F). The complete eventual periods are displayed between vertical lines, and the pre-period subsequences are before the first vertical line in panels C and F. The green nodes represent conformists and the red nodes represent anticonformists in panels B and D.

subsequences of s with τ consecutive elements that are identical to s' . If s' is the only subsequence, then the heuristic method determines the collective sequence and the corresponding cumulative sequence to be unpredictable. If there are more than one subsequences in s that are identical to s' , then the heuristic method determines the collective sequence and the corresponding cumulative sequence to be predictable.

We argue that the heuristic method faithfully determines every predictable collective sequence. Let s be a predictable collective sequence and s' be the subsequence of s consisting of the last τ elements. By definition, there exist at least one complete eventual period in the first $t + 1 - \tau$ steps of s . If there exists one eventual period in the first $t + 1 - \tau$ steps, then what follows must be in the eventual period. Hence, s' must be in the eventual period, and there exists at least one subsequence in the first $t + 1 - \tau$ steps that is identical to s' . Therefore, the heuristic method faithfully determines s to be predictable. If s is unpredictable, that is, there exists no eventual period in the first $t + 1 - \tau$ steps, then there may still be subsequences in the first $t + 1 - \tau$ steps that are identical to s' . This is because different choice patterns may have the same sum of choices. So, the heuristic method may incorrectly determine s to be predictable and underestimate the probability of cumulative sequences being unpredictable.

In Supplementary Figure 3, we show the unpredictable collective sequences that are incorrectly determined by the heuristic method as predictable ones with purple data points. In panel C, there are 75 data points with incorrect predictability for $n = 100$, 70 for $n = 50$ and 65 for $n = 150$. In panel D, there are 124 data points with incorrect predictability for $\mu = 4$ and 21 for $\mu = 12$. In panel E, there are 8 data points with incorrect predictability for $\eta = -80$ and 0 for $\eta = 8$. In panel F, there are 136 data points with incorrect predictability for $r = 30$ and 22 for $r = 70$. On average, 0.58% of the unpredictable collective sequences are incorrectly determined to be predictable by the heuristic method.

Homogeneous attributed networks

Random networks of all conformists and toric lattices with homogeneously clustered conformists and anticonformists generate predictable cumulative sequences escalating to an extreme with $\nabla P > 1$. In contrast, random networks of all anticonformists and toric lattices with homogeneously mixed conformists and anticonformists generate predictable cumulative sequences oscillating with $\nabla P = 0$.

We can deduce the cumulative sequences for the four homogeneous attributed networks displayed in Supplementary Figure 5. The toric lattice of size $m = 10$ displayed in panel B is attributed with $r = 50$, and the conformists and anticonformists are homogeneously separated into two clusters. Each anticonformist in the interior of the cluster has 8 anticonformist neighbors, and each anticonformist on the boundary of the cluster has 5 anticonformist neighbors and 3 conformist neighbors. The cluster of conformists also have the same patterns. When the initial choices are -1 for all individuals, the two clusters can not affect each other, so all anticonformists will change at every step, and all conformists will keep choosing -1 at every step. Therefore, the attributed network generates an escalating cumulative sequence displayed in panel A with $\nabla P = 50$, as the collective sequence consists of alternating 0 and -100 . The toric lattice of size $m = 10$ displayed in panel D is attributed with $r = 50$, and the individuals with different traits are homogeneously mixed such that every anticonformist has 6 conformist neighbors and 2 anticonformist neighbors, and symmetrically, every conformist has 6 anticonformist neighbors and 2 conformist neighbors. Suppose that the initial choices are -1 for all individuals. At step 1, the anticonformists will choose 1, and the conformists will keep -1 ; at step 2, the anticon-

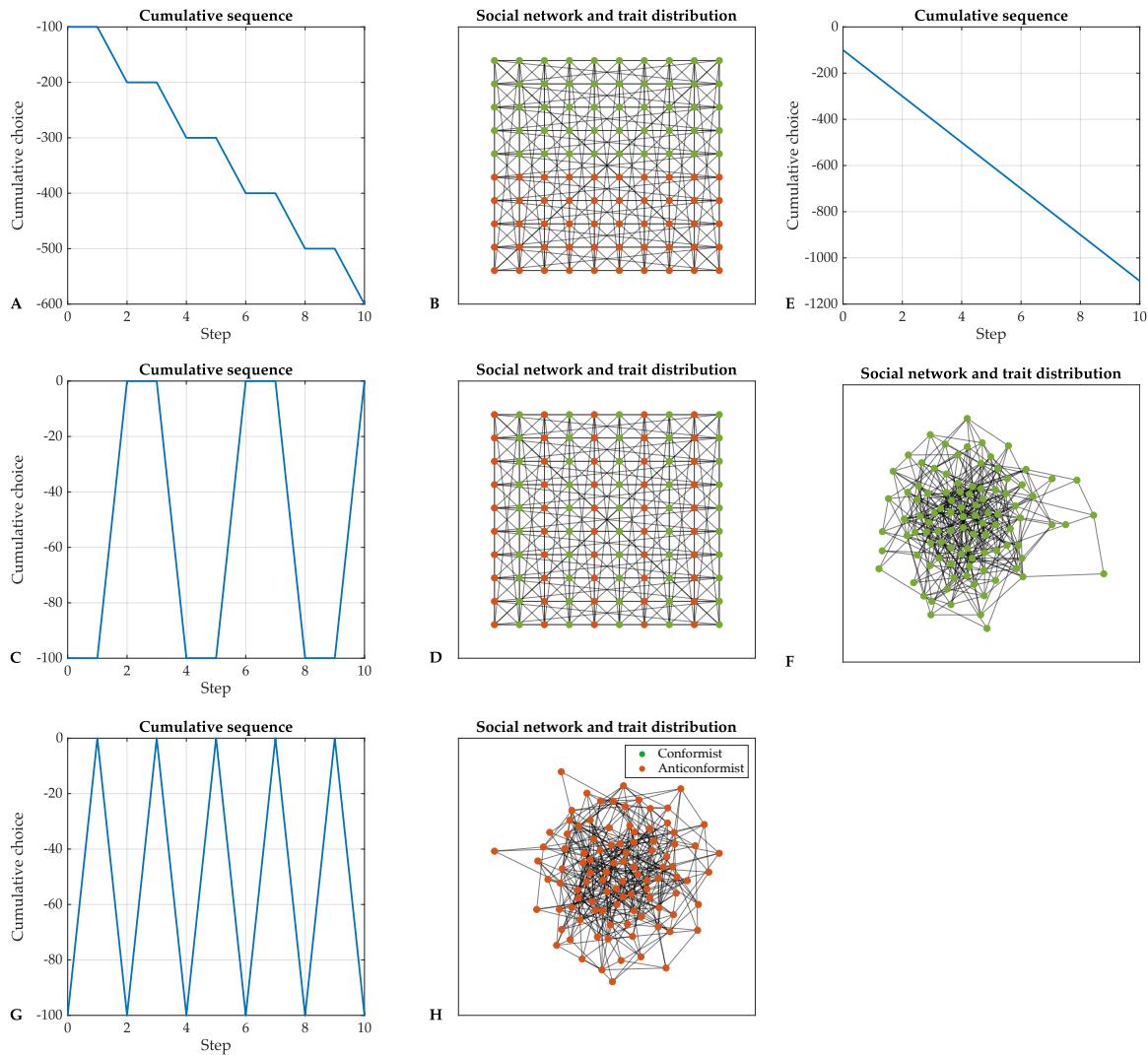
formists will keep 1, and the conformists will choose 1; at step 3, the anticonformists will choose -1 , and the conformists will keep 1; at step 4, the anticonformists will keep -1 , and the conformists will choose -1 , which is the same choice pattern as the initial choices. Thus, the attributed network generates an oscillating cumulative sequence displayed in panel C with $\nabla P = 0$. The connected random network displayed in panel F has 100 conformist individuals. When the initial choices are -1 for all individuals, all conformists will keep choosing -1 at every step, and the attributed network generates an escalating cumulative sequence displayed in panel E with $\nabla P = 100$. The connected random network displayed in panel H has 100 anticonformist individuals. When the initial choices are -1 for all individuals, all anticonformists will change their choices at every step, and the attributed network generates an oscillating cumulative sequence displayed in panel G with $\nabla P = 0$.

Results in alternative settings

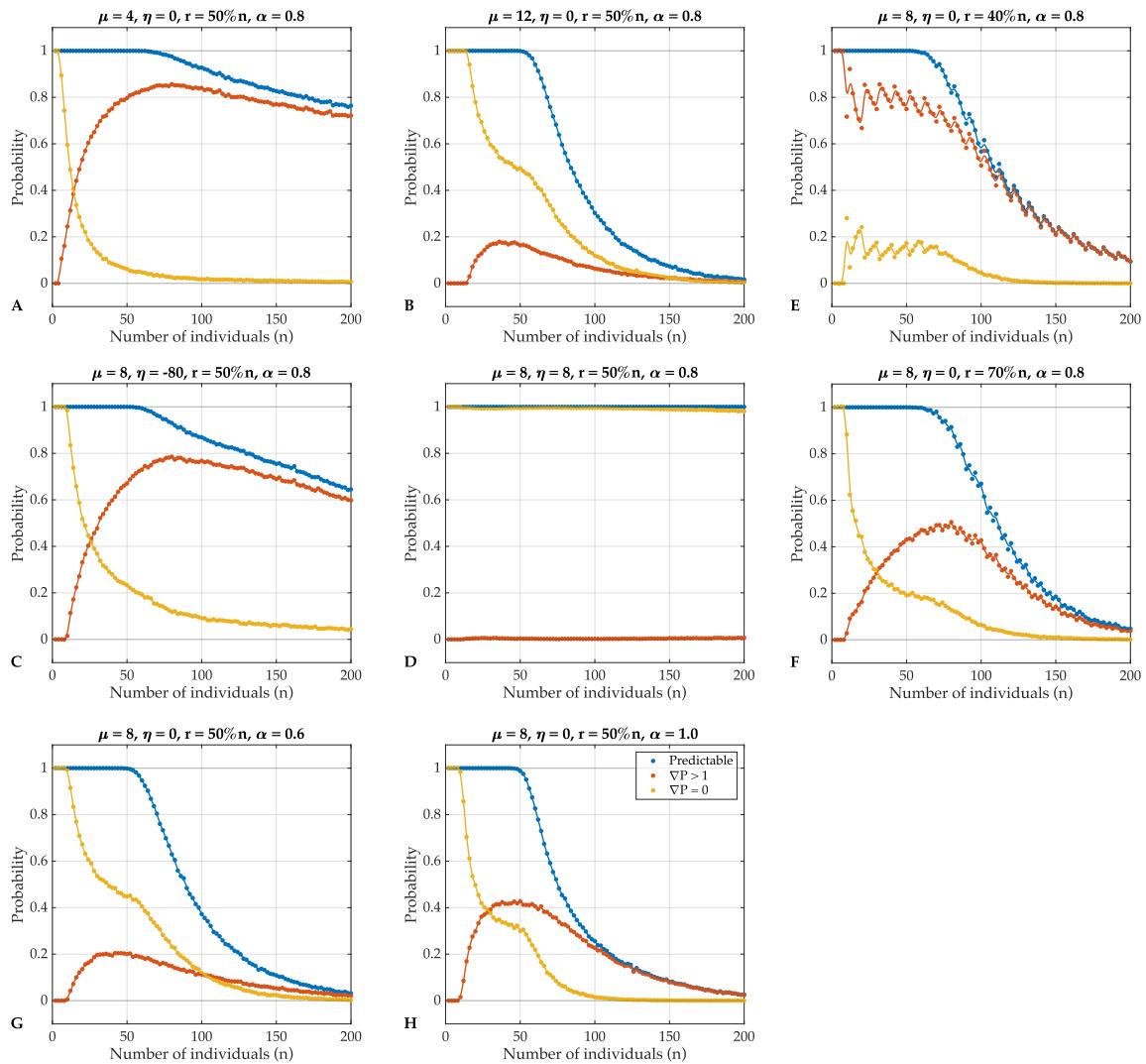
We investigate the effects of the three network topology parameters and the two trait distribution parameters on the probability of predictable, escalating and oscillating sequences in different settings. See Supplementary Figure 6 - 10. The number of individuals has a default value $n = 100$. In investigating effects of the other four parameters, we set $n = 50$ and $n = 150$. The mean degree has a default value $\mu = 8$. In investigating effects of the other four parameters, we set $\mu = 4$ and $\mu = 12$. The heterogeneity parameter has a default value $\eta = 0$. In investigating effects of the other four parameters, we set $\eta = -80$ and $\eta = 8$. The number of anticonformists has a default value of $r = 50\%n$. In investigating effects of the other four parameters, we set $r = 40\%n$ and $r = 70\%n$. The attributing parameter has a default value of $\alpha = 0.8$. In investigating effects of the other four parameters, we set $\alpha = 0.6$ and $\alpha = 1$.

We also generated random networks with parameters $n = 1912$, $\mu = 32.74$ and $\eta = 1.19$ that resembles the parameters of the real social network. We choose $\eta = 1.19$ so that the generated random networks have degree deviation near $\sigma = 55.85$. In Supplementary Figure 11, we display the relations and an attributed real social network with $r = 956$ and $\alpha = 1$ and its cumulative sequence.

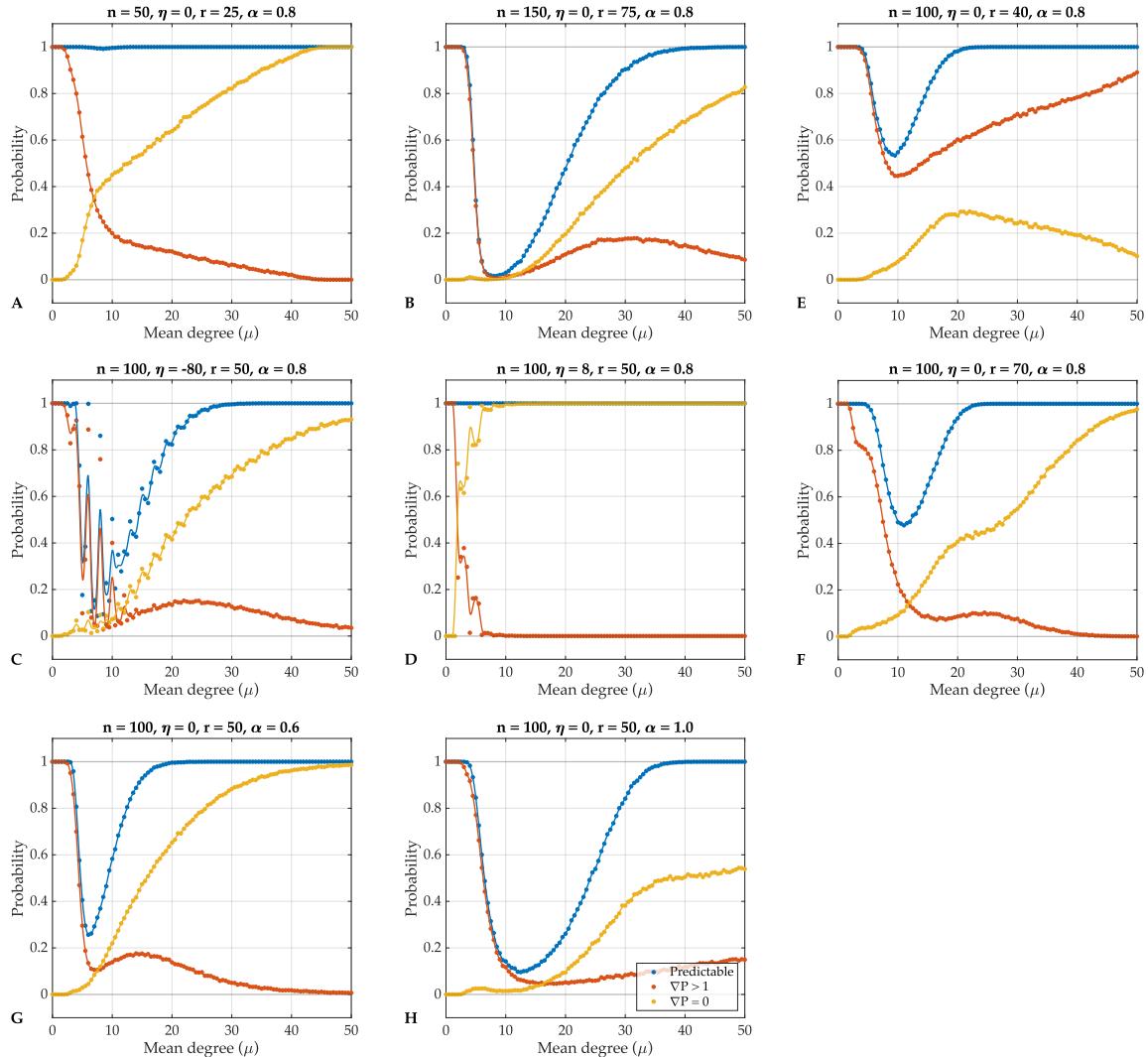
In addition, we analyze how the network topology parameters and the trait distribution parameters affect the cumulative sequences for Watts-Strogatz small-world networks. The Watts-Strogatz model has a parameter β regulating the topology of generated networks instead of the parameter η in our random graph model. We take values for the parameter β from the interval $(0, 1]$ and set the default value for β to be $\beta = 0.15$. Supplementary Figure 12 shows the relations between the parameters and the probabilities of predictable, escalating and oscillating sequences. In panel A, the number of individuals n takes value in the interval $[22, 200]$ instead of $[2, 200]$ because Watts-Strogatz small-world networks with a small number of nodes can have loops. The results are in general similar to the random networks. With the parameter β set at $\beta = 0.15$, we observe that probability of predictable sequences has the lowest value when $r > 50\%n$; see panel D of Supplementary Figure 12. This is different from the random network, whose probability of predictable sequences has the lowest value around $r = 50\%n$.



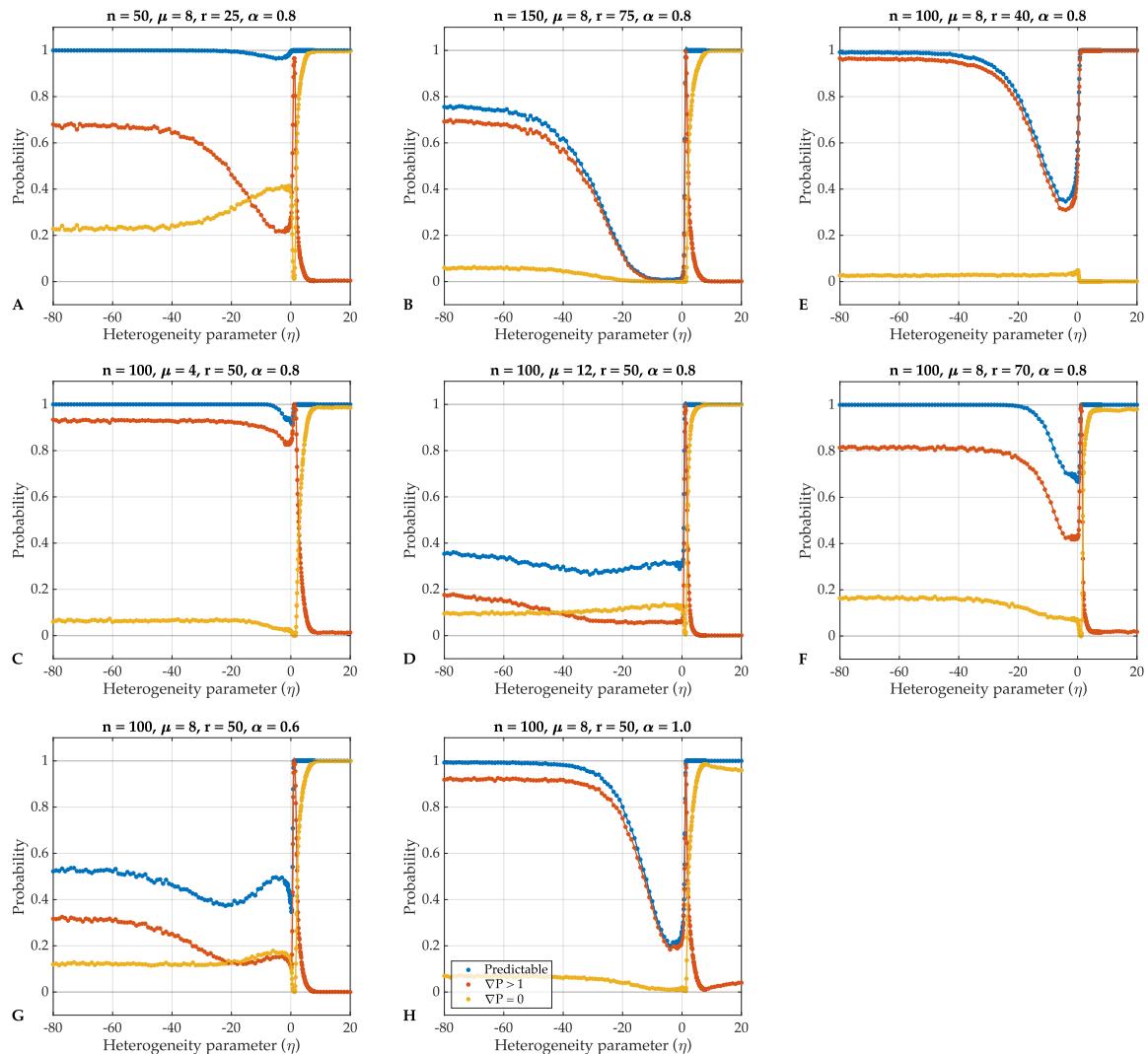
Supplementary Figure 5. Homogeneous attributed networks and the corresponding cumulative sequences. The figure shows the escalating cumulative sequence (panel A) of the toric lattice with homogeneously clustered conformists and anticonformists (panel B), the oscillating cumulative sequence (panel C) of the toric lattice with homogeneously mixed conformists and anticonformists (panel D), the escalating cumulative sequence (panel E) of the random network of all conformists (panel F), and the oscillating cumulative sequence (panel G) of the random network of all anticonformists (panel H). The initial choices for the four networks are -1 for all individuals.



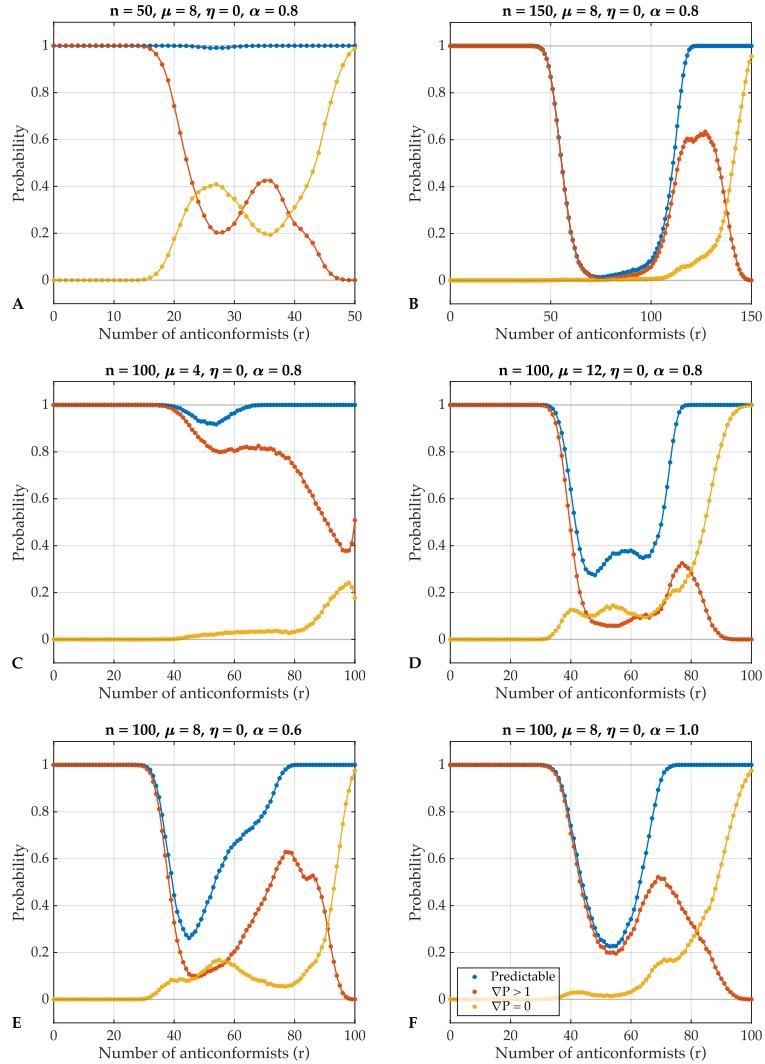
Supplementary Figure 6. Effects of the number of individuals on the probability of predictable sequences in different settings. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the number of individuals n for random networks with $\mu = 4$ (panel A) and $\mu = 12$ (panel B), networks with $\eta = -80$ (panel C) and $\eta = 8$ (panel D), networks with $r = 40$ (panel E) and $r = 70$ (panel F) and networks with $\alpha = 0.6$ (panel G) and $\alpha = 1$ (panel H). Each data point is computed with 10000 random networks and their corresponding cumulative sequences with initial choices -1 for all individuals. Smooth fitted curves are added for visualization.



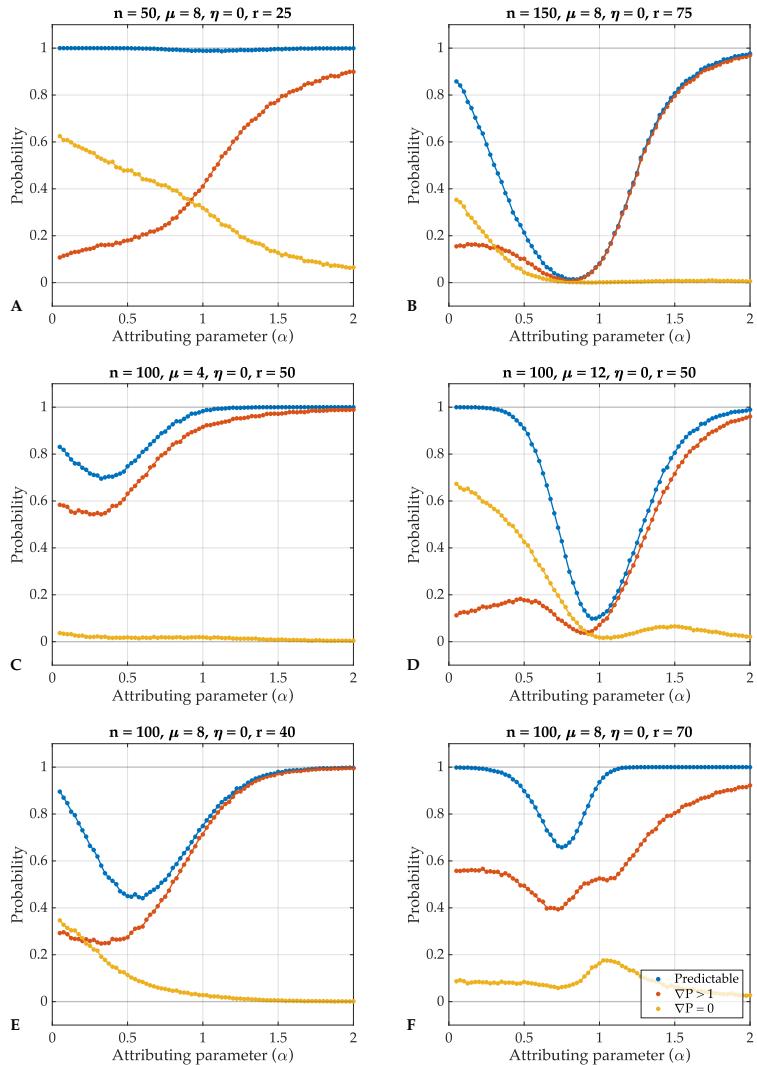
Supplementary Figure 7. Effects of the mean degree on the probability of predictable sequences in different settings. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the mean degree μ for random networks with $n = 50$ (panel A) and $n = 150$ (panel B), networks with $\eta = -80$ (panel C) and $\eta = 8$ (panel D), networks with $r = 40$ (panel E) and $r = 70$ (panel F) and networks with $\alpha = 0.6$ (panel G) and $\alpha = 1$ (panel H). Each data point is computed with 10000 random networks and their corresponding cumulative sequences with initial choices -1 for all individuals. Smooth fitted curves are added for visualization.



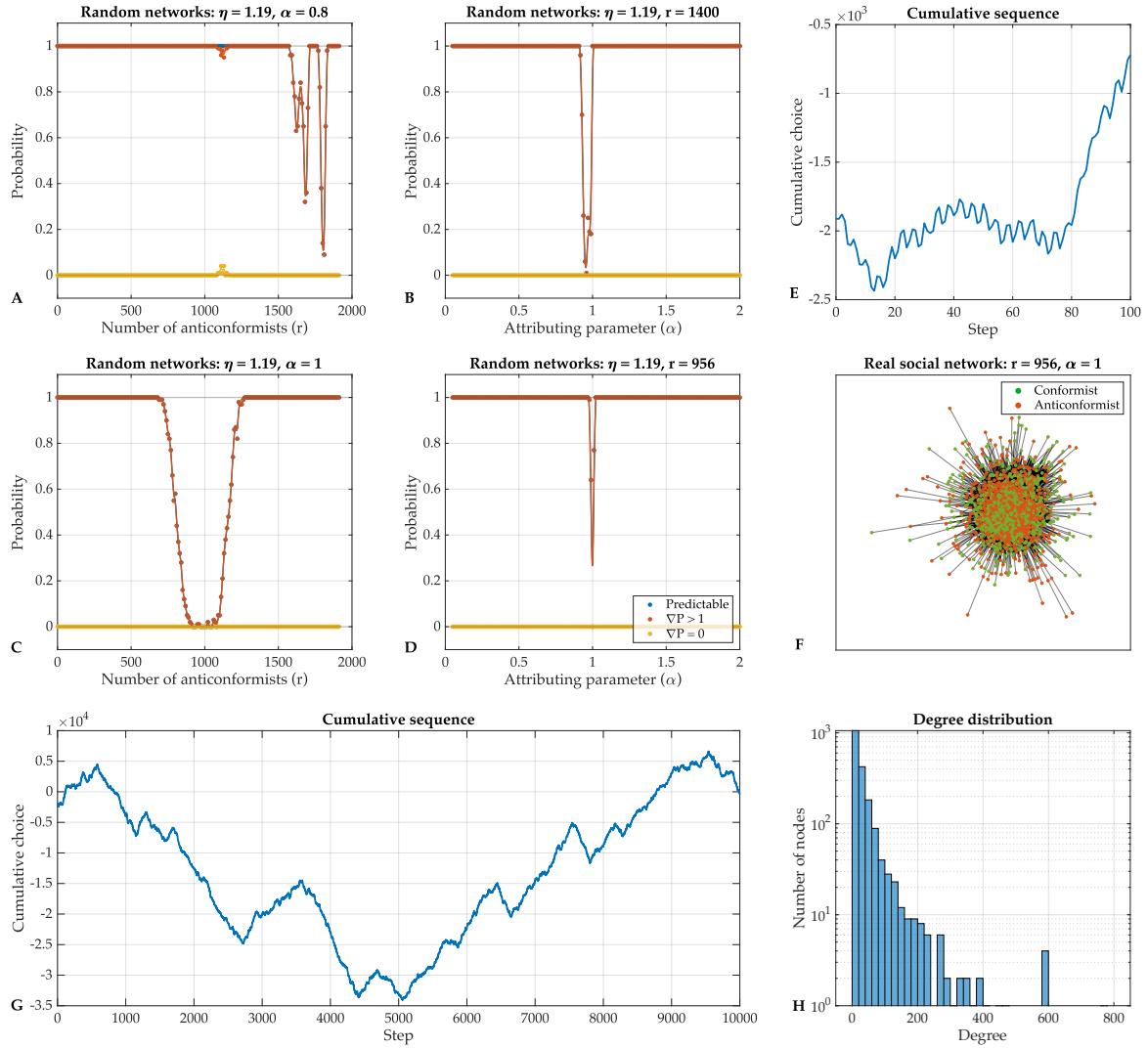
Supplementary Figure 8. Effects of the heterogeneity parameter on the probability of predictable sequences in different settings. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the heterogeneity parameter η for random networks with $n = 50$ (panel A) and $n = 150$ (panel B), networks with $\mu = 4$ (panel C) and $\mu = 12$ (panel D), networks with $r = 40$ (panel E) and $r = 70$ (panel F) and networks with $\alpha = 0.6$ (panel G) and $\alpha = 1$ (panel H). Each data point is computed with 10000 random networks and their corresponding cumulative sequences with initial choices -1 for all individuals. Smooth fitted curves are added for visualization.



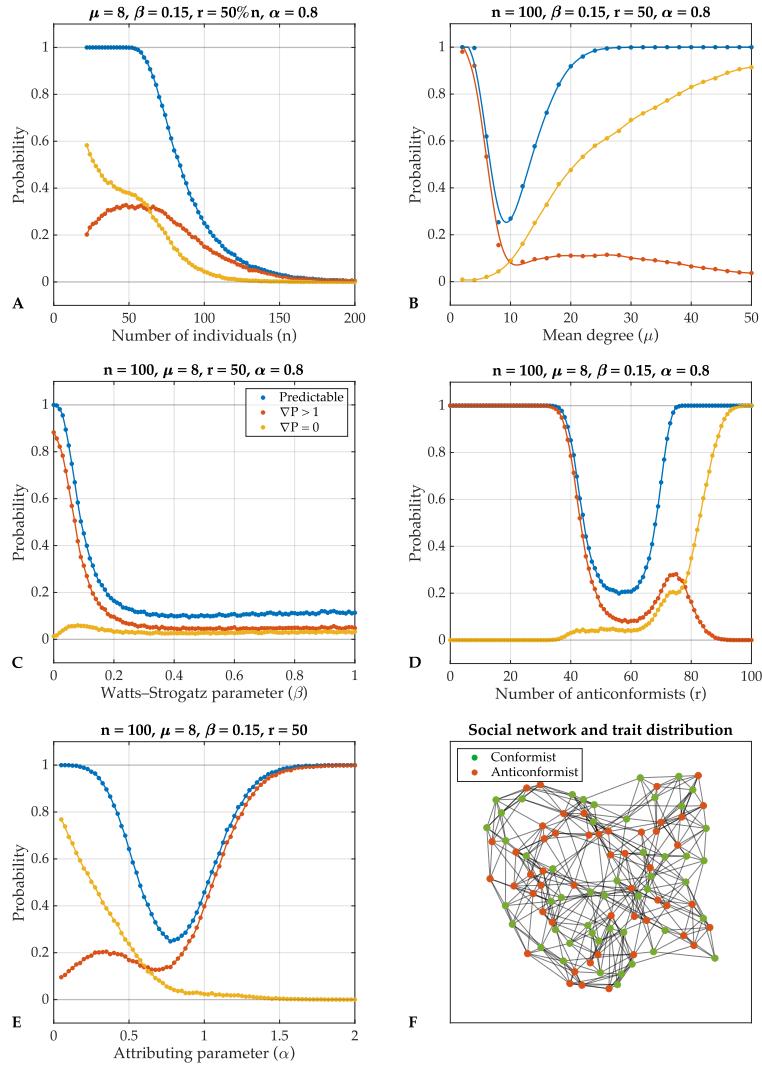
Supplementary Figure 9. Effects of the number of anticonformists on the probability of predictable sequences in different settings. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the attributing parameter α for random networks with $n = 50$ (panel A) and $n = 150$ (panel B), networks with $\mu = 4$ (panel C) and $\mu = 12$ (panel D) and networks with $r = 40$ (panel E) and $r = 70$ (panel F). Each data point is computed with 10000 random networks and their corresponding cumulative sequences with initial choices -1 for all individuals. Smooth fitted curves are added for visualization.



Supplementary Figure 10. Effects of the attributing parameter on the probability of predictable sequences in different settings. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the number of anticonformists r for random networks with $n = 50$ (panel A) and $n = 150$ (panel B), networks with $\mu = 4$ (panel C) and $\mu = 12$ (panel D) and networks with $\alpha = 0.6$ (panel E) and $\alpha = 1$ (panel F). Each data point is computed with 10000 random networks and their corresponding cumulative sequences with initial choices -1 for all individuals. Smooth fitted curves are added for visualization.



Supplementary Figure 11. Effects of the trait distribution parameters on random networks generated with parameters resembling the real social network. The figure shows the relations between the probability of predictable, escalating and oscillating sequences and the number of anticonformists r with $\alpha = 0.8$ (panel A), the attributing parameter α with $r = 1400$ (panel B), the number of anticonformists r with $\alpha = 1$ (panel C) and the attributing parameter α with $r = 956$ (panel D) for random networks with $\eta = 1.19$; each data point is computed with 100 attributions on the real social network with initial choices -1 for all individuals; smooth fitted curves are added for visualization. The trait-attributed real social network generated with $r = 50\%n = 956$ and $\alpha = 1$ is displayed in panel F, and its degree distribution is displayed in panel H. The first 100 steps and the first 10000 steps of the cumulative sequence of the attributed real social network is displayed in panel E and G respectively.



Supplementary Figure 12. Effects of network topology parameters and trait distribution parameters on the probability of predictable sequences for Watts-Strogatz small-world networks. The figure shows the relations between the probabilities of predictable, escalating and oscillating sequences and the number of individuals (nodes) n (panel A), the mean degree μ (panel B), the Watts-Strogatz parameter β (panel C), the number of anticonformists r (panel D) and the attributing parameter α (panel E) with all other parameters set at default values; each data point is computed with 10000 Watts-Strogatz small-world networks and their corresponding cumulative sequences with initial choices -1 for all individuals; smooth fitted curves are added for visualization. A Watts-Strogatz small-world network generated with $n = 100, \mu = 8, \beta = 0.15, r = 50$, and $\alpha = 0.8$ is displayed in panel F.