scientific reports

Check for updates

OPEN

Twitter communities are associated with changing user's opinion towards COVID-19 vaccine in Japan

Qianyun Wu[1], Yukie Sano[2], Hideki Takayasu[1], Shlomo Havlin[1,3]✉ & Misako Takayasu[1]✉

Despite extensive research studying the opinions on vaccination since the outbreak of COVID-19, the dynamics of temporal opinion shifts of individuals and their possible origins have been rarely studied. Here, we explore the possible influence of social interactions (retweet network) on individual's opinion shift related to vaccinations based on large-scale Twitter data in Japan. We use an opinion score which calculates the fraction of pro-, neutral and anti-vaccine tweets to measure the dynamic changes of individual opinions, and identify statistically significant communities based on retweet network. By tracking individual's dynamic opinion and its community affiliation, our study highly suggests that the opinion shifts are largely influenced by the user's Twitter community. That is, if users are within the anti-vaccine (or pro-vaccine) community, they exhibit a significantly higher likelihood of changing their position and adopt an anti-vaccine (or pro-vaccine) stance. We also find that the anti-vaccine community's influence appears to persist longer than the pro-vaccine community.

Keywords Social media, Social networks, Community detection, Opinion polarization

Vaccine is proven effective to reduce death tolls caused by COVID virus[1]. However, unsurety and unwilling to being vaccinated against COVID-19 is a major challenge worldwide for improving vaccination coverage. According to the Vaccine Confidence Survey conducted by Lancet in 149 countries[2], Japan has one of the lowest rates of vaccine confidence in the world. Indeed, Japan started first dose in April 2021, much later than most other developed countries like the US or the UK who started vaccination in Dec 2020. In January 2021, the Japanese government appointed a minister of vaccine to start promoting and rolling out vaccination[3]. Within half a year after the vaccination started, Japan achieved a vaccination rate higher than most countries (https://ourworldindata.org/covid-vaccinations). Both the first and second doses covered almost 80% of Japanese population. The booster started in December 2021 and reached a vaccination rate of over 60%, which is lower than the first and second doses. Thus, understanding the fundamental mechanisms for vaccine opinions and their shifts at individual levels in Japan could help understand the general opinion shifts and specifically in promoting vaccination.

Researchers in Japan investigated the possible reasons that affect perceptions towards the COVID-19 vaccine in Japan through surveys[4], or by analyzing social media and web data[5,6]. Some studies focused on the characteristics of the anti-vaccine accounts to understand how this minority group expands its influence[7,8]. However, most research focused on aggregated group perceptions without delving into a detailed understanding of why individuals change their opinions.

To bridge this gap, we study Twitter (now renamed as X) which offers the advantage of enabling longitudinal analysis of individual activities. Twitter is the most popular public social media platform in Japan, covering a wide range of age groups and balanced between genders[9]. Twitter is widely used by researchers studying social networks and opinion diffusion[10–12]. We collected for our analysis more than 100 million Japanese tweets and retweets related to the COVID-19 vaccine posted between January 2020 and May 2022 (for more details, see section "Dataset"). Using Twitter data, we track here individual users' opinion changes and track their long-term connections with other users based on retweets. Then we test whether and how Twitter communities are related to individual opinion shifts. (It is important to note that *retweeting* does not equate to expressing an opinion, as suggested in a survey by Metaxas et al.[13]. Rather, retweeting signifies the retweeter's interest and attention to the message. Therefore, we posit here that *original tweets* reflect individual's opinions, whereas the *retweets* represent the influence of other users.) Similar to social networks which exhibit nontrivial clustering

[1]School of Computing, Institute of Science Tokyo, Yokohama 226-8501, Japan. [2]Institute of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. [3]Department of Physics, Bar-Ilan University, Ramat-Gan 5290002, Israel. ✉email: havlins@gmail.com; takayasu@comp.isct.ac.jp

and assortative mixing[14,15], studying these Twitter communities could aid in comprehending the substructure of networks and observing the characteristics of user behaviors.

Previous research that examines vaccine opinions (not limited to COVID-19 vaccines) from a community perspective has predominantly focused on non-evolving but accumulated static networks. These studies have observed the presence of polarized pro-vaccine and anti-vaccine communities, which often results in an echo chamber effect, particularly within the anti-vaccine community[16–18]. However, we argue here that there exist primarily two important gaps in the existing body of work.

First, despite the fact that social communities and opinions co-evolve with time, prior studies have not simultaneously tracked the dynamic shifts considering both opinions and communities. Most research studying the co-evolution of opinion dynamics and group structures is based on modeling and simulations without the support of empirical data[19]. In addition, while there have been a few investigations into community changes based on empirical data[20,21], they have not elucidated how these communities are related to the dynamic shifts in individual opinions. In contrast, the present study analyzes in scales of quarter years (every 3 months) the simultaneous evolution of both individual opinions and communities during a span of 2.5 years.

Second gap, most existing studies have been focused solely on studying collective opinion changes[22,23], without tracing individual's opinion changes to comprehend the underlying mechanisms contributing to the collective opinion shifts. On the other hand, most research studying the individual opinion dynamics focus on modeling. To bridge this gap, we assess individual users' opinions using an opinion score-a floating-point variable ranging from −1 (anti) to 1 (pro). This opinion score is derived from the fractions of a user's tweets classified as pro-vaccine, neutral, or anti-vaccine by a machine learning classifier. This scoring system not only indicates stances but also assumed to quantify the strength of opinions and the degree of opinion changes, facilitating nuanced analysis of individual opinion shifts. While opinion scores are commonly utilized to model opinion dynamics related to political polarizations[24,25], only very few studies[26] have measured individual vaccine opinion changes based on large-scale data.

The structure of our paper is as follows: We begin by providing an overview of user opinion profiles, showcasing opinion transitions at both the group and individual levels across different vaccination stages. Subsequently, we identify communities and track their dynamic changes on a monthly basis to comprehend how community structures evolve over time. Finally, we combine the temporal opinion changes of individuals and community evolution to unveil the impact of communities on opinion shifts.

Results

Individual opinion shifts from pre-vaccination to the completion of the third dose

We employ a machine learning-based opinion classifier (for details, see section "Vaccine opinion classification") to extract pro-, anti-, and neutral opinions from tweets. Our primary objective is to comprehend individual opinion shifts. We introduce a variable called the opinion score, denoted as O_t^i, to measure individual opinions.

We quantify here the opinion score based on the following principles: (1) A user's opinion is measured by the fraction of their tweets labeled as pro-, neutral-, or anti-vaccine; (2) We define the opinion score O_t^i as a number ranging from −1 to 1, see Eq. (1). The sign of the score represents vaccine preference (positive for pro-vaccine and negative for anti-vaccine), while the absolute value reflects the extremity of the stance. For example, a user who exclusively posts pro-vaccine tweets will have an opinion score of 1, whereas a user who exclusively posts anti-vaccine tweets will have a score of −1. Users who post a mix of stances or a greater fraction of neutral tweets will have a score closer to 0, indicating a less extreme position. Accordingly, we use NP_t^i, NN_t^i, NA_t^i to represent the number of tweets posted by user *i* during time window *t* that are classified as pro-, neutral-, and anti-vaccine, respectively. The opinion score can be computed as Eq. (1).

$$O_t^i = \frac{NP_t^i - NA_t^i}{NP_t^i + NA_t^i + NN_t^i}. \quad (1)$$

This evaluation is modeled as a multinomial distribution, where each independent tweet exhibits one of the pro-, neutral and anti-vaccine opinions. However, this approach is susceptible to noise, particularly when the sample size is small. In the Appendix A2, Fig. A2 shows that when the tweet count is below 5, the opinion scores tend to be more noisy and less reliable. Therefore, we stipulate that the sum of pro-vaccine, neutral, and anti-vaccine tweets (NP_t^i+NN_t^i+NA_t^i) must exceed 5 to ensure statistical reliability. In the Appendix A2, we also explore the utilization of Bayesian inference to predict a user's opinion (posterior probability) based on both the actual observation and prior probability. The results obtained through Bayesian inference (as is shown in Fig. A3 and A4) do not significantly differ from those obtained using the current approach (the detailed comparison is provided in Appendix A11).

To determine the opinion category of each user, we partition the opinion score's range [-1, 1] into three evenly sized bins, considering $-1 \leq O_t^i < -0.3$ as anti-vaccine, $-0.3 \leq O_t^i \leq 0.3$ as neutral, and $0.3 < O_t^i \leq 1$ as pro-vaccine.

Trend of pro-, neutral-, anti-vaccine opinions

Figure 1a illustrates the time series of opinion fractions by count of users who tweeted more than 5 times during each time window of 1 month. It provides an overview of vaccine opinions over the span of 2.5 years, categorized into three phases: phase 1 (pre-vaccination), phase 2 (first and second doses), and phase 3 (the booster dose). Notably, during phase 1, neutral opinions dominated. Throughout phase 2, pro-vaccine opinions gradually increased and eventually surpassed neutral opinions in the third quarter of 2021. The fraction of anti-vaccine and neutral opinions increased once again between phase 2 and phase 3. The surge in anti-vaccine user fraction during this period could be attributed to a few reported cases of myocarditis, as indicated by relevant keywords.

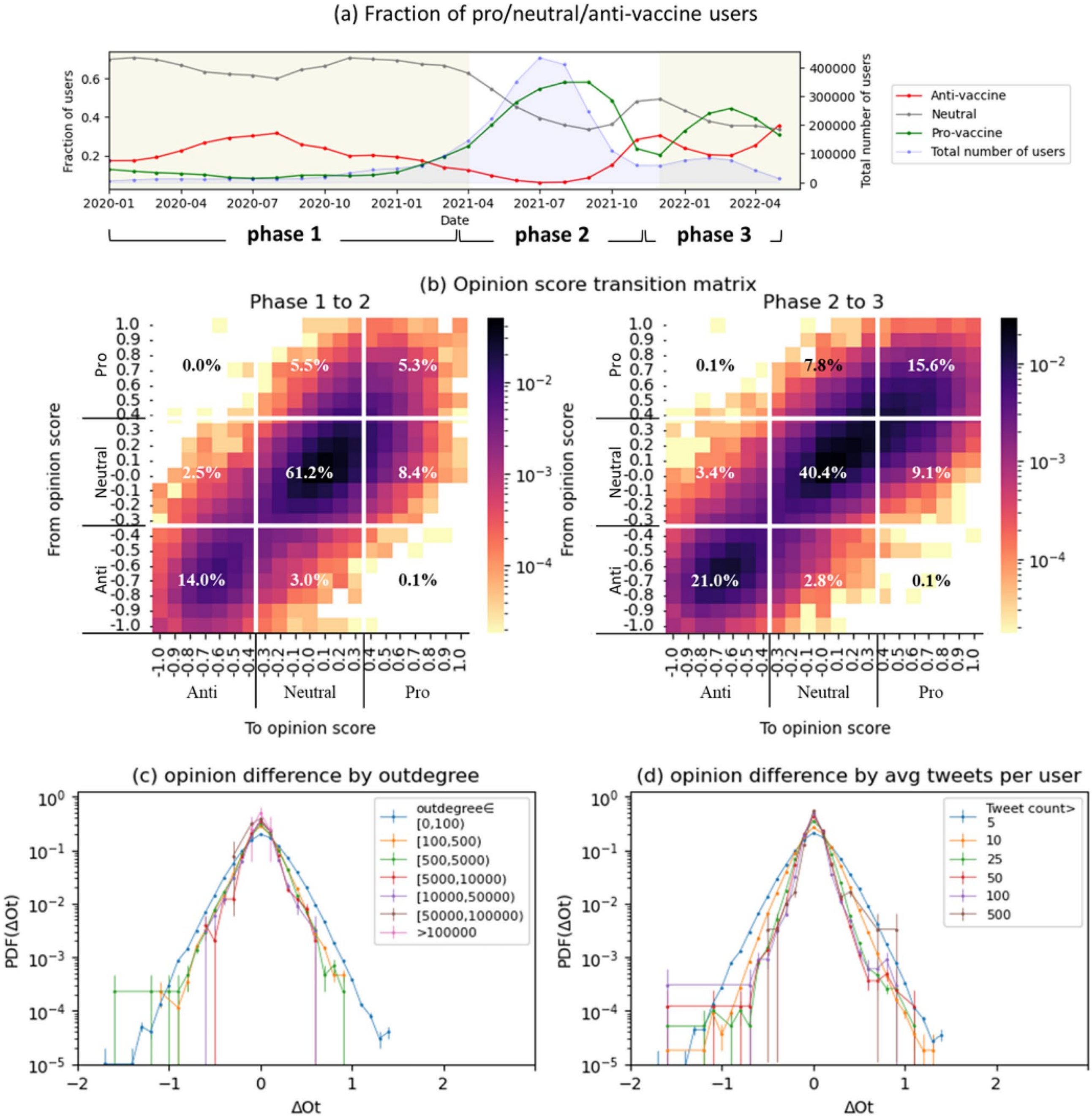

Fig. 1. User opinion profiles and the distribution of opinion transitions. (**a**) Monthly fraction of opinions evaluated by the fraction of users and total number of users. Individual users' opinions on month *t* are determined by calculating their opinion scores using Eq. (1). This calculation aggregates one month of tweets posted by each user starting from day *t* (only users with multiple tweets are considered). Users are classified as pro-vaccine if their opinion score is greater than 0.3, as anti-vaccine if their opinion score is below −0.3, and as neutral otherwise. (**b**) Heatmap illustration of opinion transitions between phase 1 (pre-vaccination), phase 2 (the first and second dose), and phase 3 (the third dose). The y-axis represents users' opinion scores in the previous phase, while the x-axis represents their opinion scores in the subsequent phase. This analysis includes only users who posted more than 5 tweets in each of the successive phases. The color intensity corresponds to the density of users, with darker shades indicating higher densities. The annotated numbers in the figure indicate the fraction of users transitioning from one opinion category to another. (**c**,**d**) Semi-log PDF plot shows the opinion difference ΔO_t^i of users with different outdegrees and tweeting activities. The errorbar is calculated by $\sqrt{p(1-p)/n}$ where *p* represents the PDF of ΔO_t^i and *n* represents sample size.

These cases may have sparked public fear and concern regarding the COVID-19 vaccine. In phase 3, the fraction of pro-vaccine users experienced another surge and became dominant again. Figure 1a also shows the monthly total number of users who tweeted over 5 times. The number of users talking about vaccination drastically surged in the second quarter of 2021 when the first dose of vaccination started.

Individual opinion shifts
Figure 1b illustrates the density of individuals transitioning their opinions from one score to another between the phases (pre-vaccination, the first/second dose and third booster dose). To ensure cohort consistency, here we include only users who tweeted more than five times in successive phases (in each phase 1 and 2, or each phase 2 and 3). In both subplots, darker colors (denser population) are observed near the diagonal, indicating that people tend to undergo moderate shifts in opinion.

As annotated in Fig. 1b, during the transitions from phase 1 to phase 2 and from phase 2 to phase 3, the majority of users maintain a neutral opinion, although the fraction in the first transition period is significantly higher than in the second (61.2% vs. 40.4%). In both periods, users holding a neutral opinion are 3 to 4 times more likely to shift to a pro-vaccine stance than to an anti-vaccine stance. Drastic shifts from a pro-vaccine to an anti-vaccine stance or vice versa are rare, constituting less than 0.1%. Between phase 2 and phase 3, a higher percentage of users maintain a pro-vaccine opinion (15.6%) compared to the transition between phase 1 and phase 2 (5.3%). Additionally, it is noteworthy that users holding an anti-vaccine opinion are unlikely to change their stance, with less than a 20% chance of doing so.

To further investigate the extent of individual opinion shift, we calculate the opinion difference for each user ($\Delta O_t^i = O_t^i - O_{t-1}^i$), where ($t$-1) represents the preceding phase and t represents the successor phase. Figure 1c and 1d depict the semi-log probability density function (PDF) of opinion difference between 3 phases for users with different outdegree and tweeting activity. We observe that users with a larger outdegree (more people retweeting this user) or a higher tweet count exhibit less opinion change, as indicated by a narrower "∧ " shape in the PDF curve. In these figures we do not differentiate the phases because the results do not show significant difference (details are provided in Appendix A3—Fig. A5).

Twitter communities based on the retweet network

We construct a network using retweet data to establish connections between users who retweeted one another. In this network, each node represents a user, and each directed link signifies a retweeting interaction. If user i retweets user j a total of k times, we create a directed link from user j to user i with a weight of k.

Six key communities
We begin by detecting communities within the static network by accumulating the entire 2.5-year dataset. This static network serves as a reference point for identifying dynamic communities. Constructing the network from 81 million retweets, we establish a network having 3 million nodes and 43 million links. Employing the Ensemble Louvain method[27,28] (see section "Dynamic community detection"), we uncover six major communities, each containing more than 5% of the total network 's nodes, collectively covering 83% of all nodes. To examine the statistical significance of these communities, we introduce a new statistical test that compares the actual communities to random ones. We demonstrate that all these six communities are statistically significant, with p-values smaller than 0.05 (see the statistical test method in the section "Significance test on individual communities").

To support the identities of these communities, we used three sets of samples:

(1) Representative Users: we selected 20 most influential users (see examples in appendix Table A4) with the highest outdegree from each of the 6 communities, obtained their user profiles and sampled 10 tweets from each user.
(2) Hashtags: we selected the 20 most frequent hashtags (see examples in Fig. 2b) from each of the 6 communities.
(3) Keywords: we selected the 20 most frequent keywords from each of the 6 communities.

To label these datasets, we employed two independent annotators. The annotators referred to previous research on vaccine communities[17,18,26], which commonly identified four types: pro-vaccine, anti-vaccine, pet, news and political communities. In cases where there was disagreement between the two annotators, a third annotator was consulted to resolve the discrepancies.

We identify the six communities with respective percentages as follows: pro-vaccine (46%), news (8%), anti-vaccine (8%), pet-hobby (8%), left-wing (7%), and right-wing (6%).

In Fig. 2a, we visualize nodes and links with weights exceeding 10 (retweeting over 10 times). The visualization clearly reveals polarization between the pro-vaccine and anti-vaccine communities, as well as between the right-wing and left-wing communities. In the Appendix A6 we provide additional information on communities, in particular details on the inter-community link matrix (Table A5) and demonstrate how these communities interact with each other throughout the pre-vaccination, first/second dose, and third dose periods (Fig. A8). From the network visualization and the inter-community interactions, we observe that the right-wing and left-wing communities interact with both the pro-vaccine and anti-vaccine communities, while the news community is closely situated between the pro-vaccine and left-wing communities. In comparison, the pet-hobby community appears to be more peripheral.

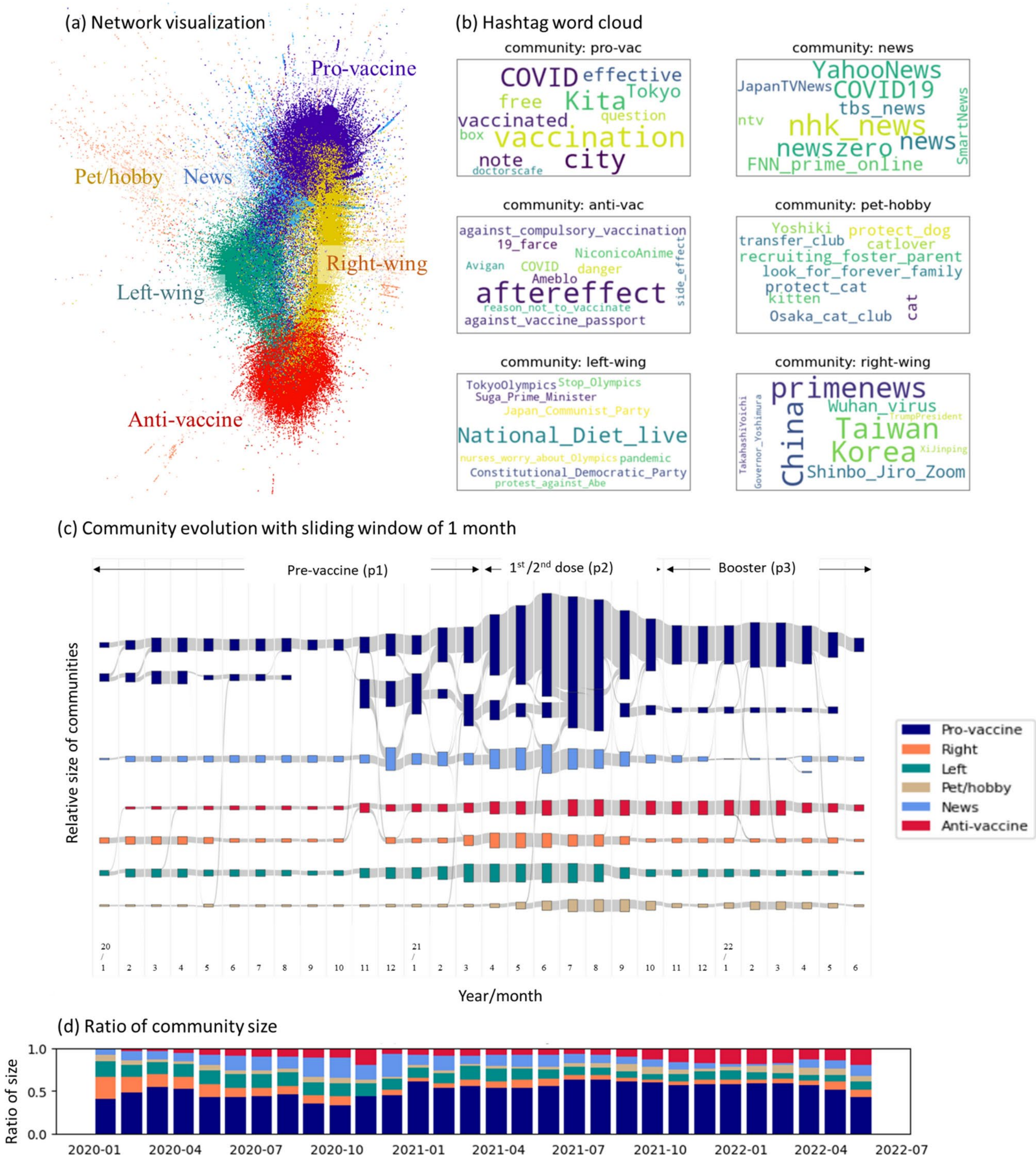

Fig. 2. Visualization of the network communities. (**a**) The interaction between the key communities. The network graph is generated using Gephi, a network visualization tool. We only include here links with weights exceeding 10, which retains 5% of the nodes and 30% of the weighted links. (**b**) Hashtag word cloud shows the 10 hashtags with the highest TF-IDF score (fraction of the frequency of a hashtag in a specific community versus its frequency across all communities). (**c**) Sankey chart illustrating the evolution of key communities when moving by a single month. Each bar symbolizes a community, with its height being proportional to the community's size. The grey links represent the fraction of users transitioning from the previous community to the succeeding community (only links involving more than 1% of users are visualized). Nodes that have been newly introduced or removed are not included in this figure. (**d**) The evolving fraction of users within each community shown in (**c**) over time.

The presence of political communities
Of particular interest is the political community's involvement in vaccine-related discussions. Through sample checks and keyword analysis, we discern that the right-wing community comprises government supporters (given that Japan's ruling party is right-wing) who advocate for the government's vaccine messages. It also encompasses nationalist groups disseminating conspiracy theories about the origin of the COVID-19 virus. The left-wing community involves politicians from opposing parties who endorse vaccines but criticize the government for alleged mismanagement, including vaccination delays. This community also comprises individuals critical of the government, suspecting hidden motives behind vaccination campaigns. Consequently, both pro-vaccine and anti-vaccine users coexist within the right-wing and within the left-wing communities. This observation contrasts with certain other countries, such as the US, where the right-wing and left-wing communities exhibit distinct vaccine leanings. For example, in the US, the right-wing (Republicans) tends to lean more towards anti-vaccine sentiments[29].

The interactions between the political communities and the pro- or anti-vaccine communities change over time. In Fig. A8 we show the interactions between communities throughout the 3 phases. When the first dose of vaccination started, both the left and the right-wing communities interacted more with the pro-vaccine community than with the anti-vaccine community, where the left-wing community exhibits a more heightened activity. However, after the third dose of vaccination began, both the left and the right-wing communities interacted much more heavily with the anti-vaccine community.

Monthly evolution of the key communities
We construct network snapshots using a window size of 3 months when sliding it in steps of 1 month. The reason for choosing a 3-month time window is that it generates stable dynamic communities without excessive overlapping (see Fig. A6). We then detect communities for each snapshot and track the communities across adjacent time windows (detailed in the section "Dynamic community detection").

The Sankey chart in Fig, 2c illustrates the evolution of communities over time. Each bar represents a community detected in each time window. The height of a bar indicates the relative size of the community, and the grey links connecting communities show the fraction of users moving from one community to another. We adopted the Louvain method to detect communities in each network snapshot, which can lead to a resolution issue[30], since the minimum detectable community size depends on the network size. As shown in Fig. fig1a, the number of users (network size) varies over time, causing the natural communities to split or merge occasionally (e.g., the navy pro-vaccine communities merged around July 2021). To overcome this problem, we compared the similarity of nodes between precursor and successor communities to understand their evolution. Additionally, we selected the 100 most influential users from each of the six communities detected in the static network as core users and tracked their community affiliation to group the sub-communities. For more details, please refer to the section "Dynamic community detection".

Keywords analysis aids in comprehending these changes. For instance, users move between the pro-vaccine and news communities in late phase 1 and phase 2, because both pro-vaccine and news communities share updates on vaccination progress. Users transit between the anti-vaccine and political communities, where political groups leverage discussions about vaccines to promote their political ideologies. A notable example is the merger of the left-wing and anti-vaccine communities in January 2020, which is probably spurred by rumors accusing the government of profiting from vaccines. Similarly, the anti-vaccine and right-wing communities merged in November 2020, probably due to the spread of conspiracy theories suggesting that COVID-19 was engineered in a laboratory, thereby casting doubt on vaccine efficacy.

In Fig. 2d, we present the evolving fraction of communities over time. Notably, the pro-vaccine community consistently maintains dominance. It begins to grow and surpasses the 50% mark since November 2020, coinciding with Pfizer's announcement of a vaccine efficacy of 90% with no significant safety concerns (https://www.businesswire.com/ news/home/20201109005539/en/). The pro-vaccine fraction experiences a slight decline during the beginning of phase 3. Conversely, the anti-vaccine community remains notably smaller. Its fraction diminishes during the second phase but rebounds in the third phase. Despite its size, the anti-vaccine community emerges as the most densely connected and highly active community across all phases (see Fig. A9). The news, left-wing, and right-wing communities exhibit relatively larger fractions in the first phase, which subsequently decrease in the second and third phases. By combining these findings with the changes in inter-community interactions (Fig. A8) it becomes apparent that not only the size fractions but also the interaction's focal points shift over time. Specifically, in phase 1, the interactions between communities are more diverse and more evenly distributed across the key communities. However, in phase 2 the focal point of interaction shifts towards the pro-vaccine community, and in phase 3 the focal point further shifts towards the anti-vaccine community.

Possible impact of communities on individual opinion shifts

Composition of opinions in each community
We first present an overall picture of opinion profiles towards vaccination in each community. Specifically, we explore the composition of opinions within each community. Using the same approach as in Fig. 1b, we assign individual users to pro-, neutral-, or anti-vaccine categories based on their opinion scores. Subsequently, we compute the fraction of users within each opinion category for every community. Figure 3a shows the distribution of opinions across communities during the first and second dose (phase 2). The results of all three phases are provided in Fig. A10.

Our observations reveal that communities possess different opinion profiles towards vaccination. The pro-vaccine and pet-hobby communities exhibit a greater presence of pro-vaccine users, followed by the news communities. In contrast, the anti-vaccine community consistently maintains a high fraction of users (above

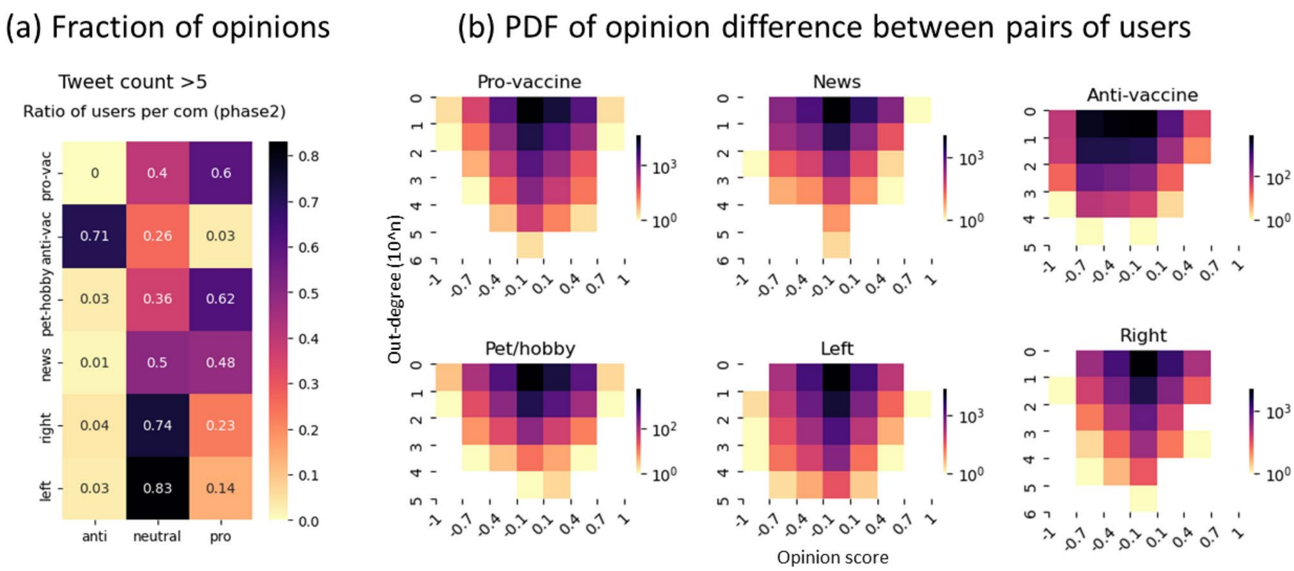

Fig. 3. Mixture of opinions in each community. (**a**) Fraction of opinions in each community across 3 phases. The heatmaps present the average user fraction associated with various opinions within each community. Each row sums to 1. The y-axis denotes the different communities, while the x-axis represents the opinion categories. (**b**) Density distribution of users based on opinion score and logarithmic outdegree. The opinion score (x-axis) is computed by taking the monthly average opinion score of a user. The logarithmic outdegree (y-axis) is derived from the monthly average outdegree of a user within the dynamic network. Both averages are computed excluding periods of user inactivity.

70%) holding anti-vaccine opinions across all three phases. Both the right-wing and left-wing communities are predominantly populated by neutral users. The composition of opinions within each community changes over time. The fraction of pro-vaccine users was the lowest in the first phase when the vaccination has not yet started, followed by a surge in the second phase and a slight drop in the third phase. Therefore, it is meaningful to study the possible influence of an individual's community on opinion shifts in a dynamic way.

To perform a more detailed analysis on the individual opinions within each community, we plot Fig. 3b the distribution of individual's opinion score O^i across varying levels of influence (measured by outdegree centrality). Here, a user's opinion score and outdegree are calculated by averaging the monthly opinion score and monthly outdegree of this user. Figure 3b not only shows the distinction of opinions between communities similar to Fig. 3a, but also the difference between users with high and low influence (outdegree). Evidently, users with higher outdegree tend to lean more towards a neutral stance. This trend is observable across all communities, except for the anti-vaccine community. Upon scrutinizing the posts shared by top influential users, we understand that influencers in pro-vaccine community (e.g. doctors, official government accounts) often disseminate scientific findings, statistical data, and official announcements, while influencers in the anti-vaccine community actively share aggressive anti-vaccine contents. The profile of top-influencing users in each community are provided in Table A4.

Probability of opinion shifts conditioned on community affiliation

In our analysis, we observed that the influencers tend to exhibit greater opinion persistence, while the receivers appear to be more variable in changing their opinions. The distinction in the distribution of opinions becomes particularly pronounced between users with an outdegree exceeding or less than 100 (see Fig. 3b). Consequently, we define those with an outdegree exceeding 100 as influencers and those with an outdegree less than 100 as receivers, then we conduct a separate examination of opinion shifts for influencers and receivers.

We measure the influence of community on opinion change using the following conditional probability. Let V_t^i be the opinion category of a user i at time t, which is determined based on the opinion score O_t^i. Similarly, we use C_t^i to denote the community that user i is affiliated to at time t. The conditional probability (Eq. 2) calculates what is the probability of a user holding an opinion V_{t+1}^i at time t+1 given that he or she held an opinion V_t^i and was in community C_t^i at time t.

$$P(V_{t+1}^i|V_t^i, C_t^i) = \frac{P(V_{t+1}^i, V_t^i, C_t^i)}{P(V_t^i, C_t^i)}. \tag{2}$$

We employ a dynamic community detection method with a 3-month window and 1-month sliding interval, and correspondingly we calculate users' opinion score with a time frame of 3 months. In other words, the temporal difference between t+1 and t corresponds to a span of 3 months. In addition, we notice that users may not be active in all time windows. For example, a user may tweet at time t and stay silent until t+2. If we only compare the opinions between consecutive time t and t+1, we may neglect such less active users. To mitigate this concern, we adjust the preceding opinion V_t^i and community C_t^i to be the most recent opinion V_k^i and community C_k^i when the user was active before t+1 (k<t+1). To track individual users' opinion changes and ensure cohort consistency, our analysis includes only users who tweeted more than five times in each of the successive time windows (e.g., users who tweeted in each Q1 and Q2 2021). Figure A16 shows the number of users included in this analysis.

Here we give a detailed example to explain how to apply Eq. (2). To calculate the conditional probability that a user holds a pro-vaccine opinion in Q2 2021, given that the user held an anti-vaccine opinion and affiliated with the anti-vaccine community in Q1 2021, we apply Eq. (2) as follows: $P(V_{t+1}^i \mid V_t^i, C_t^i) = \frac{P(V_{t+1}^i, V_t^i, C_t^i)}{P(V_t^i, C_t^i)} = \frac{N(V_{t+1}^i, V_t^i, C_t^i)/N}{N(V_t^i, C_t^i)/N} = \frac{N(V_{t+1}^i, V_t^i, C_t^i)}{N(V_t^i, C_t^i)}$, where N represents the total number of users that were active between Q1 to Q2 2021, $N(V_{t+1}^i, V_t^i, C_t^i)$ represents the number of users affiliated with the anti-vaccine community in Q1 2021 and changed their opinion from pro- to anti-vaccine between Q1 and Q2 2021, and $N(V_t^i, C_t^i)$ represents the number of users affiliated with the anti-vaccine community in Q1 2021 and held an anti-vaccine opinion at the same time.

Figure 4a illustrates the conditional probability of receivers transitioning between different opinion categories, with (colored line) or without (dotted black line) the influence from communities. For example, each point on the red line in the center-left subplot Fig. 4a(4) shows the probability that a user with a neutral opinion at time t will adopt an anti-vaccine opinion at time $t + 1$, given their affiliation with the anti-vaccine community. The line chart illustrates the quarterly trend of this probability over time. The results for influencers who hold more persistent opinions are provided in Fig. A11.

By comparing the conditional probability of opinion shifts across communities, we observe pronounced variations in their influence on altering opinions. Notably, users who are within the anti-vaccine community substantially elevate the likelihood of transitioning from neutral to anti-vaccine opinions (Fig. 4a(4)) or from pro- to anti-vaccine opinions (Fig. 4a(7)). Once a user is affiliated to the anti-vaccine community, the chance of changing his or her anti-vaccine opinion is very low (based on Fig. 4a(1)). Conversely, affiliating with the pro-vaccine community increases the probability of shifting from anti- to pro-vaccine opinions (Fig. 4a(3)) or from neutral to pro-vaccine opinions(Fig. 4a(6)). The pet-hobby and news communities also play a role in increasing the probability of pro-vaccine opinions. In contrast, the political communities exert a less influential impact on altering users' opinions to either pro-vaccine or anti-vaccine stances.

Furthermore, in Fig. 4a we observe that the conditional probability of opinion shifts varies over time. The influence of the anti-vaccine community is relatively more consistent throughout the whole period (Fig. 4a(1)(4)(7)). In contrast, the influence of the pro-vaccine community fluctuates (Fig. 4a(3)(6)(9)). In the first phase (pre-vaccination phase), the probability of changing a user's opinion towards pro-vaccine has been low. During the second phase when the vaccination campaign started[3], the pro-vaccine community's influence was much more elevated in changing users' opinions to be pro-vaccine, especially in the second quarter of 2021 when the first dose of vaccination began. However, while the influence of the pro-vaccine community further increased thereafter for users who held pro-vaccine opinions, the influence of the pro-vaccine community decays immediately after the second quarter of 2021 for users who held anti-vaccine or neutral opinions.

This influence decay is an interesting and surprising phenomenon. It could be associated with propaganda intensity. In Fig. A9d, we show the average number of tweets per influencer (with an outdegree ≥ 100). The average number of tweets posted by influencers in the pro-vaccine community peaked in the second quarter of 2021 but dropped immediately afterwards, which coincides the influence decay during the same period. Influencers in the anti-vaccine community have been comparatively more persistent and more active than those in the pro-vaccine community. The influence decay in the pro-vaccine community could also be associated with a psychological phenomenon called message fatigue[31]. Message fatigue occurs when people are repeatedly exposed to similar messages which they find overwhelming or tedium, and they become less emotionally reactive and less likely to be influenced by the messages. According to Fig. 4a, this phenomenon may occur among users inside the pro-vaccine community who hold anti-vaccine or neutral opinions. Those holding pro-vaccine opinions seem not to be affected, as the probability of remaining pro-vaccine is increasingly higher after the second quarter of 2021.

In addition, aligned with our earlier observation of moderate opinion changes, Fig. 4a demonstrates that the chance of keeping an opinion or transiting from neutral opinions to pro- or anti-vaccine (or vice versa) possesses a higher probability than switching drastically between pro-vaccine and anti-vaccine opinions. This is probably associated with the user's preference to connect with those of similar opinions. The difference between pairs of connected users is calculated using Eq. (3).

$$\Delta O_{i,j} = |O_i - O_j| \times \delta_{i,j}. \tag{3}$$

where O_i and O_j denote the opinion scores of two nodes and $\delta_{i,j}$ is a binary value which equals to 1 if the two nodes are connected. In Fig. 4b, we show the distribution of opinion differences between connected users within each community. We also show the random distribution by randomly shuffling the opinion scores of nodes while keeping the links and nodes unchanged (see details in the Appendix A9). We find that the average opinion difference based on real data is smaller than random data for all communities. This propensity may contribute to the creation of echo chambers, where individuals are exposed to limited diversity of opinions. The difference of opinion distance between actual and random connection is smallest for the pet-hobby community, suggesting that users within this community are less likely to connect based on vaccine preference, but more on other factors such as hobbies.

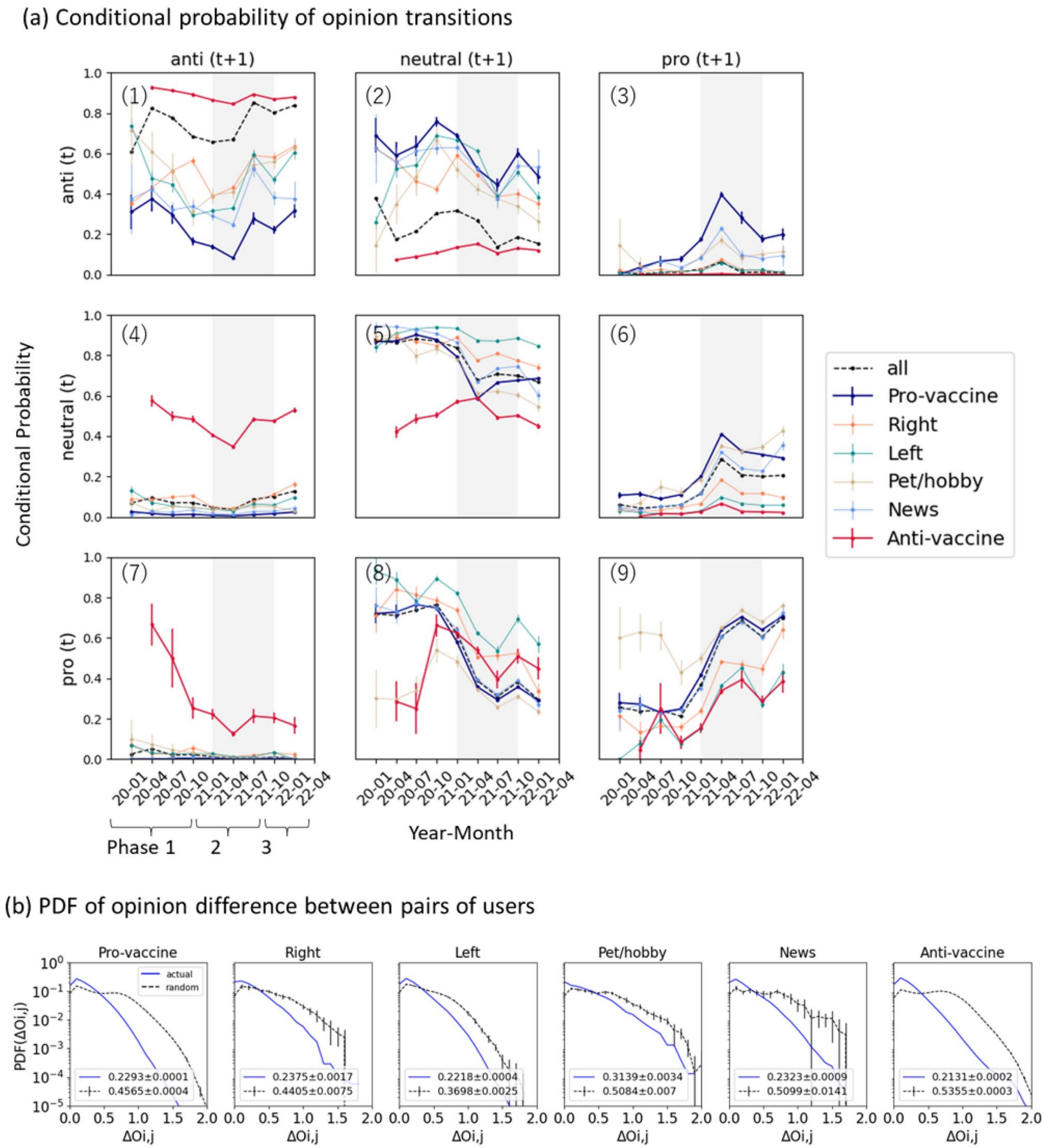

Fig. 4. Individuals' opinion transitions in each community. (**a**) Conditional probability of receiver's (users with an outdegree smaller than 100) opinion changes depending on the community affiliation. The x-axis shows time period (quarterly) and y-axis shows the conditional probability of a user changing to opinion V_{t+1}^i given that he or she was previously holding an opinion V_t^i and affiliated to community C_t^i. There are 9 sub-figures, each showing the likelihood of changing from pro-, neutral-, anti-vaccine opinions at time *t* to pro-, neutral- and anti-vaccine opinions at time *t+1*. Colored lines show the temporal conditional probability with an affiliation to a community, while the dashed black lines show the overall conditional probability regardless of communities. The error bars depend on the sample size. A greater deviation corresponds to a smaller sample size, consequently leading to lower reliability. The background is color-coded to different 3 vaccination phases. (**b**) Distribution of opinion difference between pairs of connected users $\Delta O_{i,j}$. The blue solid line depicts the semi-log PDF of opinion difference based on real data, while the black dashed line depicts the semi-log PDF of opinion difference based on randomly shuffled data. The data annotations in the legend box are the average opinion difference ± error bar (calculated by standard deviation/$\sqrt{\text{Total Data Points}}$) of real data and random data respectively.

Sensitivity analysis

In above "Results" section, we assumed three thresholds when filtering users and calculating their opinion scores:

- In each time period (quarters), only select users who posted over 5 tweets.
- Regarded users with an outdegree of larger than 100 as influencers while the rest as receivers.
- When determining vaccine stance based on the opinion score, selected ±0.3 as the threshold.

While this paper provides the rationale for selecting these thresholds, it is essential to assess how variations in these thresholds affect the results. In Appendix A2, we apply Bayesian inference to predict user opinions with smaller sample sizes as an alternative to filtering by tweet count. In Appendix A11 (Fig. A12–A15), we evaluate different threshold values for the three parameters described above. Table A7 compares the outcomes across varying thresholds, demonstrating that the main conclusions in this section remain robust.

Discussion

Our study evaluates the individual users' opinion dynamics with respect to the COVID-19 vaccine in Japan's Twitter and investigates the Twitter community's impact on opinion shifts over time. The key contribution of this paper is the introduction of a novel approach to dynamically quantify individual users' vaccine opinions and integrating this with network analysis to examine how community affiliations influence opinion changes. Applying this approach to long-term and large-scale empirical Twitter data, we uncover notable phenomena of communities influencing opinion changes towards COVID-19 vaccination, offering insights on understanding in general opinion shifts that specifically help vaccine promotions in the future.

Generally, we also find that users with a lower outdegree centrality, or a lower tweeting activity level are more likely to be influenced to change opinions. Also, most users' opinions do not shift drastically within short time (e.g. from pro-vaccine to the opposite anti-vaccine). Therefore, users holding a neutral opinion are more likely to change to a pro- or anti-vaccine. Some communities may increase the likelihood of shifting users' opinions toward pro-vaccine or anti-vaccine stances. In the following, we summarize the characteristics of key communities and evaluate their effectiveness in driving opinion shifts.

The anti-vaccine community is a relatively minor community, yet it stands out as a closely connected and highly active group comprising individuals who hold strong anti-vaccine beliefs. This community is not only active in terms of tweeting but also in engaging with political communities. Users affiliated with the anti-vaccine community have a notably higher probability of either maintaining or transitioning to anti-vaccine opinions.

In contrast, the pro-vaccine community, although dominant in terms of size, mainly consists of intermittent users. As a result, the average activity level and density are comparatively lower. Influencers within the pro-vaccine community include government accounts and medical professionals, who tend to express themselves in a more neutral manner. Nevertheless, the pro-vaccine community, together with the news community, seems to contribute to the increased probability of users shifting toward pro-vaccine opinions. This is probably because Japanese individuals have higher trust in sources like television news, newspapers, professionals, and medical workers when making vaccination-related decisions[32].

The right-wing and left-wing political communities exhibit a relatively balanced fraction of pro-vaccine and anti-vaccine users. Politicians within these communities leverage on vaccine-related topics to propagate their political agendas. However, probably because these communities tend to focus more on political matters rather than vaccination itself, their possible influence in steering people's opinions toward vaccines is somewhat diminished.

We summarize three key insights of our study. First, in the second phase, the active vaccine campaigns were carried out by the Japanese government and mass media[3]. Coincidentally, we observe that the probability of shifting to pro-vaccine opinion became higher and the ratio of anti-vaccine opinions reduced in many communities especially in the second quarter of 2021 when the campaign activity was most active. Therefore, we interpret that it is useful to involve official accounts, professionals, and news agencies in vaccine promotion efforts through social media platforms. Second, we find that the influence of the pro-vaccine community on users who hold neutral or anti-vaccine opinion is less persistent than the anti-vaccine community, which could be related to the reducing pro-vaccine campaign intensity and/or increasing message fatigue. Therefore, it might be useful for social media influencers to maintain a consistent level of tweeting pro-vaccine contents and to enhance engagement by being more active and creative in their messaging (e.g., less neutral contents). Third, the recognition that the anti-vaccine community has a strong and persistent influence on its attached users underscores the importance of actively penetrating and engaging with users within the anti-vaccine community. This implies a necessity to share accurate and credible information to counter conspiracy theories. Interventions could involve measures such as altering the Twitter algorithm or engaging with anti-vaccine influencers to correct misleading narratives[7].

While this study provides valuable insights into how communities influence vaccine opinions, several limitations should be acknowledged.

(1) The primary limitation pertains to data. Our social network construction relies on retweet data, and the Twitter API solely provides information about the original tweeter and the retweeter. However, it does not reveal the intermediary users through whom retweeters accessed the information. This leads to potential inaccuracies in the network structure.
(2) We adopted the machine learning approach to classify tweets to pro-, neutral or anti-vaccine categories. Despite its capability of processing large-scale text data, the accuracy rate (78%) and recall rate (53%)—calculated by us based on 10,000 randomly selected and manually labeled samples—may affect our ability to precisely capture users' actual opinions. Recent advancements in Large Language Models hold great poten-

tial for enhancing opinion and emotion detection tasks[33]. In the future, we aim to explore their application in opinion classification to improve accuracy.

(3) The time resolution used for analyzing the dynamic community and opinions poses a limitation. Aggregating networks and individual opinions on a quarterly basis enhances traceability of communities and improves inclusivity of less active users, but compromises the accuracy of capturing the precise timing of changes.

(4) Although we considered a cohort of the same active users active in successive quarters, we did not track the exact same pool of users throughout the entire 10-quarter (2.5-year) period. This is since only a small fraction of users remains consistently active over such a long timeframe. Future studies may overcome this limitation by leveraging larger social media data or combining longitudinal surveys to analyze more stable user cohorts over extended periods.

(5) In this study, our definition of 'neutral' encompasses a broad range of topics, including skepticism about vaccination, subjective content, and political discussions, which do not explicitly indicate vaccine preference. In future work, we will refine this category by focusing specifically on users who express skepticism about vaccination and analyzing their opinion dynamics over time.

(6) This study exclusively focuses on Japanese Twitter data. Exploring this phenomenon across various countries or languages, based on our framework, would be valuable to comprehend opinion change dynamics within diverse cultural and social contexts. Conducting a cross-country comparative study would be our future work.

Methods

Here, we provide a brief description of the dataset and an overview of the methods used for vaccine opinion classification, dynamic community detection, and community significance testing. Detailed information can be found in the Appendix.

Dataset

We use academic Twitter API (/2/tweets/search/all/) to collect the historical archive of tweets and retweets related to the COVID-19 vaccine in Japan. The search query includes the following criteria:

(1) Include the Japanese keyword " ワクチン " ("vaccine" in English). Because the keyword "COVID-19" is often omitted when people talk about the COVID-19 vaccine, here we only filter by "vaccine".

(2) Exclude the keywords that are likely spam or are related to other types of vaccine such as HPV vaccine (see detailed list of excluded words in Table A6 when obtaining COVID-19 vaccine related tweets from Twitter). We did not exclude the keyword influenza because Influenza is often mentioned in tweets related to COVID-19.

(3) Time range: Jan 2020 to May 2022. This period encompasses the pre-vaccination, first, second, and booster dose periods.

The search resulted in 41,398,876 original tweets and 81,454,507 retweets, posted by 5,621,325 tweeters and 2,927,186 retweeters respectively. Among these users, 1,657,048 users who both tweeted and retweeted are the focus of this paper, which constitutes approximately 1% of the Japanese population.

Vaccine opinion classification

In this study, we aim to classify each tweet to a pro-, neutral or anti-vaccine stance so that we can infer a user's opinion score based on the tweets that he or she posted. We define the vaccine stances following the criteria described in Table 1. It is worth noticing that the neutral stance includes not only showing uncertainty about getting vaccinated, but also a neutral statement about vaccination or topics not directly related to taking vaccine or not. In Table A8, we show sample tweets for each opinion category.

To build a supervised classifier, we prepared a set of labeled data consists of 10,000 randomly selected and manually labeled tweets. We employed two annotators to label 10,000 sample tweets. If there was a disagreement between the two annotators, we invited a third annotator to decide which label should be adopted. Then we incorporated the Word2Vec word embedding, n-grams dictionary, and user's retweet network communities as features and trained the classifier using the Support Vector Machine (SVM) model. The classifier model is

Category	Contents
Pro-vaccine	Promote vaccine, sharing the effectiveness and benefits of vaccine
	Condemn anti-vaccine contents
	Look forward to vaccination
	Plan to or have taken a vaccine (even with side effects)
Anti-vaccine	Blame vaccine
	Do not want or plan to take vaccine
	Disagree vaccine to be made compulsory
Neutral	News or official announcement (except for sharing the effectiveness of vaccines)
	Scientific statement such as what is mRNA vaccine
	Not sure whether to get vaccinated or not
	Not directly related to taking vaccine or not, such as discussions about Olympics, government, etc.

Table 1. Definition of opinion categories.

capable of classifying a Japanese tweet to the pro-, neutral or anti-vaccine category with a precision rate of 78% and a recall rate of 53%. Notably, the model excels in identifying pro- and anti-vaccine content, yielding precision rates exceeding 80% and recall rates surpassing 60%. The details of the model and its performance are provided in the Appendix A1 (Fig. A1 and Table A1).

Dynamic community detection

Previous research found that most real network systems are inherently dynamic and change over time, influencing both their topology and the processes that propagate across the network[34,35]. Communities within social networks are subject to dynamic shifts, including processes such as merging, splitting, growth, and contraction[36]. At the individual level, users exhibit dynamic behavior by remaining within, joining, or departing from communities, which signifies changes in their social interactions. Hence, the dynamic detection of communities is essential to capture these intricate behavioral changes.

As pointed out by Rossetti and Cazabet[37], instability is a key challenge inherent in dynamic community detection. The retweet network can exhibit even more significant instability due to the emergence and fading of nodes at various time points[27] due to users' retweeting activities. We adopt the following steps to mitigate the instability issue. For details, refer to the Appendix A4.

(1) Choose an appropriate aggregation time window to create network snapshots. We performed tests with various time windows ranging from 1 to 4 months, sliding by 1 month, to increase overlap from 0 to 3 months. Then we apply dynamic community detection to all options and visualize the results using the Sankey chart (see Fig. A6). Our findings indicate that a 3-month window is optimal for our dataset, offering stable and traceable communities while capturing dynamic shifts.

(2) Adopt the Ensemble Louvain method[27] to detect stable communities for each network snapshot. We choose the Louvain method for community detection because it is one of the fastest algorithms for applying to large-scale networks[38,39]. However, the Louvain method uses random sequence to choose one node at a time to decide whether to merge it with its neighbor communities, which potentially leads to unstable results. To address this, we run the Louvain method (resolution = 1.0) with different random seeds 100 times. Links connecting nodes that are clustered in the same community with a probability greater than 0.9 are retained.

(3) Combine the similarity-based approach and core-nodes based approaches to connect communities detected in adjacent time windows.

Chen et al. introduced a similarity-based approach to compare the predecessor and successor community pairs in evolving networks[40]. This approach involves evaluating the fraction of a predecessor community that is present in a successor community and comparing the sizes of the communities to identify network evolution scenarios, such as merge, split, grow, shrink, birth, and vanish. We further formulated their decision rules to simplify the algorithms (see Table A2).

However, the communities may not merge or split naturally, when applying the adopted Louvain method. The Louvain method is based on modularity maximization, which faces a resolution limit issue[30], where the minimum detectable community size depends on the total network size. This limit can obscure the visibility of small communities as networks grow. In our case where the network snapshot sizes change significantly over time, this limitation can cause natural communities to split or merge. To overcome this issue, we select the top 100 influencers (core nodes) with the highest outdegree centrality from each of the six key communities as reference points. We compute the fraction of each influencer type within each community and categorize the community based on the dominant influencer type whose fraction exceeds 50%.

Significance test on individual communities

Many community detection algorithms, such as the popular Louvain method and Infomap, do not evaluate whether the network structure is inherently likely or unlikely to generate highly clustered communities. In other words, the optimal communities detected by these algorithms may not always be unique and meaningful. An interesting example reported by Tiago[41] is a random network containing 13 nodes with degree of 20 and 230 nodes with degree of 1. Given such network settings, regardless of how nodes are randomly connected, it is highly likely to generate clustered networks. Therefore, the Louvain method will likely return results with high modularity even for random cases, which may not be always meaningful.

Some community detection algorithms such as the Stochastic Block Model (SBM)[41,42] could solve the problem. But these inference approaches are usually more computationally expensive in terms of speed and memory than the Louvain method[43–45]. In our paper, we propose a simple statistical test to evaluate the significance of communities, that can be used to evaluate individual communities detected using any algorithm. The idea is to compare the fraction of internal links between the actual and random communities. Random networks are created by randomly shuffling a network while preserving each node's degree.

The communities detected based on the actual network and random networks are not directly comparable, because the nodes in each community differ and the number of communities varies. Therefore, we first compare the actual network and the random networks (see the example in Appendix A5—Fig. A7) by examining the fraction of links within the set of nodes belonging to a community in the actual network (actual fraction A vs. random fraction B). Then, we compare the fraction of links in a random network and the random network of a random network by looking at a set of nodes belonging to a community in the random network (random fraction C vs. random's random fraction D). Here, A vs. B, as well as C vs. D are directly comparable because they are based on the same set of nodes and community. If a network's settings are prone to generate clusters, the fraction of internal links in both A and C should be much larger than B and D, causing the gap between (A-B) and (C-D) to be smaller. Therefore, the gap between (A-B) and (C-D) can suggest the significance of the communities. The detailed algorithm is provided in the Appendix A5—Table A3.

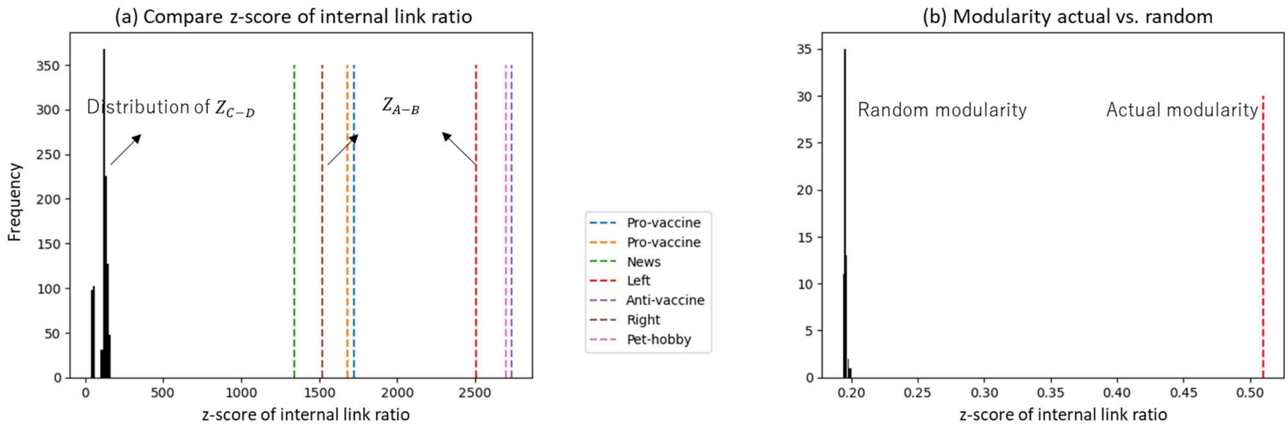

Fig. 5. Comparing the z-score of internal link ratio. (**a**) Comparing the z-score of internal link ratio. Here we compare the z-score of the actual communities versus the communities detected based on random networks. The black histogram chart depicts the distribution of Z_{C-D} while the colored dashed lines show the z-score (x-axis) of actual communities Z_{A-B}, each represent one community. (**b**) Comparing the modularity of actual and random networks. The dotted red line shows the modularity of the communities detected based on the actual network, while the black histogram shows the frequency of modularity of communities based on 100 random networks.

We apply this method to evaluate the communities detected from the Twitter retweet network (the network snapshot between 2021/1/1 to 2021/3/31) using the Louvain method. Figure 5a shows the result of comparing (A-B) and (C-D) by calculating the z-score $Z_{A-B} = \frac{A-\langle B\rangle}{\sigma(B)}$ and $Z_{A-B} = \frac{C-\langle D\rangle}{\sigma(D)}$. To create random scenarios, the network is shuffled 100 times while keeping each node's degree. The colored dashed lines show the Z_{A-B} (x-axis) of the actual communities, while the black histogram shows the distribution of Z_{C-D} for all random communities detected. We observe that the p-values (the probability that the actual z-score $\leq$ random z-score) of all actual communities are 0 based on 100 random shuffles, suggesting that all communities are significant. Additionally, in Fig. 5b, we compare the modularity between the actual and random networks. By maintaining each node's degree when randomly shuffling the networks, we ensure that the modularity is comparable between the actual and random networks. We can observe that the modularity of the actual network (red dashed line) is much larger than the random networks (black histogram chart), which agrees with the conclusion in Fig. 5a.

Data availability
The X (Twitter) data cannot be open to the public due to privacy policy, but similar data can be obtained using X (Twitter) API (https://developer.twitter.com/en/products/twitter-api). Details are provided in the paper. However, aggregated and anonymized data can be shared upon request. For data inquiries, please contact the corresponding author Misako Takayasu (takayasu@comp.isct.ac.jp).

Code availability
Original code has been deposited at Github (https://github.com/WUQIANYUN1/VaccineTwitter_Community_Opinion) and is publicly available as of the date of publication.

Received: 27 November 2024; Accepted: 21 March 2025
Published online: 05 April 2025

References

1. Polack, F. P. et al. Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
2. De Figueiredo, A., Simas, C., Karafillakis, E., Paterson, P. & Larson, H. J. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: A large-scale retrospective temporal modelling study. *Lancet* **396**, 898–908 (2020).
3. Mori, H. & Naito, T. A rapid increase in the COVID-19 vaccination rate during the Olympic and Paralympic Games 2021 in Japan. *Hum. Vaccines Immunother.* **18**, 1–2 (2022).
4. Nomura, S. et al. Reasons for being unsure or unwilling regarding intention to take COVID-19 vaccine among Japanese people: A large cross-sectional national survey. *Lancet Reg. Health West. Pac.* **14**, 100223 (2021).
5. Uehara, M. et al. Measuring concerns about the COVID-19 vaccine among Japanese internet users through search queries. *Sci. Rep.* **12** (2022).
6. Kobayashi, R. et al. Evolution of public opinion on COVID-19 vaccination in Japan: Large-scale Twitter data analysis. *J. Med. Internet Res.* **24**, e41928 (2022).
7. Miyazaki, K., Uchiba, T., Tanaka, K. & Sasahara, K. Aggressive behaviour of anti-vaxxers and their toxic replies in English and Japanese. *Human. Soc. Sci. Commun.* **9** (2022).
8. Toriumi, F., Sakaki, T., Kobayashi, T. & Yoshida, M. Anti-vaccine rabbit hole leads to political representation: The case of Twitter in Japan. *J. Comput. Soc. Sci.* (2024).
9. Report on information communication media use time and behavior in fy2022. Report (Ministry of Internal Affairs and Communications, 2023).

10. Waniek, M., Holme, P., Cebrian, M. & Rahwan, T. Social diffusion sources can escape detection. *iScience* **25**, 104956 (2022).
11. Bovet, A. & Makse, H. A. Influence of fake news in twitter during the 2016 US presidential election. *Nat. Commun.* **10**, 7 (2019).
12. Zhao, Z. et al. Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Sci.* **9**, 7 (2020).
13. Metaxas, P. et al. What do retweets indicate? Results from user survey and meta-review of research. *Proceedings of the International AAAI Conference on Web and Social Media* **9**, 658–661 (2021).
14. Newman, M. E. J. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* **68** (2003).
15. Noldus, R. & Van Mieghem, P. Assortativity in complex networks. *J. Complex Netw.* **3**, 507–542 (2015).
16. Yuan, X., Schuchard, R. J. & Crooks, A. T. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Soc. Media + Soc.* **5** (2019).
17. Cossard, A. et al. Falling into the echo chamber: The Italian vaccination debate on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* **14**, 130–140 (2020).
18. Milani, E., Weitkamp, E. & Webb, P. The visual vaccine debate on Twitter: A social network analysis. *Media Commun.* **8**, 364–375 (2020).
19. Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, L. M. & Moreno, Y. Evolutionary dynamics of group interactions on structured populations: A review. *J. R. Soc. Interface* **10**, 20120997 (2013).
20. Ojea Quintana, I., Reimann, R., Cheong, M., Alfano, M. & Klein, C. Polarization and trust in the evolution of vaccine discourse on Twitter during COVID-19. *PLoS One* **17**, e0277292 (2022).
21. Johnson, N. F. et al. The online competition between pro- and anti-vaccination views. *Nature* **582**, 230–233 (2020).
22. Cotfas, L.-A. et al. The longest month: Analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access* **9**, 33203–33223 (2021).
23. Mitra, T., Counts, S. & Pennebaker, J. Understanding anti-vaccination attitudes in social media. *Proceedings of the International AAAI Conference on Web and Social Media* **10**, 269–278 (2021).
24. Morales, A. J., Borondo, J., Losada, J. C. & Benito, R. M. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos* **25** (2015).
25. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124** (2020).
26. Mønsted, B. & Lehmann, S. Characterizing polarization in online vaccine discourse—A large-scale study. *PLoS One* **17**, e0263746 (2022).
27. Evkoski, B., Mozetič, I., Ljubešić, N. & Kralj Novak, P. Community evolution in retweet networks. *PLoS One* **16**, e0256175 (2021).
28. Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69** (2004).
29. Viswanath, K. et al. Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health* **21**, 818 (2021).
30. Fortunato, S. & Barthélemy, M. Resolution limit in community detection. *Proc. Natl. Acad. Sci.* **104**, 36–41 (2007).
31. So, J., Kim, S. & Cohen, H. Message fatigue: Conceptual definition, operationalization, and correlates. *Commun. Monogr.* **84**, 5–29 (2017). https://doi.org/10.1080/03637751.2016.1250429.
32. Hori, D., Takahashi, T., Kaneda, Y., Ozaki, A. & Tabuchi, T. The influence of information sources on intention changes to receive COVID-19 vaccination: A prospective cohort study in Japan. *Environ. Health Prev. Med.* **28**, 10–10 (2023).
33. Ziems, C. et al. Can large language models transform computational social science?. *Comput. Linguist.* **50**, 237–291 (2024).
34. Shvydun, S. & Mieghem, P. V. System identification for temporal networks. *IEEE Trans. Netw. Sci. Eng.* **11**, 1885–1895 (2024).
35. Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87**, 258701 (2001).
36. Granell, C., Darst, R. K., Arenas, A., Fortunato, S. & Gómez, S. Benchmark model to assess community structure in evolving networks. *Phys. Rev. E* **92**, 012805 (2015).
37. Rossetti, G. & Cazabet, R. Community discovery in dynamic networks. *ACM Comput. Surv.* **51**, 1–37 (2019).
38. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
39. Fogués, R. L., Such, J. M., Espinosa, A. & Garcia-Fornes, A. Bff: A tool for eliciting tie strength and user communities in social networking services. *Inf. Syst. Front.* **16**, 225–237 (2014).
40. Chen, Z., Wilson, K. A., Jin, Y., Hendrix, W. & Samatova, N. F. Detecting and tracking community dynamics in evolutionary networks. In *2010 IEEE International Conference on Data Mining Workshops* 318–327 (2010).
41. Peixoto, T. P. *Descriptive vs. Inferential Community Detection in Networks* (Cambridge University Press, 2023).
42. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137 (1983).
43. Barabási, A.-L. & Pósfai, M. *Network Science* (Cambridge University Press, 2016).
44. Peixoto, T. P. Bayesian stochastic blockmodeling. *Advances in Network Clustering and Blockmodeling* 289–332 (2019).
45. Jin, D. et al. A survey of community detection approaches: From statistical modeling to deep learning. *IEEE Transactions on Knowledge and Data Engineering* 1–1 (2021).

Acknowledgements
We thank Kenta Yamada for his suggestion on vaccine opinion classification.

Author contributions
M.T. and S.H. were the leaders of this project, designed the whole research plan and directed the writing of the manuscript. Q.W. analyzed the raw data, performed the numerical calculations, and wrote the manuscript. Y.S., H.T. and S.H. developed methods of data analysis and revised the manuscript.

Funding
This work was supported by JSPS KAKENHI Grant Number 23K28192. This work was partially supported by the "Research Project for Overcoming Coronavirus Disasters Launched" by the Institute of Innovative Research, Institute of Science Tokyo.

Declarations

Competing interests
The authors declare no competing interests

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/1

0.1038/s41598-025-95595-6.

Correspondence and requests for materials should be addressed to S.H. or M.T.