Decoding alarm signal propagation of seed-harvester ants using automated movement tracking and supervised machine learning

Xiaohui Guo, Michael R. Lin, Asma Azizi, Lucas P. Saldyt, Yun Kang, Theodore P. Pavlic, Jennifer H. Fewell

Arizona, United States

1. Spatial proximity for physical contact. We followed 7 pairs of ants that made physical contact, and measured the minimum Euclidean distance between each pair during contact (d_1 = 37 pix; d_2 = 44 pix; d_3 = 39 pix; d_4 = 37 pix; d_5 = 43 pix; d_6 = 44 pix; d_7 = 45 pix), (Fig. S1). The maximum of those distances was used as our spatial proximity criterion for physical contacts ($d_{contact}$ = 45 pix).
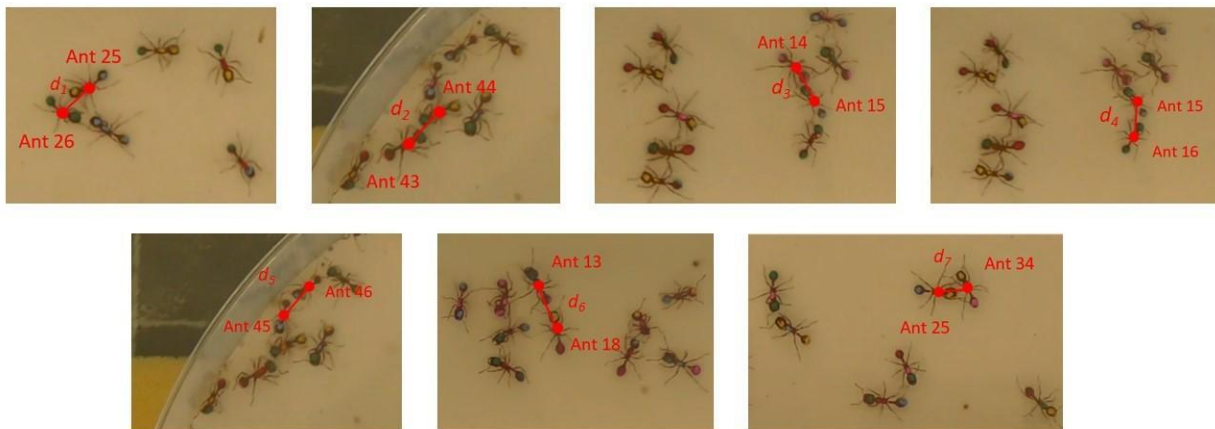

Figure S1: Snapshots for 7 samples of physical contact in spatial proximity estimation.

2. Training/Testing data. To create the training and testing datasets, we identified 16 track segments from different ants, which varied in length from 100-900 frames, with each containing a consistent movement pattern that was visually assessed to fall into only one of the three alarm states throughout the segment: *Alarmed*, *Unalarmed*$_{alert}$ or *Unalarmed*$_{calm}$ (Table S1). To cross-validate the alarm states of ants in 16 track segments an experienced rater chose, ants' alarm states were evaluated by two new raters initially. The discrepant evaluations of alarm states were chosen for two new raters, and the procedures were repeated until consistent evaluations of two raters were obtained. Specifically, three paired independent evaluations of alarm states were made by six raters in total.

Table S1: Details of the 16 track segments:

Ant ID	Tracking source	Alarm Status	Alarm Strength	Frame Start	Frame End
50	Baseline	Unalarmed$_{Calm}$	0	3001	3300
1	Baseline	Unalarmed$_{Calm}$	0	1	600
3	Baseline	Unalarmed$_{Calm}$	0	1	900
2	Baseline	Unalarmed$_{Calm}$	0	1	600
26	Alarm event	Unalarmed$_{alert}$	0.5	1	500
28	Alarm event	Unalarmed$_{alert}$	0.5	1	400
4	Alarm event	Unalarmed$_{alert}$	0.5	1	200
2*	Alarm event	Unalarmed$_{alert}$	0.5	120	300
32	Alarm event	Unalarmed$_{alert}$	0.5	1	420
27	Alarm event	Alarmed	1	101	1000
54	Alarm event	Alarmed	1	1	900
10	Alarm event	Alarmed	1	60	200
19	Alarm event	Alarmed	1	1	300
25	Alarm event	Alarmed	1	1	300
7	Alarm event	Alarmed	1	61	240
42	Alarm event	Alarmed	1	91	210

*: Ant (id = 2) in the video of the alarm event is different from the ant (id = 2) in the video of the baseline.

We then applied sliding window technique to segments, producing 6462 feature vectors in total. Each feature vector includes 5 variables: the mean frame-wise speed (MS), standard deviation of frame-wise speeds (SS), standard deviation of body axis orientations (SO), convex hall area of locomotion (AR) and mean frame-wise number of contact with neighbors (MC) over the sliding window (Fig. S2). All feature vectors within a given segment were randomly assigned to either the training or the testing set. Training data consisted of 3412 feature vectors and testing data contained 3050 feature vectors (Table S2).
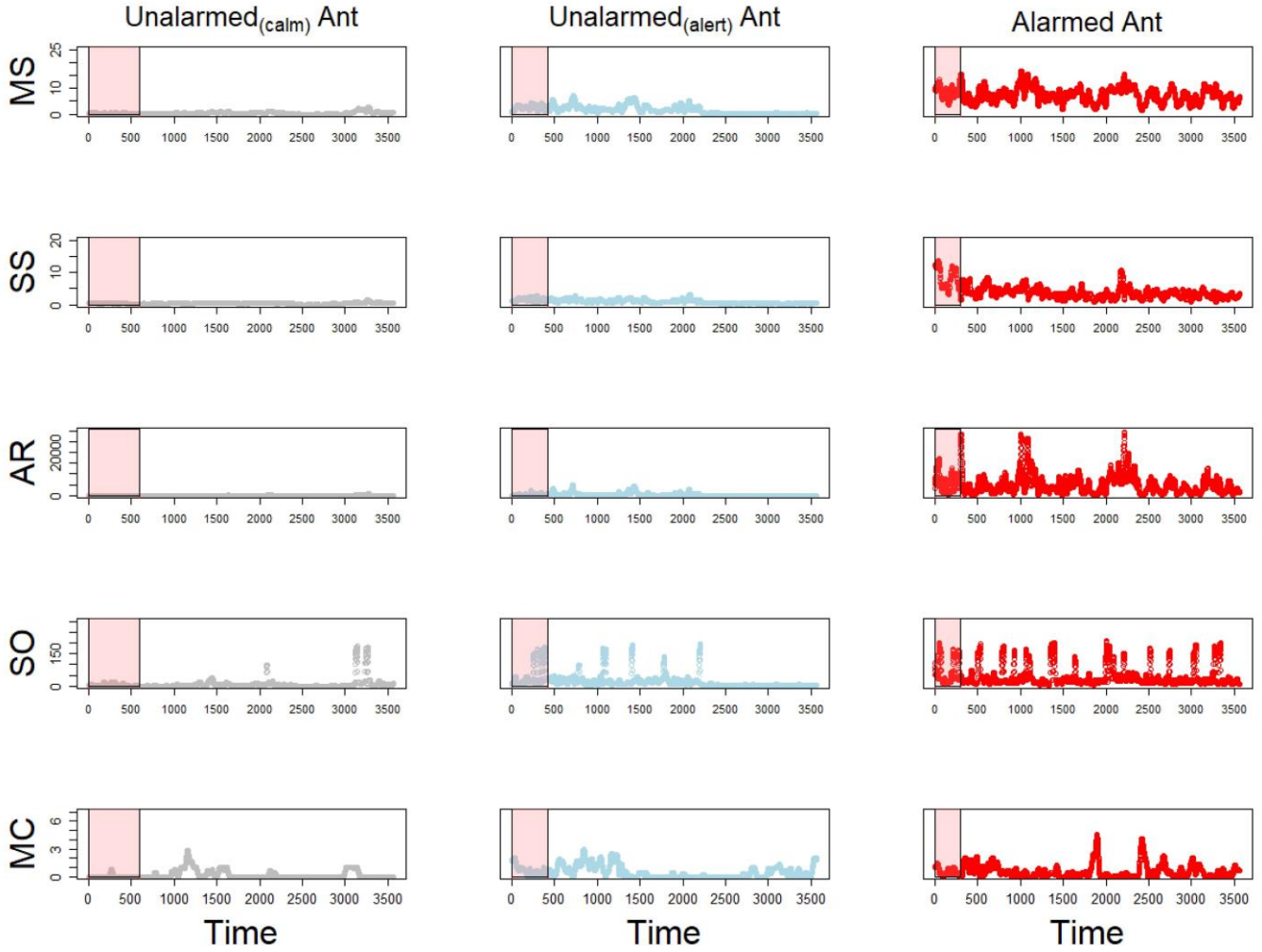
Figure S2: Features extracted from the segmental tracks. Three ants with different alarm status were given as examples (grey: *Unalarmed$_{calm}$* ant; blue: *Unalarmed$_{alert}$* ant; red: *Alarmed* ant). The alarm strength (AS) of those three ants were scored: 0 for *Unalarmed$_{calm}$* ant from frame 1 to 600; 0.5 for *Unalarmed$_{alert}$* ant from the frame 1 to 300; 1 for *Alarmed* ant from frame 1 to 150. Five feature variables over the session highlighted for training data set on three ants which were visually assessed as the *Unalarmed$_{calm}$*, the *Unalarmed$_{alert}$* ant and the *Alarmed*. MS: mean frame-wise speed; SS: standard deviation of frame-wise speeds; AR: convex hull area over the track window; SO: standard deviation of body axis orientations; MC: mean frame-wise number of contacts with neighbors.

Table S2: Amount of data used to train the machine learning model

	Alarmed (AS =1.0)	Unalarmed$_{alert}$ (AS=0.5)	Unalarmed$_{calm}$ (AS=0.0)	Total
N$_{training}$	1251	151	2010	3412
N$_{testing}$	1380	1400	270	3050
Total	2631	1551	2280	6462

3. AUC-ROC and Youden index estimations. The area under the ROC curve (AUC-ROC) was calculated (89.06%), and indicated that AUC-ROC values for the employed models are more than 80% which shows their acceptable performance. An optimal threshold for classification can be estimated by Youden index, J (J=argmax(TPR(AS) - FPR(AS))). As Youden index was equal to 0.687, we had the threshold of classification between alarmed and unalarmed status, which was equal to 74.9% (Fig.S3).

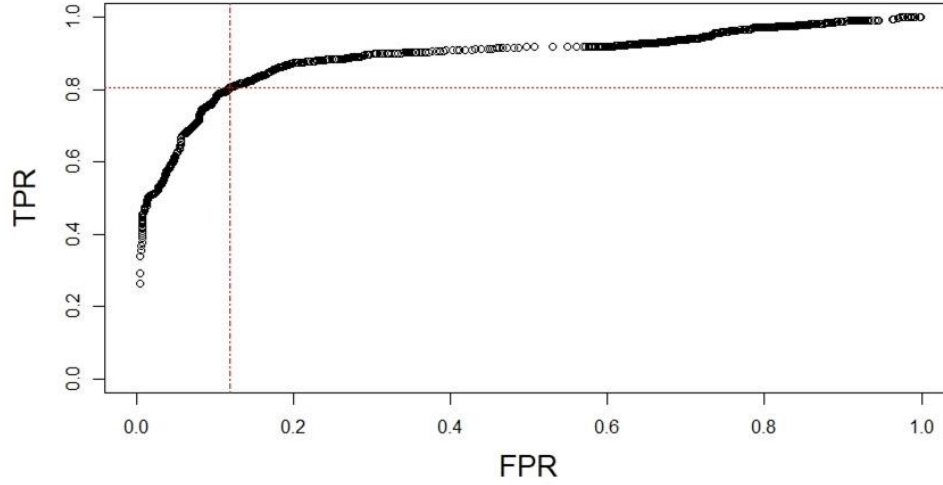**AUC=0.8906, Youden's index = 0.687, optimal-threshold = 0.749**

Figure S3. Illustration for AUC-ROC and Youden index on the ROC curve for the categorization of *Alarmed* vs *Unalarmed* status. Red dotted lines indicate the value of true positive rate (TPR) and false positive rate (FPR) as to achieve the Youden index.

4. Distance-dependent efficacy of alarm signal propagation. Immediately after the introduction of the three alarmed ants ($t=0$ sec), 5 additional ants transitioned into the alarmed state, because they came in immediate contact with initially seeded ants. We identified those 8 ants as the initially alarmed, and analyzed their impacts on nearby ants in *Unalarmed* states. For ant with id i and alarm status $s \in \{a(alarmed), u(unalarmed)\}$, A^{s_i}, set of neighboring ants is defined as

$$N(A_i^s) = \{A_j^{s'} : A_j^{s'} \in C(l(A_i^s), r)\}$$

where $l(A_i^s)$ is the location of ant and r = 135 pixel (3×45pix) is neighborhood radius, and $C(l(A_i^s), r)$ is the circle with the center $l(A_i^s)$ and radius r. Following an alarmed ant A_i^a, we find an unalarmed ant $A_j^u \in N(A_i^a)$ that meets the following criteria:

1. $\widehat{AS}(A_j^u) \leq 10\%$, that is, A_j^u is in a calm state.
2. $N(A_j^u) \subseteq \{A_i^a\}$, that is, the only potential ant in the neighborhood of A_j^u is A_i^a.
3. A_j^u has an integrated process of entering and exiting from the spatial range of a seed A_i^a, (Fig. S4).
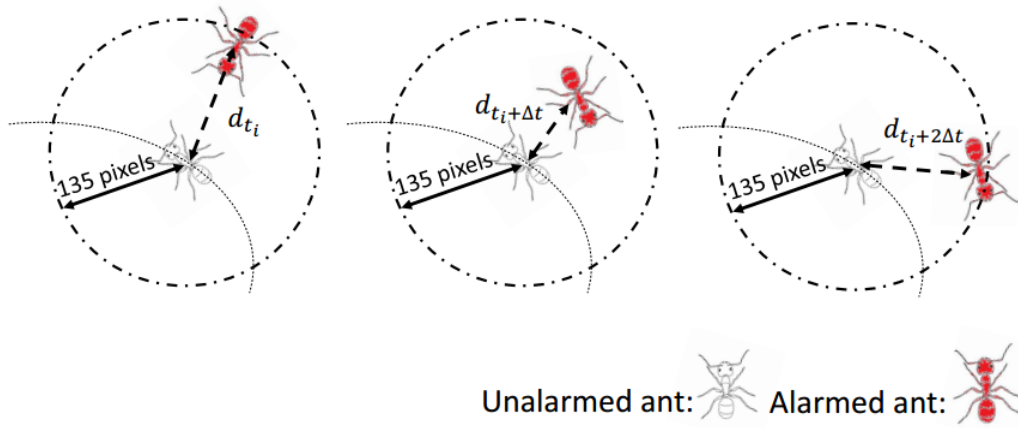
Figure S4: An illustration of movement of unalarmed ant A_j^u toward and away from alarmed neighbor A_i^a as time goes. In order to investigate the correlation between distance d and the change in alarm strength of unalarmed ant A_j^u (measured by standard deviation $\sigma_{\widehat{AS}(A_j^u)}$), we keep track of their distance at time t, d_t, and alarm strength of unalarmed ant $\widehat{AS}(A_j^u)$.

After filtering out the target alarmed and unalarmed ants, we calculate the dependent variable standard deviation of A_j^u's alarm strength ($\sigma_{\widehat{AS}(A_j^u)}$) across a time window at which she comes toward and moves away from A_i^a. The minimal distance of A_j^u away from A_i^a, d_{min}, across this time window was assigned as the independent variable.

5. Alarm signal propagation. Ants' alarm strength ($\widehat{AS}$) predicted from the Random-Forest regression model, coupled with tracking data on inter-individual distance, allowed us to identify the network pathways of alarm signal propagation indirectly by overlapping two layers of events: individual average alarm response level every second (30 frames) and its seconds-stamped contact networks. We simultaneously mapped the change in alarm state of ants engaged in contact events, based on the rule that unalarmed ants are allowed to respond to alarm recruitment with a maximum of 4 seconds' latency (Fig. S5).
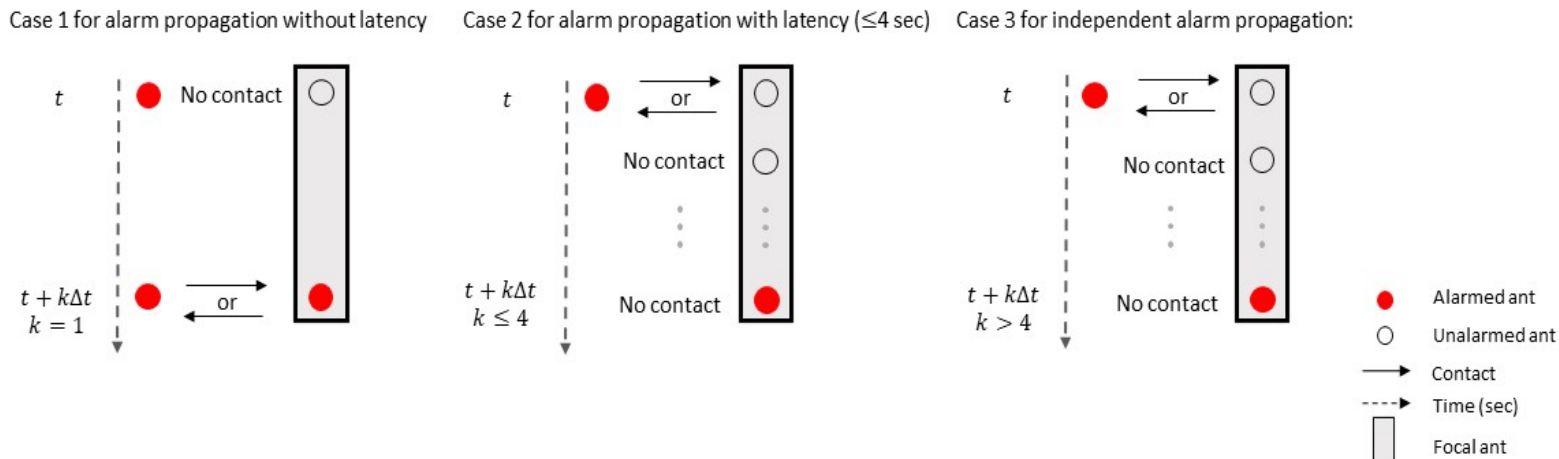
Figure S5: Networking rules of alarm-signal propagation for three cases. In cases 1 and 2, alarm propagation relies on physical contact without a latency (case 1) or with a latency less than or equal to 4 seconds (case 2) after contacting an alarmed neighbor. In case 3, alarm signal propagation is achieved independently of contact-mediated interactions.

The time-ordered and time-aggregated alarm-signal propagation networks were built by using packages of ndtv, tsna (in R 3.5.0.), networkx and teneto (in Python 3.8.3.). Physical contacts were recorded as potential pathways of alarm signal propagation whenever two ants' distance went within $d_{contact}$ = 45 pixels. Specifically, if an unalarmed ant becomes alarmed without a latency (case 1) or with a latency within 4 seconds (case 2) after alarm contact, a directional tie between the pair of ants undergoing alarm contact will be included from alarmed ant to unalarmed ant as the potential pathway of alarm propagation (Fig. S6 top). These contacts were defined as events in which previously unalarmed ants cumulatively increased their alarm strength above the threshold (0.749) for the first time and within 4 seconds' time window after contacting any alarmed ant. The weights of edges were obtained by assessing the standard deviation of $\widehat{AS}$ on unalarmed ants within a unit of time (1 second) as their contact with alarmed neighbors, which indicates the varied effectiveness of alarm contacts on alarm state transition (Fig. S6 bottom-left). If an ant became alarmed without alarm contact, we labelled this node as an independent alarm transition. The weighted highest edge to an unalarmed ant was identified as the primary pathway of alarm propagation. After pruning other low weighted edges, the static propagation network (Fig. S6 bottom-right) was formulated by aggregating the time-ordered propagation events.
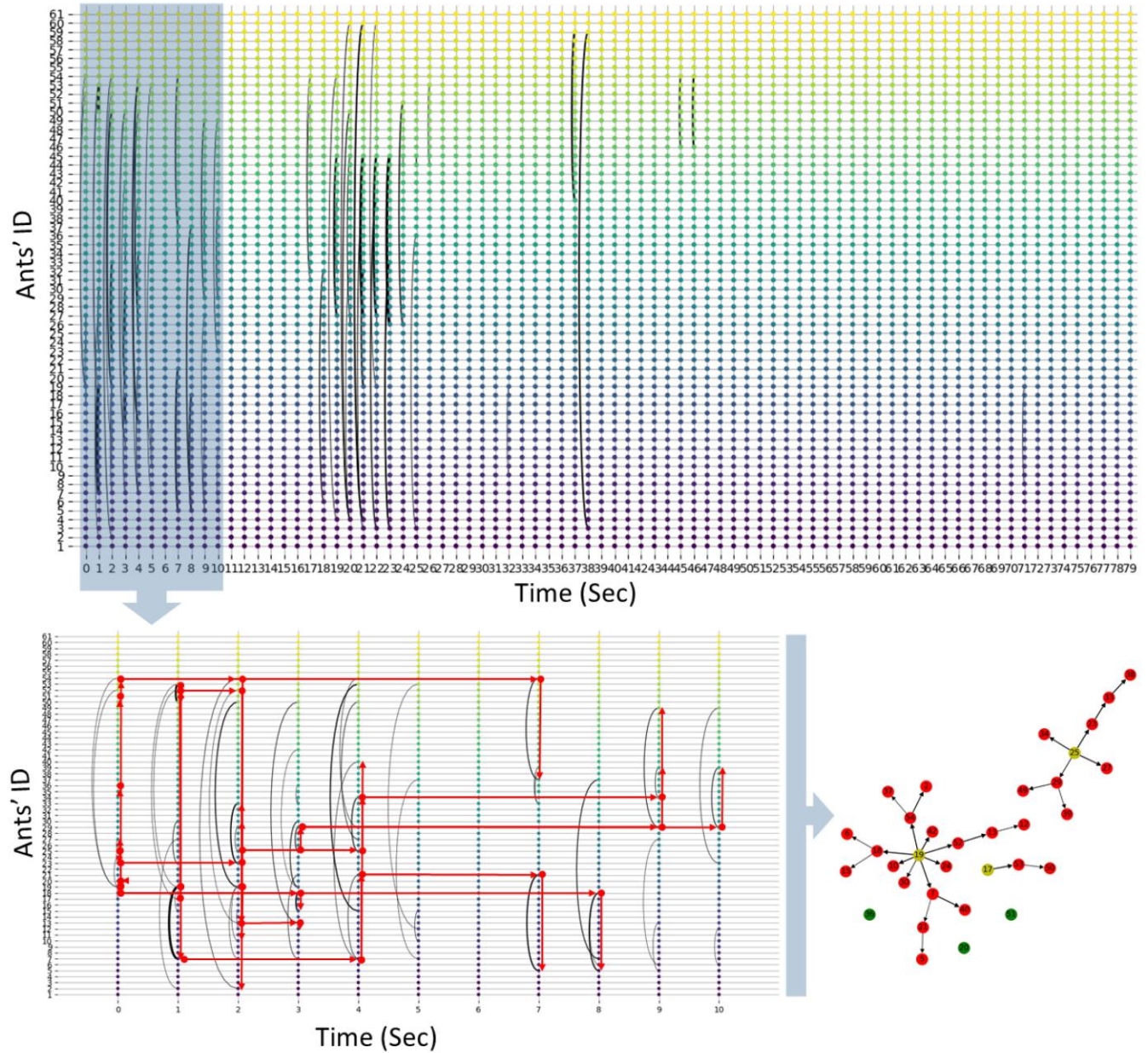
Figure S6: Alarm signal propagation networks. (Top) The time-ordered network of alarm propagation. In the network layout, y-axis indicates the id number of ants, and x-axis indicates the time in seconds. Each colored node represents an ant with the unique id number. Each weighted edge represents the varied effectiveness of an alarm contact on alarm state transition of an unalarmed ant. (Bottom-left) The time-ordered network of alarm propagation at first 10 seconds with all weighted edges. Red arrows indicated the primary pathways of alarm propagation by identifying the weighted highest edge on each unalarmed signal receiver. (Bottom-right) Time-aggregated network at first 10 seconds obtained from assessing primary pathways of alarm propagation. "Yellow" represents initially alarmed ants placed into the test arena. "Green" represents ants which become alarmed independently of contact-mediated interactions. "Red" represents ants transited to *Alarmed* via contact-mediated interactions.

Extracting and analyzing the alarm signal propagation network, we estimated the temporal dynamics of alarm contact and alarm propagation initiated by three initial alarmed ants. Within the first minute of alarm event, three initial alarmed ants contacted 55 (>90%) unalarmed ants cumulatively. Meanwhile, the proportion of alarmed ants arrived at the maximum (31.1%) at t = 7 sec and damped after.
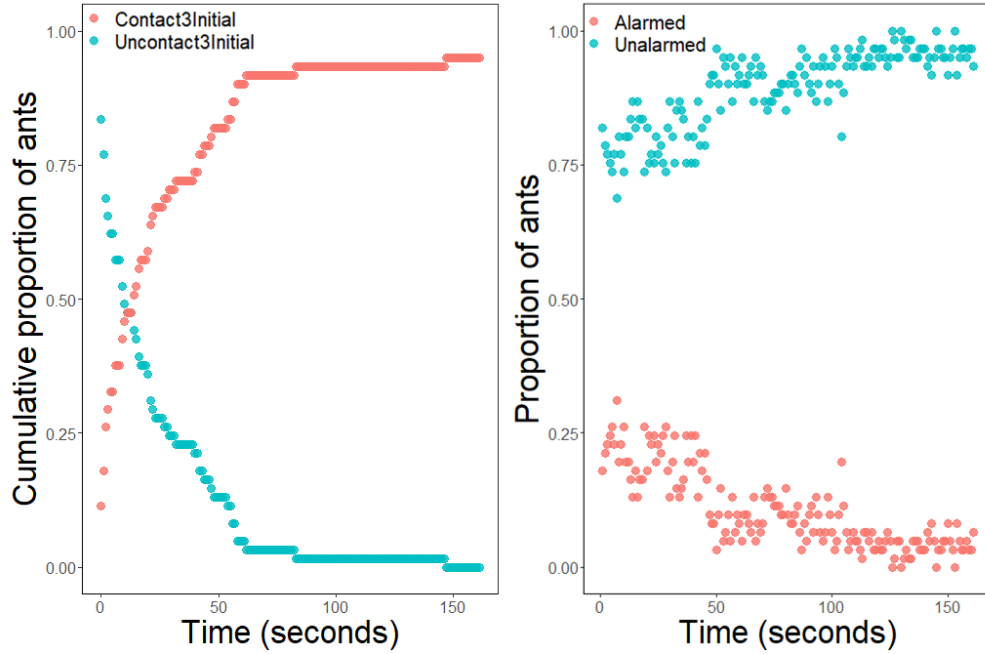

Figure S7: Temporal dynamics of alarm contact and alarm signal propagation initiated by 3 alarm seed ants. (Left) The cumulative proportion of ants contacted (red) or uncontacted (green) the 3 initial alarmed ants ever. (Right) The proportion of alarmed ants (red) and unalarmed ants (green) evaluated by the threshold of categorization (0.749).

6. Movie-S. This video shows animations of $\widehat{AS}$ across all of individual ants predicted from the RF regression model, and the alarm propagation networks (Fig. S8). It displays three animations: 1. A histogram, which tracks the distribution of $\widehat{AS}$, and the classifications after applying the threshold of categorization (0.749); 2. Colors which visualize the $\widehat{AS}$ per ant, overlapped on the live feed of ants; 3. The temporal propagation networks and nodes color-labeled by alarm recruitment events.
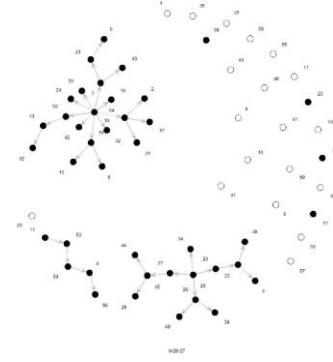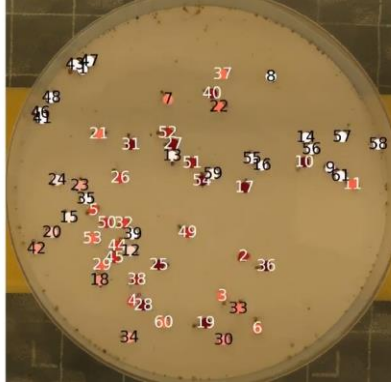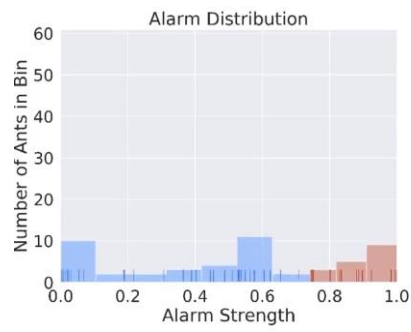
Figure S8: An snapshot of the Movie-S. (Left) A distribution of alarmed and unalarmed ants over time; (Middle) Animation by overlapping $\widehat{AS}$ per ant color-labeled with the live feed of ants; (Right) A temporal alarm propagation networks and nodes color-labeled over time.