# ARTICLE

Check for updates

# Optimising self-organised volunteer efforts in response to the COVID-19 pandemic

Anping Zhang [1,3], Ke Zhang [1,3], Wanda Li [1], Yue Wang [2], Yang Li [1✉] & Lin Zhang [1]

Crowdsource volunteering efforts have contributed significantly to pandemic response and recovery during the COVID-19 outbreak. In such efforts, individual volunteers can collaborate to achieve rapid mobilisation toward emergent community demands. In this study, we quantitively study this phenomenon using the concept of self-organisation, by proposing a data-driven framework to investigate when and how self-organisation emerged during the pandemic response and how it relates to effectiveness of volunteer organisations in general. Using activity data collected from a mobile volunteer platform in Shenzhen, China, we found that volunteers' task participation and social and task preferences show multiple phases of self-organisation in response to changing epidemic situations and centralised interventions. Simulation experiments further show that the self-organised volunteer system can become more responsive and more robust in the face of uncertain community demands with minimal centralised guidance.

[1] Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, University Town, Shenzhen, China. [2] Tsinghua University, Beijing, China. [3]These authors contributed equally Anping Zhang, Ke Zhang. ✉email: yangli@sz.tsinghua.edu.cn

## Introduction

Community participation through volunteering has been essential in the collective response to the COVID-19 pandemic, from compliance with lockdowns to providing community support in conjunction with public agencies and community organisations (Marston et al. 2020; Miao et al. 2021). To cope with the large number of demands for assistance, many countries have utilised crowdsource systems for organising volunteer efforts, including the NHS Volunteer Responders programme in the UK (Marston et al. 2020) and the CrowdSource Rescue programme in Houston, Texas, USA (Click2Houston 2021). In such systems, volunteers perform micro-tasks on a local scale and are mostly organised in a decentralised way. When properly designed, crowdsource systems can be highly effective in achieving collective goals that meet the most urgent needs of a community (Howe 2006; Besaleva and Weaver 2013; Riccardi 2016; Schimak et al. 2015), such as the targeting of help to specific groups of people when official pandemic relief organisations are overloaded. To better organise volunteer efforts in response to COVID-19 and other unexpected crises, it is important to understand how effective organisation has been achieved through the decentralised volunteer efforts during the COVID-19 pandemic, and how the crowdsource volunteer system organisation can be made more efficient.

In this study, we address the above questions about crowdsource volunteer systems using data collected from the "Anti-Pandemic Pioneers" project (a.k.a. Pioneers), a mobile platform for organising community-level volunteer activities in Shenzhen, China. Launched at the start of the COVID-19 epidemic in February 2020, this platform was quickly adopted by the community staff and volunteers in Shenzhen, with a total of 80,043 users signed up in the first year. Users can organise short-term group volunteer activities (or tasks), such as checking people's body temperature at apartment entrances, delivering packages for home-quarantined families, and helping with the crowd by posting on the platform as an organiser. They can also sign up to join any open tasks and their participation will be tracked by location-based check-ins. As the pandemic situation in Shenzhen had eased significantly by the end of 2020, this project continued to grow as a long-term platform for organising other volunteer activities, such as environmental protection and community education. We choose Pioneers for our case study as it provides a large collection of volunteer activity data during and after the pandemic. Meanwhile, the user-driven way of organisation in Pioneers is also suitable for studying collective volunteer behaviour. Unlike other crowdsource volunteer systems where the system matches volunteer resources to people who requested aid, Pioneers allows users to find volunteer activities and groups they like through exploration. It also enables proactive community members to step into leadership roles in time of need.

From the one-year user activity data collected by Pioneers, we observed recurring collective patterns in volunteer behaviours, such as participation frequencies and preferences towards different social groups and task topics. For instance, at the start of the project, ad hoc volunteer groups were quickly formed by people who frequently joined activities led by the same organiser, as shown in Fig. 1a. Increasing specialisation in certain task types can also be observed in the activity traces of these volunteer groups (Fig. 1b). Such behaviour patterns are a positive indicator of effective volunteer organisations since strong internal communication among volunteers increases volunteer retention (Bauer and Lim 2019), and specialisation leads to higher efficiency in completing designated tasks (Miao et al. 2021). However, the formation of stable behaviour patterns can be disrupted over time, often in response to external stimuli, such as a sudden change in the pandemic situation or administrative policies.

These stimuli may alter old volunteer habits and cause new gaps between societal demands and community efforts. How fast new global patterns can be re-established after such disruptions indicate the robustness of the crowdsource organisation. Thus, it is important to systematically detect each occurrence of global behaviour patterns and to analyse their dynamics and causes.

In this study, we adopted the concept of social self-organisation to explain the formation of global patterns in volunteer behaviours and examine how they change over time. Widely studied in social sciences and humanities (Bonabeau et al. 1997; Fuchs 2006), *self-organisation* is the phenomenon whereby structures appear at the global level of a system from interactions among its lower-level components (Camazine et al. 2001). In a crowdsource volunteer setting, people organise or participate in volunteer activities autonomously with simple interactions, yet they can efficiently fill in gaps between help demands and available public resources at a global scale (Marston et al. 2020). The stable behavioural patterns formed in this process are analogous to the "global structures" in the definition of self-organisation. However, previous studies do not provide a clear mathematical definition of self-organisation based on empirical observations of human behaviour.

To this end, we proposed a data-driven approach to measure self-organisation effects for several types of volunteer behaviours: volunteer participation rate, organiser choice, and an organiser's task choice. Our approach is inspired by the work of Atlan on modelling the organisation effect in information systems (Atlan 1974), where self-organisation is characterised by the process in which system entropy first increases and then decreases in the absence of an apparent external force. We introduce a more general self-organisation detection approach based on the double-exponential model (Jia and Xiaoqing 2006), which identifies different types of organisational effects. Moreover, since the self-organisation of behavioural patterns recurs over time, we developed an algorithm to detect the time intervals of the most salient self-organisation effects from time-series data. Quantitative data from the Pioneers project allowed us to correlate the self-organisation characteristics in different districts of Shenzhen with their populations and district characteristics. Additionally, we identified the causal pathway for self-organised behavioural processes among various internal and external factors using a modified causal network discovery algorithm. In retrospect, these results also reflect how well volunteers responded to different community needs in Shenzhen during the COVID-19 outbreak and subsequent period of COVID-19.

One of the major goals of this study was to determine the appropriate level of centralised supervision for such platforms and analyse how the system may react according to the volatility of volunteer demands. Due to the unavailability of data for direct comparison with other organisational schemes during the pandemic, we utilised an agent-based simulation to test three organisational schemes in a simplified crowdsource scenario, self-organised, centralised, and hybrid, under different system parameters. Our findings provide insights into ways to increase volunteer participation in crowdsourced volunteer systems. We also offer practical advice on how self-organisation can be used to improve the efficiency of volunteer organisations that respond to rapidly changing community needs.

The paper is organised as follows: In the "Methods" section, we introduce the details of our framework for self-organisation analysis of volunteer activity data. Then, in the subsequent section, we describe the key results from the Pioneers case study, including the self-organisation intervals for the three types of volunteer behaviours of interest and their causal factors. Next, we present a simulation study that compares different organisational
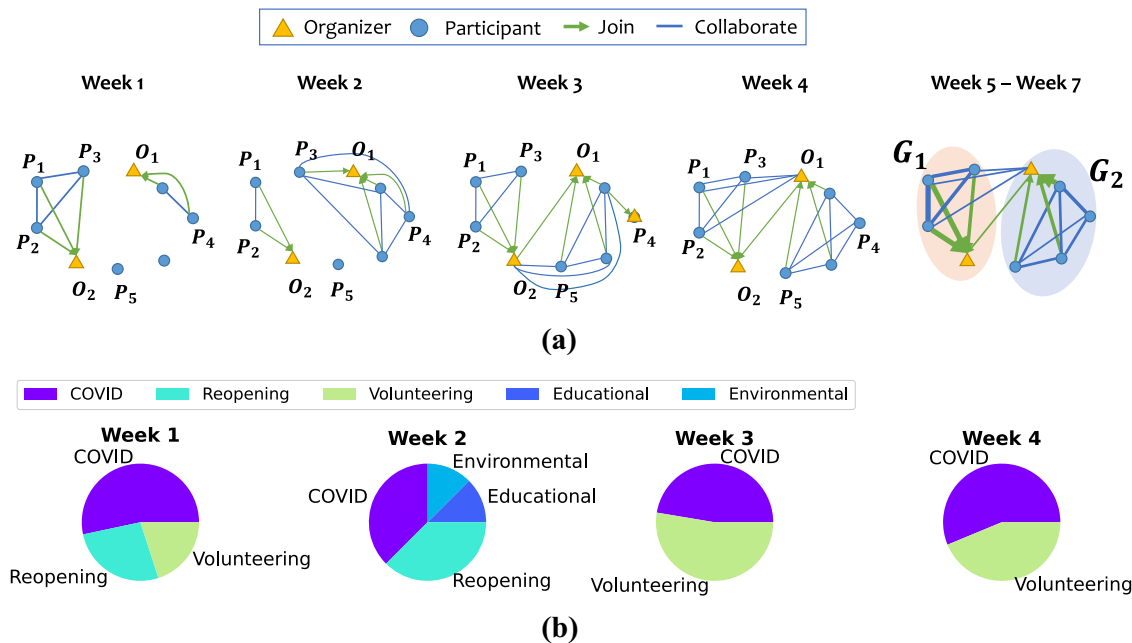
**Fig. 1 Self-organised volunteer behaviour patterns. a** A sample graph showing the social interactions of Pioneers users in different weeks. Each directed edge $P_i \rightarrow O_j$ indicates that user $P_i$ (denoted by a circle) participated in a task issued by organiser $O_j$ (denoted by a triangle) in this week. Likewise, an undirected edge indicates the collaboration between two participants in a task. In the final stage of the social graph (i.e., weeks 5–7), groups ($G_k$) are formed through the volunteering process. The edge widths are proportional to the participation frequencies of specific user pairs. The thicker an edge is, the more often the participant joined the organiser's tasks. **b** Changes in volunteer task topics posted by an activity organiser as time changes. Generally, organisers tend to focus on several specific task topics after a few attempts.
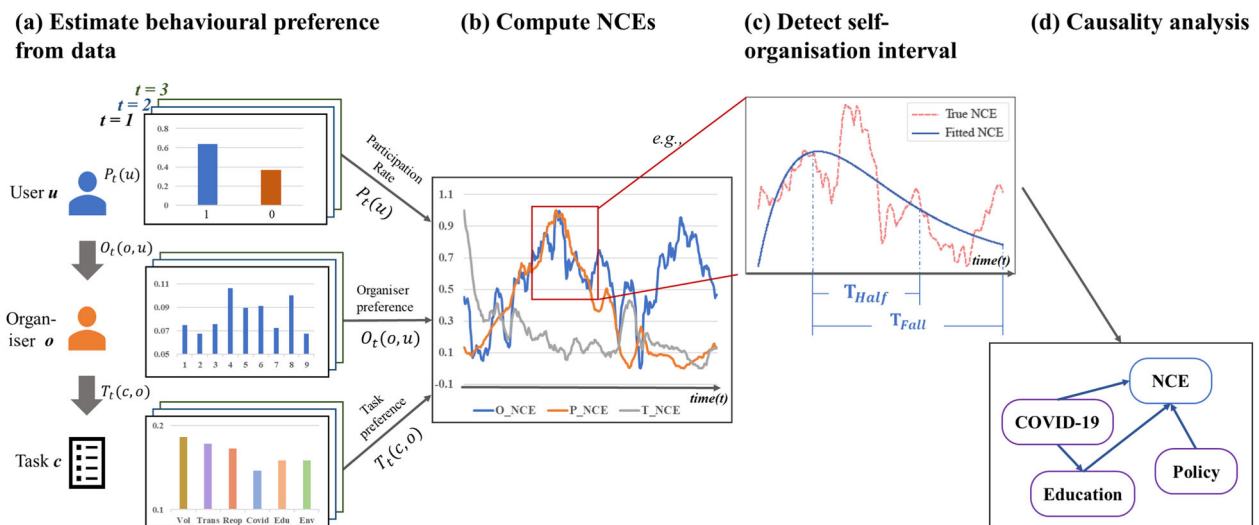


**Fig. 2 Overview of our volunteer behaviour model. a** Estimate the parameters in the volunteer behaviour model from the Pioneers data. The parameters, including users' participation rate $P_t(u)$, organiser preference $O_t(o, u)$ and task preference $T_t(c, o)$, are visualised as probability mass functions for each timestep. **b** Capture the uncertainties in volunteer behaviour by computing the normalised conditional entropy (NCE) from model parameters. **c** Detect the time intervals when the self-organisation effect exists by fitting a double-exponential model; quantify the self-organisation effect in terms of organisational speed measured by $T_{Half}/T_{Fall}$. **d** Analyse the dynamic factors that have caused self-organisation change in volunteer behaviours using time-series based causal network discovery.

schemes, and we discussed the connections between simulation and real-world data. Finally, we conclude with a discussion on the findings of our study.

## Methods

In this section, we first introduce the volunteer behaviour model and the uncertainty measure of volunteer behaviours, followed by the identification of self-organisation intervals. Finally, we introduce the method for causality analysis. An outline of our framework is illustrated in Fig. 2.

**Volunteer behaviour model**. We defined the volunteer behaviour model as a tuple $(\mathcal{U}, \mathcal{O}, \mathcal{T}, P(u), O(u, o), T(c, o))$, where $\mathcal{U}$ is the set of users, $\mathcal{O}$ is the set of organisers, $\mathcal{T}$ is the set of task categories. The following decision process determines the actions of each user and organiser:
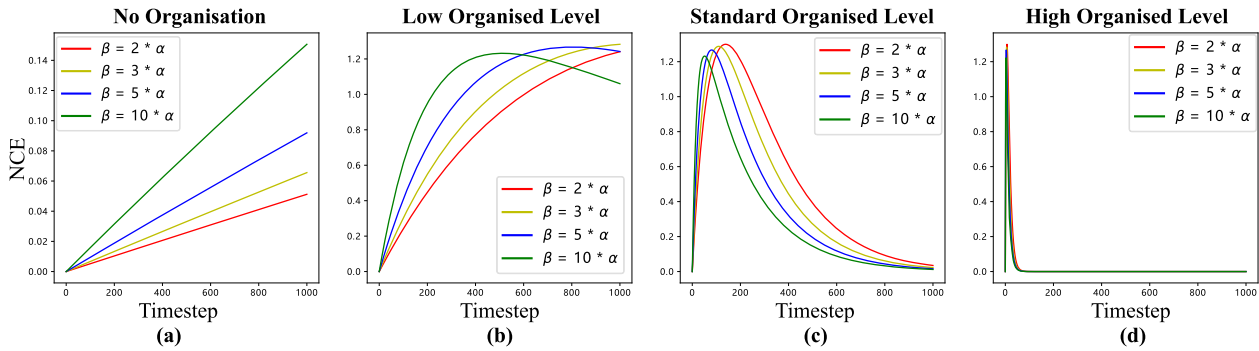
**Fig. 3 Visualisations of the double-exponential model with different values of $\alpha$ and $\beta$. a** No organisation: $\frac{1}{\alpha} > \frac{1}{\beta} > n$. **b** Low organisational level: $\frac{1}{\alpha} > n > \frac{1}{\beta}$. **c** Standard organisational level: $n > \frac{1}{\alpha} > \frac{1}{\beta}$. **d** High organisational level: $\frac{1}{\alpha} \approx \frac{1}{\beta}$.

1. In every timestep, user $u \in \mathcal{U}$ chooses to participate in a volunteer task with probability $P(u) = \Pr(X = 1 | U = u)$, where $X$ is a binary random variable indicating the event of participation. $P(u)$ is also referred as the *participation rate of $u$*.

2. When deciding which task to join, user $u$ will choose a task posted by organiser $o \in O$ with probability $O(o, u) = \Pr(O = o | U = u)$. Probability mass function $O(\cdot, u)$ is known as *the organiser preference of $u$*.

3. Each task organiser $o \in \mathcal{O}$ will post a task belonging to category $c \in \mathcal{T}$ with probability $T(c, o) = Pr(T = c | O = o)$. Probability mass function $T(\cdot, o)$ is known as the *task preference of $o$*.

As volunteer behaviours change over time, we parameterised the model using sequences of probability mass functions $P_t(u)$, $O_t(o, u)$, and $T_t(c, o)$, where $t$ represents the time index in days, as shown in Fig. 2a. Given the daily task participation records of all users $(u_t, o_t, c_t)$, the empirical distributions at each time step can be directly estimated from data using a sliding window. In the Pioneers case study, the set of task categories $\mathcal{T}$ consists of six labels including COVID-19 response, public health education, environmental protection, business reopening, public transportation, and community volunteering. These labels were learned by clustering textual task descriptions using Latent Dirichlet Allocation (Blei et al. 2003). A 14-day time window with one day offset was used to compute the model parameters from February 2020 to January 2021. We also assumed that each participant will attend no more than one task per day.

The uncertainty in volunteer behaviours can be captured via the normalised conditional entropy (NCE) (Cover 1999) of the estimated model parameters. The *NCE of organiser preference* (O-NCE) is defined in the following equation:

$$\hat{H}(O|U) = \frac{H(O|U)}{\sum_{u \in \mathcal{U}} H(O|U=u)} = \frac{-\sum_{u \in \mathcal{U}} \Pr(U=u) \sum_{o \in \mathcal{O}} O(o,u) \log(O(o,u))}{-\sum_{u \in \mathcal{U}} \sum_{o \in \mathcal{O}} O(o,u) \log(O(o,u))} \tag{1}$$

Conditional entropy $H(O|U)$ measures the uncertainty about organiser choice when user is known. The conditioning is necessary here to characterise a global behaviour pattern regardless of inter-personal variations. Also since the value of $H(O|U)$ depends on the number of organisers $|\mathcal{O}|$ observed in a time window, which may change over time, we normalised the conditional entropy by its upper bound $\sum_{u \in \mathcal{U}} H(O|U=u)$. Similarly, we defined the *NCE of task participation* (P-NCE)

and the *NCE of task preference* (T-NCE) as follows,

$$\hat{H}(X|U) = \frac{H(X|U)}{\sum_{u \in \mathcal{U}} H(X|U=u)} = \frac{-\sum_{u \in \mathcal{U}} \Pr(U=u) \left( P(u) \log(P(u)) + (1 - P(u)) \log(1 - P(u)) \right)}{-\sum_{u \in \mathcal{U}} \left( P(u) \log(P(u)) + (1 - P(u)) \log(1 - P(u)) \right)} \tag{2}$$

$$\hat{H}(T|O) = \frac{H(T|O)}{\sum_{o \in \mathcal{O}} H(T|O=o)} = \frac{-\sum_{o \in \mathcal{O}} \Pr(O=o) \sum_{c \in \mathcal{T}} T(c,o) \log(T(c,o))}{-\sum_{o \in \mathcal{O}} \sum_{c \in \mathcal{T}} T(c,o) \log(T(c,o))} \tag{3}$$

Figure 2b plot the P-NCE, O-NCE and T-NCE value at each time step in the Pioneers dataset. Their values are expected to increase as a result of external influences (such as changes in the COVID-19 situation) and to decrease in response to the emergence of self-organisation in volunteer behaviours.

**Self-organisation intervals and organisational speed**. To systematically detect whether the self-organisation effect exists within a given time interval, we fitted a double-exponential model frequently used to model lightning impulse waveforms (Jia and Xiaoqing 2006) to the NCE curves, as shown in Fig. 2c. Without the loss of generality, we let NCE($t$) represent the value of any NCE curves at the $t$th timestep with a double-exponential model as shown in Eq. (4):

$$\text{NCE}(t) = A * M * \left( e^{-\alpha t} - e^{-\beta t} \right), \tag{4}$$

where $M$ is the peak value of NCE and $A$ is a function of $\alpha$ and $\beta$, expressed as Eq. (5).

$$A(\alpha, \beta) = \frac{1}{e^{-\alpha \frac{\ln(\beta) - \ln(\alpha)}{(\beta - \alpha)}} - e^{-\beta \frac{\ln(\beta) - \ln(\alpha)}{(\beta - \alpha)}}} \tag{5}$$

Numerous studies have utilised a fitted curve's pulse length and rise time to characterise its waveform according to (Thottappillil and Uman 1993). The rise time, which measures the amount of time the pulse takes to go from 10% to 90% of the peak value, can be approximated as $\frac{1}{\alpha}$. The pulse length, which measures from 10% of the peak value to the peak and then to 50%, can be approximated as $\frac{1}{\beta}$. Using the rise time and the pulse length, we classified the NCE curve into four organisational levels: no organisation, low organisational level, standard organisational level, and high organisational level, as shown in Fig. 3. Additionally, we defined the *organisational speed* of each self-organisation interval $\eta$ by Eq. (6). $T_{\text{Half}}$ represents the duration in which NCE declines from its peak to its half value, whereas $T_{\text{Fall}}$ represents the duration from NCE peak to the end of the interval.

$$\eta = \frac{T_{\text{Half}}}{T_{\text{Fall}}} \approx \frac{\frac{1}{\alpha} - \frac{\ln(\beta) - \ln(\alpha)}{\beta - \alpha}}{n - \frac{\ln(\beta) - \ln(\alpha)}{\beta - \alpha}}. \tag{6}$$

Our study refers to a time interval as *self-organised* if its NCE is classified as either a low or standard organisation level. We call such time intervals *self-organisation intervals*. To compute such self-organisation intervals, we proposed an algorithm to greedily find the maximal set of longest non-overlapping self-organisation intervals among all time intervals within 5–90 days. We describe the details of interval detection in Algorithm S1.

**Causality network discovery**. We analysed the dynamic factors that have caused self-organisation events in volunteer behaviours to occur using time-series based causal network discovery, as illustrated in Fig. 2d. The candidate dynamic factors we considered included two groups of time series variables: internal variables and external variables. External variables track external events such as new COVID-19 cases, holidays, or systematic interventions that may influence Pioneers' volunteer behaviours. We classified external variables into five categories: school or business reopening policies, daily new COVID-19 cases in Shenzhen and China, and other ad-hoc interventions that positively or negatively affect volunteer participation. A detailed description of each external variable appears in the events listed in Table 1 and the variables are visualised in Fig. 4. Internal variables are observable quantities from

the Pioneers data, including the total number of organisers, users, and tasks in each category (Fig. 4). While we focused on identifying the most influential external factors that impact the NCE dynamics, internal variables were used as conditioning variables to avoid spurious associations.

We assumed that each self-organisation interval had stationary causal relations and computed its causal graph over all internal and external variables and the NCEs using PCMCI+, a causal network discovery algorithm based on conditional independence tests (Runge et al. 2019; Runge 2020). We modified the original PCMCI+ method to incorporate prior knowledge about the internal and external variables. Specifically, we made two assumptions regarding variables. First, we assumed that the external variables were independent from each other; Second, we assume that there were no causal links from internal or NCE variables to external variables. Each directed causality graph edge was labelled with the causal time lag and the corresponding *p*-value in the PCMCI pairwise conditional independence test. A two-sided Student's *t*-test was used to compute the *p*-value. A smaller *p*-value corresponds to a more significant causal relationship between two variables.

**Case study: self-organisation analysis of pioneers user behaviour**. In this section, we used the Pionners dataset as the case study to analyse when and how self-organisation emerged during the pandemic response on the city scale and regional scale. Following the workflow described in the methods section, we investigated how district characteristics and external variables such as the pandemic situation and centralised interventions affect the speed and effectiveness of the self-organisation, respectively.

**Characterising self-organisation in volunteer behaviour processes**. The NCE curves of participation rate, organiser preference, and task preference are all characterised by multiple modes (Fig. 5), which can be interpreted as representing multiple phases of self-organisation. The self-organisation intervals of the

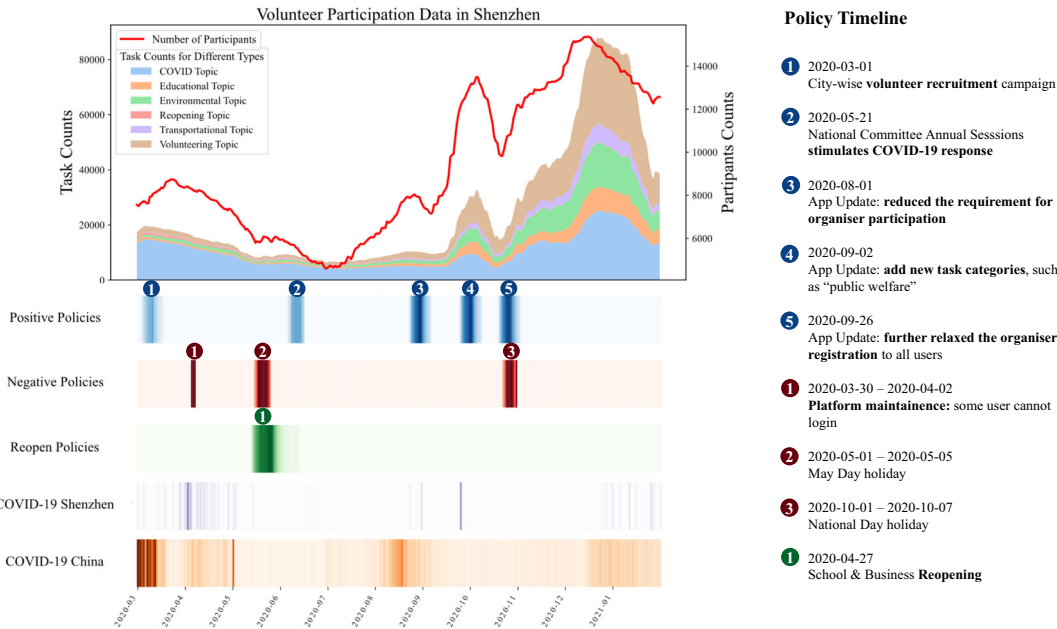| Table 1 Regional statistics of population density and registered companies. | | | | |
|---|---|---|---|---|
| | **Futian** | **Baoan** | **Pingshan** | **Guangming** |
| Population | 1,553,225 | 4,476,554 | 551,333 | 1,095,289 |
| Registered Companies | 5475 | 5994 | 440 | 1023 |
| | **Luohu** | **Nanshan** | **Yantian** | **Longgang** |
| Population | 1,143,801 | 1,795,826 | 214,225 | 3,979,037 |
| Registered Companies | 1909 | 4457 | 365 | 5761 |



**Fig. 4 Overview of the volunteer data used in this study.** Left top: The Stacked area plot shows the average number of volunteer tasks in each of the six categories (in a 2-week window) over the entire duration of the study; Red line plot shows the total number of volunteer participants per day; Left Bottom: External information used in the causality analysis. The colour shading indicates the stringency of a policy group or the number of COVID-19 cases. Details on the data can be found in the "Methods" section. Right: Timelines for positive, negative, and reopening policies.
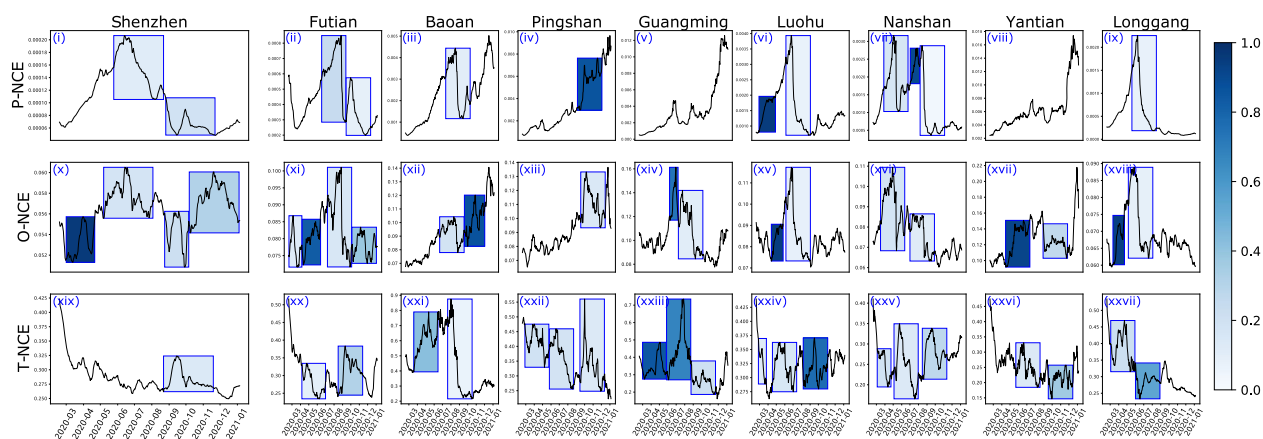
**Fig. 5 Self-organised volunteer behaviour during the COVID-19 outbreak in Shenzhen.** Maximal self-organisation intervals (boxed regions) for three types of NCE in Shenzhen and Shenzhen districts. (max duration = 90, min duration = 5.). (i–ix): NCE of task participation (P-NCE) in Shenzhen and Shenzhen districts. (x–xviii): NCE of organiser preference (O-NCE) in Shenzhen and its districts. (xix–xxvii): NCE of task preference (T-NCE) for Shenzhen and its districts.

| Table 2 Descriptions of external variables that may affect volunteer self-organisation. | | |
|---|---|---|
| **External variables** | **Variable type** | **Description** |
| Positive policies | Continuous time series data With a 14-day impulse | Policies that may positively affect the number of participates, such as the national committee annual sessions and some app updates (Increasing task organisers and task types) |
| Negative policies | | Policies that may negatively affect the number of participates, such as holidays |
| Reopen policies | | Policies related to school and company reopening |
| COVID_CHN | Discrete time series data | The number of COVID-19 daily confirmed cases in China |
| COVID_Shenzhen | | The number of COVID-19 daily confirmed cases in Shenzhen |

three types of NCEs computed using the entire Shenzhen data are highlighted in the first column of Fig. 5. Darker shading indicates higher organisational speed (as defined in Eq. (6)), which characterises the rate of NCE decrease after the peak. The self-organisation intervals for P-NCE were from May 25 to August 22 and from August 26 to November 22. Four self-organisation intervals for O-NCE were identified: March 1 to April 21, May 7 to August 3, August 24 to October 6, and October 7 to January 4 (2021). For T-NCE, the only self-organisation interval that was detected was from August 22 to November 19. These results show that volunteers' social group preference was less stable than the other two behavioural preferences, as O-NCE had more self-organisation intervals relative to P-NCE and T-NCE. Notably, Shenzhen's first COVID-19 outbreak (from March to May, 2020) coincided with the first O-NCE self-organisation interval. The correlation coupled by a relatively high organisational speed indicate that stable volunteer groups were established more quickly during the height of the pandemic than in other periods of the year. Additionally, we expected the generation of only a single self-organisation interval by T-NCE, as a large increase in task diversity only happened once when the platform expanded its activity topic selection from COVID-19 to include a variety of community services in the third quarter of 2020 as shown in Fig. 5. We will delve deeper into what caused each self-organisation phase in volunteer behaviours in later sections.

**Regional differences in self-organisation dynamics.** In addition to the city-scale self-organisation effects, significant regional variations existed in the NCE dynamics and the self-organisation intervals. Figure 5 and Fig. S1 show self-organisation intervals' counts (Fig. 5), organisational speed (blue shading in Fig. 5), distribution (Fig. S1 a(i)–a(iii)), and clustering results (Fig. S1

b(i)–b(iii)). These factors were computed using activity records from each of Shenzhen's eight districts. We discovered that regional volunteer behaviour was significantly influenced by population density (ShenzhenNews, Accessed 10 Aug 2021) and district type (Regulation, Accessed 10 Aug 2021), according to the regional statistics listed in Table 2.

Districts with a higher residential population had a slower rate of self-organisation formation in participation behaviour than did other districts. As shown in Fig. 5, of the districts that exhibited self-organisation, the two most populated districts, Baoan (4.47 M) and Longgang (3.97 M) were characterised by lower organisational speed which reflect slower responses to new volunteer demand (Fig. 5 iii, ix). Baoan, in particular, has nearly nine times the population of Pingshan (0.55 M), and its organisational speed (0.11) was 88% lower than that of the less populated Pingshan district (0.86). On the other hand, the residential population did not have a clear association with the number of self-organisation intervals.

Additionally, various district types showed different self-organisation behaviour in terms of participation rate and task preference, reflecting different community needs. For instance, districts with similar concentrations of businesses including Baoan (5994 companies) and Longgang (5761 companies), or Pingshan (440 companies) and Guangming (1023 companies), had similar self-organisation intervals in T-NCE, as illustrated in Fig. S1 (b(iii)). Based on P-NCE we found that commercial districts (Baoan and Longgang) experienced a self-organisation interval with low organisational speed during the period in which companies were reopening (from May to July, 2020), whereas non-commercial districts (Pingshan and Guangming) did not have self-organisation intervals during this period. By late 2020, the volunteer task types in commercial districts such as Longgang and Baoan appeared to have become stable, as few
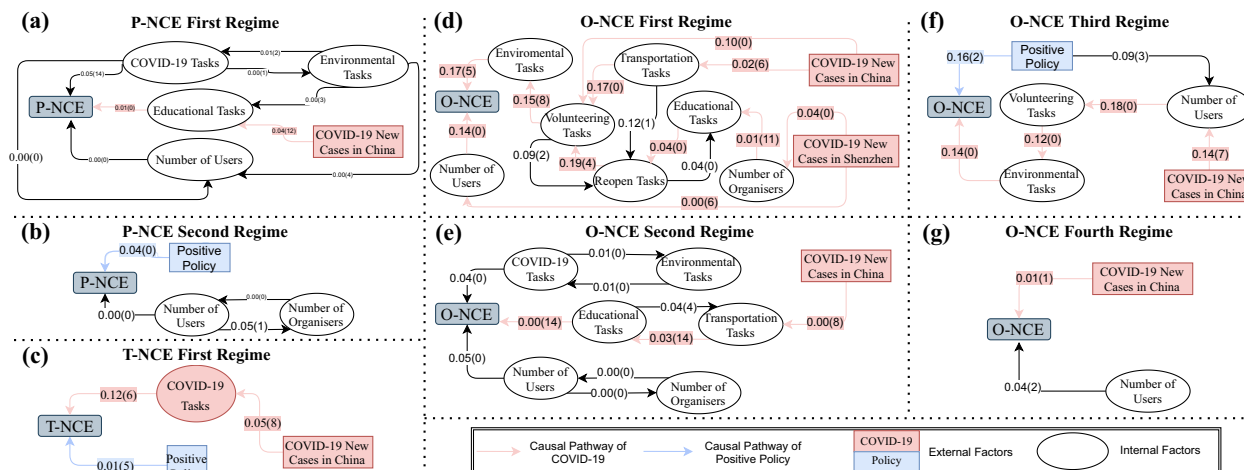
**Fig. 6 Causality graph for participation rate NCE, organiser preference NCE (O-NCE), and task preference NCE (T-NCE). a, b** Causality graph for two P-NCE regimes. **c** Causality graph for the T-NCE regime. **d–g** Causality graph for four O-NCE regimes.

T-NCE self-organisation intervals occurred in such districts after that period.

*What factors caused the self-organisation dynamics in Shenzhen?.* Besides the static analysis in regional comparisons, we further analysed the dynamic factors that have caused self-organisation events in volunteer behaviours to occur using modified PCMCI+ method (Runge 2020). The self-organisation intervals detected (Fig. 5i, x, xix) define two causal regimes for P-NCE, four regimes for O-NCE, and one regime for T-NCE. The resulting causal graphs are shown in Fig. 6. For clarity of presentation, we removed the variables and edges that did not lie on a direct path toward NCEs. Figures S2–S8 demonstrate the full causal graphs. We discovered from most regimes that external variables such as COVID-19 daily new cases and centralised interventions (positive policies) affected volunteer behaviour in Shenzhen through internal variables such as user and organiser count, the number of educational tasks, and COVID-related tasks.

The change in the epidemic situation was the most important external factor affecting the stability of participation rate, organiser, and task preferences. This finding is evident since the number of COVID-19 daily new cases acts as an external variable in most regimes, including the first regime (May 25 to August 22, 2020) in P-NCE, the first (March 1 to April 21, 2020), the second (May 7 to August 3, 2020), and the fourth (October 7, 2020 to January 4, 2021) regimes in O-NCE, and the unique regime (August 22 to November 19, 2020) in T-NCE. Except for the fourth regime in O-NCE, where COVID-19 directly affected O-NCE ($p = 0.013$) with a one-day lag, COVID-19 affected NCEs via various internal intermediary factors, such as user and organiser count and different topic tasks. Specifically, COVID-19 affected O-NCEs in most regimes via various internal intermediary factors, which include user and organiser count and different topic tasks. On the one hand, COVID-19 affected T-NCE through COVID-related tasks. On the other hand, it affected O-NCE and P-NCE through educational tasks in the first regime in P-NCE and the first two regimes in O-NCE. Note that COVID-19 had no significant impact on volunteer participation in the second P-NCE regime (August 26 to November 22, 2020) when the pandemic was mostly under control.

Another crucial external variable that influenced both participation and task preference of volunteers was systematic interventions, represented by positive policies. It affected the second regime (August 26 to November 22, 2020) in P-NCE, the third regime (August 24-October 6, 2020) in O-NCE, and

the unique regime (August 22 to November 19, 2020) in T-NCE. Unlike COVID-19 cases, positive policies had a direct causal link to all NCEs. Specifically, positive policies affected P-NCE and T-NCE with no time lag and O-NCE with a two-day lag. Moreover, all affected regimes occurred between August and November, during which period policies included application updates such as lowering of organiser registration requirements and increased numbers of task types. We conclude that systematic interventions affected task preference ($p = 0.007$) and participation rate ($p = 0.040$) in descending order of statistical significance.

## Simulation analysis of self-organisation effectiveness

*Simulation setup.* To understand the advantages and limitations of self-organised crowdsource systems, we designed a simulation implementing a minimal volunteer crowdsource scenario that can exhibit self-organisation effects in a controlled environment. The simulated environment is a 10-by-10 grid world, where each grid cell represents a task with a specific value (Fig. 7a). The task value measures a level of task importance and is set to change at frequency $\Delta$ to reflect the changing community demand. Each simulation experiment emulated the actions of 150 agents (volunteers) for a total of 100 time steps given 100 tasks in total. For simplicity, we assumed that all tasks needed the same number of people to complete. To prevent unbalanced volunteer allocation, each grid cell (task) can be taken by at most $k = 3$ agents at a given time step. To simulate changes in the environment, the grid values changed twice, at the 30th and 70th time steps. Specifically, grid cells were divided into two groups: low-value cells (0, 5) and high-value cells (5, 10). Each time the value changed, the high-value group shifted to the right, as illustrated in Fig. S9. At each time step of the simulation, agents representing individual volunteers can participate in a task by moving to the corresponding cell. The collective goal of the agents is to participate in tasks that maximise combined values.

We defined three task selection strategies based on different organisational schemes in real-world crowdsourced systems. A *self-organised scheme* is one in which individuals make independent decisions about which task to participate. We simulated this scheme by having agents choose tasks randomly while giving preference to high task values in their trajectories. As agents explored new tasks, they correspondingly learned new values, analogous to the way people gain experience in real life.
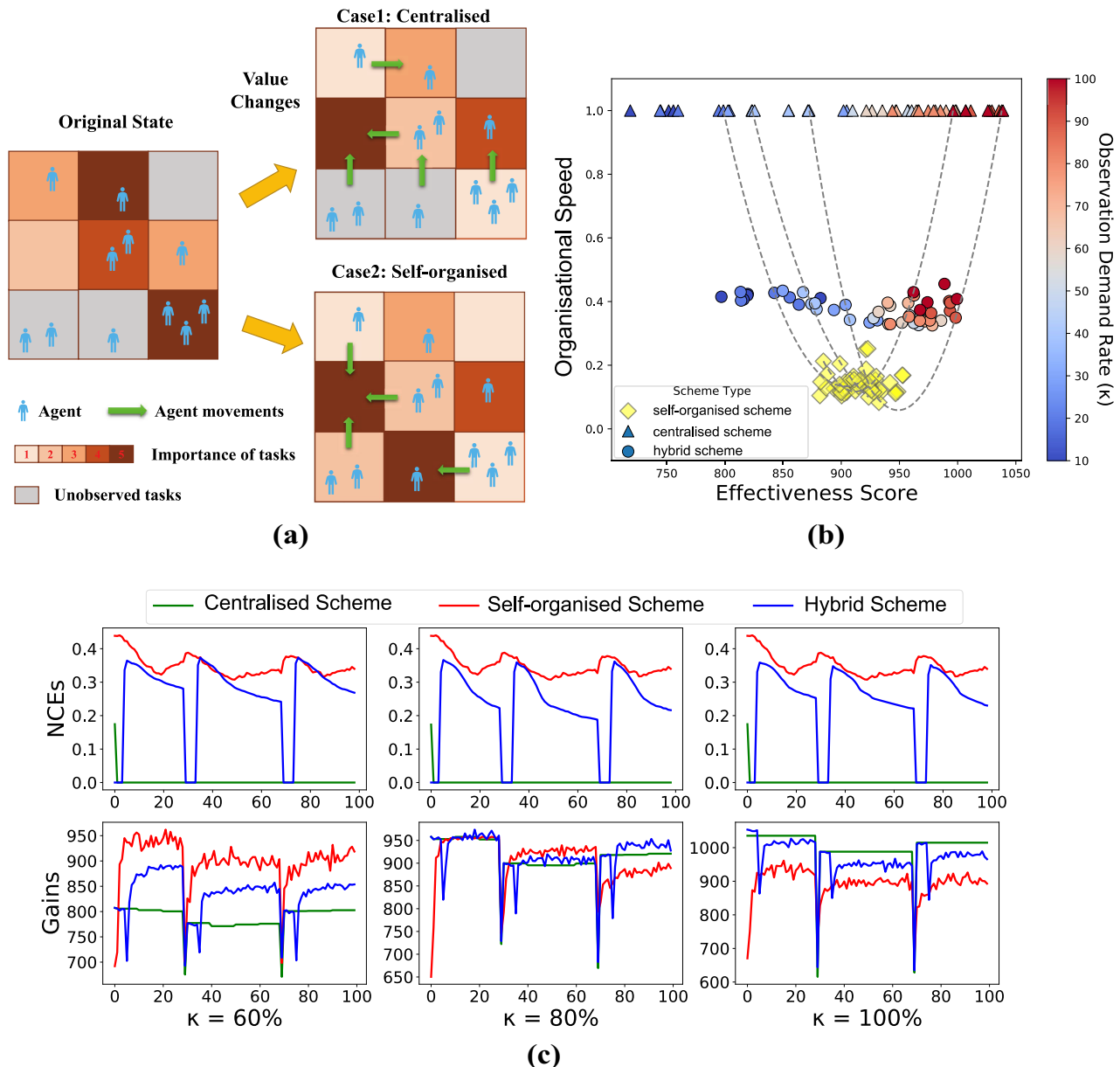
**Fig. 7 Schematic diagram and experimental results of grid simulation. a** Schematic diagram of a simplified version 3-by-3 grid simulation. After task value changes from the original state, agents are assigned optimally to the observed top value tasks in a centralised step (case 1). In contrast, agents move to the task with the highest value nearby in a self-organised step (case 2). **b** Comparisons of NCEs and effectiveness scores under different observable demand rates (κ). **c** Comparison among organisational speeds and effectiveness scores in different schemes. Each dashed line connects groups with the same random values and observable demand rates.

A *centralised scheme* is one in which individuals entirely follow some external assignment given by a decision maker, who symbolises a centralised system or a leadership board. As in the real-world case where local community demands are difficult to learn precisely, a decision maker can only observe a fraction κ of the task values (a.k.a. *the observable demand rate*) and assign agents optimally to the observable tasks. A *hybrid scheme* is in between self-organised and centralise scheme. Under this scheme, the system executes the centralised strategy for τ time steps whenever the task value changes and then switches to the self-organised strategy. The parameter τ adds a trade-off between the centralised guidance and volunteers' independent decisions. In both self-organised and hybrid schemes, we assumed that the maximum distance an agent could travel at each time step was at most two cells, which reflects the physical constraints of volunteer

participation. See Figs. S10 and S11 for the comparison of agent distribution and organisational speed among different schemes.

*Is self-organisation effective for the successful organisation of a crowdsourced volunteering system?.* We compared these organisation schemes under the various simulation settings shown in Table 3. Each simulation was evaluated based on two system-level metrics: *organisational speed $O_s$* and *organisational effectiveness $E_s$*. Similar to the analysis of the Pioneers data, we captured the dynamics of agents' task selection behaviour using the NCE of selecting an agent to whom a task had been given. The organisational speed $O_s$ of schemes $s$ was defined as the average organisational speed $\eta$ (Eq. 6) over all detected self-organised intervals. By definition, the centralised scheme maintained maximum organisation speed, with $O_s = 1$. Additionally, system

**Table 3 Average organisational speed and organisational effectiveness under different simulation configurations over ten trails, with standard deviation in parentheses.**

| $\kappa$ | Schemes | Change Frequency $\Delta$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | **2** | | **3** | | **5** | |
| | | *Speed* | *Effectiveness* | *Speed* | *Effectiveness* | *Speed* | *Effectiveness* |
| 60% | Centralised | 1.00 (0.00) | 810.94 (23.37) | 1.00 (0.00) | 775.30 (22.46) | 1.00 (0.00) | 774.75 (22.23) |
| | Hybrid | 0.37 (0.04) | 890.11 (22.37) | 0.45 (0.03) | 841.21 (19.60) | 0.42 (0.12) | 823.10 (17.83) |
| | Self-organised | 0.16 (0.03) | **957.78 (15.74)** | 0.14 (0.04) | **911.69 (13.38)** | 0.25 (0.09) | **908.73 (12.65)** |
| 80% | Centralised | 1.00 (0.00) | 971.86 (21.49) | 1.00 (0.00) | 920.10 (20.40) | 1.00 (0.00) | 916.69 (18.46) |
| | Hybrid | 0.31 (0.04) | **987.30 (14.14)** | 0.42 (0.06) | **934.31 (14.85)** | 0.47 (0.17) | **923.35 (14.05)** |
| | Self-organised | 0.16 (0.03) | 957.78 (15.74) | 0.14 (0.04) | 911.69 (13.38) | 0.25 (0.09) | 908.73 (12.65) |
| 100% | Centralised | 1.00 (0.00) | **1070.30 (17.21)** | 1.00 (0.00) | **1022.00 (13.90)** | 1.00 (0.00) | **1016.73 (13.71)** |
| | Hybrid | 0.33 (0.02) | 1030.29 (14.93) | 0.42 (0.03) | 984.35 (11.44) | 0.39 (0.07) | 979.65 (12.01) |
| | Self-organised | 0.16 (0.03) | 957.78 (15.74) | 0.14 (0.04) | 911.69 (13.38) | 0.25 (0.09) | 908.73 (12.65) |

| $\kappa$ | Schemes | Centralised Steps $\tau$ | | | |
|---|---|---|---|---|---|
| | | **1** | **5** | **10** | **15** |
| 60% | Hybrid | *Speed* 0.36 (0.04) | 0.45 (0.03) | 0.49 (0.08) | 0.51 (0.09) |
| | | *Effectiveness* **843.35 (21.14)** | 841.21 (19.60) | 833.51 (19.46) | 823.90 (18.21) |
| 80% | | *Speed* 0.29 (0.02) | 0.42 (0.06) | 0.45 (0.05) | 0.47 (0.07) |
| | | *Effectiveness* **936.10 (12.69)** | 934.31 (14.85) | 932.06 (13.25) | 928.77 (15.49) |
| 100% | | *Speed* 0.34 (0.02) | 0.42 (0.03) | 0.47 (0.09) | 0.54 (0.11) |
| | | *Effectiveness* 982.20 (12.32) | 984.33 (11.44) | 987.82 (11.83) | **991.20 (12.78)** |

Top: Comparison among three organisation schemes under different observable demand rates $\kappa$ and change frequency $\Delta$. Bottom: The impact of centralised steps $\tau$ in the hybrid scheme under different $\kappa$.

effectiveness $E_s$ evaluates the overall task completion performance. It was defined as the expected total task values that agents completed at each time step during the simulation.

One of the main findings from the simulation is that the self-organised scheme always had the slowest organisational speed $O_s$, followed by the hybrid scheme. For centralised and hybrid schemes, the effectiveness $E_s$ decreased rapidly as the observable demand rate $\kappa$ decreased (see Fig. 7c and Table 3). Meanwhile, the variance of $E_s$ for the self-organised scheme was the smallest when $\kappa$ was less than 80%. Only when $\kappa$ was greater than 80% were the centralised and hybrid effectiveness scores higher than those of the self-organised scheme (visualisation results are shown in Figs. S12–S16). This was the case when true community need was easily accessible to the decision maker, who was able to make globally optimal task assignments whenever demand changed. Overall, when $\kappa$ was unknown, the hybrid system was the most robust choice. We further demonstrated the optimality of the hybrid system under these conditions by performing 50 simulations with randomised task values and under various $\kappa$ values ranging from 0 to 1 (Fig. 7b). We found that the self-organised scheme had the smallest variance in effectiveness $E_s$ for a given value of $\kappa$, but its organisation speed $O_s$ has the largest variance. The result for the centralised scheme was the opposite, while the hybrid scheme demonstrated a balance between the two objectives.

As the hybrid scheme makes a trade-off between organisational effectiveness and speed, it is crucial to consider the level of centralised guidance the hybrid scheme should use. In Table 3, we showed that as the number of centralised steps ($\tau$) increased, the centralised intervention increased organisation speed $O_s$, particularly when $\kappa$ was less than 100%. Meanwhile, effectiveness $E_s$ decreased when $\kappa$ was 60% but increased when $\kappa$ was 100%. These observations suggest that a successful crowdsource volunteer system should adjust the supervision strength to its users according to the amount of knowledge of the underlying community demand. Table 3 shows how the organisational speed

and system effectiveness were affected by the frequency of demand change ($\Delta$). As $\Delta$ increased, the effectiveness of the system in all three schemes decreased. These results imply that most systems cannot adapt to the pace of the external changes under the rapid change in volunteer demand.

**Connection between simulation and real-world cases.** The organisation of Pioneers resembles a hybrid scheme because some activity organisers are also official community leaders responsible for the COVID-19 response and other community projects. Our simulation demonstrates that the extent of the centralised organisation in this hybrid scheme had a significant impact on the organisational speed of the system. To see this effect in real data, we used case studies from groups of densely populated neighbourhoods in the same district to show the difference in organiser preference NCE when different supervision strengths were in place. Various levels of supervision strengths were placed during different stages of the COVID-19 outbreak for a range of task types. At the beginning of the pandemic (First 3 months from 2020 Feb), supervision strength on COVID-19 tasks was low since Pioneers had just started operation, more time is required for a community to fully mature and understand their needs. However, in the later stages of the pandemic (Last 3 months to 2020 Dec), more pandemic control policies and command chains had been established, making the volunteer organisation more centralised at this point. Furthermore, educational tasks (e.g., community events delivering the latest public health information) are often led by official community leaders; thus, their organisation is more centralised than other tasks, such as sustainability activities.

In the first case study (Fig. 8a, b), we compared the O-NCE of neighbourhoods in COVID-related tasks during the early (March to May, 2020) and late (October, 2020 to January, 2021) stages of the outbreak. We observed that O-NCE was greater during the early stages of the outbreak than during the later stages. In the
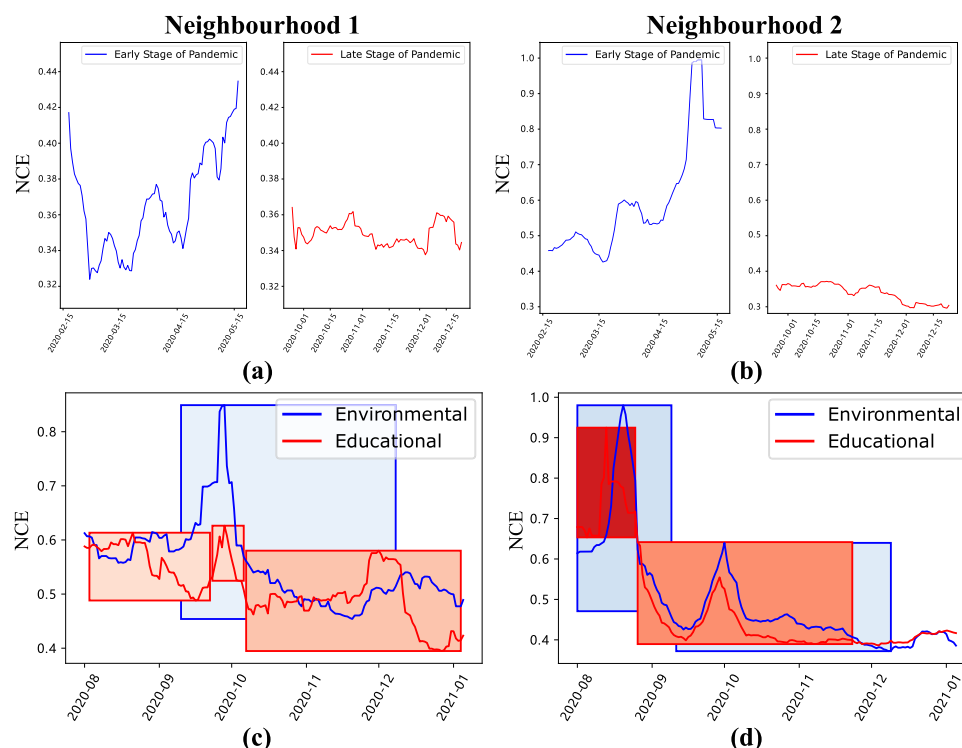
**Fig. 8 Case Study: changes in organiser preference NCE under different supervision strengths. Colour shading in intervals indicates the magnitude of organisational speed. a, b** Comparison of O-NCEs at different COVID-19 outbreak stages in the same neighbourhood. **c, d** Comparison of O-NCEs for environmental tasks (blue) and education tasks (red) in the same neighbourhood.

second case study (Fig. 8c, d), we compared the neighbourhood O-NCE and self-organisation intervals of environmental and educational tasks over the same period. The average organisational speed for the self-organised intervals for the O-NCE of educational tasks (0.56, 0.16) was faster than those for the O-NCE of environmental tasks (0.16, 0.07). These two experiments show that as supervision strength increases, the extent of the centralised organisation increases in the hybrid system, resulting in a lower NCE and higher organisational speed, which corroborates the findings of our simulation.

## Discussion

Using 1 year of data from the Anti-Pandemic Pioneers platform in Shenzhen, China, we critically analysed the effect of self-organisation during the pandemic using a novel entropy-based identification method. Furthermore, we simulated a volunteer behaviour model to compare different organisational schemes to develop better crowdsource systems. From the aforementioned results, we arrived at several major findings in light of volunteer organisation and pandemic response.

**Interpretation of regional differences in self-organisation**. In Shenzhen, self-organisation was observed through the gradual stabilisation of user participation, organiser, and task preferences without strict intervention from a centralised decision maker. Our regional results reveal that population density and district type had a significant impact on regional volunteer behaviour. Specifically, districts with a larger population appeared to have slower self-organisation formation due to lower organisational speed in response to the pandemic, whereas districts with more businesses responded rapidly during the company reopening period as concluded from comparisons of self-organisation intervals. This phenomenon may be explained by the fact that location-based

volunteering is closely related to the physical social networks in each community, which can vary in size and structure. Specifically, the larger the population, the more complex the community's social network. Consequently, in a community social network that contains more nodes and edges, more communications is necessary for group actions or reaching consensus (Carley et al. 2002; Friedkin 2006), leading in a slower organisation. Additionally, as the characteristics of two districts become similar, so does their social network structures, resulting in homogeneous volunteer behaviour.

**The role of centralised intervention in crowdsource volunteering systems**. The behaviour of volunteers during the COVID-19 pandemic was undoubtedly affected by the nature of this extreme crisis. Self-organisation intervals were detected during the COVID-19 outbreak in Shenzhen (March to May, 2020 and December 2020 to January 2021). We discovered that volunteer groups self-organised from chaos to stability in response to changes in the pandemic situation. Our causality findings further indicate that in most cases, the daily number of new COVID-19 cases affected volunteer self-organisation indirectly by initially affecting the number of users and organisers or the distribution of task types. Volunteers gradually formed self-organisation against the pandemic situation as more users participated in tasks or organisers posted COVID-related tasks. The causality analysis on Pioneers' data also demonstrated that centralised interventions significantly impacted volunteers' participation rate and preferences for organisers and tasks, especially between August and November. For example, in September, the platform administrators expanded the types of tasks available from COVID-19-related tasks to general volunteering. As the number of users, organisers, and tasks increased to their maxima, the centralised intervention significantly impacted the Pioneers' system. The expansion of task

types exemplified how centralised intervention results in high user engagement on the Pioneers' platform. In our grid simulation system, the observable demand rate which represents the level of centralised intervention on the Pioneers was shown to significantly affect the system's effectiveness. Therefore, it is crucial to consider the extent to which centralised interventions can exert control over a crowdsource volunteer system.

**Building a robust volunteer system using self-organisation**. Using agent-based simulation, we discovered that the hybrid scheme is the most robust organisational scheme for a crowd-sourced volunteering platform, since hybrid schemes can maintain their effectiveness under conditions of uncertain and frequently changing social demands. When the observable demand rate is higher than 80%, the system gains under the hybrid system are greater than those under the self-organised scheme even after the centralised guidance ceases. This phenomenon reveals that agents can spontaneously find more valuable tasks after a few steps of external guidance. When the observable demand rate is lower than 60%, the system can benefit more under a hybrid scheme than under a centralised scheme. Self-organisation provides the hybrid system with a unique self-adjusting ability. In addition, regarding different frequencies of environmental change, proper and timely directions from the decision maker are essential for the system to adapt to new situations. For a hybrid system, the extent to which the centralised leader intervenes in the organisation is critical. As shown in Fig. S16, even though the centralised guidance lasts only a single time step, the gains of the hybrid scheme are higher than the gains of the self-organised scheme. We conclude that in a crowdsource volunteering system, the decision maker should provide timely guidance based on known demands to encourage volunteers to collaborate more efficiently. Nevertheless, the decision maker should not intervene too strongly to allow users so as to self-adapt to unknown community needs.

**Future works and conclusions**. Our study has several limitations that should be addressed in subsequent research. For instance, we considered all available variables in the causality study, such as COVID-19 cases and policies; however, some unknown and undetectable causal factors may still affect self-organisation. For example, we did not have access to personal communication between volunteers that may have affected organiser and task selection preferences. Therefore, additional social interaction data could potentially increase the confidence in causality analyses results. Additionally, our grid simulation is a simplified abstract model that simulates the task selection process and demonstrates the optimal organisational scheme, with a presumption that individuals choose tasks independently. In order to make the simulation even more realistic, we are working on mechanisms to distinguish organisers from participants and to model more complex volunteer interactions.

In conclusion, our study demonstrates the potential of self-organisation in volunteer platforms, even in the face of a pandemic. Individuals can adjust and organise themselves spontaneously in order to complete tasks. Effective interventions can enhance the success of a system during the process of self-organisation. Additionally, self-organisation enables the system to achieve a stable state more quickly, resulting in increased overall system benefits. We hope that our research can help volunteer organisations make better use of the self-organisation effect in collective human behaviour to increase organisational effectiveness under uncertain community demands.

**References**
Atlan H (1974) On a formal definition of organisation. J Theor Biol 45:295–304
Bauer S, Lim D (2019) Effect of communication practices on volunteer organisation identification and retention. Sustainability 11:2467
Besaleva LI, Weaver AC (2013) Applications of social networks and crowdsourcing for disaster management improvement. 2013 International conference on social computing. 213–219. https://doi.org/10.1109/SocialCom.2013.38
Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
Bonabeau E, Theraulaz G, Deneubourg J-L, Aron S, Camazine S (1997) Self-organisation in social insects. Trends Ecol Evol 12:188–193
Camazine S, Deneubourg JL, Franks NR, Sneyd J & Bonabeau E (2001) Self-organisation in biological systems. Princeton University Press, Princeton
Carley KM, Lee J-S, Krackhardt D (2002) Destabilising networks. Connections 24:79–92
Click2Houston KP (2021) CrowdSource Rescue helps deliver food to 2300+ vulnerable Houstonians amid coronavirus outbreak. Retrieved from https://www.click2houston.com/food/2020/03/25/crowdsource-rescue-helps-deliver-food-to-2300-elderly-immunocompromised-locals-amid-coronavirus-outbreak/. Accessed 10 Aug 2021.
Cover TM (1999) Elements of information theory. Wiley, Hoboken
Friedkin NE (2006) A structural theory of social influence, vol. 13. Cambridge University Press, Cambridge
Fuchs C (2006) The self-organisation of social movements. Syst Pract Action Res 19:101–137
Howe J (2006) The rise of crowdsourcing. Wired Mag 14:1–4
Jia W, Xiaoqing Z (2006) Double-exponential expression of lightning current waveforms. In: The 2006 4th Asia-Pacific conference on environmental electromagnetics 320–323. https://doi.org/10.1109/CEEM.2006.257962
Marston C, Renedo A, Miles S (2020) Community participation is crucial in a pandemic. Lancet 395:1676–1678
Miao Q, Schwarz S, Schwarz G (2021) Responding to COVID-19: community volunteerism and coproduction in China. World Dev 137:105128
Regulation SA Regional distribution of enterprises report. Regional distribution of enterprises report. Retrieved from http://amr.sz.gov.cn/xxgk/qt/ztlm/sjfb/sjfl/qyfb/. Accessed 10 Aug 2021
Riccardi MT (2016) The power of crowdsourcing in disaster response operations. Int J Disaster Risk Reduct 20:123–128
Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D (2019) Detecting and quantifying causal associations in large nonlinear time series datasets. Sci Adv 5:eaau4996
Runge, J. (2020, August). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Conference on Uncertainty in Artificial Intelligence (pp. 1388–1397). PMLR.
Schimak G, Havlik D, Pielorz, J (2015) Crowdsourcing in crisis and disaster management–challenges and considerations. Int Symp Environ Softw Syst. https://doi.org/10.1007/978-3-319-15994-2_5
ShenzhenNews Shenzhen 7th census bulletin. Retrieved from http://www.sznews.com/news/content/2021-05/17/content_24221359.htm. Accessed 10 Aug 2021
Thottappillil R, Uman MA (1993) Comparison of lightning return-stroke models. J Geophys Res 98:22903–22914

**Ethical approval**
Not applicable as this study did not involve human participants

**Informed consent**
Not applicable as this study did not involve human participants

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1057/s41599-022-01127-2.

**Correspondence** and requests for materials should be addressed to Yang Li.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.