Kinetics of competing social contagions: Symmetry breaking and equilibrium indeterminacy

Teruyoshi Kobayashi
Department of Economics, Kobe University, Kobe, Japan
Center for Computational Social Science, Kobe University, Kobe, Japan

Complex contagion on social networks has been explained as a collective outcome of threshold behaviors in which influence from local neighbors may trigger a global cascade. However, it is largely unexplored what behavioral rules best describe individuals’ optimization and how spreading dynamics emerge from them. Here, we develop a microfounded general threshold model that enables us to analyze the collective dynamics of individual behavior in the propagation of competing information/technologies (i.e., “social memes”). The analysis reveals that the popularities of competing memes in systems of finite size are indeterminate since the propagation process follows a saddle path, leading to symmetry breaking. It suggests that the virality of social memes may not be attributed to their intrinsic attractiveness but rather to randomness in network structure.

New technologies, information and political opinions occasionally spread globally through social ties among individuals. The dynamical processes of complex contagion have been extensively studied to understand whether and to what extent a “social meme” (e.g., a particular information/technology) spreads on a social network [1–9]. However, it is common in reality that multiple memes coexist and are competing/complementing each other. Examples of competing memes include “format wars” (e.g., VHS vs Betamax, Blu-ray Disc vs HD-DVD, etc) [10, 11], the propagation of political opinions (e.g., democratic/republican) [12–17], and vaccination behavior (i.e., pro- and anti-vaccination) [18–20]. There are also complementary memes, the acquisition of one reduces the cost of acquiring the other (e.g., learning Python and R).

Here, we develop a generalized threshold model of global cascades that allows us to analyze the propagation dynamics of competing/complementing memes. Our model is microfounded in the sense that individual behavior is fully optimized; individuals maximize coordination with neighbors. Therefore, any stationary state of the dynamical process, if it exists, is interpreted as a collective outcome of the individuals’ strategic choices, namely, a Nash equilibrium [21–25].

In the model, there are two types of information/technologies, respectively labeled as a and b, which we collectively call "memes". The two types of memes can be either complementary, exclusive or neutral. Each individual decides whether to accept a or b, or both (called the bilingual option, denoted by ab), depending on the popularity of each meme among local neighbors [25–26]. Let \( s \in \{0, a, b, ab\} \) \( \equiv S \) be a strategy (or a choice) of an individual where \( s = 0 \) indicates the status-quo (i.e., neither meme is accepted). In an infinitesimal time interval \( dt \), randomly selected individuals in the network update their strategies (i.e., asynchronous update [3][27]) to maximize the total payoffs of coordination games. The payoff matrix for a bilateral coordination game is presented in Tab. I.

TABLE I. Payoff matrix of a coordination game. We have \( a, b > c > 0 \), where the two memes are complements (resp. exclusive) when \( \tilde{c} < c \) (resp. \( \tilde{c} > c \)).

|       | 0  | a  | b    | ab  |
|-------|----|----|------|-----|
| 0     | 0, 0 | 0, -c | 0, -c | 0, -2c |
| a     | -c, 0 | a - c, a - c | -c, -c | a - c, a - 2c |
| b     | -c, 0 | -c, -c | b - c, b + c | b - c, b - 2c |
| ab    | -2c, 0 | a - 2c, a - c | b - 2c, b - c | a + b - 2c, a + b - 2c |

Each element of the payoff matrix shows the returns for the corresponding strategy pair. For instance, the pair \((-c, 0)\) in the \((a, 0)\)th element of the matrix indicates that individuals \( i \) and \( j \) receive payoffs \(-c \) and \( 0 \), respectively. \( a \) and \( b \) are the benefits of coordinating with neighbors in adopting strategies \( a \) and \( b \), respectively, and \( c \) denotes the net cost of accepting a meme, where we assume that \( a, b > c > 0 \). \(-2\tilde{c}\) in the bottom row represents the net cost of adopting the bilingual option \( ab \) that will be incurred when individuals accept both of the memes. \( \tilde{c} \) may be larger or less than \( c \), depending on the extent to which the two memes are complementary or exclusive. If \( \tilde{c} \gg c \), then \( ab \) will no longer be a plausible option since the two memes are prohibitively exclusive (e.g., democratic and republican). In contrast, when \( \tilde{c} \) is low enough, \( ab \) would be preferred to \( a \) and \( b \), because accepting a meme reduces the cost of accepting the other (e.g., MacBook and iPhone).

Neighbors’ states are represented by \( \mathbf{m} = (m_0, m_a, m_b, m_{ab})^\top \), where \( m_s \) denotes the number of neighbors adopting strategy \( s \in S \). Note that we have \( \sum_{s \in S} m_s = k \) for nodes with degree \( k \). The total payoff value of a player is given by the sum of the payoffs obtained by playing \( k \) bilateral games. Let \( v(s, \mathbf{m}) \) denote the total payoff of a player adopting strategy \( s \in S \) and facing the neighbors’ strategy profile \( \mathbf{m} \). For

* kobayashi@econ.kobe-u.ac.jp
a given $\mathbf{m}$, the total payoff of each strategy leads to
\[
\begin{align*}
  v(0, \mathbf{m}) &= 0, \\
  v(a, \mathbf{m}) &= -ck + aM_a, \\
  v(b, \mathbf{m}) &= -ck + bM_b, \\
  v(ab, \mathbf{m}) &= -2ck + aM_a + bM_b,
\end{align*}
\]
where $M_a \equiv m_a + m_{ab}$ (resp. $M_b \equiv m_b + m_{ab}$) denotes the total number of neighbors accepting meme $a$ (resp. $b$), including bilinguals. The optimal strategy $s^*$ is then expressed as a function of $\mathbf{m}$:
\[
s^*(\mathbf{m}) = \arg \max_{s \in S} v(s, \mathbf{m}).
\]

In a time interval $dt$, a randomly chosen fraction $dt$ of $N$ individuals updates their strategies following Eq. (5). It is assumed that the initial states are kept unchanged for nodes with $k = 0$ since isolated nodes do not have a chance to play coordination game.

The optimal strategy $s^*$ for each individual is given by the following threshold rules (see Sec. S1 in Supplemental Material):
\[
s^* = \begin{cases} 
  a & \text{if } \frac{M_a}{k} > \theta_a, \frac{M_b}{k} < (1 - \lambda)\theta_b \text{ and } \frac{M_b}{M_a} > \theta_b, \\
  b & \text{if } \frac{M_b}{k} > \theta_b, \frac{M_a}{k} < (1 - \lambda)\theta_a \text{ and } \frac{M_a}{M_b} < \theta_a, \\
  ab & \text{if } \frac{M_a}{k} > (1 - \lambda)\theta_a, \frac{M_b}{k} > (1 - \lambda)\theta_b \text{ and } \theta_a + \frac{M_b}{k} > \theta_b(2 - \lambda), \\
  0 & \text{otherwise},
\end{cases}
\]
where $\theta_a \equiv c/a \in (0, 1)$, $\theta_b \equiv c/b \in (0, 1)$, and $\lambda \equiv 2(1 - \frac{c}{c}).$ $\lambda$ captures the degree of complementarity between $a$ and $b$ where $\lambda > 0$ (resp. $\lambda < 0$) indicates that the two memes are complementary (resp. exclusive). When $\lambda = 0$, they are mutually neutral; the virality of a meme is independent of the popularity of the other. In the analysis, we focus on a reasonable range of parameter values such that Nash equilibria of bilateral games are $(0, 0)$, $(a, a)$, $(b, b)$ and $(ab, ab)$. In fact, this assumption sets natural constraints for the threshold values: $\lambda < 1$, $(1 - \lambda)\theta_a < 1$ and $(1 - \lambda)\theta_b < 1$ (see Sec. S2 for a derivation). Note that if $M_b = 0$ (resp. $M_a = 0$), then the threshold rules reduce to the single threshold condition in the binary-state cascade model à la Watts [1]: $m_a/k > \theta_a$ (resp. $m_b/k > \theta_b$).

In the threshold model of cascades with two types of memes, any of the three strategies $\{a, b, ab\}$ may spread globally, and the shares of each strategy in the stationary state, denoted by $\{\rho^s\}$, generally vary depending on the payoff parameters and network structure. This type of spreading process is considered as a multistate dynamical process, for which the approximate master equations (AMEs) method [3, 28] has been used to analytically calculate the dynamical paths and the stationary state [23, 29] (see Sec. S3 for the description of the AME equations). Depending on the inherent attractiveness (i.e., $a$ and $b$), the degree of complementarity $\lambda$ and the mean degree $z$, there are three phases as to which strategy is dominant in the steady state (Figs. 1 and S1). The figures for dominant regions confirm that the AME solutions (shaded) well predict the corresponding simulation results (lines). We find that the cascade region [12] within which we have $1 - \rho^b > 0$ is mostly covered by the combined dominant region (Fig. S2), meaning that a strategy often dominates the others once a global cascade occurs.

While it is natural that the attractiveness parameters $a$ and $b$ explain the differences in popularity between $a$ and $b$, the following question still remains: what happens when the two memes are equally attractive (i.e., $a = b$) yet exclusive? When $a = b$, the predicted shares of $a$ and $b$ obtained by the AMEs are always identical by construction since there is no intrinsic difference between the two (black line in Fig. 2a). Simulated average cascade sizes also closely follow the corresponding theoretical values (circle and cross in Fig. 2a).

It should be noted, however, that the observed match between theory and simulation does not necessarily mean that the AMEs correctly describe the simulated dynamical processes. In fact, symmetry breaking occurs in the propagation of competing memes, in which either of $a$ and $b$ will incidentally dominate the other regardless of the perfect symmetry of their intrinsic attractiveness. We find that there are three phases: i) $\rho^a, \rho^b > 0$ and $\rho^{ab} = 0$, ii) $\rho^a, \rho^{ab} > 0$, and $\rho^b > 0$, and iii) $\rho^a = \rho^b = 0$ and $\rho^{ab} > 0$ (Fig. 2). As we see in Fig. 2a, the stationary values of $\rho^a$ and $\rho^b$ are nearly 0.5 in phase i (i.e., $\lambda < -1$), but this does not indicate that each of the strategies $a$ and $b$ is adopted by 50% of the population. The average values for simulated $\rho^a$ and $\rho^b$ are nearly 0.5 because the chance of $a$ or $b$ being a dominant strategy (i.e., $\rho^a \approx 1$ or $\rho^b \approx 1$) is close to 0.5 (Fig. 2b, left). The popularity of
each meme is either 0% or 100%, exhibiting no diversity in the stationary state.

In phase ii, where \(-1 \lesssim \lambda \lesssim -0.7\), we would have a different set of diffused strategies: \{a, ab\}, \{b, ab\} and \{ab\} (Fig. 2b, middle). The memes are neither too complementary nor too exclusive, and this is the only phase in which a strategy diversification may be observed. In phase iii, where \(\lambda > -0.7\), the two memes are not strongly mutually exclusive, so the only strategy that would be adopted in the stationary state is ab (Fig. 2b, right). This suggests that the intrinsic symmetry between the two types of memes (i.e., \(a = b\)) is maintained only in phase iii while symmetry is likely to be broken in the other phases.

To understand the fundamental mechanics behind the observed symmetry breaking and diversification, we draw phase diagrams based on a mean-field (MF) approximation using random \(z\)-regular networks (i.e., the degree distribution \(p_k = \delta_{k,z}\)), for which it is assumed that the states of neighbors are independent of each other \(\mathbb{F}\). In the MF method, the evolution of \(\rho^s\) for each \(s \in S\) is described by the following differential equation \(\mathbb{F}\): \(\mathbb{F}\):

\[
\dot{\rho}^s = -\sum_{s' \neq s} \rho^s \sum_{m|z=m} M_z(m, \rho) F_m(s \rightarrow s') + \sum_{s' \neq s} \rho^{s'} \sum_{m|z=m} M_z(m, \rho) F_m(s' \rightarrow s),
\]

where \(\rho \equiv (\rho^a, \rho^b, \rho^{ab})^T\), and \(M_z(m, \rho)\) is the multinomial distribution given by

\[
M_z(m, \rho) = \frac{z!}{m_0!m_a!m_b!m_{ab}!} (\rho^a)^{m_0} (\rho^b)^{m_b} (\rho^{ab})^{m_{ab}}.
\]

The first term in Eq. (7) captures the rate at which a node changes its strategy from \(s\) to \(s'\) (\(\neq s\)), and the second term denotes the rate at which a node newly employs strategy \(s\). Note that this is a system of four differential equations (\(|S| = 4\)), but it is sufficient to use three of them because there is an obvious constraint \(\sum_{s \in S} \rho^s = 1\).

Fig. 3 presents phase diagrams in the \(\rho^a-\rho^b\) space for three different values of \(\lambda\), respectively representing the phases i–iii defined above. Note that the theoretical equilibrium (indicated by point A) is saddle-path stable in all the three cases, but the diagrams differ in the size of the region in which \(\rho^{ab} > 0\) (shaded in gray). When the two memes are highly exclusive (\(\lambda = -2\), Fig. 3a), there is no chance for strategy ab to gain popularity, so \(\rho^{ab} = 0\) for any combination of \((\rho^a, \rho^b)\). In simulation with networks of finite size (\(N = 10^4\)), the saddle-path equilibrium indicated by the MF/AME method, \((\rho^a, \rho^b) = (0.5, 0.5)\), is not practically reachable; simulated paths of \((\rho^a, \rho^b)\) converge to \((0, 1)\) or \((1, 0)\) once they deviate from the stable balanced path: \(\rho^a(t) = \rho^b(t)\) for all \(t \geq 0\) (red dotted in Fig. 3b, bottom). In principle, the symmetric MF/AME solution would correspond to a “simulated” equilibrium for infinitely large networks with no structural fluctuations. However, any finite-size networks are generally not free from finite-size effects, so it is not guaranteed that \(\rho^a(t) = \rho^b(t)\) for all \(t \geq 0\).

In phase ii, there arises an area in which \(\rho^{ab} > 0\) (Fig. 3b). This suggests that the feasible region of \((\rho^a, \rho^b)\) (i.e., \((\rho^a, \rho^b) : \rho^a + \rho^b + \rho^{ab} \leq 1\)) gradually shrinks as \(\rho^{ab}\) increases as long as the current state of \((\rho^a, \rho^b)\) is in the gray-shaded area. Both cascades with and without symmetry breaking are observed (Fig. 3b, bottom). In the latter case, both \(\rho^a\) and \(\rho^b\) initially increase and then begin to decrease as the feasible region shrinks in accordance with a rise in \(\rho^{ab}\). In phase iii, we always have...
\( \rho_{ab} > 0 \) (Fig. 3). This indicates that any path of \((\rho^a, \rho^b)\) will move toward the origin at some point in time as the popularity of \(ab\) increases. Therefore, \(s = ab\) will be the only diffused strategy in equilibrium. Fig. 3 shows that the time to reach convergence in a simulated cascade follows a heavy-tailed distribution when symmetry breaking always occurs (i.e., in phase I), while otherwise, a spreading process promptly reaches an equilibrium.

In the existing models of social contagion, it is often assumed that the current response of an individual is determined independently of the history of his/her responses [3, 29]. This suggests that the popularity of each strategy may rise and fall, depending on the local and global states at each point in time. On the other hand, one of the implications provided by the standard binary-state cascade models [1, 2] is that the cascading process is always monotonic; individuals’ states are never be reverted from active to inactive since state reversion is not optimal (see [23] for a proof).

In the model shown above, individuals’ choices are fully reversible where the past strategies do not affect the current optimal choice (Eq. S15), and this is a reason why either of the social memes could dominate the other and polarization does not occur. Such a reversible decision making, however, would be practically infeasible when the switching cost is high (e.g., switching from Mac to Windows, from democrat to republican, etc). To investigate such irreversible dynamics, we introduce a parameter \(q \in [0, 1]\) representing the degree of irreversibility; \(q = 0\) and \(1\) respectively correspond to the fully reversible (i.e., the baseline model) and the fully irreversible cases. Here, “fully irreversible” (i.e., \(q = 1\)) means that only the following five switching patterns are allowed: \(0 \rightarrow a, 0 \rightarrow b, 0 \rightarrow ab, a \rightarrow ab, \) and \(b \rightarrow ab\). Thus, once a meme is accepted, there is no possibility that the meme will be abandoned (i.e., \(a \rightarrow 0, a \rightarrow b, ab \rightarrow b, \) etc). The specification of the response function with irreversibility is explained in section S4.

Let \(G_s\) be a function that represents the right-hand side of the MF equation (7): \(\dot{\rho}^s = G_s(\rho)\). A stable (resp. unstable) equilibrium is defined as an equilibrium at which \(\dot{\rho}^s = 0\) for all \(s \in S\) and the maximum eigenvalue of the Jacobian of vector \(G = (G_0, G_a, G_b, G_{ab})^\top\) is non-positive (resp. positive). We find that introducing a partial irreversibility (i.e., \(q < 1\)) does not change the dynamical process qualitatively; there are still two unstable symmetric equilibria, \((\rho^a, \rho^b) = (0, 0)\) and \((0.5, 0.5)\) (red circles in Fig. 4b), and two stable asymmetric equilibria, \((0, 1)\) and \((1, 0)\) (blue circles in Fig. 4b). When \(q\) is small enough, symmetry breaking always occurs as in the fully reversible case (Fig. 4). Note, however, that the greater the degree of irreversibility \(q\), the longer the time to convergence for \(q < 1\) (Fig. S4).

When the two memes are highly exclusive (such that \(\rho_{ab}(t) = 0\) for all \(t\)) and the strategies are fully irreversible, the saddle equilibrium disappears. Instead, there arises a continuum of stable equilibria such that \(\rho^a + \rho^b = 1\) (Fig. 4b). This indicates that equilibrium is indeterminate in irreversible dynamics, even analytically. Indeed, the simulated equilibria are widely and continuously distributed, representing polarization where \(\rho^a \gg 0, \rho^b \gg 0\) and \(\rho_{ab} \approx 0\) [15, 16] (Fig. 4b). This is intuitive given the constraint that strategies are never reverted once they are adopted; the popularity of a meme
at a given point in time is fully dependent on the past trajectory \(\{\rho^a(t)\}\), while such history dependency is less relevant for reversible cascades since any path will converge to one of the two stable equilibria irrespective of the differences in trajectory (Fig. 4a). Since the possible patterns of strategy updates are limited, the time to convergence is minimized at \(q = 1\) (Fig. S4). We also find that in phases ii and iii, it is more likely for the bilingual option to be dominant as \(q\) rises due to the monotonic nature of irreversible dynamics (Figs. S5 and S6).

We presented a general model of complex social contagion with multiple social memes based on a game-theoretic foundation. While the average popularity of each meme can be well approximated by the AME and MF methods, averaging is not appropriate when the spreading dynamics are described as a saddle path. The equilibrium instability and indeterminacy should be taken into account in understanding the dynamics of real-world social contagions, and I hope that the present work will stimulate further research in this direction.

T. K. acknowledges financial support from JSPS KAKENHI 19H01506 and 20H05633.

[1] D. J. Watts, Proc. Natl. Acad. Sci. USA 99, 5766 (2002).
[2] J. Gleeson and D. Cahalane, Phys. Rev. E 75, 56103 (2007).
[3] J. P. Gleeson, Physical Review X 3, 021004 (2013).
[4] A. Nematzadeh, E. Ferrara, A. Flammini, and Y.-Y. Ahn, Phys. Rev. Lett. 113, 088701 (2014).
[5] C. D. Brummitt and T. Kobayashi, Phys. Rev. E 91, 062813 (2015).
[6] T. Kobayashi, Phys. Rev. E 92, 062823 (2015).
[7] L. Böttcher, J. Nagler, and H. J. Herrmann, Phys. Rev. Lett. 118, 088301 (2017).
[8] J. P. Gleeson and M. A. Porter, in Complex Spreading Phenomena in Social Systems (Springer, 2018) pp. 81–95.
[9] S. Unicomb, G. J. Iñiguez, J. Kertész, and M. Karsai, Phys. Rev. E 100, 040301 (2019).
[10] M. A. Cusumano, Y. Mylonadis, and R. S. Rosenbloom, Bus. Hist. Rev. 66, 51 (1992).
[11] N. Anscombe, Nat. Photonics 2, 412 (2008).
[12] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, EPJ Data Sci. 1, 1 (2012).
[13] P. T. Metaxas and E. Mustafaraj, Science 338, 472 (2012).
[14] E. Ferrara, Information Sciences 418, 1 (2017).
[15] V. V. Vasconcelos, S. A. Levin, and F. L. Pinheiro, J. R. Soc. Interface 16, 20190196 (2019).
[16] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Phys. Rev. Lett. 124, 048301 (2020).
[17] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, Proc. Natl. Acad. Sci. 118 (2021).
[18] N. F. Johnson, N. Velásquez, N. J. Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, and Y. Lupu, Nature 582, 230 (2020).
[19] R. Prieto Curiel and H. González Ramírez, Sci. Rep. 11, 1 (2021).
[20] P. Adepoju, Nat. Medicine 27, 1122 (2021).
[21] S. Morris, Rev. Econ. Stud. 67, 57 (2000).
[22] M. O. Jackson and L. Yariv, Phys. Rev. Lett. 124, 048301 (2020).
[23] M. O. Jackson and Y. Zenou, in Handbook of Game Theory with Economic Applications, Vol. 4 (Elsevier, 2015) pp. 95–163.
[28] J. P. Gleeson, Phys. Rev. Lett. 107, 068701 (2011).
[29] P. G. Fennell and J. P. Gleeson, SIAM Rev. 61, 92 (2019).
[30] P. G. Fennell, https://github.com/peterfennell/multi-state-SOLVER
Supplemental Material

“Kinetics of competing social contagions: Symmetry breaking and equilibrium indeterminacy”

Teruyoshi Kobayashi

S1. OPTIMAL CHOICE OF STRATEGY

Based on the payoffs of each strategy [1–4], an individual optimally selects a strategy \( s \) if \( v(s, m) > v(s', m) \) for all \( s' \neq s \).

(i) \( s^* = a \) if \( v_a > v_0, v_a > v_b, \) and \( v_a > v_{ab} \):

\[
-ck + a(m_a + m_{ab}) > 0, \\
-ck + a(m_a + m_{ab}) > -ck + b(m_a + m_{ab}), \\
-ck + a(m_a + m_{ab}) > -2\tilde{c}k + a(m_a + m_{ab}) + b(m_b + m_{ab}),
\]

where \( v \) is shorthand for \( v(s, m) \). In the same manner, we have the following conditions for \( s^* = b \) and \( ab \):

(ii) \( s^* = b \) if \( v_b > v_0, v_b > v_a, \) and \( v_b > v_{ab} \):

\[
-ck + b(m_b + m_{ab}) > 0, \\
-ck + b(m_b + m_{ab}) > -ck + a(m_a + m_{ab}), \\
-ck + b(m_b + m_{ab}) > -2\tilde{c}k + a(m_a + m_{ab}) + b(m_b + m_{ab}),
\]

(iii) \( s^* = ab \) if \( v_{ab} > v_0, v_{ab} > v_a, \) and \( v_{ab} > v_b \):

\[
-2\tilde{c}k + a(m_a + m_{ab}) + b(m_b + m_{ab}) > 0, \\
-2\tilde{c}k + a(m_a + m_{ab}) + b(m_b + m_{ab}) > -ck + a(m_a + m_{ab}), \\
-2\tilde{c}k + a(m_a + m_{ab}) + b(m_b + m_{ab}) > -ck + b(m_b + m_{ab}).
\]

By employing the definitions \( M_a \equiv m_a + m_{ab}, \ M_b \equiv m_b + m_{ab}, \lambda \equiv 2(1 - \tilde{c}/c), \theta_a \equiv c/a \) and \( \theta_b \equiv c/b \), we can immediately rewrite the above inequalities as the threshold conditions shown in Eq. (6). When there are “tie” strategies (i.e., \( v_s = v_{s'}, \) for \( s \neq s' \)), we randomly select a strategy among the tie strategies.

S2. CONSTRAINTS FOR \( \lambda \)

Since we focus on a situation in which the pure strategy Nash equilibria for each bilateral game are given by \((0, 0), (a, a), (b, b)\) and \((ab, ab)\), the payoff of strategy \( s \) must be the highest if the opponent’s strategy is \( s \). We have the following conditions for each of these strategy pairs to be attained as a Nash equilibrium:

(i) For the strategy pair \((0, 0)\) to be a Nash equilibrium, we need to have \(-2\tilde{c} < 0. \) Since \( \lambda = 2(1 - \tilde{c}/c) \), it indicates that

\[
\lambda < 2. \tag{S10}
\]

(ii) For the strategy pair \((a, a)\) to be a Nash equilibrium, we need to have \( a - c > a - 2\tilde{c}. \) It follows that

\[
\lambda < 1. \tag{S11}
\]

Note that the condition for the pair \((b, b)\) is the same.
(iii) For the strategy pair \((ab, ab)\) to be a Nash equilibrium, we need to have \(a + b - 2c > a - c\) and \(a + b - 2c > b - c\) (Recall that \(a - c > 0\) and \(b - c > 0\)). It follows that

\[
(1 - \lambda)\theta_a < 1 \quad \text{and} \quad (1 - \lambda)\theta_b < 1.
\]

Given the conditions \([S10] - [S12]\), \(\lambda\) must satisfy \(\lambda < 1\), \((1 - \lambda)\theta_a < 1\), and \((1 - \lambda)\theta_b < 1\).

**S3. AME EQUATIONS**

Here, we describe the spreading process of competing memes based on the AME method. Let \(\rho^*_{k,m}\) denote the fraction of \(k\)-degree nodes belonging to the \((s,m)\) class (i.e., \(k\)-degree nodes adopting strategy \(s\) and facing the neighbor profile \(m\)). Using the AME formalism, the evolution of \(\rho^*_{k,m}\) is given by \([3, 28, 29]\):

\[
\dot{\rho}^*_{k,m} = - \sum_{s' \neq s} F_m(s \to s') \rho^*_{k,m} - \sum_{r \in S, r' \neq r} m_r \phi_s(r \to r') \rho^*_{k,m} + \sum_{s' \neq s} F_m(s' \to s) \rho^*_{k,m} + \sum_{r \in S, r' \neq r} (m_{r'} + 1) \phi_s(r' \to r) \rho^*_{k,m - e_r + e_{r'}}.
\]

(S13)

for \(s \in S\), where \(\phi_s(r \to r')\) denotes the probability that a neighbor of a node adopting strategy \(s\) changes its strategy from \(r\) to \(r'\):

\[
\phi_s(r \to r') = \frac{\sum_k p_k \sum_{|m| = k} m_s \rho^*_{k,m} F_m(r \to r')}{\sum_k p_k \sum_{|m| = k} m_s \rho^*_{k,m}}.
\]

(S14)

\(p_k\) denotes the degree distribution, and the response function \(F_m(s \to s')\) describes the rate at which individuals change their strategy from \(s\) to \(s'\) for a given neighbors’ profile \(m\):

\[
F_m(s \to s') = \begin{cases} 1 & \text{if } s' = s^*(m), \\ 0 & \text{otherwise}, \end{cases}
\]

(S15)

where \(s^*(m)\) is the optimal strategy defined by Eq. \([5]\). The expected fraction of individuals adopting strategy \(s \in S\) leads to \(\rho^* = \sum_k p_k \sum_{|m| = k} \rho^*_{k,m}\), where \(\sum_{|m| = k}\) denotes the sum over all combinations of \(\{m_s\}\) such that \(\sum_{s \in S} m_s = k\).

There are four factors that change \(\rho^*_{k,m}\) over time in Eq. \([S13]\). Individuals will *leave* the \((s,m)\) class if \(i)\) their strategy changes from \(s\) to \(s'(\neq s)\) (the first term) or \(ii)\) their neighbor profile changes from \(m\) to \(m'(\neq m)\) (the second term). Individuals will *enter* the \((s,m)\) class if \(iii)\) their strategies newly change from \(s'(\neq s)\) to \(s\) (the third term) or \(iv)\) the neighbors’ profile shifts from \(m'(\neq m)\) to \(m\) (the fourth term). The expression \(m - e_r + e_{r'}\) in the fourth term denotes the neighbor profile that has \(m_{r'} + 1\) in the \(r'\)-th element and \(m_r - 1\) in the \(r\)-th element.

The denominator of Eq. \([S14]\), \(\sum_k p_k \sum_{|m| = k} m_s \rho^*_{k,m}\), represents the expected number of \((s)-(r)\) edges. Since the expected number of \((s)-(r)\) edges that shift to \((s)-(r')\) in an infinitesimal interval \(dt\) is given as \(\sum_k p_k \sum_{|m| = k} m_s \rho^*_{k,m} F_m(r \to r')dt\), the probability of a \((s)-(r)\) edge shifting to a \((s)-(r')\) edge, denoted by \(\phi_s(r \to r')dt\), is obtained as the ratio of the two, leading to Eq. \([S14]\). The AME solution is calculated using Matlab codes provided in \([30]\).
S4. RESPONSE FUNCTION WITH IRREVERSIBILITY

The irreversibility parameter \( q \in [0, 1] \) denotes the rate at which a strategy will not be reverted. For each combination of \((s, s')\), the response function with irreversibility constraints, denoted by \( \tilde{F}_m(s \to s') \), is given in Table S1.

| \( s' \) | \( s = 0 \) | \( s = a \) | \( s = b \) | \( s = ab \) |
|-------|---------|---------|---------|---------|
| 0     | \( F_m(0 \to 0) \) | \( F_m(0 \to a) \) | \( F_m(0 \to b) \) | \( F_m(0 \to ab) \) |
| a     | \( (1 - q)F_m(a \to 0) \) | \( 1 - \sum_{s \neq a} \tilde{F}_m(a \to s) \) | \( (1 - q)F_m(a \to b) \) | \( F_m(a \to ab) \) |
| b     | \( (1 - q)F_m(b \to 0) \) | \( (1 - q)F_m(b \to a) \) | \( 1 - \sum_{s \neq b} \tilde{F}_m(b \to s) \) | \( F_m(b \to ab) \) |
| ab    | \( (1 - q)F_m(ab \to 0) \) | \( (1 - q)F_m(ab \to a) \) | \( (1 - q)F_m(ab \to b) \) | \( 1 - \sum_{s \neq ab} \tilde{F}_m(ab \to s) \) |

For nodes with \( s = 0 \), there is no constraint in updating their strategy. For nodes with \( s = a \) (resp. \( s = b \)), shifting to \( s' = b \) (resp. \( s' = a \)) or \( s' = 0 \) is restricted, for which the transition probability is multiplied by a factor of \((1 - q)\). For nodes with \( s = ab \), any state change is restricted. Note that the unconstrained response function is recovered if \( q = 0 \), while the response is fully irreversible if \( q = 1 \).
FIG. S1. Dominant strategy in the stationary state. See the caption of Fig. 1 for details.

FIG. S2. Theoretical and simulated cascade region. (a) AME solution and (b) simulation. Color indicates the value of $\rho^a + \rho^b + \rho^{ab} = 1 - \rho^0$, which will be significantly larger than $\rho^a(0) + \rho^b(0) + \rho^{ab}(0) = 0.06$ if a global cascade occurs. See the caption of Fig. 1 for the parameter values.
FIG. S3. Distribution of the time to convergence. (a) Dominant regions as in Fig. 1b. A symbol denotes a point at which a complementary cumulative distribution function (CCDF) of convergence time is generated in panel b. (b) CCDF of the time to convergence. Each symbol denotes a particular parameter combination indicated in panel a. When $\theta_a/\theta_b = 1$ and $\lambda = -1.5$ (i.e., in phase i), at which symmetry breaking always occurs, simulated time to convergence follows a heavy-tailed distribution. We conduct 10,000 simulations on Erdős-Rényi networks with $z = 4$ and discard simulation runs that did not reach convergence by $t = 10,000$. See the caption of Fig. 4 for the other parameter values.

FIG. S4. Time to convergence and the degree of irreversibility $q$. Error bar denotes one standard deviation while circle denotes the average. See the caption of Fig. 4 for the details of simulation.
(a) Reversible dynamics, $q = 0, \lambda = -0.9$

(b) Irreversible dynamics, $q = 1, \lambda = -0.9$

FIG. S5. Phase diagrams of (a) reversible and (b) irreversible dynamics for $\lambda = -0.9$. See Fig. 3 for a detailed description of the phase diagrams. $q$ denotes the extent to which a strategy is irreversible (i.e., $q = 0$ and $q = 1$ represent fully reversible and irreversible cases, respectively). Red cross denotes the MF solution. In panel (b), due to the presence of irreversibility constraints, the popularity of ab increases faster than in the case of $q = 0$, which shrinks the feasible region of $(\rho^a, \rho^b)$ faster along with it.
FIG. S6. Phase diagrams of (a) reversible and (b) irreversible dynamics for $\lambda = 0$. See Figs. 3 and S5 for the description of the phase diagrams.