

Balanced-imbalanced transitions in indirect reciprocity dynamics on networksKoji Oishi ^{*}*Department of International Politics, Aoyama Gakuin University, Tokyo, Japan
and Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-8471, Japan*Shuhei Miyano *Department of Applied Physics, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo,
Bunkyo-ku, Tokyo 113-8656, Japan*Kimmo Kaski *Department of Computer Science, Aalto University School of Science, P.O. Box 15500, Espoo, Finland
and The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom*Takashi Shimada [†]*Mathematics and Informatics Center and Department of Systems Innovation, Graduate School of Engineering, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

(Received 21 April 2021; accepted 28 July 2021; published 12 August 2021)

Here we investigate the dynamics of indirect reciprocity on networks, a type of social dynamics in which the attitude of individuals, either cooperative or antagonistic, toward other individuals changes over time based upon their actions and mutual monitoring. We observe an absorbing state phase transition as we change the network's link or edge density. When the edge density is either small or large enough, opinions quickly reach an absorbing state, from which opinions never change anymore once reached. In contrast, if the edge density is in the middle range, the absorbing state is not reached and the state keeps changing, thus being active. The result shows an effect of social networks on spontaneous group formation.

DOI: [10.1103/PhysRevE.104.024310](https://doi.org/10.1103/PhysRevE.104.024310)**I. INTRODUCTION**

Social relations are characterized by being of either a cooperative (friendly) or an antagonistic (unfriendly) nature. However, it can also happen that friends become foes or foes become friends [1], i.e., an individual network link can turn from positive to negative or negative to positive. Hence understanding the temporal evolution of such behavior-changing processes is essential for getting deeper insight into the functions of real social networks.

To gain such insight, agent-based modeling approaches have turned out to be versatile. One of the earlier models on how people change their attitude towards others was based on indirect reciprocity [2,3], which is commonly observed in human behavior and is closely related to the evolution of cooperation in human society [4]. In this model, people change their attitude (of either liking or disliking) through their action and observation of the actions of others. A typical example of such dynamics (action rules and norms) is as follows: First, people are cooperating with only those they like. Second, they get to like those who cooperate with those whom they like and do not cooperate with those whom they dislike.

In spite of the importance of indirect reciprocity, the dynamics of social networks induced by indirect reciprocity has not yet been fully understood. Previous studies have focused

on the case of fully connected networks (i.e., systems in which all people know each other) [5,6]. It was found that indirect reciprocity can result in a split of the society into clusters, only within which people are cooperative. However, the real social networks are usually not fully connected. Therefore, we focus our attention on the dynamics of indirect reciprocity in such networks of agents, and we examine whether agents split into fixed clusters as in the case of fully connected networks. In this study, among several variations of the indirect reciprocity dynamics, we focus on the Kandori assessment rule [7]. Comparing to another type of indirect reciprocity model in which an agent can get to like those who cooperate with those who he dislikes, the Kandori rule is said to be more strict because an agent dislikes the people in the same case. And this rule is suggested to be one of the most efficient and robust rules to promote cooperation [3,7].

This paper is organized such that after this Introduction, in Sec. II we describe the model of indirect reciprocity. Then we first present the simulation results in Sec. III and then develop the mean-field analysis for the model in Sec. IV. Finally, in Sec. V we summarize the results and discuss their implications.

II. THE MODEL

Let us consider a nondirected network of agents $G = (V, E)$, where $V = 1, \dots, N$ is the set of agents or network nodes, and E is the set of links or network edges $e_{ij} \in E$ between agents i and j , meaning that they know each other. Here

^{*}oishi@sipac.aoyama.ac.jp[†]shimada@sys.t.u-tokyo.ac.jp

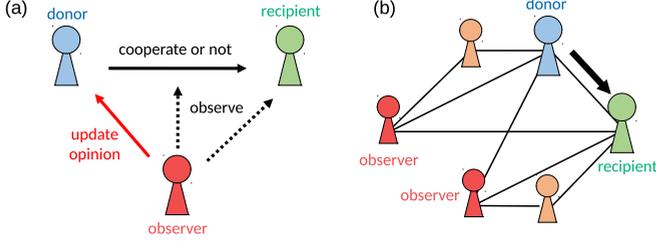


FIG. 1. Schematics of the model. (a) Donor, recipient, and a third-party observer. (b) Observers in a network. Agents not labeled either donor, recipient, or observers do not observe the interaction between the donor and recipient.

the structure of the network is assumed not to change in time, i.e., being static, while the agents have their opinion about their neighbors and change it in time, i.e., being dynamic. We denote the opinion of the agent i about the agent j by σ_{ij} . If the agent i likes the agent j at time step t , then $\sigma_{ij}(t) = 1$, but if the agent i dislikes j , then $\sigma_{ij}(t) = -1$. The opinions do not need to be reciprocal, so that i may dislike j even if j likes i .

The time evolution of the model is set in such a way that at each time step, two neighboring agents are randomly chosen, one as the donor and the other as the recipient. The donor cooperates with the recipient if the donor likes the recipient, while the donor does not cooperate if the donor dislikes the recipient. The action of the donor, either cooperating or not cooperating, is observed by the donor (him- or herself), the recipient, and the common neighbors of the donor and the recipient, i.e., the third party called the observer, as depicted in Fig. 1(a). Each observer of the donor-recipient pair updates her/his opinion about the donor according to a given rule, which is called the assessment rule.

In this study, we adopt the Kandori assessment rule [7]: observers get to like the donor if the observers like the recipient and the donor cooperates with the recipient or when observers dislike the recipient and the donor does not cooperate, while the observers get to dislike the donor otherwise. Therefore, the opinion of an observer k about the donor is updated as follows:

$$\sigma_{kd}(t+1) = \sigma_{kr}(t)\sigma_{dr}(t), \quad (1)$$

where d and r denote the donor and the recipient at time step t , respectively. Agents other than the donor, recipient, and the observers do not update their opinions [Fig. 1(b)]. The agents' initial opinions are drawn independently at random from an even distribution of opinions, i.e., $\{-1, +1\}$, where -1 corresponds to an antagonistic or unfriendly opinion and $+1$ to a cooperative or friendly opinion (i.e., liking or disliking, respectively).

III. NUMERICAL RESULTS

In this section, we investigate the dynamics of indirect reciprocity based on the Kandori assessment rule on Erdős-Rényi random graphs, where each agent is linked to another agent with the probability p . In the following, the ensemble averages are taken over independently generated networks. Note that the number of links can be different across samples even with the same probability p , due to stochastic fluctuations.

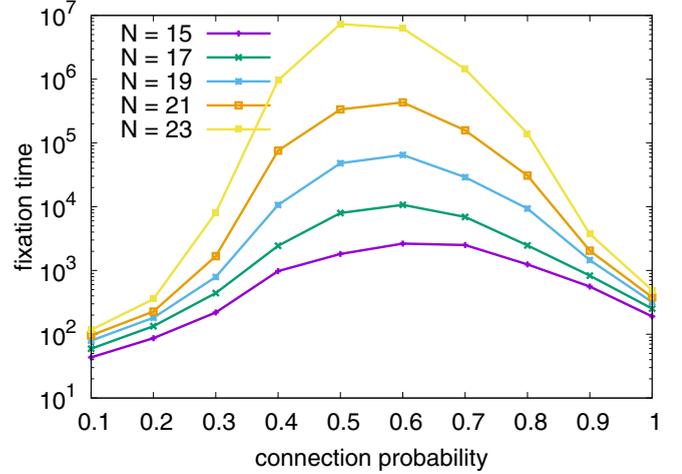


FIG. 2. Average fixation time as a function of the connection probability for five different system sizes (N) and for 100 independent runs, each up to 10^9 time steps. All these 100 samples reached absorbing states, after which we took the sample average of the fixation time.

First we focus on the question of whether the opinions of agents eventually become fixed, as was found in the case of fully connected networks [5]. Following the update rule [i.e., Eq. (1)], the condition that opinions do not change anymore is

$$\sigma_{kd} = \sigma_{kr}\sigma_{dr} \quad (2)$$

for all the network edges $e_{dr} \in E$ and all the observers k , i.e., d , r , and common neighbors of d and r . This condition is equivalent to

$$\sigma_{ii} = 1 \quad \text{for all } i \in V, \quad (3)$$

$$\Theta_{ij} \equiv \sigma_{ij}\sigma_{ji} = 1 \quad \text{for all } e_{ij} \in E, \quad (4)$$

$$\Phi_{ijk} \equiv \sigma_{ij}\sigma_{ik}\sigma_{jk} = 1 \quad \text{for all triads } (i, j, k), \quad (5)$$

where Θ_{ij} and Φ_{ijk} stand for the edge balance and triad balance, respectively. The first condition [Eq. (3)], meaning everyone regards her/himself well, is quickly fulfilled. The second condition [Eq. (4)] means that the mutually connected agents have the same opinion as each other, i.e., like or dislike. The third condition [Eq. (5)] asks if all the triads are balanced in the sense that friends of your friends and enemies of your enemies are your friends.

Because the system cannot move anymore after reaching a configuration that fulfills these conditions, we call all such configurations an absorbing state. And because the system always has an absorbing state (e.g., $\forall i, j \sigma_{ij} = 1$), the question is whether and when the system goes to the absorbing state. In Fig. 2 we depict the average fixation time, which is defined as the number of time steps taken until Eqs. (3)–(5) are satisfied. When the connection probability p is high or low, the fixation time is relatively short, while the time for the opinions getting fixed or fixation time turns out to be much longer with p in the middle range and it rapidly increases even with a small increase of the system size N . This means that if the network is not fully connected and the system size or the number of

agents is moderate, i.e., $N > 100$, their opinion cannot split into fixed clusters in a realistic timescale.

The next question is how the opinion of an agent evolves before it reaches the absorbing state. To investigate the time development, we introduce an order parameter imbalance \mathcal{I} as

$$\mathcal{I} = \begin{cases} 1 - \frac{\sum_{i,j,k} \Phi_{ijk}}{6N_{\text{triad}}} & (N_{\text{triad}} > 0), \\ 0 & (N_{\text{triad}} = 0), \end{cases} \quad (6)$$

where N_{triad} is the number of agent triads.

As we show here, with some reasonable assumptions, the imbalance is equal to 0 in the absorbing states and it is expected to be 1 in the case of random opinions. Therefore, the purpose of this order parameter is to evaluate how far the system is from the absorbing state and how near to the random state. We note that there might be other quantities with similar properties to the imbalance, to serve as an order parameter. Let us examine Eq. (1) to see how the order parameter works. First, the update of the self-image of the donor at time step t is as follows:

$$\sigma_{id}(t+1) = [\sigma_{dr}(t)]^2 = 1. \quad (7)$$

It means that the agents like themselves and never change their opinion once they take an action as a donor. Therefore, the first condition [Eqs. (3)] is quickly satisfied for all the agents. Next, the update of the opinion of the recipient on the donor is

$$\sigma_{rd}(t+1) = \sigma_{rr}(t)\sigma_{dr}(t). \quad (8)$$

After some transient time, i.e., $t \gg 1$, where $\sigma_{rr} = 1$ as we have just discussed above, it reduces to

$$\sigma_{rd}(t+1) = \sigma_{dr}(t). \quad (9)$$

Then if the edge e_{ij} does not have any common neighbors, once i takes an action to j or vice versa, then $\sigma_{ij} = \sigma_{ji}$ is realized and they never change their opinion. On the other hand, if the edge e_{ij} has common neighbors, the relation $\sigma_{ij} = \sigma_{ji}$ is not always maintained in the long run, because σ_{ij} may change when the agent i observes the agent j 's action to a common neighbor. Therefore, a nontrivial condition for the absorbing states is

$$\sigma_{ij} = \sigma_{ji} \quad (10)$$

for all the edges e_{ij} having common neighbors. Finally, the update of the opinion of common neighbors cannot reduce from Eq. (1), therefore another nontrivial condition for a fixed state is

$$\sigma_{ij}\sigma_{ki}\sigma_{kj} = 1. \quad (11)$$

Summing up these arguments, $\mathcal{I} = 0$ is equivalent to the condition for the opinion being in a fixed state, under the assumption $\sigma_{ii} = 1$ and $\sigma_{ij} = \sigma_{ji}$ for all the edges e_{ij} not having common neighbors, which are quickly satisfied.

In Fig. 3 we show the time development of the imbalance for the network of size $N = 100$. For the connection probability $p < 0.01$ or $p > 0.99$, the imbalance goes quickly down to zero. In contrast, the imbalance remains positive if $0.01 < p < 0.99$. This means that in the middle range values of the edge density, the system relaxes to a nonabsorbing stationary state. In these nonabsorbing stationary states, the

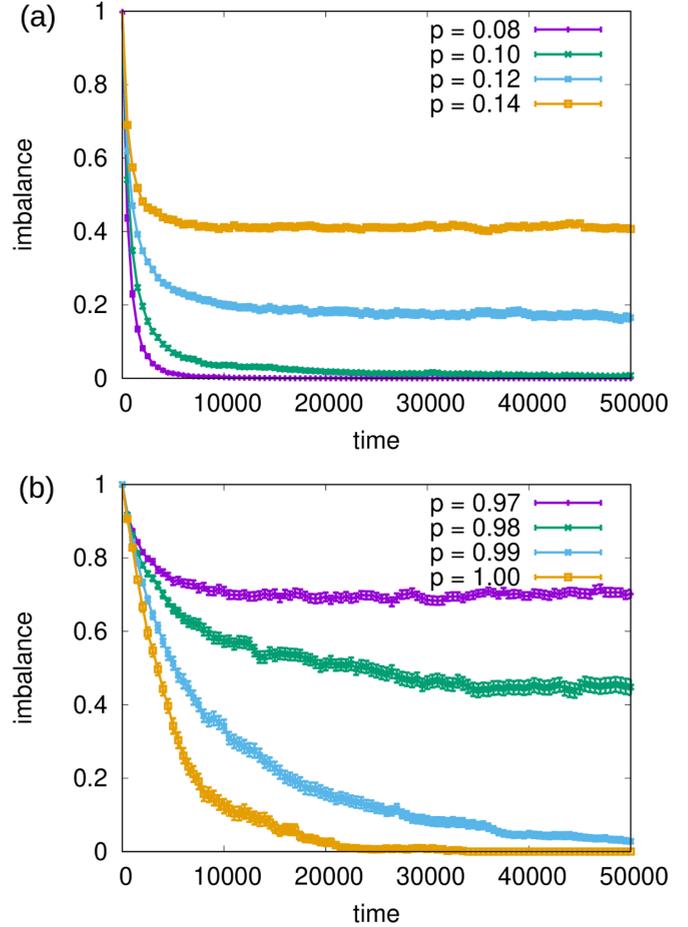


FIG. 3. Time development of imbalance, averaged over 100 samples, in the case of (a) sparse and (b) dense random networks with $N = 100$ agents. Error bars represent standard error.

imbalance fluctuates around a finite value. This observation implies that even when the action in one time step increases the number of balanced triads, the newly balanced triads make other balanced triads more likely to become imbalanced. With the current dynamics of our model, we have not observed the system reaching jammed states, which some related models are known to have [12]. Therefore, we do not exclude the possibility that the nonabsorbing states include jammed states, from which absorbing states are not reachable.

We also investigate the effect of the system size N on the time development of imbalance. When the network is sparse [Fig. 4(a)], the time-series of imbalance \mathcal{I} scales as $\mathcal{I}(t, p, N) \sim \mathcal{I}(t/N^2, pN^2, 1)$, but when it is dense [Fig. 4(b)] it scales as $\mathcal{I}(t, 1-p, N) \sim \mathcal{I}(t/N^2, (1-p)N, 1)$. In both cases, the time development becomes N^2 times slower when the system size gets larger.

As the next step, we examine more closely how the stationary imbalance depends on the edge density. In Fig. 5 we show the time-average and temporal fluctuation of imbalance after allowing the network of agents to run for $5 \times N^2$ time steps of relaxation. Note that $5 \times N^2$ time steps is not enough for fixation to absorbing states even for much smaller network sizes (Fig. 2), but it is enough for the system to relax to stationary states (Fig. 4). We have also confirmed that

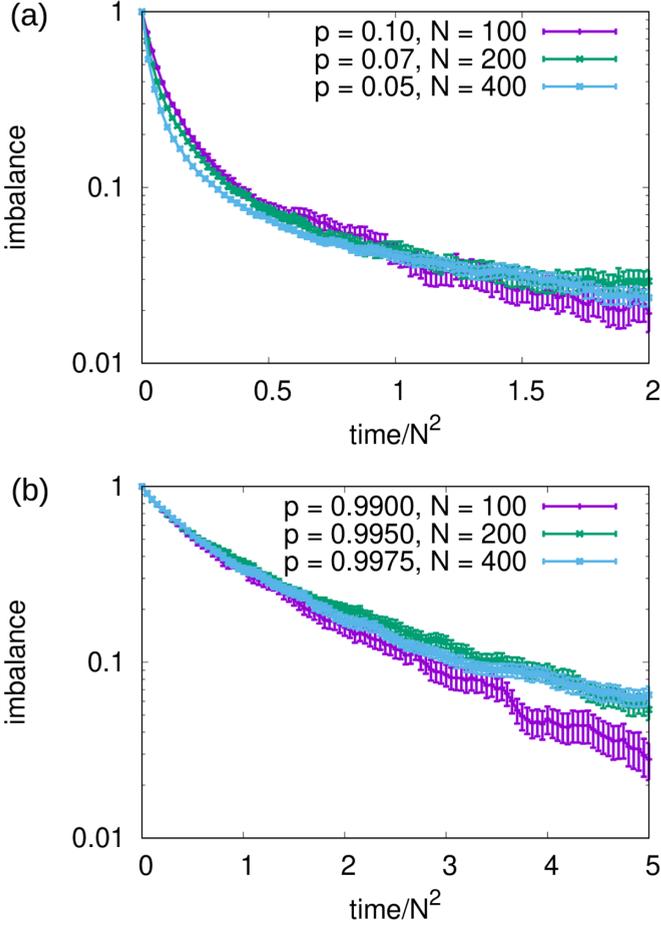


FIG. 4. Imbalance as a function of time. The effect of the system size on the time development. The imbalance is averaged over 100 samples. Time-series for different p and N are compared with (a) $pN^2 = 1$ and (b) $(1-p)N = 1$.

the increase of simulation time 10–30 times longer does not change the following results qualitatively. We observe that the stationary imbalances for the different system sizes are scaled well by Np^2 in the sparsely connected networks and by $N(1-p)$ in the densely connected networks, as depicted in Figs. 5(c) and 5(e). When $Np^2 < C_L \sim O(1)$ or $N(1-p) > C_U \sim O(1)$, i.e., when the network is quite sparse or quite dense, the stationary imbalance is equal to zero and the system reaches absorbing states. While both of them rapidly increase, the system remains more random with larger fluctuation as $Np^2(1-p)$ exceeds 1. The temporal fluctuations show peaks at around the same scaled boundary area. This means that the systems remain far from absorbing states for the moderate edge density, while for more sparse or dense edge densities they show large fluctuations. However, when the edge density is quite sparse or dense and p exceeds certain values, the system reaches absorbing states.

These observations indicate that the edge density p causes transitions between the absorbing phase in which the system goes to the absorbing state and the active phase in which the system does not relax to the absorbing state [8,9]. Note that, as N increases, the absorbing regions of p (basins) get narrower: lower transition point $p_* \sim N^{-1/2}$ and higher transition point

$p^* \sim 1 - 1/N$ (Fig. 6). Furthermore, a more detailed analysis of the phase transition behavior with the order of the transition, transition point, and possible critical exponents is beyond the scope of this study. Instead, we will next focus on the mean-field analysis.

IV. MEAN-FIELD ANALYSIS

In this section, we show that the transitions between the absorbing and active phases at $Np^2(1-p) \sim 1$ are consistent with a mean-field approximation for the indirect reciprocity dynamics. For our analysis, we introduce tetrad balance (Fig. 7) for each tetrad (four-node clique) as

$$\Psi_{ijkl} \equiv \sigma_{ik}\sigma_{il}\sigma_{jk}\sigma_{jl} = \Phi_{ikl}\Phi_{jkl}, \quad (12)$$

in addition to the edge balance Θ_{ij} and triad balance Φ_{ijk} already defined in Eqs. (4) and (5). Note that the tetrad balance is by definition invariant under the exchange of the first two suffixes and the latter suffix pair:

$$\Psi_{ijkl} = \Psi_{jikl} = \Psi_{jilk}. \quad (13)$$

For the mean-field analysis, we use the averages of these quantities:

$$\Theta \equiv \langle \Theta_{ij} \rangle = \frac{\sum_{i>j} \Theta_{ij}}{N_{\text{edge}}}, \quad (14)$$

$$\Phi \equiv \langle \Phi_{ijk} \rangle = \frac{\sum_{i,j,k} \Phi_{ijk}}{6N_{\text{triad}}} \quad (= 1 - \mathcal{I}), \quad (15)$$

$$\Psi \equiv \langle \Psi_{ijkl} \rangle = \frac{\sum_{i>j} \sum_{k>l} \Psi_{ijkl}}{6N_{\text{tetrad}}}, \quad (16)$$

as the order parameters of indirect reciprocity dynamics, where N_{edge} , N_{triad} , and N_{quad} are the number of edges, triads, and tetrads in the system.

As shown in the Appendixes, a mean-field approximation on the indirect reciprocity dynamics yields the dynamical equations of the order parameters in the following closed form:

$$\frac{d\Theta}{dt} = \frac{1}{N_{\text{edge}}} [(1 - \Theta) + T(\Phi - \Theta)], \quad (17)$$

$$\frac{d\Phi}{dt} = \frac{1}{6N_{\text{triad}}} [T\{2(1 - \Phi) + (\Theta - \Phi)\}] + \frac{1}{6N_{\text{triad}}} [Q(\Psi - \Phi) + R(\Phi^2 - \Phi)], \quad (18)$$

$$\frac{d\Psi}{dt} = \frac{1}{6N_{\text{tetrad}}} R\{1 - \Psi + 2(\Phi - \Psi)\} + \frac{1}{6N_{\text{tetrad}}} 2(S_1 + S_2)\Phi(\Phi - 1). \quad (19)$$

Here T , Q , R , S_1 , and S_2 are the numbers of subgraphs on which the balance quantities are affected by an interaction between a pair of donors d and r , i.e., T for triads of d , r , and a third-party observer o [Fig. 8(a)]; Q for tetrads of d , r , and third-party observers o and p [Fig. 8(b)]; R for trusses of d , r , and a third-party observer o , and a nonobserving neighbor n [Fig. 8(c)]; S_1 for tetrads of d , a third-party observer o , and two nonobserving neighbor n and m [Fig. 8(d)]; and S_2 for tetrads of d , two third-party observer o and p , and a nonobserving

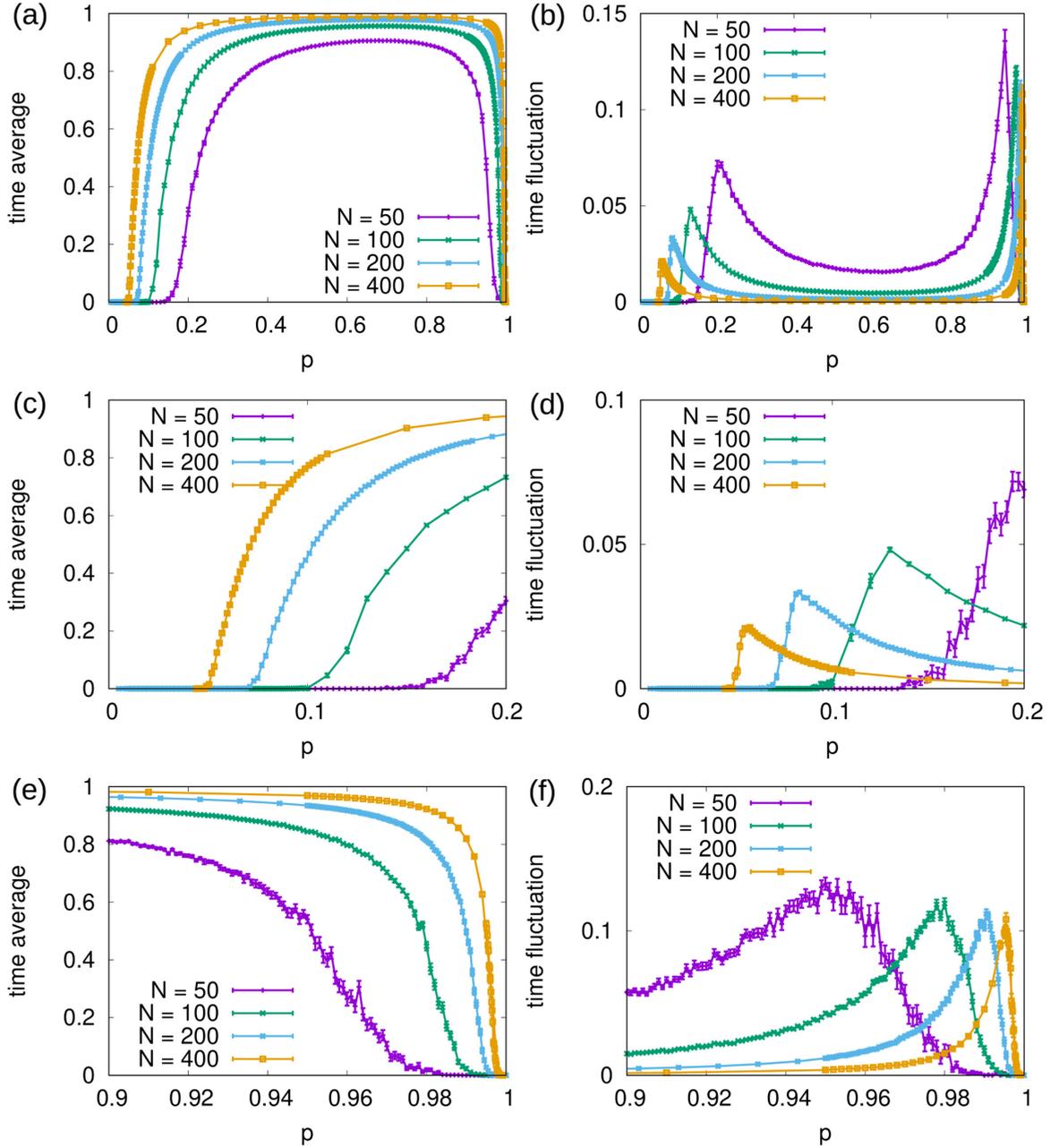


FIG. 5. (a) Time-average and (b) temporal fluctuation of stationary imbalance on random networks. Time-average and temporal fluctuation are calculated over 100 time points for $5 \times N^2$ to $10 \times N^2$ time steps. Lines represent sample averages, and error bars represent standard error over 100 different time-series for $N \leq 200$ and 50 time-series for $N = 400$. Panels (c), (e) and (d), (f) show the zoomed-in parts of (a) and (b) around the lower and upper transition points, respectively.

neighbor n [Fig. 8(e)]. A truss is a four-node graph in which all but one pair of nodes are connected. A nonobserving neighbor for a donor-recipient pair is a node that is connected to the recipient and to a third-party observer but not to the donor. These equations tell that nonobserving neighbors are driving the system toward imbalance as subgraphs with nonobserving neighbors [Figs. 8(c)–8(e)] always decrease triad and tetrad balance, i.e., the terms for R , S_1 , and S_2 in Eqs. (18) and (19) are always negative. This is because nonobserving neighbors do not change their opinion while other observers may change their opinion about the donor, which is on average likely to make their triads or tetrads imbalanced.

For our present case of an Erdős-Rényi random graph with $N \gg 1$, the expected numbers of the subgraphs of a donor-recipient pair are

$$\begin{aligned}
 T &\simeq Np^2, \\
 Q &\simeq \frac{1}{2}N^2p^5, \\
 R &\simeq \frac{1}{2}N^2p^4(1-p), \\
 S_1 &\simeq \frac{1}{2}N^3p^7(1-p)^2, \\
 S_2 &\simeq \frac{1}{2}N^3p^8(1-p).
 \end{aligned} \tag{20}$$

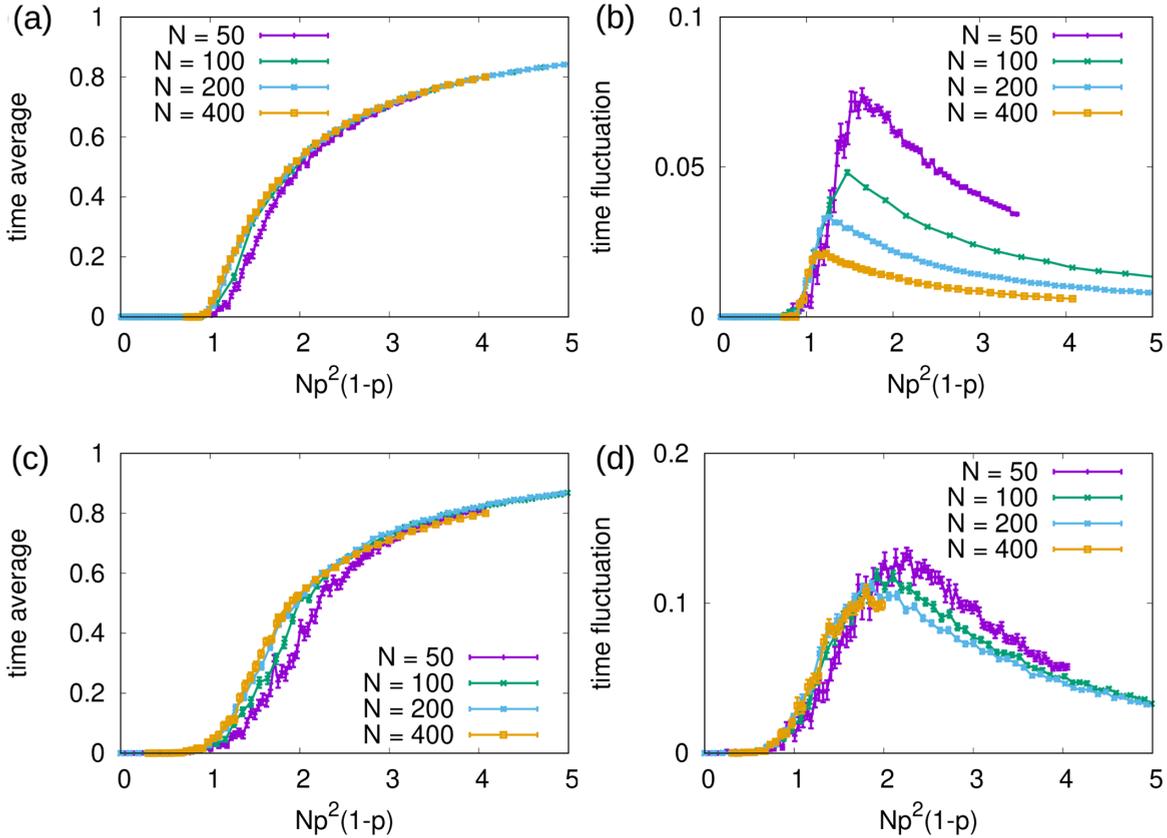


FIG. 6. Dependency of stationary imbalance on $Np^2(1-p)$. (a) Time-average and (b) temporal fluctuation of stationary imbalance in the sparse edge density regions, $Np^2 \sim O(1)$. As for Fig. 5, time-average and temporal fluctuation are calculated over 100 time points for $5 \times N^2$ to $10 \times N^2$ time steps, and lines represent sample averages and error bars represent standard error for 100 time-series for $N \leq 200$ and 100 time-series for $N = 400$. (c) Time-average and (d) temporal fluctuation of stationary imbalance in the dense edge region, $N(1-p) \sim O(1)$.

By solving $\frac{d\Theta}{dt} = \frac{d\Phi}{dt} = \frac{d\Psi}{dt} = 0$, the condition for $\Theta = \Phi = \Psi = 1$ to be the only stable fixed point of Eqs. (17), (18), and (19) is

$$3 \left[\left(\frac{2+T}{1+T} \right) - 3\lambda \right] + Tp(1-2\lambda) > 0, \quad (21)$$

where $\lambda \equiv R/T = Np^2(1-p)$ is the average number of nonobserving neighbors of each triad. In the dense regime where $T \gg 1$, the second term dominates and hence the condition reduces to $\lambda < 1/2$. In the sparse regime where $Tp \ll 1$, the first term dominates. Since $(2+T)/(1+T)$ is bounded between 1 and 2, there is a value $2/3 \leq \lambda_c \leq 1$ such

that $\lambda < \lambda_c$ causes the system to be absorbed. This means that the number of nonobserving neighbors should not exceed a

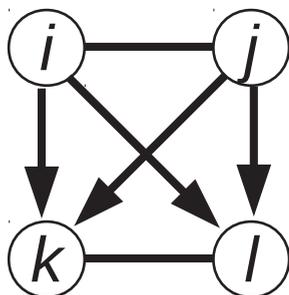


FIG. 7. Tetrad balance.

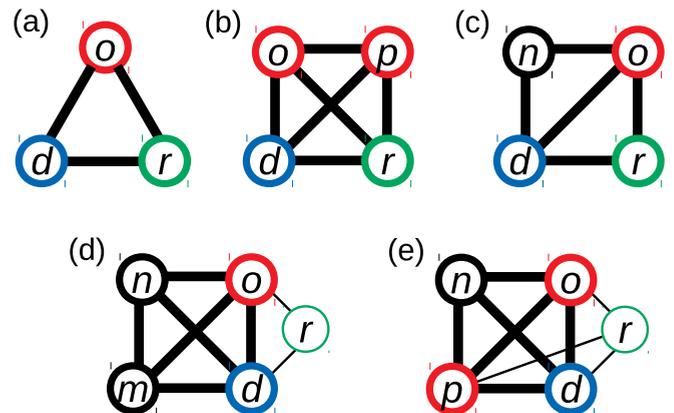


FIG. 8. The subgraphs affected by the game between the donor d and recipient r , which includes third-party observers o and p , and nonobserving neighbors n and m . (a) Triads of d , r , and o . (b) Tetrads of d , r , o , and p . (c) Trusses of d , r , o , and n . (d) Tetrads of d , o , n , and m . (e) Tetrads of d , o , p , and n . Note that the recipients in (d) and (e) are shown only to illustrate the difference between third-party observers and nonobserving neighbors, and they are not included in the focal tetrads.

certain value for the system to be in the absorbing phase. These results from the mean-field analysis are consistent with the simulation results, in that the system has the absorbing phase for the sparse and dense edge density regions, $Np^2(1-p) < 1$, while the active phase is in the middle of them.

The mean-field Eqs. (17)–(19) provide us with further understanding of the opinion dynamics, including the timescales of the relaxations of the order parameters (see the Appendix D in detail). In the sparsest regime, $Np = c \sim O(1)$, in which triangles are formed but the density is too small to have percolated clusters of those, the asymptotic forms of the equations tell us that an autonomous relaxation of Θ to 1 and the subsequent relaxation of Φ to 1 take place fast. After these fast relaxations, the relatively slow relaxation of Ψ takes place as follows:

$$\frac{d\Psi}{dt} = \frac{1 - \Psi}{cN^3}, \quad (22)$$

meaning that Ψ is also driven to an ordered value of 1.

In the regime with more links: $Np^2 \sim O(1)$, which corresponds to the point around the percolation of triangles, Θ and Φ relax to equilibrium values, which are 1 if $c < \sqrt{2}$ and less than 1 if $c > \sqrt{2}$. This process is followed by the slower and independent relaxation of Ψ to 0. In more densely connected systems in which $Np = O(N)$ with $p < 1 - O(N^{-1})$, the asymptotic form of the dynamical equations yields the relaxation time of all the order parameters to be of the same order, and giving rise to a para-phase equilibrium $(\Theta_*, \Phi_*, \Psi_*) = (0, 0, 0)$.

In the most densely connected regime in which $1 - p = c/N$, the relatively fast processes lead to relaxations of $\Theta_* \rightarrow \Phi$ and $\Phi_* \rightarrow \Psi$. The dynamics of Ψ after these fast relaxation follows:

$$\frac{d\Psi}{dt} \sim \frac{(1 - \Psi)[1 - (1 + c)\Psi]}{2N^2}, \quad (23)$$

which corresponds to the equilibrium

$$\Theta_* = \Phi_* = \Psi_* = \frac{1}{1 + c}. \quad (24)$$

This again tells us that the system shows a transition from the para-phase $(\Theta_*, \Phi_*, \Psi_*) = (0, 0, 0)$ to the ordered phase $(\Theta_*, \Phi_*, \Psi_*) = (1, 1, 1)$ at around $1 - p \sim N^{-1}$, in the limit $N \rightarrow \infty$.

V. DISCUSSION

In this study, through numerical simulations and mean-field analysis, we found that, as a result of the indirect reciprocity, the density of social networks drastically changes the friendship and enmity structure. In contrast with complete or fully connected networks [5], in which agents split into two fixed clusters and cooperate only within their own cluster, their relations (who likes or dislikes whom) keep changing in a wide range of network density. The friendship and enmity structure was found to get fixed only if the network is sparse or dense enough, i.e., $p^2(1-p) < 1/N$.

A similar absorbing phase transition in friendship and enmity networks was observed in Heider's structural balance models [10–20]. The structural balance models assume that agents either mutually like or dislike each other and change

their opinion to increase their triad balance. A structural balance model on random networks at certain edge densities shows a phase transition from an absorbing phase in sparse networks to an active phase in dense networks [13]. In contrast, the present indirect reciprocity model allows agents to have different opinions of each other and assume the action and its observation as the reason behind the changes of their relation. This in turn results in changes of the triad balance. Furthermore, when we increase the edge density, the indirect reciprocity model shows an additional phase transition from active to absorbing phase for the higher density region, in addition to the absorbing-to-active transition in the lower density region, as in the structural balance model.

The transition in the lower density region can have a significant implication for a variety of complex networks. It is common that we study, either theoretically or empirically, sparse networks with average degree $k \sim O(1)$, which corresponds to $p \sim O(N^{-1})$ in our model. Therefore, the lower transition point $p_* \sim 1/\sqrt{N}$ is not unrealistically low, and the transition can be relevant for various large networks such as online social networks [21,22].

Moreover, the transition in the higher density region also has negligible implications. The value of the higher transition point $p^* \sim 1 - 1/N$ means that the agent networks need to be almost fully connected to be in the absorbing phase, in the sense that agents need to know all the other agents except for at most one agent. For large networks, e.g., for those with millions of nodes, the requirement is likely too strict and the transition in the higher density is unlikely to be observed. However, important examples of signed or like-dislike social networks include those of moderate size. For example, international relations were recently analyzed as signed networks [23–26], while the networks usually consist of some hundreds of nodes. Another important example of signed social networks of moderate size and high density is who likes or dislikes whom in an organization (e.g., firms, schools) [27].

There are several questions yet to be answered in future studies. One is indirect reciprocity dynamics in real (or more realistic models of) social networks. Real social networks are not considered random, in contrast to the present model, and the network characteristics other than edge density, such as degree distribution or community structure, may largely affect the indirect reciprocity dynamics and friendship-enmity structure of our society. Yet another interesting issue is the character of the observed transition. Although this is beyond the scope of the present study, it would be desirable to investigate whether the transitions are continuous or not, and if continuous, whether they belong to the directed percolation universality class [8,9].

ACKNOWLEDGMENTS

K.O. acknowledges support from JSPS KAKENHI Grants No. JP18H03621, No. JP19H01457, No. JP20J01060, and No. JP20K04995. T.S. was partly supported by JSPS KAKENHI Grant No. 18K03449. K.K. acknowledges support from EU HORIZON 2020 INFRAIA-1-2014-2015 program project (SoBigData) Grants No. 654024, INFRAIA-2019-1 (SoBigData++) No. 871042, and NordForsk's Nordic

Programme for Interdisciplinary Research project ‘‘The Network Dynamics of Ethnic Integration 2021–2024.’’ K.K. also acknowledges the Visiting Fellowship at The Alan Turing Institute, UK.

K.O., S.M, T.S., and K.K. conceived and designed the research; K.O. and S.M performed computational experiments; K.O. and T.S. conducted mean-field analysis; K.O., T.S., and K.K. wrote the paper.

APPENDIX A: UPDATES OF THE EDGE BALANCE

In the following, we denote the values of variable x before and after an update at a time step as x and x' , respectively. Let d , r , and o be the donor, the recipient, and a common neighbor of them at the time step (i.e., d , r , and o form a triad). Then only the following opinions are updated after the time step:

$$\sigma'_{rd} = \sigma_{dr}, \quad (\text{A1})$$

$$\sigma'_{od} = \sigma_{or}\sigma_{dr}. \quad (\text{A2})$$

Therefore, the edge balances that can be updated to a different sign after the time step are those on (d, r) and (d, o) edges:

$$\begin{aligned} \Theta'_{dr} &= \sigma'_{dr}\sigma'_{rd} = \sigma_{dr}^2 = 1, \\ \Theta'_{do} &= \sigma'_{do}\sigma'_{od} = \sigma_{do}(\sigma_{or}\sigma_{dr}) = \Phi_{dor}. \end{aligned} \quad (\text{A3})$$

APPENDIX B: UPDATES OF THE TRIAD BALANCE

Update rules of the triad balance are more complicated. The first triad to be considered is the one formed by the donor d , the recipient r , and the observer o [Fig. 8(a)]. While the triad balances Φ_{dro} and Φ_{dor} are kept under the opinion change, the updated sign of the other four quantities depends on the opinions before the update:

$$\begin{aligned} \Phi'_{odr} &= \sigma'_{od}\sigma'_{or}\sigma'_{dr} \\ &= (\sigma_{or}\sigma_{dr})\sigma_{or}\sigma_{dr} = 1, \\ \Phi'_{ord} &= \sigma'_{or}\sigma'_{od}\sigma'_{rd} \\ &= \sigma_{or}(\sigma_{or}\sigma_{dr})\sigma_{dr} = 1, \\ \Phi'_{rod} &= \sigma'_{ro}\sigma'_{rd}\sigma'_{od} \\ &= \sigma_{ro}\sigma_{dr}(\sigma_{or}\sigma_{dr}) = \Theta_{ro}, \\ \Phi'_{rdo} &= \sigma'_{rd}\sigma'_{ro}\sigma'_{do} \\ &= \sigma_{dr}\sigma_{ro}\sigma_{do} = \Phi_{dro}. \end{aligned} \quad (\text{B1})$$

The second types of triads we should consider are those formed by the donor d and the two observers of the time step knowing each other, o and p [Fig. 8(b)],

$$\begin{aligned} \Phi'_{opd} &= \sigma'_{op}\sigma'_{od}\sigma'_{pd} \\ &= \sigma_{op}(\sigma_{or}\sigma_{rd})(\sigma_{pr}\sigma_{dr}) = \Phi_{opr}, \\ \Phi'_{odp} &= \sigma'_{od}\sigma'_{op}\sigma'_{dp} \\ &= (\sigma_{or}\sigma_{dr})\sigma_{op}\sigma_{dp} = \Psi_{odpr}, \\ \Phi'_{pod} &= \Phi_{por}, \\ \Phi'_{pdo} &= \Psi_{pdor}, \end{aligned} \quad (\text{B2})$$

and because of the symmetry between o and p .

The third and last triads one must take into account involve a nonobserving neighbor n , who is connected to observer o

and the donor d but not to the recipient r [Fig. 8(c)]. Because of the flip of σ_{od} which takes place if $\Phi_{odr} = -1$, the following triad balances are updated:

$$\begin{aligned} \Phi'_{nod} &= \Phi_{odr}\Phi_{nod}, \\ \Phi'_{odn} &= \Phi_{odr}\Phi_{odn}, \\ \Phi'_{ond} &= \Phi_{odr}\Phi_{ond}. \end{aligned} \quad (\text{B3})$$

APPENDIX C: UPDATES OF TETRAD BALANCE

We have seen that the edge balance after a time step is determined by the triad balance, and the updates of the triad balance can be described by the edge, triad, and tetrad balance. So we next consider the update rules of the tetrad balance. Updated tetrads are divided into those that include the recipient and two third-party observers, and those that include three third-party observers. Note that all updated tetrads include the donor. For the latter type, we consider tetrads of d , r and o , p [Fig. 8(b)]. Then Ψ_{opdr} , Ψ_{rpdo} , and $\Psi_{rod p}$ are updated:

$$\begin{aligned} \Psi'_{opdr} &= \sigma'_{od}\sigma'_{or}\sigma'_{pd}\sigma'_{pr} \\ &= (\sigma_{or}\sigma_{dr})\sigma_{or}(\sigma_{pr}\sigma_{dr})\sigma_{pr} \\ &= (\sigma_{or}\sigma_{dr}\sigma_{pr})^2 = 1, \\ \Psi'_{rpdo} &= \sigma'_{rd}\sigma'_{ro}\sigma'_{pd}\sigma'_{po} \\ &= \sigma_{dr}\sigma_{ro}(\sigma_{pr}\sigma_{dr})\sigma_{po} \\ &= (\sigma_{pr}\sigma_{po}\sigma_{ro})\sigma_{dr}^2 = \Phi_{pro}, \\ \Psi'_{rod p} &= \Phi_{opr}. \end{aligned} \quad (\text{C1})$$

For the former type, we consider tetrads of d and o , p , q [Fig. 8(d)]:

$$\begin{aligned} \Psi'_{opqd} &= \sigma'_{oq}\sigma'_{od}\sigma'_{pq}\sigma'_{pd} \\ &= \sigma_{oq}(\sigma_{or}\sigma_{dr})\sigma_{pq}(\sigma_{pr}\sigma_{dr}) \\ &= (\sigma_{oq}\sigma_{or}\sigma_{pq}\sigma_{pr})\sigma_{dr}^2 = \Psi_{opqr}, \\ \Psi'_{pqod} &= \Psi_{pqor}, \\ \Psi'_{qopd} &= \Psi_{qopr}. \end{aligned} \quad (\text{C2})$$

Then, we take a mean-field treatment of the exact update rules considered above by replacing the quantities on the right-hand sides of the equations by the average at that time:

$$\Theta_{ij} \sim \Theta, \quad \Phi_{ijk} \sim \Phi, \quad \Psi_{ijkl} \sim \Psi, \quad (\text{C3})$$

which result in Eqs. (17)–(19).

APPENDIX D: ASYMPTOTIC BEHAVIOR OF MEAN-FIELD DYNAMICS

In the sparsest regime, $Np = c \sim O(1)$, in which triangles are formed but the density is small to have percolated clusters of those, Eqs. (17)–(19) read as follows:

$$\frac{d\Theta}{dt} \sim \frac{c}{N}(1 - \Theta), \quad (\text{D1})$$

$$\frac{d\Phi}{dt} \sim \frac{c}{6N}[(1 - \Phi) + (\Theta - \Phi)], \quad (\text{D2})$$

$$\frac{d\Psi}{dt} \sim \frac{(\Theta - 1)\Psi}{2N^2} + \frac{c[(1 - \Psi) + (\Phi - \Psi)]}{2N^3}. \quad (\text{D3})$$

This means that the autonomous relaxation of Θ to 1 and the subsequent relaxation of Φ to 1 take place fast. And after reaching these equilibrium values, the leading term of the relatively slow relaxation of Ψ becomes

$$\frac{d\Psi}{dt} = \frac{1 - \Psi}{cN^{\frac{3}{2}}}, \quad (\text{D4})$$

meaning that Ψ is also driven to an ordered value 1.

In the regime with more links, $Np^2 = c \sim O(1) > 1$, which corresponds to the point around the percolation of triangles, the asymptotic dynamical equations read as follows:

$$\frac{d\Theta}{dt} \sim \frac{(1 - \Theta) + c(\Phi - \Theta)}{\sqrt{c}N^{\frac{3}{2}}}, \quad (\text{D5})$$

$$\frac{d\Phi}{dt} \sim \frac{(1 + \Theta - 2\Phi) - 3c\Phi(1 - \Phi)}{6\sqrt{c}N^{\frac{3}{2}}}, \quad (\text{D6})$$

$$\frac{d\Psi}{dt} \sim -\frac{(1 - \Theta)\Psi}{2N^2}. \quad (\text{D7})$$

The first two equations correspond to relatively fast dynamics, which has equilibrium at $\Theta_* = \frac{2c+3}{(c+1)^2}$, $\Phi_* = \frac{c+2}{c(c+1)}$ other than the trivial one, $(\Theta_*, \Phi_*) = (1, 1)$. Therefore, these order parameters relaxes to a not-fully-ordered value (< 1) if $c > \sqrt{2}$. The slower relaxation of Ψ is a decay to 0, though it can be very slow when $c < \sqrt{2}$ and hence Θ decays faster to the equilibrium $\Theta_* = 1$.

In more densely connected systems in which $Np = O(N)$ with $p < 1 - O(N^{-1})$, the asymptotic forms of the dynamical equations are

$$\frac{d\Theta}{dt} \sim \frac{p}{N}(\Phi - \Theta), \quad (\text{D8})$$

$$\frac{d\Phi}{dt} \sim \frac{p}{6N}[-(1 - p)\Phi(1 - \Phi) + p(\Psi - \Phi)], \quad (\text{D9})$$

$$\frac{d\Psi}{dt} \sim -\left(\frac{p^2(1 - p)}{2N}\right)\Psi(1 - \Psi). \quad (\text{D10})$$

This yields a para-phase-equilibrium $(\Theta_*, \Phi_*, \Psi_*) = (0, 0, 0)$.

In the most densely connected regime in which $1 - p = c/N$, the asymptotic forms of the dynamical equations are

$$\frac{d\Theta}{dt} \sim \frac{\Phi - \Theta}{N}, \quad (\text{D11})$$

$$\frac{d\Phi}{dt} \sim \frac{\Psi - \Phi}{6N}, \quad (\text{D12})$$

$$\frac{d\Psi}{dt} \sim \frac{(1 - \Psi)(1 - c\Psi) - (1 - \Theta)\Psi}{2N^2}. \quad (\text{D13})$$

Under this dynamics, the relatively fast processes lead to relaxations of $\Theta_* \rightarrow \Phi$ and $\Phi_* \rightarrow \Psi$. Therefore, we can replace Θ in the last equation by Ψ to obtain

$$\frac{d\Psi}{dt} \sim \frac{(1 - \Psi)[1 - (1 + c)\Psi]}{2N^2}. \quad (\text{D14})$$

From this we have the equilibrium

$$\Theta_* = \Phi_* = \Psi_* = \frac{1}{1 + c}, \quad (\text{D15})$$

which tells that this system shows a transition from the para-phase $(\Theta_*, \Phi_*, \Psi_*) = (0, 0, 0)$ to the ordered phase $(\Theta_*, \Phi_*, \Psi_*) = (1, 1, 1)$ at around $1 - p \sim N^{-1}$, in the $N \rightarrow \infty$ limit. Together with the results above, this system has an ordered state as the absorbing state in the regime $p < O(N^{-\frac{1}{2}})$ and $1 - p < O(N^{-1})$, and otherwise it goes to a para-phase.

-
- [1] N. M. Harrigan, G. J. Labianca, and F. Agneessens, *Soc. Netw.* **60**, 1 (2020).
- [2] M. A. Nowak and K. Sigmund, *Nature (London)* **393**, 573 (1998).
- [3] H. Ohtsuki and Y. Iwasa, *J. Theor. Biol.* **231**, 107 (2004).
- [4] M. A. Nowak, *Science* **314**, 1560 (2006).
- [5] K. Oishi, T. Shimada, and N. Ito, *Phys. Rev. E* **87**, 030801(R) (2013).
- [6] A. Isagozawa, T. Shirakura, and S. Tanabe, *arXiv:1603.04945*.
- [7] M. Kandori, *Rev. Econ. Stud.* **59**, 63 (1992).
- [8] H. Hinrichsen, *Adv. Phys.* **49**, 815 (2000).
- [9] S. Lübeck, *Int. J. Mod. Phys. B* **18**, 3977 (2004).
- [10] F. Heider, *J. Psychol.* **21**, 107 (1946).
- [11] D. Cartwright and F. Harary, *Psychol. Rev.* **63**, 277 (1956).
- [12] T. Antal, P. L. Krapivsky, and S. Redner, *Phys. D* **224**, 130 (2006).
- [13] F. Radicchi, D. Vilone, S. Yoon, and H. Meyer-Ortmanns, *Phys. Rev. E* **75**, 026106 (2007).
- [14] F. Radicchi, D. Vilone, and H. Meyer-Ortmanns, *Phys. Rev. E* **75**, 021118 (2007).
- [15] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, *Phys. Rev. Lett.* **103**, 198701 (2009).
- [16] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, *Proc. Natl. Acad. Sci. (USA)* **108**, 1771 (2011).
- [17] T. H. Summers and I. Shames, *Europhys. Lett.* **103**, 18001 (2013).
- [18] V. A. Traag, P. Van Dooren, and P. De Leenheer, *PLoS One* **8**, e60063 (2013).
- [19] R. Nishi and N. Masuda, *Europhys. Lett.* **107**, 48003 (2014).
- [20] S. Wongkaew, M. Caponigro, K. Kułakowski, and A. Borzi, *Eur. Phys. J. Spec. Top.* **224**, 3325 (2015).
- [21] J. Leskovec, D. Huttenlocher, and J. Kleinberg, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, USA, 2010), pp. 1361–1370.
- [22] M. Szell, R. Lambiotte, and S. Thurner, *Proc. Natl. Acad. Sci. (USA)* **107**, 13636 (2010).
- [23] Z. Maoz, L. G. Terris, R. D. Kuperman, and I. Talmud, *J. Polit.* **69**, 100 (2007).
- [24] T. C. Warren, *J. Peace Res.* **47**, 697 (2010).
- [25] P. Doreian and A. Mrvar, *J. Soc. Struct.* **16**, 1 (2015).
- [26] W. Li, A. E. Bradshaw, C. B. Clary, and S. J. Cranmer, *Sci. Adv.* **3**, e1601895 (2017).
- [27] H. J. Park, S. D. Yi, D. J. Kim, and B. J. Kim, *Phys. A* **461**, 647 (2016).