PAPER

Cluster Structure of Online Users Generated from Interaction Between Fake News and Corrections

Masaki AIDA[†a)], *Fellow*, Takumi SAKIYAMA[†b)], Ayako HASHIZUME[††c)], *Nonmembers*, *and* Chisa TAKANO[†††d)], *Member*

SUMMARY The problem caused by fake news continues to worsen in today's online social networks. Intuitively, it seems effective to issue corrections as a countermeasure. However, corrections can, ironically, strengthen attention to fake news, which worsens the situation. This paper proposes a model for describing the interaction between fake news and the corrections as a reaction-diffusion system; this yields the mechanism by which corrections increase attention to fake news. In this model, the emergence of groups of users who believe in fake news is understood as a Turing pattern that appears in the activator-inhibitor model. Numerical calculations show that even if the network structure has no spatial bias, the interaction between fake news and the corrections creates groups that are strongly interested in discussing fake news. Also, we propose and evaluate a basic strategy to counter fake news.

key words: fake news, reaction-diffusion system, activator-inhibitor system, Turing pattern

1. Introduction

Social networking service (SNS) has deeply penetrated into our social life, and online user dynamics now greatly influence user behavior in both online societies and the real world. In particular, fake news has had a great impact on the real world, and it is said that 70% of claims about Trump during the 2016 US presidential election were "false" or "mostly false" information [1]. Since it is impossible to eradicate fake news from the world, it is necessary to build a mechanism that is less susceptible to the adverse effects of fake news on the premise that modern society will have to coexist with fake news. For that purpose, it is necessary to understand the mechanism by which fake news spreads across online social networks (OSNs) and to establish an effective countermeasure.

1.1 Problem Description

The current solution to fake news is to proactively promote

the correct information that denies the content of fake news. Certainly, this seems to be intuitively valid. However, in reality, the results are not so direct. A case has been reported in Japan where the information issued to correct fake news did not work properly [2]–[4].

The details of the matter are as follows. First, in late February 2020, at the beginning of the COVID-19 epidemic, a claim that toilet paper was out of stock was posted on SNS. The claim was that shortages of toilet paper were imminent as the production area of toilet paper, in China, was negatively impacted by COVID-19. In reality, there was sufficient inventory and production capacity in Japan, so stock shortages would never occur with normal demand. For this reason, this post was non-factual fake news. At this point, the fake news itself did not receive much attention and did not trigger the hoarding of toilet paper. To counter this fake news, corrections were disseminated basically saying that toilet paper was not in short supply and that the original report was fake news. Many users who received this correction predicted that many other people would try to hoard toilet paper due to the fake report, which stimulated a sense of crisis. In addition, the spread of the corrections attracted the attention of users to the fake report, and as a result, the spread of fake news was actually promoted. As a result, the dissemination of the corrections to the fake report stimulated the hoarding of toilet paper, ironically resulting in a shortage of toilet paper.

This case reinforced the difficulty of countering fake news in that the dissemination of correct information does not always lead to good results. In particular, since corrections can activate fake news, it has been suggested that it is necessary to consider the interaction between fake news and corrections in order to properly evaluate the effect of any correction being considered. In [2], this case is analyzed by the data-scientific approach and the authors concluded that the corrective information promoted unnecessary purchasing behavior. On the other hand, this paper aims to understand the mechanism of generating user clusters that discuss fake news actively from the viewpoint of a theoretical model.

They are summarized in Table 1. Most simple position is (a) in Table 1. It describes the process by which a user who was initially unaware of fake news becomes aware of fake news through the propagation of information. In this position, the spread of fake news is often described by using the analogy of infectious disease models [5]–[7]. However, the toilet paper case mentioned above was not caused by

Table 1 Basic position of this research.

(a) Do users know fake news?	YES	NO
(b) Assuming that users know both fake news and corrections	Believe in fake news	Believe in correction
(c) Assuming that users know both fake news and corrections	User behavior is influenced by both	User behavior is NOT affected by either

knowing or not knowing fake news. It is the position of (b) to deal with the conflict of different information. When both fake news and its corrections propagate on OSNs, the user is in the position of which information to believe and act. In this position, the spread of different information is often described by competing contagion processes [8]–[10]. However, in the toilet paper example, the hoarding users did not believe the fake news. The hoarders tried to buy toilet paper before stocks were exhausted by other users, even though they knew the information was fake news. From this point of view, it is not important which information the users believe in. It is the position of (c) to consider whether the user's behavior is influenced by fake news, and this is the viewpoint of this paper.

This paper proposes a model describing the diffusion of fake news on a social network using the activator-inhibitor approach of reaction-diffusion systems. Also, based on the proposed model, this paper demonstrates that the interaction between fake news and corrections triggers the clustering of nodes with a high interest in fake news. In addition, this paper discusses a sufficient condition that prevents the generation of these user clusters.

1.2 Research Objectives

Many models have been proposed to describe the process by which fake news spreads over OSNs. They can be classified into models based on the user's own characteristics [11], [12] and models based on the environment surrounding the user [13]–[16]. However, since these models deal with the spread of one type of information, they fail to describe the interaction between fake news and corrections. Therefore, it is necessary to consider the characteristics of fake news and corrections and develop an information spreading model that considers their interaction.

This paper is an extended version of [18]. In this paper, we consider the individual characteristics of fake news and corrections and their interaction and propose a model in which the two types of information are intertwined and spread across OSNs. The proposed model is based on an activator-inhibitor model of the reaction-diffusion system. This model shows that even if the network structure itself is not initially biased at all, the interaction between fake news and the corrections may result in the formation of OSN clusters that believe in fake news. The evaluation in [18] assumed the state in which fake news and corrections were randomly distributed. In this paper, we evaluate the situation in which fake news is already widely distributed but not seen as particularly well noticed before the correction spread from specific nodes. This follows the actual toilet paper shortage

case. The result can explain the mechanism whereby the corrections amplify the influence of fake news.

Also, in addition to [18], we consider the sufficient condition to prevent the amplification of the influence of fake news and evaluate a basic strategy to counter fake news.

2. Preliminary

2.1 Reaction-Diffusion System

Reaction-diffusion systems are mathematical models that describe phenomena in which multiple objects are distributed in a certain space and diffuse in the space while interacting with each other. The models can be written as reaction-diffusion equations. For example, consider two types of objects in a two-dimensional Euclidean space whose coordinates are written as (x, y). Let the density functions representing the spatial distribution of objects a and b at time t be $a(x, y; t)$ and $h(x, y; t)$, respectively. The reaction-diffusion equations are given as follows:

$$
\begin{aligned}
\frac{\partial a(x, y; t)}{\partial t} &= f(a, h; t) + \kappa_a \triangle a(x, y; t), \\
\frac{\partial h(x, y; t)}{\partial t} &= g(a, h; t) + \kappa_h \triangle h(x, y; t),
\end{aligned}
\tag{1}
$$

where, $f(a, h; t)$ and $g(a, h; t)$ in the first terms on the right hand sides are called reaction terms. Since the reaction terms are functions of $a(x, y; t)$ and $h(x, y; t)$, they describe the effect of the interaction between the two types of objects. In particular, because we consider local interactions that are locally determined at each point (x, y), the reaction terms are determined by the information at the point (x, y) given by

$$
\begin{aligned}
f(a, h; t) &= f(a(x, y), h(x, y); t), \\
g(a, h; t) &= g(a(x, y), h(x, y); t).
\end{aligned}
\tag{2}
$$

The second terms on the right-hand side are called the diffusion term, they describe how the two types of objects diffuse spatially. $\triangle$ denotes the Laplacian

$$
\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},
$$

and κ_a and κ_h are diffusion coefficients that describe the diffusion intensity of each object. The larger the diffusion coefficient, the faster is the object diffuses.

The reaction-diffusion system is widely applied in not only the field of chemistry but also other fields such as biology, ecology, and physics.

2.2 Activator-Inhibitor Model

The activator-inhibitor model, which is a type of reaction-diffusion system, has been applied in network engineering [19], [20]. The application examples utilize the Turing pattern that is characteristic of the activator-inhibitor model. The Turing pattern is a kind of spatial pattern that is spontaneously generated and is known to explain the generation
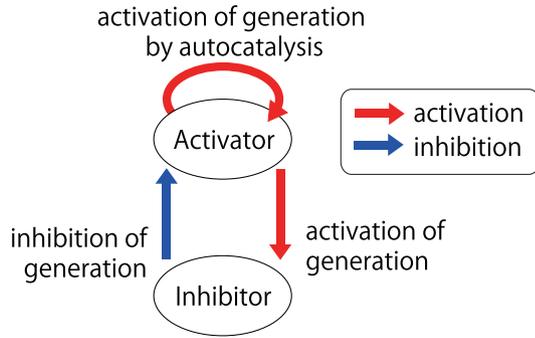
Fig. 1 Basic mechanism of the activator-inhibitor system.

of the geometric patterns that can appear on the surfaces of animal bodies.

As a simple and natural way of giving the reaction terms, we can use the following linear model:

$$f(a, h; t) := c_a\, a(x, y; t) - s_a\, h(x, y; t)$$
$$- \mu_a\, a(x, y; t) + \rho_a, \qquad (3)$$
$$g(a, h; t) := c_h\, a(x, y; t) - \mu_h\, h(x, y; t) + \rho_h, \qquad (4)$$

where c_a, s_a, μ_a, ρ_a, c_h, μ_h and ρ_h are positive constants.

In the activator-inhibitor model, $a(x, y; t)$ is called the activator and $h(x, y; t)$ is called the inhibitor. The intent of this naming can be understood by looking at the structure of the reaction terms in (3) and (4). The reaction terms can be explained qualitatively as follows. The presence of the activator acts to promote the increase of both the activator and the inhibitor at that location, while the presence of the inhibitor acts to suppress the increase of the activator at that location. The first terms on the right hand sides of (3) and (4), and the second term on the right hand sides of (3) represent these effects. The third term on the right-hand sides of (3) and the second terms on the right-hand sides of (4) represent the reaction of self-decay, and the fourth term on the right-hand sides of (3) and the third term on the right-hand sides of (4) represent the rate of natural generation. Figure 1 illustrates these mutual relationships.

Next, with regard to the diffusion terms of the reaction-diffusion equations (1), the activator-inhibitor model requires the following relationship

$$\kappa_h > \kappa_a, \qquad (5)$$

meaning that the inhibitor spreads faster and more widely than the activator.

In the activator-inhibitor model, when the reaction term and the diffusion term interact appropriately, the following situations occur.

- The activator present at a position generates more activator and the inhibitor at the same position.
- At that time, the inhibitor quickly diffuses around the position and suppresses the increase of the activator in the surroundings.
- When the activator is suppressed in the surroundings, not only the activator but also the growth of the inhibitor
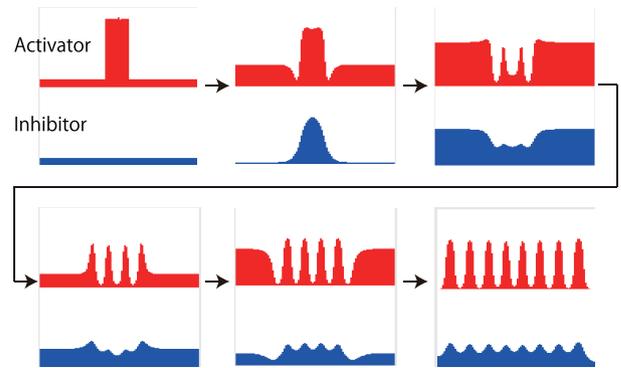

Fig. 2 Turing patterns generated by the activator-inhibitor system.

is suppressed.
- When the inhibitor decreases, both the activator and the inhibitor proliferate by the action of the activator, and the generated inhibitor diffuses to its surroundings more quickly than the activator.

By repeating this process, the spatial distributions of the activator and the inhibitor become uneven and appear as geometric patterns in the space. This is called the Turing pattern. Figure 2 shows an example of the evaluation of the behavior of the activator and the inhibitor by the activator-inhibitor model in one-dimensional space [21]. We can recognize that a periodic pattern appears in the spatial distributions of both the activator and the inhibitor with the passage of time.

3. Activator-Inhibitor Model for Fake News and Corrections

3.1 Basic Concept

Based on the case of toilet paper hoarding shown in the Introduction, we propose a mathematical model to understand the mechanism by which the spread of both fake news and the corrections results in toilet paper hoarding. In preparation, let us consider the basic properties of fake news and the corrections and their interaction. We assume the following properties as a starting point for the model.

- The spread of the corrections fuels users' anxiety that toilet paper may be hoarded by many people contrary to the original purpose of the corrections. For this reason, the corrections not only convey their original intent but also attract attention to the fake news.
- The spread of fake news makes the content of fake news known, and it becomes difficult to spread the corrections that deny it.
- In general, fake news deals with topics that are of interest to ordinary users and therefore tend to spread faster over OSNs than regular news [17].

The first item in the bullet point list shows that the spread of the correction has a role in inspiring attention to not only the correction itself but also the fake news. The second item shows the property that the spread of fake news suppresses
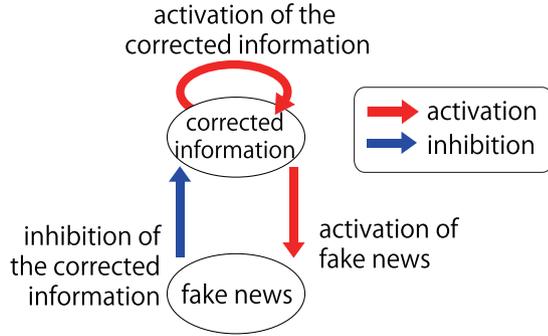
Fig. 3   Interaction between fake news and their corrections.

the activity of correction, and the last bullet point means that the diffusion coefficient of fake news is larger than the diffusion coefficient of the correction. Based on the above properties, the following correspondence can be assumed in the activator-inhibitor model.

- activator: corrections
- inhibitor: fake news

Figure 3 shows the structure of this relationship corresponding to that of Fig. 1.

3.2   Activator-Inhibitor Model on Networks

In this section, we propose an activator-inhibitor model that describes the interaction between fake news and corrections in OSNs, based on the concept shown in Sect. 3.1.

Consider undirected graph $G(V, E)$ with n nodes representing OSN structure, where $V = \{1, 2, \ldots, n\}$ is the set of nodes and E is the set of undirected links. In terms of correspondence with OSN, node $i \in V$ represents a user, and undirected link $(i, j) \in E$ represents the relationship between user i and user j.

Originally, the strength of relationships between users in OSNs is asymmetric, so it is natural to model the structure of OSNs with a directed graph. However, in this paper, we would like to discuss the phenomenon that information dispersal becomes biased even though the network structure itself is not biased at all. For this reason, we dare to introduce a highly symmetric network model represented by an undirected graph. As a side note, it is easy to extend the network model to directed graphs and to discuss information spreading on them, as there are no technical problems in doing so. However, it is worth noting that directed graphs yield asymmetric effects on the networks.

We assume that each link in $G(V, E)$ has a link weight that represents the strength of the relationship; let the weight of undirected link $(i, j) \in E$ between node i and node j be $w_{ij} > 0$. For convenience, the weight of the node pair $(i, j) \notin E$ is assumed to be $w_{ij} = 0$.

The Laplacian matrix L representing the structure of $G(V, E)$ is defined as

$$L := D - A, \tag{6}$$

where A is the adjacency matrix defined as

$$A := \left[w_{ij} \right]_{1 \le i, j \le n},$$

and D is the weighted degree matrix defined as

$$D := \mathrm{diag}(d_1, \ldots, d_n).$$

Here, d_i is the weighted nodal degree of node i defined as

$$d_i := \sum_{i=1}^{n} w_{ij}.$$

The Laplacian matrix L can be used to describe the diffusion phenomenon on the network. Let $x_i(t) \, (\ge 0)$ be the state quantity of node i at time t, and the state vector $x(t)$ be defined as

$$x(t) := {}^t(x_1(t), \ldots, x_n(t)).$$

The diffusion equation on the network can, by using the Laplacian matrix, be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = -L \, x(t).$$

When considering the activator-inhibitor model on the network, the diffusion term appearing in the activator-inhibitor model can be expressed by using the Laplacian matrix. The activator-inhibitor model on the network is expressed as follows. Let $a_i(t)$ and $h_i(t)$ be the amounts of activator and inhibitor at node i at time t, respectively. Next, we define two types of state vectors,

$$\begin{aligned} a(t) &:= {}^t(a_1(t), \ldots, a_n(t)), \\ h(t) &:= {}^t(h_1(t), \ldots, h_n(t)), \end{aligned} \tag{7}$$

which have the amounts of activators and inhibitors for all nodes as elements. Then, the equations that describe the activator-inhibitor model on the network can be written as

$$\begin{aligned} \frac{\partial}{\partial t} a(t) &= F(a(t), h(t)) - \kappa_a \, L \, a(t), \\ \frac{\partial}{\partial t} h(t) &= G(a(t), h(t)) - \kappa_h \, L \, h(t). \end{aligned} \tag{8}$$

Here, $F(a(t), h(t))$ and $G(a(t), h(t))$ are functions that express the reaction terms, and $-\kappa_a \, L \, a(t)$ and $-\kappa_h \, L \, h(t)$ are the diffusion terms. By following (3) and (4), the functions representing the reaction terms are introduced as

$$\begin{aligned} F(a(t), h(t)) &:= C_a \, a(t) - S_a \, h(t) - M_a \, a(t) + R_a \, \mathbf{1}, \\ G(a(t), h(t)) &:= C_h \, a(t) - M_h \, h(t) + R_h \, \mathbf{1}. \end{aligned} \tag{9}$$

Here, C_a, C_h, M_a, M_h, R_a, and R_h are all diagonal matrices defined as

$$\begin{aligned} C_a &:= c_a \, I = \mathrm{diag}(c_a, \ldots, c_a), \\ S_a &:= s_a \, I = \mathrm{diag}(s_a, \ldots, s_a), \\ M_a &:= \mu_a \, I = \mathrm{diag}(\mu_a, \ldots, \mu_a), \end{aligned}$$

$$M_h := \mu_h I = \text{diag}(\mu_h, \ldots, \mu_h),$$
$$R_a := \rho_a I = \text{diag}(\rho_a, \ldots, \rho_a),$$
$$C_h := c_h I = \text{diag}(c_h, \ldots, c_h),$$
$$R_h := \rho_h I = \text{diag}(\rho_h, \ldots, \rho_h),$$

where I is an $n \times n$ unit matrix, and 1 is an n-dimensional vector $1 := {}^t(1, \ldots, 1)$. The fact that these matrices are diagonal matrices reflects the assumption that the interaction between the activator and the inhibitor is a local interaction at each node.

In the above model, for the sake of simplicity and higher symmetry, all the above diagonal matrices are set to be proportional to the identity matrix. This corresponds to all nodes having the same parameter value. More generally, we can use a diagonal matrix where each node has a different parameter value.

4. Experimental Evaluation for User Cluster Generation

This section performs numerical experiments of the activator-inhibitor model that describes the interaction between fake news and the corrections on a homogeneous network that does not have a cluster structure. The purpose of the experimental evaluations is to show that even if there are no clusters in the network structure itself, clusters of users that are strongly influenced by fake news might be formed on OSNs. Therefore, we conduct experiments on network models with extremely high symmetry.

The experimental environment is described as follows. We use a two-dimensional lattice network model consisting of 10,000 nodes in a 100×100 configuration, with a periodic boundary condition that yields a torus topology (Fig. 4). All links are weighted at 1. Other parameters are summarized in Table 2. The parameter values were determined with reference to the Reaction-Diffusion System Simulator [22] published on the website of Professor Shigeru Kondo's Laboratory, Osaka University. In addition, the values of $a_i(t)$ and $h_i(t)$ are restricted to the range of $[0, 1]$. Figures 5 and 6 show the initial conditions for two different conditions. The dark part indicates where the activator value is large, and the light part indicates where the activator value is small.

Figure 5 shows the situation that the correction of the fake news starts to spread from the center nodes after the fake news has spread widely in advance. The left panel is the initial distribution of the correction which exists only at the center nodes; the right panel is the initial distribution of the fake news that has been uniformly distributed. This situation mirrors the actual toilet paper case. The fake news was widely distributed in advance but not particularly well noticed; the correction was subsequently spread from specific nodes. On the other hand, Figure 6 shows the situation that both the fake news and its correction have been widely distributed by mass media, many users have seen and been influenced by both pieces of information, more or less. Both the left and right panels are the initial distributions of the correction and the fake news that have been randomly
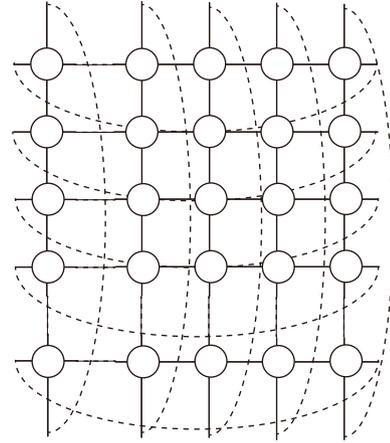

Fig. 4 100×100 2-dimensional lattice network model with torus boundary.

Table 2 Parameter setting for the activator-inhibitor model.

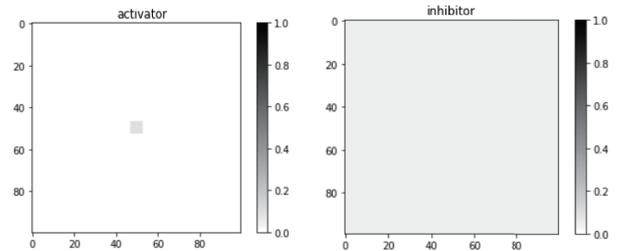| c_a	s_a	μ_a	ρ_a	κ_a	c_h	μ_h	ρ_h	κ_h
0.08	0.08	0.03	0.0015	0.02	0.11	0.12	0.0005	0.5

Fig. 5 The initial condition of the distributions of the strengths of the correction (left) and the fake news (right): the correction starts to propagate from the center nodes, and the fake news has been uniformly distributed in advance.
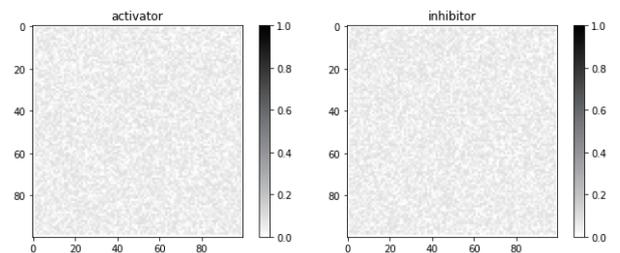

Fig. 6 The initial condition of the distributions of the strengths of the correction (left) and the fake news (right): both are randomly distributed between $[0.0, 0.1]$.

distributed in the range of $[0, 0.1]$.

Figure 7 shows the distributions of the correction and the fake news for the initial condition of Fig. 5 after sufficient time has passed. The difference in the intensity of the correction (the value of activator) directly corresponds to the difference in the intensity of the fake news (the value of inhibitor). That is, where the intensity of the correction is large, the intensity of the fake news is also large. Conversely, where the intensity of the correction is small, the intensity of
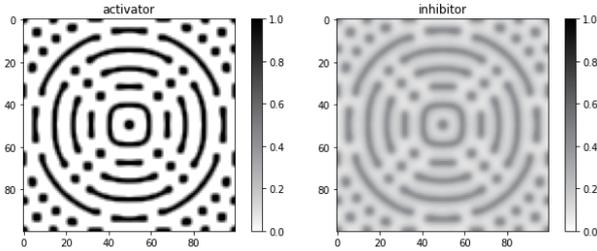
Fig. 7 Turing patterns generated by the interaction between fake news and the correction: the distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 5.

Fig. 8 Turing patterns generated by the interaction between fake news and the correction: the distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 6.
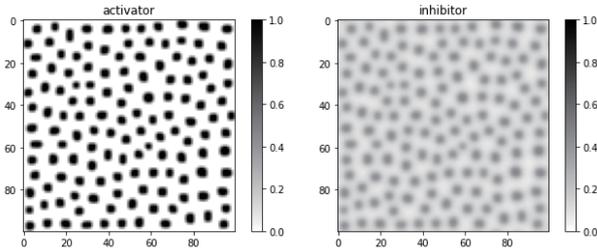
the fake news is also small. From the correspondence with fake news, we find that fake news is activated in areas where the correction is actively discussed, while fake news does not attract attention in places where the correction did not receive attention. The important point of this experimental result is that the correction did not uniformly reduce the influence of fake news, but triggered differentiation yielding areas where fake news is strongly noticed and parts where it is not.

Figure 8 shows the distributions of the correction and the fake news for the initial condition of Fig. 6 after sufficient time has passed. The Turing patterns appearing on the network are different from those of Fig. 7, but we have similar results in that the correction did not uniformly reduce the influence of fake news, but triggered the generation of clusters of users actively influenced by the fake news.

Since the network model used in the experiment is a network model with very high symmetry (a homogeneous undirected graph with lattice topology) and periodic boundary conditions, the network structure has no location-specific properties. Therefore, the polarization that appears in the experimental results is the heterogeneity created by the interaction of fake news and the correction. In this way, the activator-inhibitor model can qualitatively explain the phenomenon in which the corrections cause a situation in which fake news receives more attention from users.

5. Condition for Preventing User Cluster Generation

In order to prevent the generation of user clusters activated by the interaction of fake news and corrections, it is necessary to prevent the Turing pattern corresponding to the user clusters

from occurring. In order to find the solution, it is necessary to understand what kind of conditions cause the Turing pattern. Here, based on the reaction term (9) of the activator-inhibitor model, we discuss Turing instability, which is a necessary condition for the Turing pattern to occur.

From reaction term (9), the ith components ($1 \le i \le n$) of $\boldsymbol{F}(\boldsymbol{a}(t), \boldsymbol{h}(t))$ and $\boldsymbol{G}(\boldsymbol{a}(t), \boldsymbol{h}(t))$ are expressed, respectively, as

$$f(a_i(t), h_i(t)) = c_a\, a_i(t) - s_a\, h_i(t) - \mu_a\, a_i(t) + \rho_a,$$
$$g(a_i(t), h_i(t)) = c_h\, a_i(t) - \mu_h\, a_i(t) + \rho_h. \tag{10}$$

The necessary conditions for the Turing pattern to occur include two factors: uniform solutions are stable if there are no diffusion effects, otherwise they become unstable.

First, when there is no diffusion (assuming the diffusion coefficients are $\kappa_a = \kappa_h = 0$), the spatially-uniform distribution needs to be a stable steady solution of (1). Since there is no diffusion, it is not necessary to consider spatial uniformity, so it is sufficient to derive the conditions for stabilizing the steady-state solution. In (8), by setting $\kappa_a = \kappa_h = 0$, relations of the ith components are extracted as

$$\frac{\partial}{\partial t}\, a_i(t) = f(a_i(t), h_i(t)),$$
$$\frac{\partial}{\partial t}\, h_i(t) = g(a_i(t), h_i(t)). \tag{11}$$

If small deviations from the steady-state solution $\bar{a}$ and $\bar{h}$ are Δa and Δh, respectively, f and g are expanded as

$$f(\bar{a} + \Delta a, \bar{h} + \Delta h)$$
$$= f(\bar{a}, \bar{h}) + \frac{\partial f(\bar{a}, \bar{h})}{\partial a}\, \Delta a + \frac{\partial f(\bar{a}, \bar{h})}{\partial h}\, \Delta h + \cdots,$$
$$g(\bar{a} + \Delta a, \bar{h} + \Delta h)$$
$$= g(\bar{a}, \bar{h}) + \frac{\partial g(\bar{a}, \bar{h})}{\partial a}\, \Delta a + \frac{\partial g(\bar{a}, \bar{h})}{\partial h}\, \Delta h + \cdots.$$

Substituting the above into (11), and applying the steady-state solution condition

$$\frac{\partial \bar{a}}{\partial t} = f(\bar{a}, \bar{h}) = 0,$$
$$\frac{\partial \bar{h}}{\partial t} = g(\bar{a}, \bar{h}) = 0,$$

we obtain

$$\frac{\partial(\bar{a} + \Delta a)}{\partial t} = \frac{\partial(\Delta a)}{\partial t}$$
$$= \frac{\partial f(\bar{a}, \bar{h})}{\partial a}\, \Delta a + \frac{\partial f(\bar{a}, \bar{h})}{\partial h}\, \Delta h + \cdots,$$
$$\frac{\partial(\bar{h} + \Delta h)}{\partial t} = \frac{\partial(\Delta h)}{\partial t}$$
$$= \frac{\partial g(\bar{a}, \bar{h})}{\partial a}\, \Delta a + \frac{\partial g(\bar{a}, \bar{h})}{\partial h}\, \Delta h + \cdots.$$

By linearizing the above equations, we obtain

$$\frac{\partial}{\partial t} \begin{pmatrix} \Delta a \\ \Delta h \end{pmatrix} = \begin{bmatrix} \dfrac{\partial f(\bar{a}, \bar{h})}{\partial a} & \dfrac{\partial f(\bar{a}, \bar{h})}{\partial h} \\ \dfrac{\partial g(\bar{a}, \bar{h})}{\partial a} & \dfrac{\partial g(\bar{a}, \bar{h})}{\partial h} \end{bmatrix} \begin{pmatrix} \Delta a \\ \Delta h \end{pmatrix}. \tag{12}$$

The condition that the steady-state solution ${}^{t}(\bar{a}, \bar{h})$ is stable is that all eigenvalues of the matrix on the right side of (12) are negative. The following two equations can be obtained as conditions for this.

$$\frac{\partial f(\bar{a}, \bar{h})}{\partial a} + \frac{\partial g(\bar{a}, \bar{h})}{\partial h} < 0, \tag{13}$$

$$\frac{\partial f(\bar{a}, \bar{h})}{\partial a} \frac{\partial g(\bar{a}, \bar{h})}{\partial h} - \frac{\partial f(\bar{a}, \bar{h})}{\partial h} \frac{\partial g(\bar{a}, \bar{h})}{\partial a} > 0. \tag{14}$$

Next, we consider the conditions for destabilizing the steady-state solution due to the presence of diffusion. When the effect of the diffusion term (general diffusion coefficient κ_a, $\kappa_h > 0$) is added, the spatial uniformity of the solution must be broken and the uniform distribution is no longer a stationary solution of (1). First, since the Laplacian matrix $\boldsymbol{L}$ that appears on the right side of (8) is a symmetric matrix, it can, by using an appropriate orthogonal matrix $\boldsymbol{P}$, be diagonalized as

$${}^{t}\boldsymbol{P}\,\boldsymbol{L}\,\boldsymbol{P} = \boldsymbol{\Lambda} := \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Here, due to the nature of the Laplacian matrix, it is $\lambda_i \geq 0$ for all i ($1 \leq i \leq n$). Diagonalizing (8) using this yields

$$\frac{\partial}{\partial t} \left({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t) \right) = \boldsymbol{F}({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t), {}^{t}\boldsymbol{P}\,\boldsymbol{h}(t)) - \kappa_a\, \boldsymbol{\Lambda}\left({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t) \right),$$
$$\frac{\partial}{\partial t} \left({}^{t}\boldsymbol{P}\,\boldsymbol{h}(t) \right) = \boldsymbol{G}({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t), {}^{t}\boldsymbol{P}\,\boldsymbol{h}(t)) - \kappa_h\, \boldsymbol{\Lambda}\left({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t) \right). \tag{15}$$

By introducing the components of the vectors $\left({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t) \right)$ and $\left({}^{t}\boldsymbol{P}\,\boldsymbol{h}(t) \right)$ respectively as

$$\left({}^{t}\boldsymbol{P}\,\boldsymbol{a}(t) \right) := {}^{t}(\tilde{a}_1(t), \ldots, \tilde{a}_n(t)),$$
$$\left({}^{t}\boldsymbol{P}\,\boldsymbol{h}(t) \right) := {}^{t}(\tilde{h}_1(t), \ldots, \tilde{h}_n(t)),$$

we can extract the relationships of the ith components as

$$\frac{\partial}{\partial t} \tilde{a}_i(t) = f(\tilde{a}_i(t), \tilde{h}_i(t)) - \kappa_a\, \lambda_i\, \tilde{a}_i(t),$$
$$\frac{\partial}{\partial t} \tilde{h}_i(t) = g(\tilde{a}_i(t), \tilde{h}_i(t)) - \kappa_h\, \lambda_i\, \tilde{h}_i(t). \tag{16}$$

Following the same procedure as (12), we consider small variations $\Delta\tilde{a}_i$ and $\Delta\tilde{h}_i$ around $\tilde{a}_i(t)$ and $\tilde{h}_i(t)$ corresponding to spatially-uniform solutions. By linearizing it, we obtain

$$\frac{\partial}{\partial t} \begin{pmatrix} \Delta\tilde{a}_i \\ \Delta\tilde{h}_i \end{pmatrix}$$
$$= \begin{bmatrix} \dfrac{\partial f(\tilde{a}_i, \tilde{h}_i)}{\partial a} - \kappa_a \lambda_i & \dfrac{\partial f(\tilde{a}_i, \tilde{h}_i)}{\partial h} \\ \dfrac{\partial g(\tilde{a}_i, \tilde{h}_i)}{\partial a} & \dfrac{\partial g(\tilde{a}_i, \tilde{h}_i)}{\partial h} - \kappa_h \lambda_i \end{bmatrix} \begin{pmatrix} \Delta\tilde{a}_i \\ \Delta\tilde{h}_i \end{pmatrix}. \tag{17}$$

In order for diffusion to destabilize a spatially-uniform steady-state solution, the matrix appearing on the right-hand side of (17) must have at least one positive eigenvalue. In the range of $\lambda_i \leq 0$, as the sufficient conditions for the existence of at least one positive eigenvalue in the above matrix, we obtain

$$\kappa_h \frac{\partial f(\tilde{a}, \tilde{h})}{\partial a} + \kappa_a \frac{\partial g(\tilde{a}, \tilde{h})}{\partial h} > 0, \tag{18}$$

$$\left(\kappa_h \frac{\partial f(\tilde{a}, \tilde{h})}{\partial a} + \kappa_a \frac{\partial g(\tilde{a}, \tilde{h})}{\partial h} \right)^2$$
$$- \kappa_a\, \kappa_h \left(\frac{\partial f(\tilde{a}, \tilde{h})}{\partial a} \frac{\partial g(\tilde{a}, \tilde{h})}{\partial h} - \frac{\partial f(\tilde{a}, \tilde{h})}{\partial h} \frac{\partial g(\tilde{a}, \tilde{h})}{\partial a} \right) > 0. \tag{19}$$

The conditions (13)–(14) and (18)–(19) are known as conditions for Turing instability.

Next, we consider the concrete conditions that are derived from Turing instability conditions (13)–(14) and (18)–(19). From (3) and (4), we obtain

$$\frac{\partial f}{\partial a} = c_a - \mu_a, \quad \frac{\partial f}{\partial h} = -s_a, \quad \frac{\partial g}{\partial a} = c_h, \quad \frac{\partial g}{\partial h} = -\mu_h.$$

We first consider the conditions for parameters other than the diffusion coefficients κ_a and κ_h. From (13) and (18), we ontain

$$0 < c_a - \mu_a < \mu_h. \tag{20}$$

Also, from (14), we obtain

$$(c_a - \mu_a)\, \mu_h < s_a\, c_h. \tag{21}$$

Next, we consider the conditions related to diffusion coefficients κ_a and κ_h. After satisfying the conditions (20)-(21), in order to satisfy (19) and (5), the following condition is required:

$$\kappa_a < \kappa_h \frac{1}{\mu_h^2} \Bigg[-(c_a - \mu_a)\, \mu_h + 2\, s_a\, c_h$$
$$- 2\sqrt{s_a^2\, c_h^2 - s_a\, c_h\, (c_a - \mu_a)\, \mu_h}\,\Bigg].$$

Therefore, in order to block the generation of the Turing pattern, the diffusion coefficient κ_a of the activator should be set as

$$\kappa_a \geq \kappa_h \frac{1}{\mu_h^2} \Bigg[-(c_a - \mu_a)\, \mu_h + 2\, s_a\, c_h$$
$$- 2\sqrt{s_a^2\, c_h^2 - s_a\, c_h\, (c_a - \mu_a)\, \mu_h}\,\Bigg]. \tag{22}$$

6. Experimental Evaluation of the Condition for Preventing User Cluster Generation

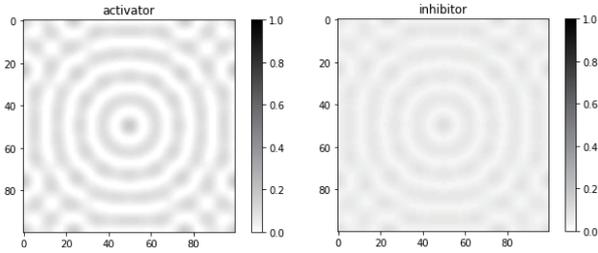The lower limit of the diffusion coefficient κ_a of the activator

Fig. 9 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 5 with $\kappa_a = 0.057$.
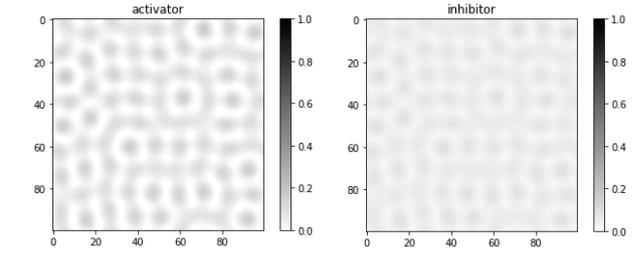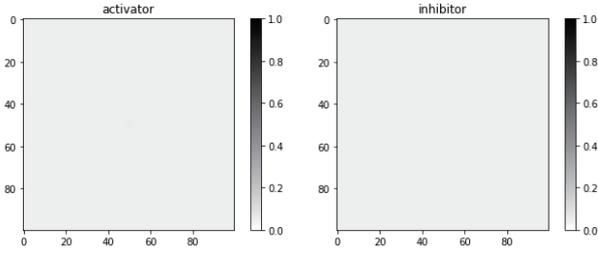

Fig. 10 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 5 with $\kappa_a = 0.058$.

Fig. 11 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 5 with $\kappa_a = 0.059$.

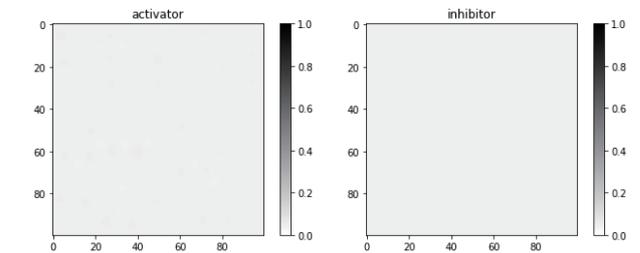
Fig. 12 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 6 with $\kappa_a = 0.057$.

Fig. 13 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 6 with $\kappa_a = 0.058$.
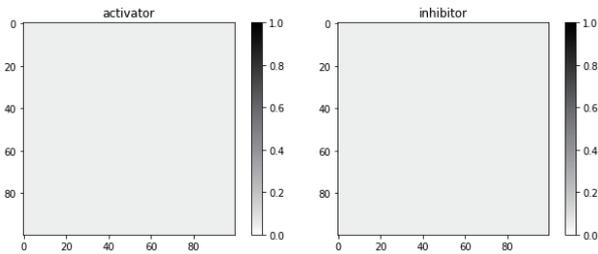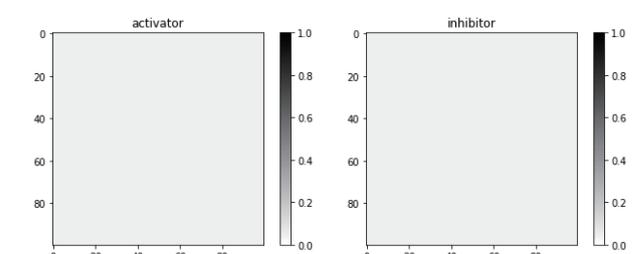

Fig. 14 The distributions of the strengths of the correction (left) and the fake news (right) for the initial condition of Fig. 6 with $\kappa_a = 0.059$.

given by (22) gives sufficient conditions to prevent the Turing pattern from occurring. From the values in Table 2, the lower limit of κ_a in the experiments of this paper is determined as follows:

$$\kappa_a \simeq 0.058. \tag{23}$$

In this section, we investigate the occurrence of Turing patterns around this value and consider the effectiveness of condition (22) in countering fake news.

Figures 9–11 show the distributions of correction (left panels) and fake news (right panels) for the diffusion coefficient $\kappa_a = 0.057$, 0.058, and 0.059 after sufficient time has passed, for the initial condition illustrated in Fig. 5. This initial condition describes the situation in which the correction is sent from the central node after the fake news has spread widely. The result of Fig. 9 is when the diffusion coefficient $\kappa_a = 0.057$ does not satisfy condition (22), and the partial emergence of the Turing pattern can be recognized. On the other hand, the results of Figs. 10 and 11 correspond to the case where condition (22) is satisfied, and the Turing pattern has not emerged.

Next, we show similar evaluations for a different initial

condition. Figures 12–14 show the distributions of correction (left panels) and fake news (right panels) for the diffusion coefficient $\kappa_a = 0.057$, 0.058, and 0.059 after sufficient time has passed, for the initial condition illustrated in Fig. 6. The result of Fig. 12 is when diffusion coefficient $\kappa_a = 0.057$ does not satisfy condition (22), and the slight emergence of the Turing pattern can be recognized. On the other hand, the results of Figs. 13 and 14 correspond to the case where condition (22) is satisfied, and the emergence of the Turing pattern is blocked.

From these results, it can be seen that condition (22), which blocks the Turing pattern, gives a highly accurate threshold value for the presence or absence of the Turing pattern; it is a sufficient condition. It can also be seen that its effectiveness does not depend on the initial conditions.

The value of κ_a is related to the speed at which the correction spreads. It might be generally difficult to know the values of the parameters κ_h, c_a, μ_a, s_a, c_h, and μ_h that appear on the right-hand side of condition (22). Therefore, it might be difficult to determine the value on the right side of condition (22), and it may not be possible to know the lower limit of κ_a. However, there is no doubt that the occur-

rence of the Turing pattern can be suppressed as long as κ_a is increased until it exceeds a certain value. For this reason, there is no doubt that these results set the direction of countermeasures against fake news. Moreover, the conditions do not depend on the initial state of the network, and in theory, they do not depend on network topology, so the versatility and utility of these results are expected to be high.

The structures of existing online social networks are, in general, asymmetrical and might be another factor generating user clusters. Asymmetrical structures may be due to, in part, directed links, differences in link weights, and the original clustered structure of the social network. Therefore, if we want to know a general method to prevent the generation of user clusters, we should investigate the applicability of the condition for avoiding user cluster generation to the general asymmetric networks; however, this issue is out of the scope of the paper.

7. Conclusion

In this paper, we proposed a model that explains the occurrence of the phenomenon in which the information issued to correct fake news on OSNs does not always yield the expected effect, and may, in fact, worsen the influence of fake news. By using the activator-inhibitor model, which is a kind of reaction-diffusion system, and making the inhibitor correspond to fake news and the activator correspond to the corrections, we were able to well replicate the characteristics and interactions of both. Numerical experiments showed that even if there is no bias in the network structure itself, the interaction between fake news and the corrections may trigger the generation of user clusters that are strongly influenced by fake news. In addition, by using linear stability analysis, we derived the sufficient condition for the diffusion coefficient of the correction that prevents the generation of user clusters affected by fake news. Although it is a sufficient condition, experimental results show it gives a sufficiently accurate threshold to the generation of user clusters. In addition, the condition is independent of the initial condition of the distributions for the fake news and its correction. If the network structure has lower symmetry, user clusters can be generated for other reasons related to the asymmetry of the structure. Asymmetric structures can be generated by directed links, differences in the link weights, originally biased structures of the social network, and so on. Therefore, preventing user clusters requires the consideration of various factors. However, even if we eliminate the issues raised by asymmetric structures, user clusters are still generated by the mechanism discussed in this paper. In the future, based on the proposed activator-inhibitor model, we will consider effective and strategic methods for transmitting corrections to counter fake news that are applicable to general network structures.

Acknowledgments

References

[1] A. Guess, B. Nyhan, and J. Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign," European Research Council, 2018.

[2] R. Iizuka, F. Toriumi, M. Nishiguchi, M. Takano, and M. Yoshida, "Impact of correcting misinformation on social disruption," PLoS ONE, vol.17, no.4, e0265734, 2022. doi.org/10.1371/journal.pone.0265734

[3] F. Toriumi, Study Group on Platform Service, Ministry of Internal Affair and Communications, Japan, April 2021. www.soumu.go.jp/main_content/000745175.pdf

[4] Nikkei Business Online, business.nikkei.com/atcl/gen/19/00002/022801115/

[5] A. Kucharski, "Study epidemiology of fake news," Nature, vol.540, 525, 2016.

[6] P. Khurana and D. Kumar, "Sir model for fake news spreading through Whatsapp," Proc. 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), Jaipur, India, 2018.

[7] M. Maleki, E. Mead, M. Arani, and N. Agarwal, "Using an epidemiological model to study the spread of misinformation during the black lives matter movement," arXiv, arXiv:2103.12191, 2021.

[8] Q-H. Liu, L-F. Zhong, W. Wang, T. Zhou, and H.E. Stanley, "Interactive social contagions and co-infections on complex networks," Chaos, vol.28, no.1, 013120, 2018.

[9] B. Min and M.S. Miguel, "Competing contagion processes: Complex contagion triggered by simple contagion," Scientific Reports, vol.8, 10422, 2018.

[10] B. Min and M.S. Miguel, "Competition and dual users in complex contagion processes," Scientific Reports, vol.8, 14580, 2018.

[11] D.C. Brody and D.M. Meier, "How to model fake news," arXiv, arXiv:1809.00964, 2018.

[12] L. Cheng, R. Guo, K. Shu, and H. Liu, "Causal understanding of fake news dissemination on social media," arXiv, arXiv:2010.10580, 2020.

[13] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," PLOS ONE, 2018.

[14] H. Peng, A. Nematzadeh, D.M. Romero, and E. Ferrara, "Network modularity controls the spread of information diffusion," Phys. Rev. E, vol.102, 052316, 2020.

[15] Q.F. Lotito, D. Zanella, and P. Casari, "Realistic aspects of simulation models for fake news epidemics over social networks," Future Internet, vol.13, no.3, 76, 2021.

[16] L. Hébert-Dufresne, S.V. Scarpino, and J.-G. Young, "Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement," Nat. Phys., vol.16, pp.426–431, 2020.

[17] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol.359, no.6380, pp.1146–1151, 2018.

[18] M. Aida and A. Hashizume, "Activator-inhibitor model for describing interactions between fake news and their corrections," Complex Networks & Their Applications X, R.M. Benito, C. Cherifi, H. Cherifi, E. Moro, L.M. Rocha, and M. Sales-Pardo, eds., pp.54–65. Springer, Cham, 2022.

[19] G. Neglia and G. Reina, "Evaluating activator-inhibitor mechanisms for sensors coordination," Proc. International Conference on Bio-Inspired Models of Network, Information and Computing Systems

(BIONETICS 2007), pp.129–133, 2007.

[20] N. Wakamiya, K. Hyodo, and M. Murata, "Reaction-diffusion based topology self-organization for periodic data gathering in wireless sensor networks," Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems, pp.351–360, 2008.

[21] M. Aida, Distributed Control and Hierarchical Structure in Information Networks, Information Network Science Series, vol.3, Corona Publishing, 2015 (in Japanese).

[22] Reaction-Diffusion System Simulator, Kondo Laboratory, Osaka University. www.fbs.osaka-u.ac.jp/labs/skondo/simulators/rd_new.html

Chisa Takano received a B.E. degree in Telecommunication Engineering from Osaka University, Japan, in 2000, and a Ph.D. in System Design Engineering from Tokyo Metropolitan University, Japan, in 2008. In 2000, she joined the Traffic Research Center, NTT Advanced Technology Corporation. From April 2008 to March 2020, she was an Associate Professor at the Graduate School of Information Sciences, Hiroshima City University. She has been a Professor of the Graduate School of Information Sciences, Hiroshima City University since April 2020. Her research interests are computer networks, distributed systems and social networks. She received the IEICE Young Researchers' Award in 2003. She received the Information Network Research Awards from the IEICE in 2004, 2012, 2015 and 2019. She is a member of IEEE, IEICE and IPSJ.

Masaki Aida received his B.S. degree in Physics and M.S. degree in Atomic Physics from St. Paul's University, Tokyo, Japan, in 1987 and 1989, respectively, and his Ph.D. in Telecommunications Engineering from the University of Tokyo, Japan, in 1999. In April 1989, he joined NTT Laboratories. From April 2005 to March 2007, he was an Associate Professor at the Faculty of Systems Design, Tokyo Metropolitan University. He has been a Professor of the Graduate School of Systems Design, Tokyo Metropolitan University since April 2007. His current interests include analysis of social network dynamics and distributed control of computer communication networks. He received the Best Tutorial Paper Award and the Best Paper Award of IEICE Communications Society in 2013 and 2016, respectively, and IEICE 100-Year Memorial Paper Award in 2017. He is a fellow of IEICE, a senior member of IEEE, and a member of ACM and ORSJ.
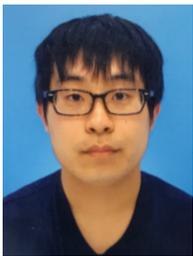
Takumi Sakiyama received his B.E. degree in Computer Science from Tokyo Metropolitan University, Japan, in 2022. Currently, he is a graduate student of the Graduate School of Systems Design, Tokyo Metropolitan University.

Ayako Hashizume received Ph.D. in Kansei Science from Tsukuba University, Japan, in 2011. She is an associate professor at Hosei University, Japan. From 2012 to 2018, she was an assistant professor at Tokyo Metropolitan University, Japan. Her research interests include psychology of emotion and Kansei engineering in relation to HCD (human centered design) and UX (user experience), especially user surveys for service design. She was awarded the kids design prospective competition 2008 from the Japan Kids Design Association, the research prize 2013 from the Funai Foundation for Information Technology, and the best book award 2020 and 2021 from Japan Society of Kansei Engineering. She is also a member of ISO/TC 159/SC 4.