Supplemental Information

Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion

Appendix A: Experimental methods for spontaneous fast-start experiments

Groups of 150 ± 4 juvenile golden shiners (*Notemigonus crysoleucas*) were allowed to school freely in a 2.1×1.2 m experimental tank, in 4.5 – 5 cm deep water. Measures were taken to ensure the experimental arena was acoustically and visually isolated from external stimuli. Two layers of sound insulation were placed under the tank to provide acoustic isolation, and the tank was enclosed in a tent made of featureless white sheets, within an environmental chamber, within which no persons were present during filming. We chose to work with this species because they display highly coordinated schooling behavior [1, 2]. Juvenile fish approximately 5 cm in length were purchased from I. F. Anderson Farms (http://www.andersonminnows.com), and kept in a climate controlled laboratory space for two months before they were used in experiments. There were approximately 1,000 fish in total, housed in seven separate 20 gallon tanks, at a density of approximately 150 fish per tank. The tank water was conditioned, de-chlorinated, oxygenated, and filtered continuously. Fifty percent of tank water was exchanged twice per week. Nitrates, nitrites, pH, saline and ammonia levels were tested weekly. The room temperature was controlled at 16 °C, with 12 hours of light and 12 hours of dark, using dawn-dusk simulating lights. The fish were filmed overhead at 30 frames per second, using a Sony EX1, at a resolution of 1920×1080 pixels. All fish were fed flake food three times per day, and were filmed in the experimental tank 2–4 hours after feeding. The 138 spontaneous startle events were collected from 5 different free swimming groups of 150 (± 4) fish, with each group composed of the members of a randomly selected tank. Each group was recorded for 53 minutes. The experimental procedure is identical to that of Katz et al. [1], where more details can be found. All experimental procedures were approved by the Princeton University Institutional Animal Care and Use Committee.

Stereotypy of startle response in golden shiners

Like many fish, golden shiners exhibit a startle escape response to the presence of certain stimuli [3]. There are several types of startle response and fast-start motion (e.g. C and S starts) exhibited by fish, but for this work we focused on a simple characterization of response given by the sudden acceleration of an individual, deviating well outside the norm of normal swimming behavior. Fast-starts are known to be induced through vision or sound alone, or by multi-modal combinations of unexpected stimuli [4]. We identified 138 instances of spontaneous startle events, in which a single individual within the group performs an initiating fast-start in the absence of any known cue. Such spontaneous startles were rare, occurring, per fish, approximately every 3.3 hours. This rate will likely depend on how secure the fish feel in the environmental arena. We tracked a window of 60 frames (2 seconds) around the startle event itself, where the tracking window was determined by the length of time between initiator and last responder. These fish are capable of very rapid accelerations, and sustain speeds of about 50 cm/s (Figure S1B) for a short time after the onset of a startle. We observe in our data, in both spontaneous and triggered alarms, that the cascade propagates outwards at about 100 cm/s (Figure 1C, Figure S3A). Given that our schools, to a rough approximation, measure 100 cm along the long axis when occupying the polarized state, it would only take about 1 second for the alarm to reach every individual, were the alarm sustained. Since the false alarms we study never propagate to the entire school, 1.6 seconds is long enough to observe the entire cascade.

In 71 spontaneous startle events, one or more individuals are classified as responding to the initiator. We most often observed cascade sizes of 1 or 2 individuals, with the alarm only rarely spreading to a large fraction of the school (Figure 1B). We classified startles and responses using a speed and turning rate threshold of two standard deviations above the baseline swimming speed and turning rate (this baseline varies from event to event). Initiators and first responders have speeds and turning rates higher than the school baseline (Wilcoxon rank sum test, initiators $p < 0.00001$; first responders $p < 0.00001$). The later responders tend to be slower, however, making it necessary to compare to the school's average swimming speed to identify those fish which are responding.

In many cases individuals have been found to be unable to distinguish, from the behavior of conspecifics, the reason for sudden behavioral change (i.e. whether a threat is present or absent [5]). This is especially true for organisms like fish whose startle behavior is relatively stereotyped [3]. Variability in the magnitude of the initial startle is not predictive of resulting cascade size, supporting the hypothesis that neighboring fish process an observed startle event as a binary behavior, and cannot distinguish between real alarms and false alarms, which we test explicitly below. Qualities such as the maximum speed (likelihood ratio test, $\chi^2(df = 1, N = 138) = 2.04, p = 0.15$), maximum acceleration (likelihood ratio test, $\chi^2(df = 1, N = 138) = 2.89, p = 0.09$), maximum turning rate (likelihood ratio test, $\chi^2(df = 1, N = 138) = 1.61, p = 0.21$), and maximum turning acceleration (likelihood ratio test, $\chi^2(df = 1, N = 138) = 2.67, p = 0.11$) are not good predictors of the influence of the initiator, where the magnitude quantities were fitted to the cascade sizes using Poisson regression because the dependent variable (cascade size) is a count variable. Response latency of first responder increases with distance from initiator, consistent with [6], Figure S1A. The difference between the swimming speeds of startlers, responders, and non-responders can be seen in Figure S1B. The change in speed over time characteristic of a startle is shown in Figure S1C, where we plot the mean normalized speed and standard error of all observed initial startle events, relative to time of maximum speed. The startles are characterized by a rapid acceleration followed by a slow deceleration.

The spatial extent of the behavioral cascade was calculated by measuring, for each frame, the largest distance between locations of all responses that occurred in that frame and the location of the cascade initiation. We averaged over all behavioral cascades in the data, and plotted the mean ± the standard error (Figure 1C). The speed of propagation was calculated by measuring the difference in spatial extent per frame, for each behavioral cascade in the data. Again, we averaged over all cascade events and plotted the mean ± the standard error (Figure 1D).

Golden shiners have been found to swim in both ordered and disordered states. Ordered states are characterized by high polarization (all fish swimming in a common direction), or high rotation (where fish are locally polarized, but are rotating around a common center), while disordered swimming is characterized by both low rotation and low polarization [2]. Groups of 150 fish only very rarely enter the disordered regime, predominantly exhibiting polarized or rotating motion [2]. We find that initiators occur non-randomly in the group structure. For example, we find that initiators tend to be found closer to the group boundary than non-responders (Wilcoxon rank sum test, $p < 0.0001$ all group states), and in polarized states, initiators are more likely to be found closer to the group front than non-responding fish (Wilcoxon rank-sum test, $p = .008$, polarized groups). However, locations of initiators don't tend to happen in more or less dense regions than non-responding fish (initiators: Wilcoxon rank sum test, $p = .09$, all group states). See Figure S4 for details. We found evidence that cascades propagate slightly further in polarized groups than in rotating groups, a weak, but statistically significant increasing trend ($R = 0.23$, $p = 0.02$).

Response to induced fast-start behavior Considering the inherently stereotyped nature of fast-start behavior, and the lack of sensitivity of responders to the magnitude of the initiating startle (see above), we expect that spontaneous fast-starts are responded to in the same way as induced fast-starts (i.e. those that are triggered in direct response to an aversive stimulus). To test directly this hypothesis, and to validate our approach, it is necessary to compare social contagion resulting from spontaneous and triggered evasion maneuvers. However, to avoid confounding asocial and social factors in the response, we must employ a stimulus that triggers fast-start in focal individuals but is not perceived by potential responders.

Initially we attempted to employ a laser mounted on a series of servo motors positioned 2 meters above the tank. Custom-designed software allowed us to aim quickly at any location in the tank and then to turn on and off the laser, providing a rapid and relatively localized 'blink' of light. While this approach was successful at inducing startles, and could be easily controlled, we found that it was not a truly local stimulus since light was scattered off both the water surface and the base of the tank. Furthermore we could not exclude the possibility that the necessary motion of the laser mounting may have been perceptible to the fish below.

Following this attempt we focused on developing a device that could provide local tactile (mechanosensory) stimulation to induce fast-start evasion behavior. We designed the monofilament perturbative stimulus illustrated in Figure S2 that allows us to automatically raise a thin filament (made out of polyethylene) across a channel of dimensions 122 cm $\times$ 30 cm within our experimental arena, using a HiTec HS-5965MG high-speed programmable digital servo to control motion. Following computer-controlled initiation by an experimenter, the filament was raised from the base of the tank to a height of 3 cm within the water body, before lowering to the base again, with the full procedure taking less than 200ms. To remove potential acoustic stimuli caused by this motion, the design allowed the servo motors controlling the motion to be placed outside the tank (Figure S2A) and teflon was applied to the areas where the filament was in contact with the tank structure to minimize friction, and thus minimize sound production. A hydrophone was placed in the tank allowing us to validate quiet operation and generally effective sound insulation.

The striking of fish by the monofilament induced fast-starts, whereas releasing the mechanism in close proximity to fish without physical contact resulted in no response (Figure S3B), thus demonstrating that perturbations were effectively localized. We triggered startle responses in schools of 150 fish, recording the motion of individuals at 150 frames per second, using a Prosilica GX-1050 (with a 9mm F1.4 c-mount lens, and resolution of 1024 $\times$ 768). The fish were allowed to habituate for one hour prior to triggering startles. Following the habituation time, startles were triggered at each of the three monofilament locations (see Figure S2B), over the next three hours, with a minimum of five minutes between triggering events. In total, we collected 67 triggered startle events. Shiners maintain a density of 0.02 fish/cm^2 in the experimental channel, matching what we observe in free swimming schools (Figure S4).

The trajectories of all fish were obtained via automated tracking (see Appendix B, below). The timescale of individual response to triggered fast-start evasion (Figure S3C, for response to spontaneous and induced fast-startles) and the speed of resulting waves of evasion (Figure S3A, for response to spontaneous and induced fast-starts) were indistinguishable, providing evidence that our fish do not differentiate between spontaneous fast-starts and those triggered by exposure to an aversive stimulus. We performed linear regressions on the speed of propagation for both spontaneous and triggered startle events, compared the regression coefficients (slope and intercept) with t-tests, and found that the slopes and intercepts for both types of startles do not differ, with p-values of $p = 0.97, p = 0.49$, respectively.

Since the triggered startles were conducted at a frame rate of 150 fps, compared to a frame rate of only 30fps for the spontaneous startle data, we smoothed the triggered startles using a 5-frame smoothing window, and plotted both speed profiles in Figure S3C. Smoothing was necessary in order to be able to compare the maximum fish swimming speeds between datasets.

Appendix B: Description of tracking software

Automated tracking of large groups of fish is a challenging problem due to their non-rigid body structure, as well as partial and total occlusions of individuals when viewed from above, generated by fish swimming over and under one another. Occlusions occur in our experiments approximately twice a minute per individual for free-swimming groups of 150 fish, which is enough to be problematic for accurate tracking. The SchoolTracker computer vision software used in this project (developed by HSW) was created to reliably solve this problem. SchoolTracker is comprised of three modules: detection, tracking, and track linking. The detection module takes video recorded from a high-resolution camera mounted above the experimental arena (providing an overhead view of the school) and identifies possible fish body positions and orientations. This step is made difficult by the presence of occlusions, which is why most image segmentation methods fail. Instead, the detection module looks for local features such as corners and line segments, and systematically estimates first fish head positions from the corners, then body position and orientations from the optimal assignment between line segments and estimated head positions. It then determines the statistical threshold values based on the distributions of head intensity or fish image intensity difference and removes the detection outliers based on this threshold.

Given a set of detections for each frame, the tracking module identifies which detections from one frame to the next belong to the same individual, forming a track for that individual over time. The multiple object tracking algorithm we use is able to track the complex motion of each fish, and resolve oc-

clusions (overlapping bodies) by matching against a library of learned robust body images. The algorithm first estimates the motion states of each individual via two types of adaptive alpha-beta filters (a form of simplified Kalman filter [7], but modified for tracking fish in this instance), and then associates the predicted future states of the individuals with detections in the next frame by minimizing the sum of association costs, which in our implementation was defined by the difference of head positions, body angle and corresponding learned body images. This is formulated as a linear assignment problem, and is solved by the data association algorithm that is similar to what we proposed previously in [8]. It scales very well when handling a large number of fish as it was designed to deal with detection error, temporal disappearance due to short-term occlusion or missed detections.

Due to some (rare) long-term occlusions (e.g. when two fish completely overlap for many frames), the tracks constructed by the tracking module may be broken into shorter segments. Since the number of fish in the experimental arena is constant throughout the experiment, the track linking module is used to piece together track segments via combinatorial optimization, to construct coherent tracks that span the length of the full video. Upon inspection, manual relinking is performed to correct errors. By itself, the automated system recovers over 95% of tracks without misidentifications, and over 97% of tracks without missing frames, minimizing the amount of post-tracking correction necessary (accuracy evaluated on 10 randomly selected, 120 frame video segments of 150 free-swimming golden shiners recorded at 30 fps).

The SchoolTracker software provides a simple graphical user interface to streamline the process of parameter tuning and track verification/editing, so that the final results accurately reflect the positions and orientations of each individual fish in the school over the whole length of the video.

Appendix C: Description of vision software

Visual, acoustic, touch, and lateral line sensory modalities may all contribute to fast-start behaviors. Due to the importance of vision to schooling in Golden shiner fish [9] and the relatively small contribution of the lateral line [10], we (CRT) developed a method for approximating the planar field of view of each individual in a group when individuals are restricted to an effectively planar environment as in this experiment (restricted depth). After tracking, the positions and orientations of each individual were used as the first step in estimating the posture of each individual's body. Posture estimation is trivial when fish are non-overlapping: there is a null-cline in the image Laplacian that traces the midline of the fish's body. Points along this null cline are sparsely sampled using a kd-tree ($k = 2$) data structure, and linked using a greedy energy minimizing line following method, where the energy of a line is minimized by reducing the distance between subsequent points, and minimizing the angle of curvature between adjacent line segments. Finally, a cubic basis spline is fit to the selected points to approximate the midline curve, while the flanks of the individual are estimated by searching for boundary pixels (after thresholding to remove the background) perpendicular to the tangent of the curve, and fitting a linear function to their distance relative to the curve (similar to [11]; here we use the Thiel-Sen estimator for robustness). This gives an approximation of the body of each individual summarized by the coefficients of the basis spline and the slopes and intercepts of the left and right flanks.

Posture estimation is substantially complicated by the occurrence of partial occlusions. This is why skeletonizing after thresholding out the background fails to work well for mid-line estimation, as clumps of individuals yield highly irregular skeletons. The method above is robust to this, but may still fail when individuals cross while their bodies are curved. To counter this, information about body posture estimation on previous frames is used to predict the posture of an individual in the future using a simple Kalman filter.

After the body posture for every individual in the school has been characterized, an estimation of the location of each eye, for each individual, is made using the current heading of the fish, the position of the fish's head, and the estimated flanks. From the midpoint position of the head, the transition points from body to background on either side of the individual (perpendicular to the heading) are determined. Estimates are improved using a Kalman filter to smooth over brief occlusions and possibly noisy detections. Line segments encoding the shape of every body are stored in a simple spatial data structure (in this case a 2d hash table) for fast look-up based on spatial coordinates. For each eye of each individual, rays are cast (at a resolution of $\pi \times 10^{-3}$ radians per ray) from that position and the point of first intersection with other geometry in the frame is recorded (distance and identity). The total number of rays from an eye intersecting each individual in the scene is recorded as the angular area subtended by that individual on the focal individual's eye. The circular mean of the rays intersecting the other body is recorded as that individual's angular position relative to the focal individual. An example showing line of sight and the angular area occupied by one focal individual in the field of view of each member of a group is shown in Movie S1.

Appendix D: Features used in model selection

Preliminary analysis allowed us to eliminate the Voronoi neighborhood as a predictive feature in our regression models. The Voronoi tessellation of the group was determined by computing the dual of the Delauney triangulation of individuals' positions [12]. The Voronoi neighborhood alone is a good predictor of first responders, as it contains similar information to other features such as metric and topological distance. However, when it is included in models with these other feature variables, the Voronoi neighborhood loses significance (logistic regression, likelihood ratio test, $p = 0.15$). This is reasonable, as the Voronoi neighborhood is a binary variable, and doesn't contain as much information as metric distance or angular area.

In our main model selection procedure, we included 12 feature variables, motivated by the literature [13, 14], and listed in Table S1. The features each belong to one of two categories: absolute or comparative. Absolute features are those that contain no implicit information about what other neighbors are or are not doing. Examples of absolute features are metric distance, relative position, relative heading, and subtended angular area. Comparative features do contain implicit information about other neighbors. Some examples of comparative features are ordinal variables, such as topological distance and ranked angular area, and relative features, like relative metric distance ($\mathrm{MD}_{\mathrm{min}}/\mathrm{MD}(i)$) and relative angular area ($\mathrm{A}(i)/\mathrm{A}_{\mathrm{max}}$).

Although there is some collinearity in the predictor variables, as is to be expected since some predictors such as metric and topological distance are good proxies for each other, the correlations are not strong enough to merit any predictor's exclusion from the model selection. We inspect the variance inflation factors (VIF) between pairs of predictor variables.

$$VIF = \frac{1}{1 - R^2} \tag{S1}$$

A good rule of thumb [15] is that VIFs less than about 10 are acceptable. Metric distance and topological distance are the most collinear, with a VIF = 2.55, so we can justify including all pairs of predictors without being overly redundant.

Log-transformed versions of features were also included as sensory perception is often on a log-scale [16]. We included log-transformed versions of features for which taking the log improved the maximum likelihood of that individual feature predicting the data (Figure S5). Along with the log-transforms, two more transformations were needed, in order to change relative angular position (θ), and loom (L) into forms that made sense for a logistic regression model, such that a monotonic relationship could be expected between the independent and dependent variables. We transformed θ into $cos(\theta - \phi)$, where ϕ was the phase shift resulting in the optimal log-likelihood for the feature (Figure S6A). Similarly, we selected $|L|$ (absolute value of loom) instead of L, because $|L|$ had a higher log-likelihood (Figure S6B). Neither of these features ended up performing well in the model selection procedures.

Multi-model inference

For each feature subset, a generalized linear model (GLM) with a logistic link function is fit to predict the probability that a witness of the startle event would be likely to be the first responder. We use a multi-model inference technique to find the subset of the 12 features and their log-transforms used in our analysis that best fit our data. The important features are not simply the features used in the "best" model, as this is likely subject to overfitting given that we are testing every possible feature combination, and thus could vary depending on the exact dataset used. Instead, the relative importance of each feature, ω_f, is estimated following [17] as

$$\omega_f = Z^{-1} \sum_i \delta_i^f \exp(\mathrm{BIC}_{\min} - \mathrm{BIC}_i) \tag{S2}$$

$$\mathrm{BIC}_i = l(\hat{\theta}_i) - \frac{k_i}{2} \log(N) + O(1) \tag{S3}$$

$$\approx \log P(s_1 | s_0, M_i) \tag{S4}$$

Where BIC is Schwartz's Bayesian Information Criterion, Z is a normalization constant, $l(\hat{\theta}_i)$ is the maximum likelihood estimation for the parameters of model i, k_i is the number of features in model i, and δ_i^f is an indicator function that is 1 if model i contains feature f, and zero otherwise. The BIC of each model i is compared against the minimum BIC score obtained by any model, $\mathrm{BIC}_{\min}$, such that features found in models close to $\mathrm{BIC}_{\min}$ are given more weight than features in models significantly greater than $\mathrm{BIC}_{\min}$. $P(s_1|s_0, M_i)$ represents the probability of a fish being a first responder (s_1), given a startle by a neighbor fish (s_0), using model M_i.

Results from model selection on the 12 untransformed candidate features and 7 log-transformed candidate features, when ranked by feature weight, place log Euclidean distance, ranked angular area, log ranked angular area, and log topological distance as the top four candidates for inclusion in any model (in that order, Figure S7). However, one important consideration is the extent to which different features exclude each other. Although two features may have similar relative importance scores, and thus seem to both be necessary for a good model, the presence of both in the same model may be redundant with one essentially excluding the other. This is the case for log metric and log topological distances (Figure S8); one is sufficient for inclusion in a model, and of the two log metric distance has a higher relevance score (the relative feature importance for metric distance is four times greater than for topological distance, Figure S7). Similarly, log ranked angular area and ranked angular area mutually exclude each other. The remaining features have relevance scores that are even lower, more than ten times lower than the top feature, so we exclude them from the final model. We calculate this joint conditional importance as follows:

$$C_{fg} = \frac{\sum_i \delta_i^f \delta_i^g \exp(\mathrm{BIC}_{\min} - \mathrm{BIC}_i)}{\sum_i \delta_i^f \exp(\mathrm{BIC}_{\min} - \mathrm{BIC}_i)} \tag{S5}$$

Taking this into account, as well as the relative ranking of best models (Figure S7B), we find that a model that includes ranked angular area and log metric distance is supported best by the data.

We use this pairwise mapping of sensory input to behavioral response to build the interaction networks. While the majority of strong connections in these networks occur within a small local neighborhood, there also exist many long range weaker connections, connecting more distant parts of the network (Figure S9). An example of the dynamic network is shown in Movie S2.

L_1-penalized logistic regression for feature selection

An alternative to the multi-model inference above is to fit a logistic regression model with all variables present, but with an L_1 norm imposed on the regression problem to both prevent overfitting and favor sparse models. We employ a logistic function to model the probability that an individual will be the first to respond to an initial startle, based on the ensemble of features quantified about the initiator.

To enforce that only non-redundant features are used, θ (the vector of coefficients weighting the variables present in the logistic function modeling $P(s_1|s_0)$, where s_0 represents the initiator of the startle event, and s_1 is the first responder) is chosen such that the maximum likelihood estimate of the model is balanced against the L_1 norm of θ, with the strength of the penalty governed by λ. By varying λ, the tradeoff between model complexity (the number of features with non-zero coefficients) and the predictive power of the model can be explored, with higher values of λ yielding less complex (fewer non-zero coefficients) models of $P(s_1|s_0)$. The goal is to find

$$\max_\theta \sum_{i=1}^{N} \log P(s_1^{(i)} | s_0^{(i)}) - \lambda ||\theta||_1 \tag{S6}$$

where i indexes all the candidate initial startle/first responder pairs in the data, and $||\cdot||_1$ is the L_1 norm.

Choosing a model based on $\lambda_{\min}$ (the value of λ at which the binomial deviance of the model is minimized) gives a set of features with nonzero coefficients that represent the best balance between model performance and the number of features used by the model. A suggested heuristic to be even more conservative about overfitting is to use the smallest $\lambda_{1\mathrm{SE}} < \lambda_{\min}$ that is within 1 standard error of $\lambda_{\min}$. Results from this L_1 regularization are shown in Figure S10.

Interaction terms. To ensure that we are not missing an important synergistic effect from feature pairs, we separately consider models which include interaction terms from the top 4 performing features. Only the top individually performing features were included because using all pairwise interaction terms, let alone all interaction terms, from our full feature

set would be computationally prohibitive. For pairwise interactions, the number of terms grows with the square of the number of features, while for all possible interactions the number grows exponentially with the number of features. In an L_1 regression, none of the interaction terms were selected as important (Figure S11), leading us to conclude that interaction effects from pairs of features are not necessary in the model.

Appendix E: Group spatial measures

Local density. We draw a circle of radius R around the focal fish, and count the number of neighboring fish falling in that circle. We divide this number by the area of the circle that falls within the group boundary, to get rid of unwanted edge effects.

$$\rho_i = \frac{\sum_{j=1}^{N}(d_{ij} < R)}{C_i \cap A_{\text{group}}}, \tag{S7}$$

Where d_{ij} is the distance from fish i to fish j, R is the local density radius, C_i is the circle of radius R around fish i, and A_{group} is the area of the group. To make sure the boundary density is not being affected by having a tiny number of fish fall within the local radius, R, we define this local radius as the distance to the Nth nearest neighbor, thereby ensuring that at least N neighbors fall within R. An example of local density is shown in a sample fish group in Figure S12A.

Distance from group boundary. We detected the group boundary using an alpha-shape method. A boundary edge is drawn between any two fish, P_i and P_j in the group where a disk of radius $1/a$ can be placed such that it contains all points in the group and P_i and P_j are on its boundary. We use the alpha shapes implementation available on the Matlab file exchange platform [18], with a radius of 25 cm. Once we have identified the boundary individuals, we define a fish's minimum distance from the boundary as the minimum distance to any point on a line connecting two consecutive boundary fish. Fish located on the boundary have a minimum distance to the boundary of zero. An example of distance from group boundary is shown in a sample fish group in Figure S12B.

Distance front-back. We calculate the heading direction of the group as the mean heading of all fish. We then measure the distance front-back as the shortest distance from each fish's position to the line drawn through the group center of mass along the group heading. A fish has a distance front-back of 0 if there are no other fish further behind the group center of mass. Likewise, a fish has a distance front-back of 1 if no other fish are further ahead of the group center of mass.

$$d_i = (\bar{r_i} - \bar{r}_{\text{COM}}) \cdot |\bar{H}| \tag{S8}$$

Where $\bar{H}$ is the mean group heading, r_i is the fish position, and r_{COM} is the group center of mass position. An example of distance front-back is shown in a sample fish group in Figure S12C.

Influential regions (as estimated by the local clustering coefficient) tend to be found near the front and near the boundary of groups (Fig. S12D, Fig. 3E,F main text).

Polarized/Rotational states. We calculate polarized and rotational states following Tunström et al. [2]. The polarization of the group is a value between 0 and 1, where 0 is completely unpolarized, and 1 is completely polarized. The polarization is the average of all individual headings (u_i). A group is categorized as polarized if it has a value of $P > .65$ and $R < .35$, and it is categorized as rotating if $P < .35$ and $R > .65$.

$$P_{\text{group}} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} u_{xi}^2 + u_{yi}^2} \tag{S9}$$

The amount of rotation around the group center of mass position is calculated as follows:

$$R_{\text{group}} = \frac{1}{N}\left|\sum_{i=1}^{N} u_i \times r_i\right| \tag{S10}$$

Where a group is completely rotational if $R = 1$.

Positioning of individuals within group. On a 10 minute video for 20 verified correct tracks, we asked what is the typical duration of individual spatial fidelity within the group, with respect to the group boundary. We find that the residency time of any particular position in the group is relatively short (about 25 seconds) given the decay in the autocorrelation function for distance from the boundary (Figure S13A,B), and all 20 individuals of the random subset selected spend time on both the boundary and in the interior of the group, thus individual motion in golden shiner schools is relatively fluid.

Appendix F: Spatial structure of the network

Although many global network measures that exist in the literature have been useful in characterization of these networks, such as efficiency measures and centrality measures, we find that such global metrics are not as applicable to complex contagion processes as they are to simple contagion processes. These global measures are also useful when information often reaches the entire network, unlike in our data where information spread is most often localized to a small portion of the network, although it can spread extensively. Instead we focus our investigation on local network measures. We include distributions of some of these local measures in Figure S14.

A simple measure of path redundancy is the local clustering coefficient, which compares the number of closed triangles (where all three neighbors are connected) to the number of total possible triangles (all pairs of neighbors with the focal node). We use the weighted directed clustering coefficient from [19] as a simple extension of the clustering coefficient concept to weighted and directed graphs, which is defined as the following;

$$C_i = \frac{1}{2}\frac{\sum_j \sum_k (w_{ij} + w_{ji})(w_{ij} + w_{ki})(w_{jk} + w_{kj})}{[d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}]} \tag{S11}$$

$$d_i^{in} = \sum_{j \neq i} a_{ji}, \quad d_i^{out} = \sum_{j \neq i} a_{ij} \tag{S12}$$

$$d_i^{tot} = d_i^{in} + d_i^{out} \tag{S13}$$

$$d_i^{\leftrightarrow} = \sum_{j \neq i} a_{ji} a_{ij} \tag{S14}$$

Where w_{ij} is an entry in the weighted adjacency matrix, and a_{ij} is an entry in the binary adjacency matrix. We found that some characteristic network properties are spatially structured. Most notably, the local clustering coefficient tends to be found in regions of local density, near the boundary. In polarized groups, fish with high clustering coefficients tend to be found near the front of the group (Figure S15, Figure 3, main text).

Additionally, we find that there exist regions of local clustering in the fish schools (Figure S16). We calculate the correlation function for fluctuations in clustering coefficient as a

function of distance. The correlation length, which is the distance beyond which there tends to be no similarity between pairs of clustering coefficients, averaged over all nodes in all networks in our dataset is 32 cm.

$$\psi(r) = \frac{\sum_{i,j} \delta(r < d_{ij} < r + \epsilon)(CC_i - |CC|) \cdot (CC_j - |CC|)}{\sum_{i,j} \delta(r < d_{ij} < r + \epsilon)} \tag{S15}$$

Where $|CC|$ is the average value of local clustering coefficient over all nodes in the network, so that we are measuring fluctuations from the mean value, δ is an indicator function that selects pairs of nodes i and j such that the distance between them is between r and $r+\epsilon$. The denominator is just the normalization by number of node pairs which are this distance r apart.

Exterior visual field. Access to visual information outside the group is important for obtaining direct visual information regarding predators. We evaluate this exterior visual field of each fish by measuring the fraction of its field of view which extends to the tank boundaries without intersecting with a neighboring fish. As expected, we find that the fraction of the field of view extending outside the group is much higher for fish located on the group boundary, as well as those near the group front and group rear (Figure 3E,F main text, Figure S17).

Fitting local clustering coefficient model of cascade size

We determine how well the local clustering coefficient of the initiator describes the resulting cascade size by fitting a generalized linear model with log link function (Poisson regression because our dependent variable is a count variable), to the data using the local weighted clustering coefficient of the initiator as a predictor variable. We find that this model is highly significant (likelihood ratio test, $\chi^2 = 78.1, (df = 1, N = 138), p < 0.00001$). To test whether this effect is truly due to the network structure, or results from other correlated quantities, such as local density, out-degree, and distance from group boundary, we estimate the relative contribution of each feature in two ways (Figure S18). First we evaluate the relative importance of each feature using L_1 regularization. We find that, at the point of minimum deviance and minimum deviance plus 1/2 standard error, the local clustering coefficient has the highest relative weight. Second, we evaluate each variable's unique contribution to the explained variance of the cascade size distribution [20]. The local clustering coefficient has the highest unique contribution to the 69% of variance explained by the total model. The sum of the values in Figure S18B is smaller than 69% due to correlations among the variables. The strong performance of clustering coefficient in these models suggests that the interconnectedness of a node's neighbors is essential to understanding the way in which alarm cascades flow through a group.

Along with fitting models to the cascade size data, we also plot the mean and standard error of cascade size as a function of clustering coefficient. We assign the data to logarithmically spaced bins, of sizes such that all but the very smallest contain four or more data points, and calculate the mean and standard error of the cascade sizes in each bin. We find that most of the mean values fall within the 95% confidence interval for the generalized linear model fit (blue shaded region, Fig. 3C, main text).

Alternative versions of weighted, directed local clustering coefficient. The generalization from the binary undirected clustering coefficient to weighted directed clustering coefficient ($WDCC$) is not entirely obvious. We use the most widely accepted definition [19] for the majority of the analysis. However, one drawback to this version is that it does not distinguish between in-connections and out-connections, but rather sums the in-strength and out-strength. For the purposes of considering influential and susceptible individuals this may not be the best version to use. Only the out-connections of a focal node, and not the in-connections will have bearing on that node's influence. Likewise, only the in-connections will be relevant to a node's susceptibility. For this reason we propose two slightly modified versions of the weighted clustering coefficient, defined as follows.

$$C_i^{out} = \frac{1}{2} \sum_j \sum_k (w_{ji})(w_{ki})(w_{jk} + w_{kj}) / [d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}] \tag{S16}$$

Where C_i^{out} is equivalent to (10), except for removing the in-connections from the focal node.

$$C_i^{in} = \frac{1}{2} \sum_j \sum_k (w_{ij})(w_{ik})(w_{jk} + w_{kj}) / [d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}] \tag{S17}$$

Where C_i^{in} is also equivalent to (10), now removing the out-connections from the focal node and only keeping the in-connections. Using the modified C^{out} results in a slightly better model when fit to the cascade size data (likelihood ratio test, $\chi^2 = 94.7, (df = 1, N = 138), p \leq 0.00001$, compared to $\chi^2 = 78.1, (df = 1, N = 138), p \leq 0.00001$ for the unmodified $WDCC$), as well as a slight improvement when we look at responder susceptibility (see the following section), (likelihood ratio test, $\chi^2 = 57.8, (df = 1, N = 5502), p \leq 0.00001$, compared to $\chi^2 = 57.7, (df = 1, N = 5502), p \leq 0.00001$). The difference between in-connections and out-connections tends to be relatively small compared to the weight of the connections themselves, which is why the effect of using these clustering coefficient variations is not tremendous.

Alternative network measures. We compare the local clustering coefficient to network measures that have been used in different contexts to identify influential nodes [21, 22, 23]. A summary of the results is included in Table S2. In unweighted networks, the k-core is the set of nodes which have at least k-connections [24]. A node's k-core number is the largest k-core to which it belongs. As a definitive weighted version of the k-core hasn't yet been established, we approximate a weighted k-core by averaging k-core numbers of each node in unweighted networks created by binarizing the weighted network at different threshold values. We also normalize the k-core numbers such that they fall between 0 and 1, to account for networks that have been binarized using higher thresholds having a smaller number of connections, and thus a smaller number of cores. The unweighted k-core number of initiator nodes is negatively correlated with resulting cascade size. This is likely due to the fact that central nodes tend to have higher k-core numbers, and these central nodes experience more reaction inhibition, as a result of having more neighbors. The model using weighted k-core number of initiator nodes is not significant, (GLM fit, likelihood ratio test, $p = 0.179$).

We also consider the weighted betweenness centrality of the initiator nodes as predictors of cascade size, given by:

$$g(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{S18}$$

Where σ_{st} is the number of shortest paths passing from node s to node t, and $\sigma_{st}(i)$ is the subset of those paths which also

pass through node i. Nodes with high betweenness centrality are less likely to be influential. This gives more support to the hypothesis that cascades in our system spread through fractional contagion. Nodes with high betweenness centrality are located in the center part of the group, where they are more inhibited from responding due to having a larger number of neighbors.

The edge weight variation is calculated as the variance in edge connections per node, as a measure of how homogeneous or heterogeneous a certain node's connections are. This measure is also a negative predictor of cascade size, but the significance is most likely explained by spatial effects (central nodes tend to have more edge-weight variation).

Susceptibility

We find that the same network properties that determine an individual's influence also determine its susceptibility, where we define susceptibility as an individual's likelihood of responding to an observed startle event. This concept of susceptibility results from looking at responses from an individual's perspective, while influence results from looking at responses from an initiator's perspective. An individual will be more likely to respond (is more susceptible) if it is strongly connected to the initiator (short path length), and if it has neighbors which are strongly connected to each other (high clustering coefficient). The mechanism by which shortest paths are important is simple to understand: shortest paths represent most probable paths. In fact, in our data we find that, even when controlling for Euclidean distance, the shortest path length retains significant power for predicting responders (logistic regression, (df = 2, N = 5502), $p < 0.00001$). We also find that a node is more susceptible if it is in a highly clustered region, that is, if it is in a region with strong interconnections between neighbors. A node that is a member of a highly clustered region can take advantage of observing multiple responders. In a simple contagion process, these multiple ties would be redundant, but in complex contagion, they are helpful in instigating a response. Here we define susceptibility as the likelihood of a fish responding given that it observes the initiator.

We measure the likelihood of late responders (all responders after the first responder) as a function of clustering coefficient (controlling for in-degree, local density, distance from boundary, and edge-weight connecting observing fish to startled fish), and find that nodes with a higher $WDCC$ are more likely to respond to an observed startle event than those with a lower $WDCC$ (logistic regression, (df = 5, N = 5502), $p < 0.00001$). We consider only subsequent responders (after the first) because these are the fish that have the chance to benefit from being in a highly clustered region by observing multiple responders. When we repeat the same analysis for just the first responder, we find that the clustering coefficient (interconnectedness of its neighbors) is not important (logistic regression, (df = 5, N = 5502), $p = 0.50$). This is because the first responder does not experience multiple exposures, and thus cannot be affected by a clustered neighborhood, thus we only expect to see the complex propagation effect in individuals responding after the first responder. Both perspectives (influence and susceptibility to influence) support the main point; that redundant paths amplify information, by asking on the one hand, "given the local neighborhood of an initiator, how many fish are likely to respond?", and on the other hand, "given the local neighborhood of any potential responder, how likely is a response from that fish?".

We use k-fold cross-validation to prevent overfitting when evaluating influence and susceptibility effects. This was done by partitioning the data into $k = 10$ subsets, training a model on $k - 1$ subsets, and testing it on the remaining subset left out of the training sample. This procedure is repeated k times in total, withholding a different test subset each time. We plot the values predicted from the training model on the test data and compare with observed data (Figure S19). The range of predicted values falls within the 95% confidence interval for the best-fit model using all the data. This gives us confidence that we are not overfitting to the current dataset, and that we will observe the effect in a new set of data.

Appendix G: Simple, numeric, and fractional models

We simulated complex and simple propagation on our estimated fish interaction networks, using both numeric thresholds and fractional thresholds. Activation using numeric thresholds requires an absolute number of neighbor activations, while activation using fractional thresholds requires that a fraction of neighbors become activated. Because our data only contain cascades initiated by a single fish, we only consider cascade simulations seeded by one node. We activate a seed node i at time t_0. In the following time-step each neighbor, j, of i has a chance of observing i as active, with probability w_{ij}, which is the strength of the link between the two nodes. Node j itself becomes active when the number or fraction of observed active nodes exceeds a threshold, for numeric and fractional contagions, respectively. To compensate for the increased barrier to activation with an increasing threshold, we include an amplification parameter C, which represents the number of chances each node has to observe a neighbor's activation in each time-step t (alternatively the number of coin-flips weighted by w_{ij}, where at least one positive result is needed to connect nodes i and j). The simulation continues until no more activations occur. Seeding the cascade with only a single node presents a problem for homogeneous-threshold based contagions, however, because any threshold greater than 1 requires that more than one fish begin in the active state. Consequently, we assign thresholds randomly on nodes in the graph, selected uniformly (although our results do not depend on this form of distribution, as we will show in the following section) from integers $[1, \phi_{\mathrm{nmax}}]$. For example, for an average threshold of 3, there will be equal numbers of nodes with thresholds 1, 2, 3, 4, or 5. Some nodes with threshold 1 are necessary to allow the cascade to propagate beyond the seed node. Similarly for fractional thresholds, we distribute thresholds uniformly from $(0, \phi_{\mathrm{fmax}}]$ on the graph.

The simulation proceeds as follows, continuing over t time-steps, until no more activations occur. Only one activation (A_i changes state from 0 to 1) is allowed per node, per simulation, and any node that becomes active during the course of the simulation remains active for the remainder.

$$A_i(t) = 1 \text{ if } \begin{cases} A_i(t-1) = 1 \\ \sum_{j=1}^{N} \delta_{i,j} > \phi_i \end{cases} \tag{S19}$$

$$\delta_{i,j} = (w_{ij} A_j(t-1) > \min(\mathbf{U_j})) \tag{S20}$$

Where U_j is a vector of C_f or C_n uniform random numbers, N is the total number of fish in the school, and ϕ_i is a threshold defined by the type of contagion used in the model.

$$\phi_{i,\mathrm{simple}} = 1, \quad \phi_{i,\mathrm{numeric}} = n_i, \quad \phi_{i,\mathrm{fraction}} = n_i \sum_j M_{ij} \tag{S21}$$

Where n_i is the particular activation threshold of node i, M_{ij} is the binary adjacency matrix, and $\sum_j M_{ij}$ is the number of neighbors of node i.

We estimate the prior distributions of parameters as follows. Clearly, $\phi_{\text{simple}} = 1$, because simple contagions only require one connection to propagate. The average node in our graph has 30 visible neighbors, so we allow ϕ_{nmax} to be drawn from the uniform distribution [1,30]. ϕ_{fmax} is drawn from the uniform distribution (0,1], so that on average the number of neighbors required for activation at the highest value of ϕ_{fmax} equals the number required for the highest value of ϕ_{nmax}. The coin-flip parameter, C, is intended to compensate for impedance to propagation due to higher thresholds. Preliminary analysis reveals that across all values of ϕ_{nmax} and ϕ_{fmax}, the optimal value for C_n is twice the value of ϕ_{nmax}, and C_f is twice the value of $\phi_{\text{fmax}} \times 30$. So we set C_n to be $2 \times \phi_{\text{nmax}}$, and C_f to $2 \times \phi_{\text{nmax}} \times 30$ for the remainder of our analysis (Figure S20).

To evaluate the performance of the three contagion models–simple, numeric, or fractional–on our derived interaction networks, we seed the propagation simulation with the initiator nodes (individuals) observed from the cascade data, and evaluate the results using two scoring metrics. First, we find the model that best predicts the identities of the responder nodes. We calculate the maximum likelihood over a random selection of parameter values, drawn from the prior distributions, averaged over 10 parameter sets of 500 parameters each. For each set of parameters for each node we averaged 50 iterations of the simulation to remove any effect from the random distribution of thresholds on the graph. We compare our three candidate model performances on our fish network to the performances on randomized versions of the network. The randomization is achieved by swapping the identities of responder nodes with the same number of randomly selected nodes in the group. The randomized models provide a baseline for comparing the simple, numeric, and fractional contagion models, and all three candidate models at a minimum clearly better predict responder identity than the baseline (Figure S21A).

Next, we consider an alternative scoring metric, where a model is evaluated by how accurately it predicts the resulting cascade length. The best fit model for predicting the identities of the responders is also the best fit model for predicting the size of the cascade (Figure S22). The best fit threshold for the fractional complex model is approximately $\phi_f = 0.25$, with agreement between both scoring metrics. Best fit values for both scoring metrics are included in Table S3. While the numeric and fractional models perform similarly using the first scoring metric, the numeric model is much worse at predicting cascade size, with a negative correlation persisting for most of the threshold range. In the fractional model, however, we see a strong positive correlation between predicted cascade size and experimentally observed cascade size. This is likely due to the difference in the two contagion types near the group boundary. The numeric threshold will predict smaller cascades from the boundary than from the bulk, due to boundary nodes having fewer neighbors. In fractional contagion, cascades from the boundary will be larger relative to the bulk, for the same reason (boundary nodes have fewer neighbors).

We purposefully chose to use basic models to allow us general insights into complex and simple propagation events evolving on our networks, and to determine if there exists support for one contagion mechanism over another. Further work is needed to fully understand the interplay between the structure of the network and the type of contagion mechanism by simulating cascades on specific types of networks.

We also calculate the marginal likelihood, which is the integration of the likelihoods over the prior distribution of ϕ, and find qualitatively the same results. The marginal likelihood is calculated because a reasonable model should not depend on fine-tuning of parameters (Figure S21B) [25].

In order to determine which of the considered models best supports the finding that the final cascade size increases with the local clustering coefficient of the initiator, we simulate the fractional, numeric and simple threshold models initiated from randomly selected nodes throughout the network, using the best-fit parameters for ϕ and C (Figure 3D, main text). When controlling for out-degree and local density (these are both correlated with clustering coefficient, and can affect the final cascade size), we find that simulated cascade size correlates strongly with local clustering coefficient only in the fractional threshold model. In the numeric model, cascade size is slightly correlated positively with clustering coefficient, but not as strongly as in the fractional model. In the simple model, cascade size is essentially independent of clustering coefficient, when controlling for out-degree and local density. These results are robust to the choice of response threshold distribution (Figure S23A,B), as we describe in the next section.

Alternative response threshold distributions We consider alternative threshold distributions in this section, since we do not know the form of variation arising from sensory noise, or from individual variability in sensitivity to visual cues, for example. We explore the effects on simulated cascade size when we relax the constraint of a uniform distribution of response thresholds in the group, to see if the effects we observe are consistent with a wider range of threshold distributions. We now set the response thresholds of each individual by drawing from a Gaussian distribution with mean $\phi_{n,f}$ and standard deviation $\sigma_{n,f}$, while restricting the distribution of thresholds to be greater than zero.

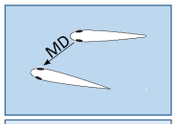
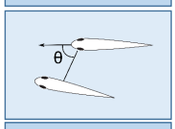
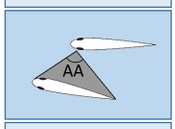
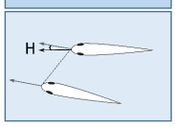
We repeat the analysis described above, this time using a Gaussian distribution for response thresholds, to find values for $\phi_{n,f}$ and $\sigma_{n,f}$ which best predict responder identities and experimental cascade size. In Figure S24, we plot model performance as heatmaps with the variation in response threshold on the x-axis, and the mean response threshold on the y-axis (this is a 2-dimensional version of Figure S22). We find that, for fractional contagion, models that best predict both cascade size and responder identities have low values of mean and standard deviation (Figure S24A,B). For numeric contagion, models that best predict responder identities also have low values of mean and standard deviation (Figure S24C), however numeric contagion is a poor predictor of cascade size, as for much of parameter space simulated cascade size is negatively correlated with experimental cascade size (Figure S24D), for spatial reasons we described in the preceding section (see Figure S22D). We also find that the maximum likelihood is greatest for fractional contagion models irrespective of the type of distribution (Figure S25A and Figure S21 for normal and uniform distributions, respectively).

Next, we chose parameters from the optimal region of phase space, for the three types of contagion ($\phi_f = 0.05$, $\sigma_f = 0.06$; $\phi_n = 1.5$ $\sigma_n = 1$; $\phi_s = 1$, $\sigma_s = 0$), and simulated cascades initiated by every node, in 20 randomly selected networks. We confirm our results from Fig. 3D (main text); that cascade size varies most strongly with clustering coefficient when using fractional contagion, as compared with numeric or simple contagion (Figure S25 B). Note that the mean value of the cascade size for numeric contagion here is higher than in Figure 3D, main text. This is because the parameters that best

fit the data for the Gaussian distribution result in a slightly higher average cascade size, for numeric contagion.

Finally, we performed a sensitivity analysis to the width of the Gaussian distribution, by selecting a mean threshold value and varying the standard deviation of the distribution from which individual thresholds are chosen. We selected mean values for the distributions to be small enough to allow responses to propagate, even for small values of σ, as well as resulting in consistent values between fractional and numeric contagion. Since the average fish in the group has 30 visible neighbors, having a mean numeric threshold (ϕ_n) of 3 corresponds to a mean fractional threshold (ϕ_f) of 0.1. From 10 randomly selected networks in our data we simulated a cascade initiated from every node in the network, and calculated the resulting cascade size as the average from five trials, where each node's threshold is drawn independently in each trial. For both fractional and numeric contagion as σ increases the average cascade size increases, because larger values of σ allow for more individuals with lower thresholds, facilitating propagation. We find that simulated cascade size scales much more strongly with fractional contagion than with numeric contagion, invariant to the width of the distribution from which response thresholds are drawn (Figure S23A,B).

1. Katz Y, Tunstrm K, Ioannou CC, Huepe C, Couzin ID (2011) Inferring the structure and dynamics of interactions in schooling fish. Proc Natl Acad Sci USA 108(46):18720–18725.
2. Tunstrm K, Katz Y, Ioannou CC, Huepe C, Lutz MJ, Couzin ID (2013) Collective States, Multistability and Transitional Behavior in Schooling Fish. PLoS Comput Biol 9(2)e1002915.
3. Domenici P, Blake RW (1997) The kinematics and performance of fish fast-start swimming. J Exp Biol. 200(8):1165–1178.
4. Eaton RC (1984) Neural mechanisms of startle behavior (Springer).
5. Lima SL (1995) Collective detection of predatory attack by social foragers: fraught with ambiguity? Anim Behav 50(4):1097–1108.
6. Domenici P, Batty RS (1997) Escape behaviour of solitary herring (Clupea harengus) and comparisons with schooling individuals. Mar Bio 128(1):29–38.
7. Blackman S, Popoli R (1999) Design and Analysis of Modern Tracking Systems (Artech House).
8. Wu H, Zhao Q, Zou D, Chen Y (2011) Automated 3D trajectory measuring of large numbers of moving particles. Opt Express 19(9):7646–7663.
9. Strandburg-Peshkin A, et al. (2013) Visual sensory networks and effective information transfer in animal groups. Curr Biol 23(17):709–711.
10. Burgess JW, Shaw E (1981) Effects of acoustico-lateralis denervation in a facultative schooling fish: a nearest-neighbor matrix analysis. Behav Neural Biol. 33(4):488–497.
11. Fontaine E, et. al. (2008) Automated visual tracking for studying the ontogeny of zebrafish swimming. J Exp Biol 211(8):1305–1316.
12. Watson DF (1981) Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes. Br Computer Soc 24(2):167–172.
13. Dill LM, Holling C, Palmer LH (1997) In Animal groups in three dimensions: how species aggregate 207–224. New York: Cambridge University Press.
14. Partridge BL, Pitcher T, Cullen JM, Wilson J (1980) The three-dimensional structure of fish schools. Behav Ecol Sociobiol 6(4):277–288.
15. Chatterjee S, Hadi AS (2000) Regression analysis by example. John Wiley & Sons.
16. Dehaene S (2003) The neural basis of the Weber-Fechner law: a logarithmic mental number line. Trends Cogn Sci 7(4):145–147.
17. Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19(2):101–108.
18. Lundgren J (2010) Alpha shape of 2D/3D point set. MATLAB Central File Exchange, retrieved August 2013. http://www.mathworks.com/matlabcentral/fileexchange/28851-alpha-shapes.
19. Fagiolo G (2007) Clustering in complex directed networks. Phys Rev E 76(2):026107.
20. Meyers LS, Gamst G, Gaurino AJ (2006) Applied multivariate research: Design and interpretation. (Sage).
21. Barthelemy M (2004) Betweenness centrality in large complex networks. Eur Phys J B 38(2):163–168.
22. Kitsak M, et al. (2010) Identification of influential spreaders in complex networks. Nature Phys 6(11):888–893.
23. Pei J, Andrade J, Muchnik L, Zheng Z, Makse HA (2014) Searching for superspreaders of information in real-world social media. Sci Rep 4:http://dx.doi.org/10.1038/srep05547.
24. Dorogovtsev SN, Goltsev AV, Mendes JFF (2006) K-core organization of complex networks. Phys Rev Lett 96(4):040601.
25. Mann RP, et. al. (2012) Multi-scale inference of interaction rules in animal groups using Bayesian model selection. PLoS Comput Biol 8(1):e1002308.

	Absolute features	
	Metric distance (MD, $\log(MD)$)	Euclidean distance from focal fish to neighboring fish
	Speed (S)	$d(MD)/dt$
	Angular position (θ)	Relative angular position of neighboring fish from direction of heading of focal fish
	Angular speed ($d\theta$)	$d\theta/dt$
	Angular area ($AA, \log(AA)$)	Angular area subtended on the retina of the focal fish by neighboring fish
	Loom ($L, \log(L)$)	dAA/dt
	Heading (H)	Difference between heading of focal fish and heading of neighboring fish
	Heading change (dH)	dH/dt

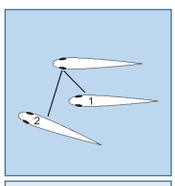
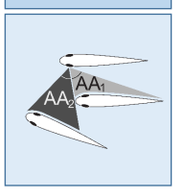
	Comparative features	
	Topological distance ($TD, \log(TD)$)	Ranked Euclidean distance from focal fish to neighbor fish
	Relative metric distance ($relMD, \log(relMD)$)	Euclidean distance to from focal fish to nearest neighbor divided by Euclidean distance to neighbor fish.
	Ranked angular area ($AR, \log(AR)$)	Rank of neighbor sorted by size of subtended angular area on retina of focal fish
	Relative angular area ($relA, \log(relA)$)	Angular area of neighbor fish on retina of focal fish divided by maximum angular area observed by focal fish

Table S1. Description of all features used in model selection.

Network feature	Coefficient	χ^2	p value
Local weighted clustering coefficient	0.37	78.1	<0.0001
K-core	-0.11	9.05	0.0026
Weighted K-core	0.05	1.81	0.179
Weighted betweenness	-0.35	52.4	<0.0001
Edge weight variation	0.14	15.4	0.0008

Table S2. Comparison of model performance using alternative network measures.

Complex fractional model

Parameter	Mean	SE
ϕ_{fmax}; identities	0.2314	0.0172
ϕ_{fmax}; cascade size	0.2438	0.0270

Complex numeric model

Parameter	Mean	SE
ϕ_{nmax}; identities	8.900	0.3797
ϕ_{nmax}; cascade size	24.57	0.5102

Table S3. Optimal parameter values for both evaluation metrics and both contagion types. Optimal parameter values do not coincide exactly with marginal likelihood parameter values. The optimal values were obtained by finding the location of the maximum log-likelihood in each of the 10 simulation runs, while the marginal likelihood values in Figure S22 are obtained by averaging every log-likelihood value from each value of $\phi_{\max}$ on the x-axis. The marginal likelihood parameters are best on average, while the optimal parameter values are best overall. The difference between them is accounted for by there being more spread in the data near the optimal parameter values.

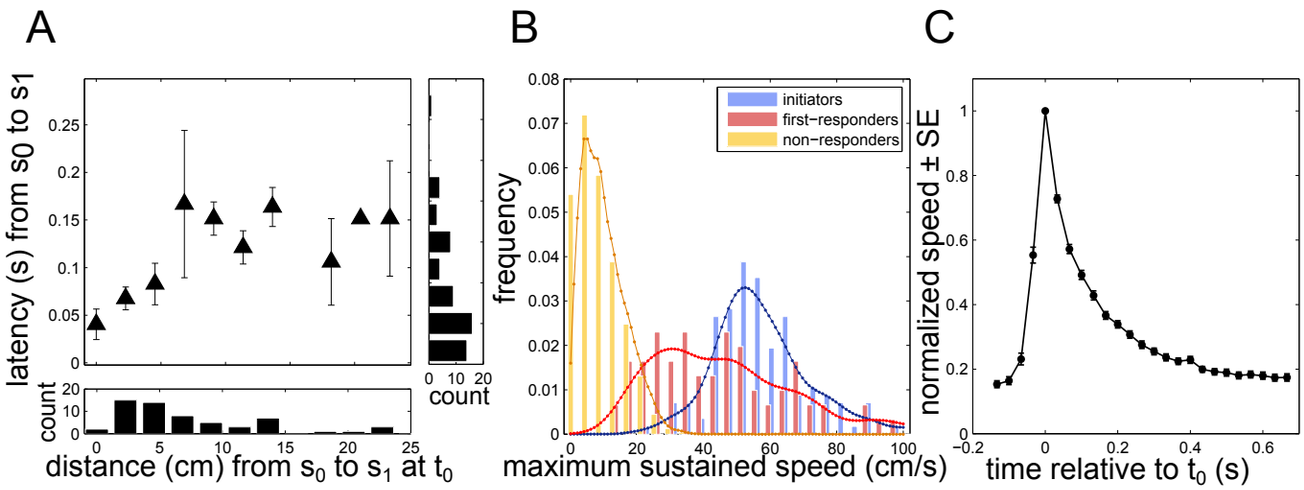


Fig. S1. (**A**) The latency between the initial startle and first response grows with distance between the initiator and first responder. (**B**) The distribution of speeds of initiators (blue), first responders (red), and non-responders (yellow). Data are shown in bars, while the kernel-smoothed distribution is overlaid in dotted lines. Initiator and first responder speeds are significantly higher than non-responder speeds. (**C**) Average shape of speed-profile for startle response (centered on zero and standardized so that max speed is one, and occurs at t = 0. Curve is created by averaging over all initiators, with error bars showing the standard error).

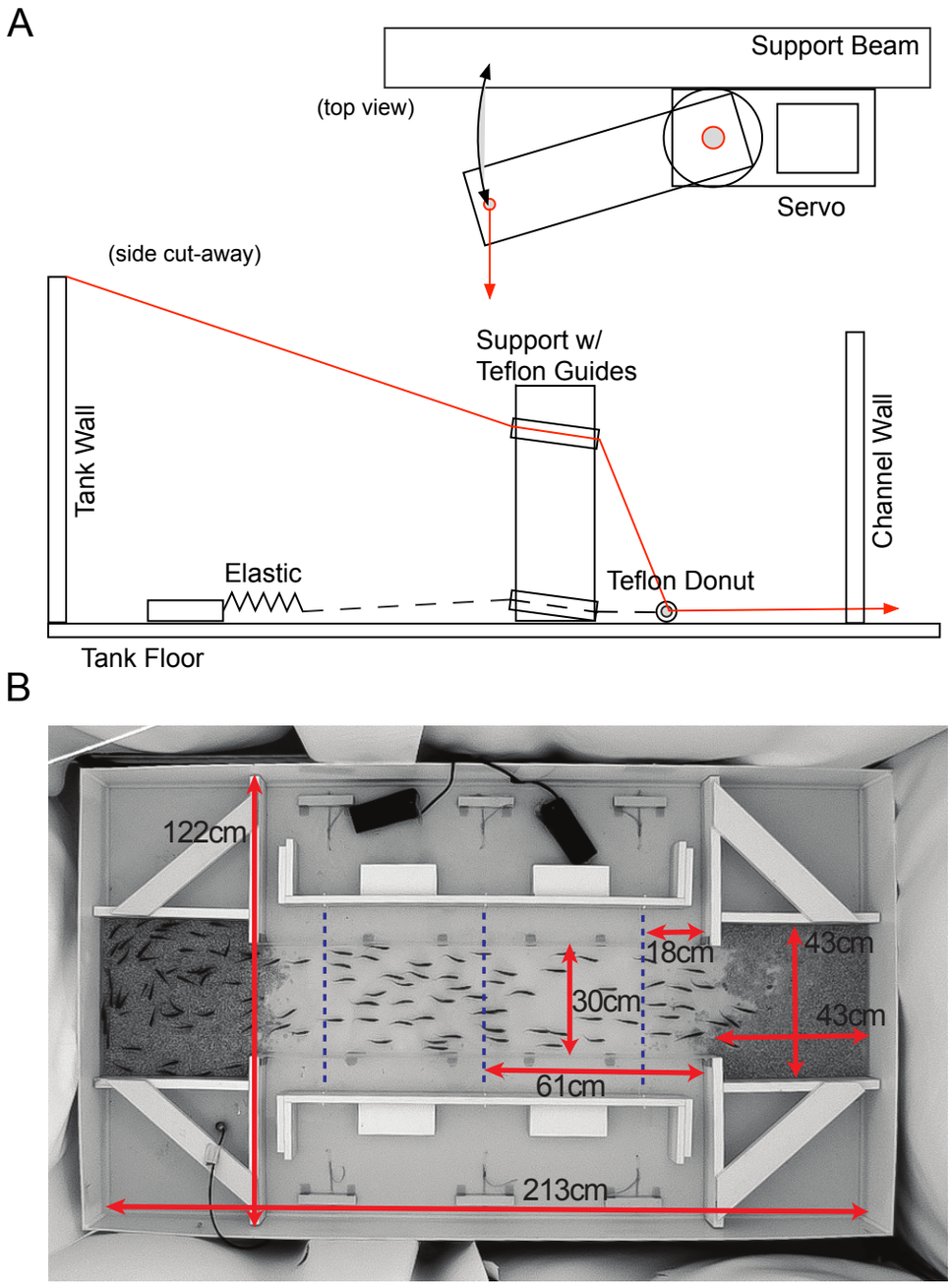


Fig. S2. Experimental setup for triggered startles (**A**) The monofilament is raised by a servo, pictured from above in top view, which is fixed outside the light tent. Here we used a HiTec HS-5965MG high-speed programmable digital servo, with a 10cm arm. The monofilament then runs through the white support structure, through the a white teflon return donut, through the channel wall, and then through the same structures in reverse order on the other side—shown in side cut-away. The white teflon donut is attached to another filament and an elastic. When the servo is triggered the wire raises, stretching the elastics, which in turn pull the monofilament back to the channel floor when the servo relaxes. (**B**) All channel experiments were conducted in the above experimental tank. The fish were constrained a region of the 213cm × 122cm tank consisting of two 43cm × 43cm gravel-bottomed shelters and a 122cm × 30cm channel connecting the shelters. Through trial and error we found that 30cm was the widest we could make the channel while discouraging the school from entering a torus or milling state. The channel had clear internal walls with clear monofilaments running along the bottom of the tank at three different positions—designated here by dashed blue lines. These monofilaments could be quickly raised and lowered to strike the underside of passing fish. The light tent which can be seen hanging around the boarder of the tank, provides both even illumination and separation between the fish and the experimenter. Unseen, there are two layers of sound insulation below the tank to prevent disturbances, such as closing doors and flushing toilets from disturbing the fish.

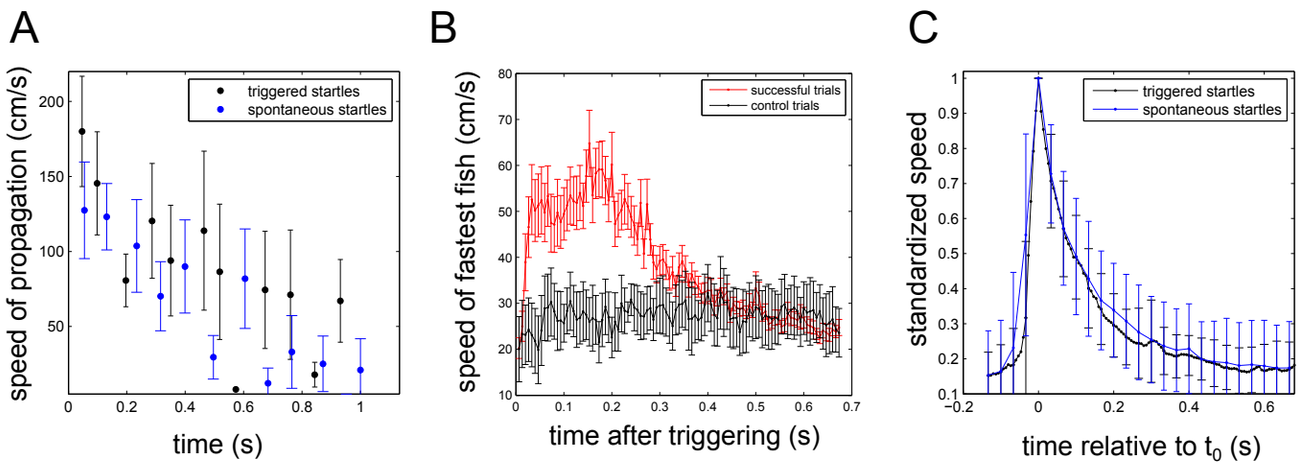


Fig. S3. Triggered behavioral cascades (**A**) Average speed of cascade wave propagation, calculated as the change in the spatial extent of the cascade per second, averaged over all triggered cascade events. Compare to blue curve, for speed of spontaneous behavioral cascade. (**B**) Comparison of response speeds of successful trials (red) to control trials (black). Speed profiles of fastest fish in school following alarm triggering are averaged over all events, and plotted for both successful and control trials. No significant increase in speed is observed for control trials, and speed in successful trials was significantly higher than in control trials (Wilcoxon rank-sum test, $p < 0.00001$). (**C**) Average speed profile for triggered startle response (centered on zero so that the max speed is one, and occurs at t = 0. Curve is created by averaging over all initiators. Compare to blue curve, for spontaneous startles. Data points in (A), (B), and (C) show the mean, with error bars for ± the standard error.

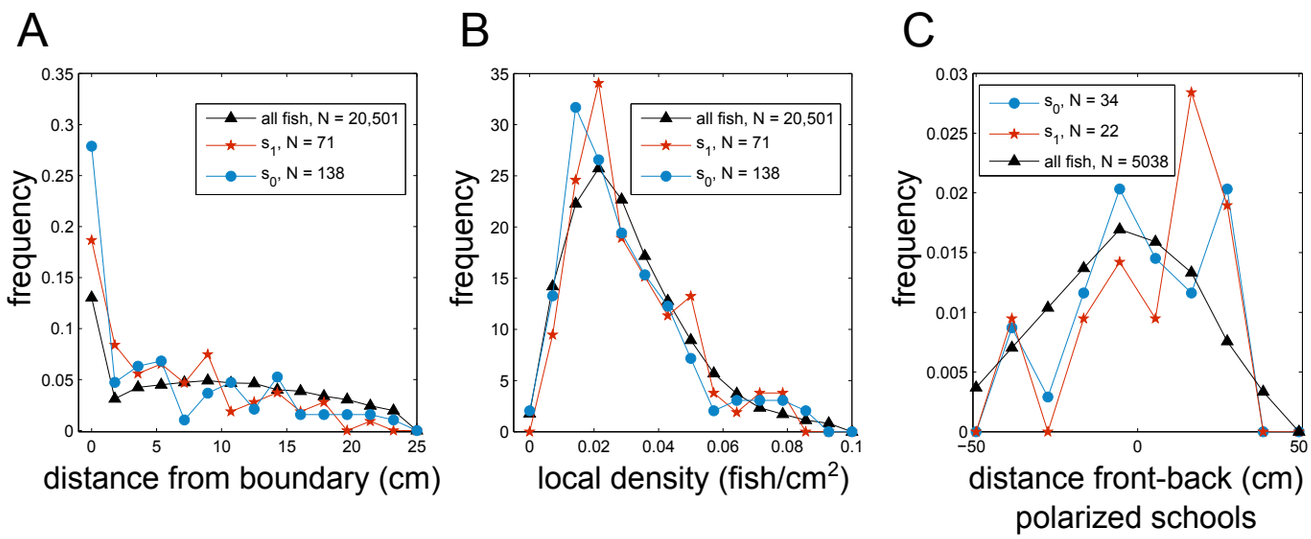


Fig. S4. Distributions of initiator, first responder, and non-responder positions in relation to (**A**) distance from group boundary, (**B**) local density, and (**C**) distance front-back (in polarized schools).

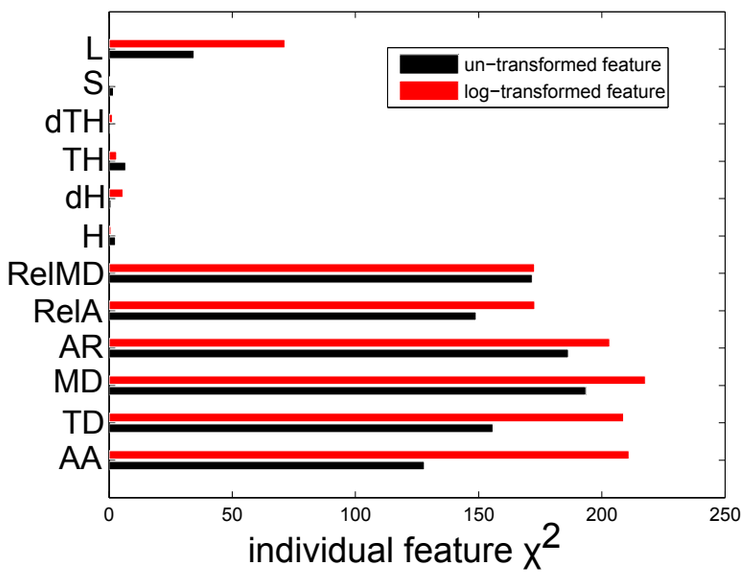


Fig. S5. Comparison of log-transformed features to un-transformed features. Some features have improved individual performance when using the log-transformed version. We use both log-transformed and un-transformed features in our model selection procedures. We include the log-transformed version of a feature if it has significant predictive power on its own, and if using the log-transform results in an improvement in individual feature performance.

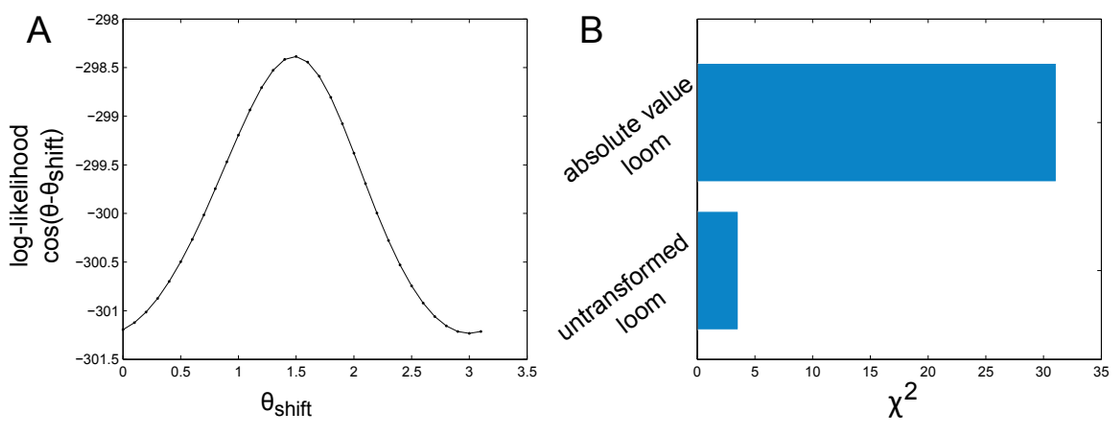


Fig. S6. Transforms of feature variables, fitted to first responder data. (**A**) Cosine-transformed and shifted relative angular position. The best shifted value for angular position is approximately $\pi/2$. (**B**) Comparison of un-transformed loom to absolute value of loom. The absolute value of loom out-performs the un-transformed loom.

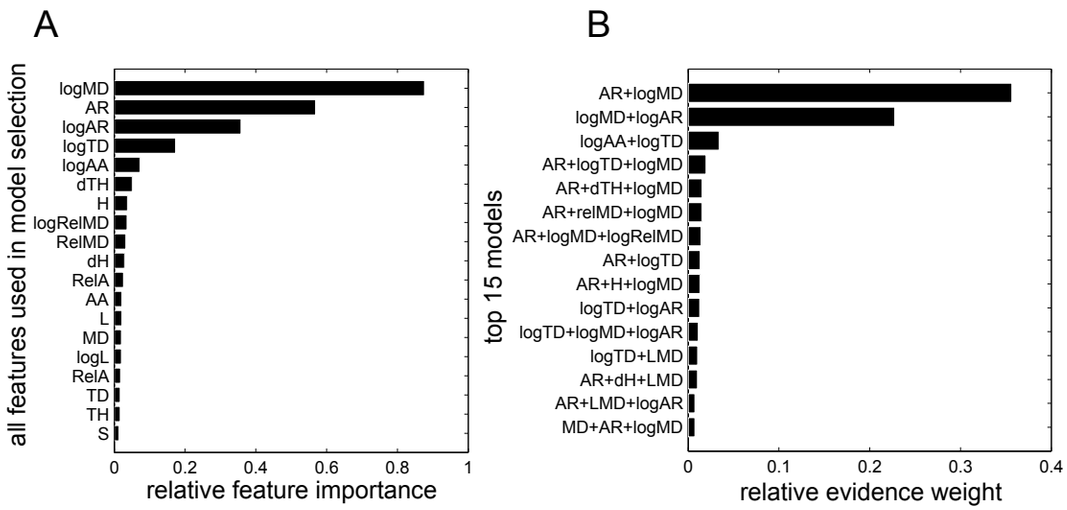


Fig. S7. Results from multi-model inference. (**A**) Relative importance of every feature used in model selection. (**B**) Relative evidence weight for top the 15 models.

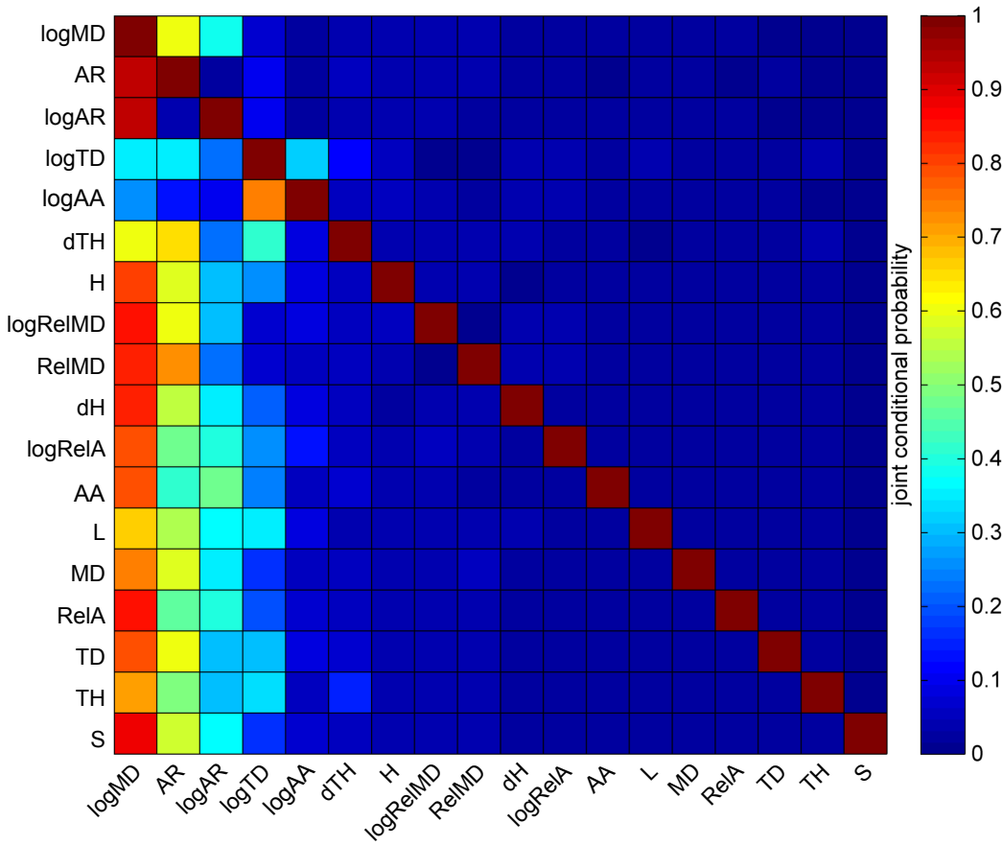


Fig. S8. Feature pair conditional probabilities. This plot represents the likelihood, weighted by the relative evidence weight of the model, that the column variable appears in a model given that the row variable is in the model. Some variables exclude each other, meaning that they are not frequently found in the same highly weighted models, even though they may show up separately in highly weighted models. This would be expected for variables that contain redundant information. Topological distance and metric distance exclude each other here, meaning that they contain similar information, and only one should be included in the final model.

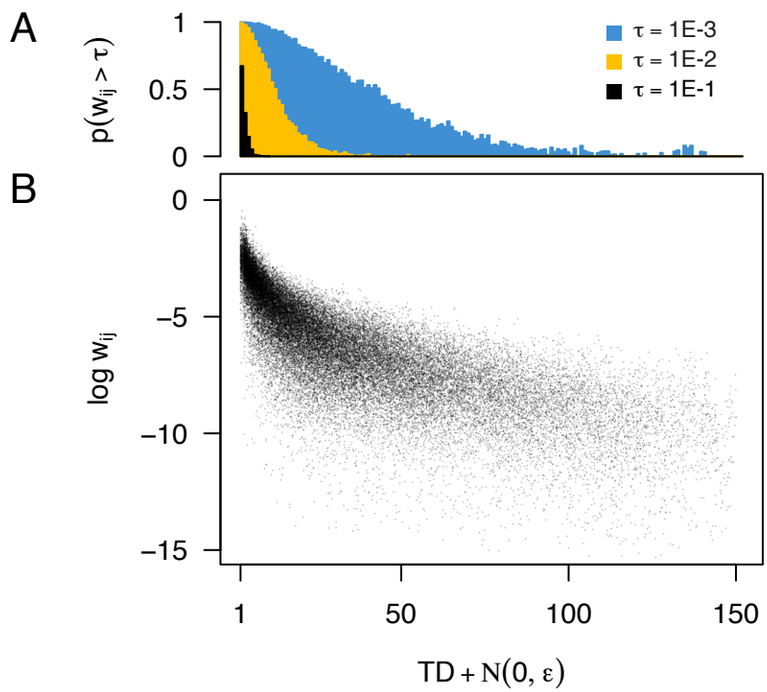


Fig. S9. Sensory network edge weight (w_{ij}) as a function of topological distance (TD) between i and j, showing a lattice-like emphasis on spatially near-neighbors, but with many weak, long-range (topologically far) connections. (**A**) shows a histogram of the (normalized) frequency that an edge connecting i to j is greater than a threshold τ (see legend), as a function of the topological distance between i and j. High probability edges (black) are highly concentrated at low topological distances, while a large number of low probability edges (blue) can still be found at much larger distances. In (**B**), the (log-transformed) edge weights are plotted as a function of topological distance (plus a small amount of Gaussian noise to prevent points from otherwise obscuring each other at each discrete interval) to show the full distribution of points summarized in (**A**). Data in both panels are from 10 networks subsampled uniformly at random from the set of all networks (one per trial).

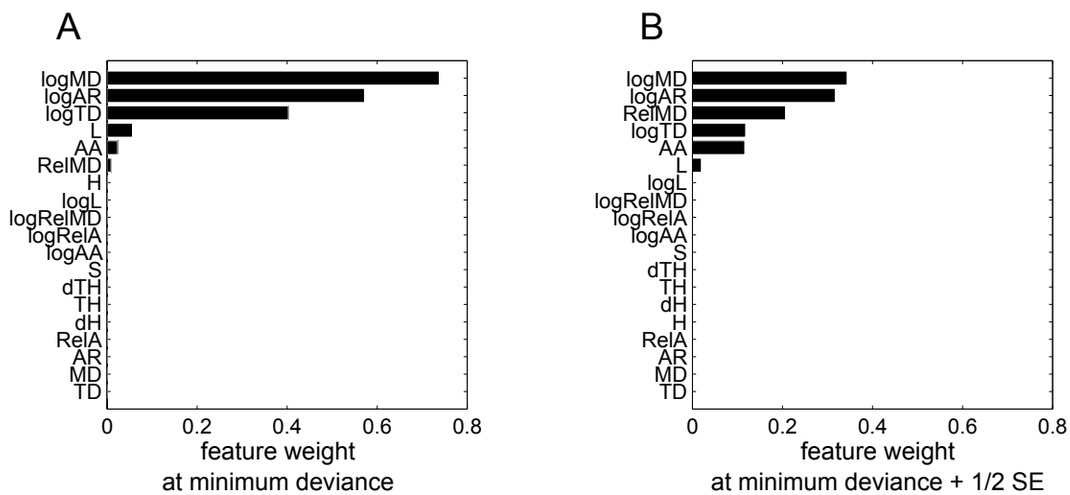


Fig. S10. Results from L_1 regularization. (**A**) Feature weights for the value of λ with minimum binomial deviance ($\lambda_{\min}$), and (**B**) feature weights for $\lambda > \lambda_{\min}$ at 1/2 standard error greater than the binomial deviance for $\lambda_{\min}$. The top two selected features are in agreement with the results of the multi-model inference method.

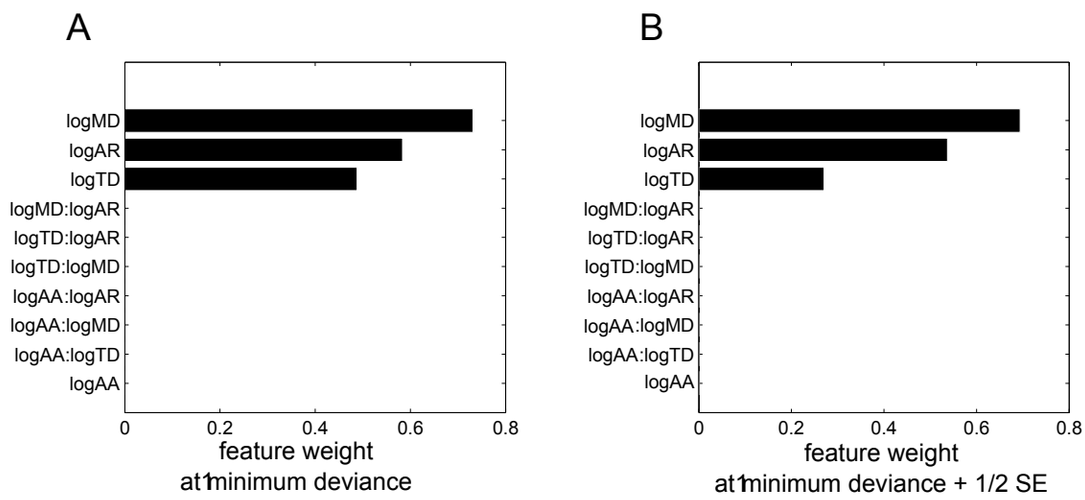


Fig. S11. L_1 regularization for interaction terms from top four features selected from multi-model inference. None of the interaction terms have nonzero feature weights. (**A**) Feature weights for the value of λ with minimum binomial deviance ($\lambda_{\min}$), and (**B**) feature weights for $\lambda > \lambda_{\min}$ at 1/2 standard error greater than the binomial deviance for $\lambda_{\min}$.

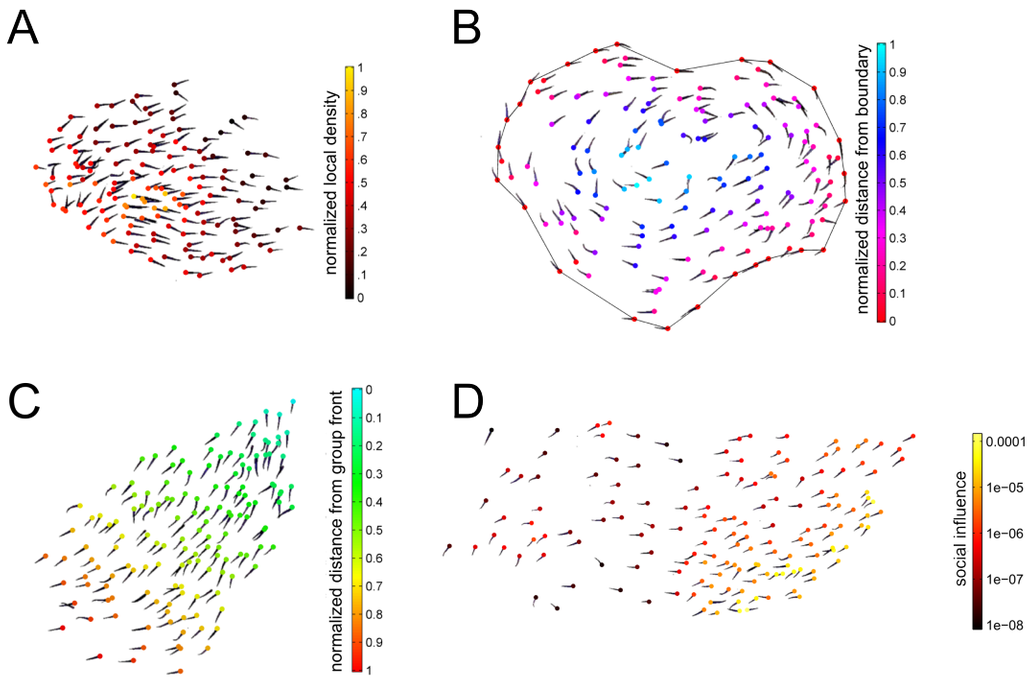


Fig. S12. Illustration of spatial measures for four sample groups. (**A**) local density, (**B**) distance from group boundary, (**C**) distance from the group front, and (**D**) social influence (as estimated by the local clustering coefficient).

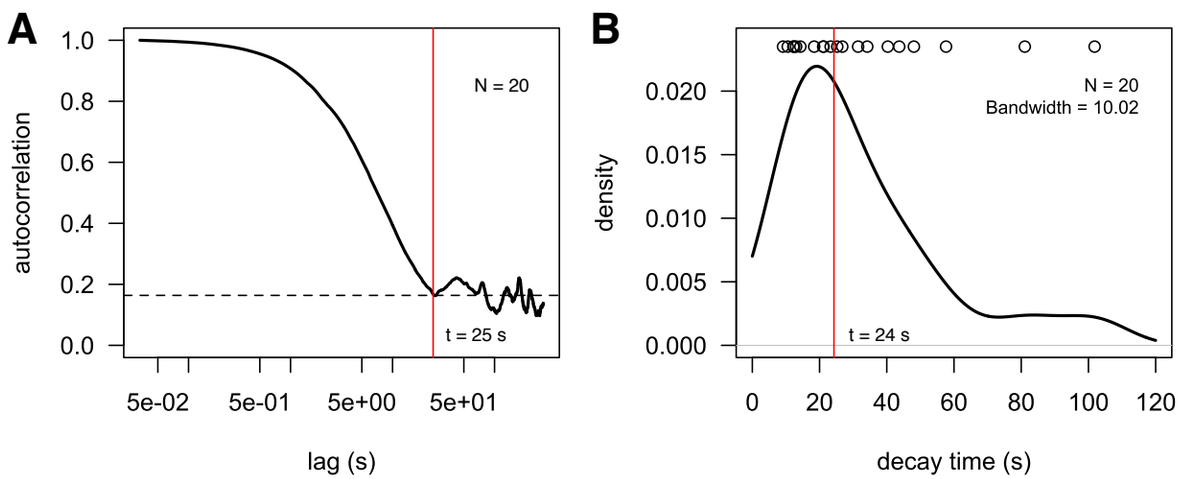


Fig. S13. Lin-log plot of distance from boundary (**A**) autocorrelation function (ACF) for 20 individuals in a group of 150, with random reshuffled ACF baseline (dashed line). Red vertical line marks the decay time (minimum lag in autocorrelation function needed to reach a correlation indistinguishable from noise). In (**B**), a density plot shows the decay time for 20 individuals, with the median decay time shown in red.

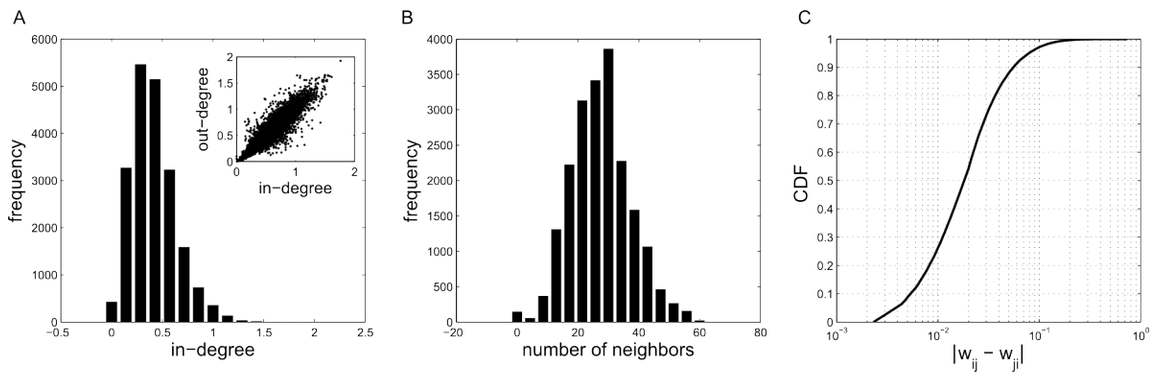


Fig. S14. Descriptive network statistics. (**A**) Frequency of in-degree; inset: in-degree and out-degree are correlated, but not perfectly so. (**B**) Frequency of number of neighbors (based on a binary threshold of AA > 0.02 radians). (**C**) Cumulative distribution function of the difference between incoming and outgoing edge weights. Although many edge weights have similar in and out values, there are enough differences that symmetrizing the network and treating edges as undirected is not justified in general.

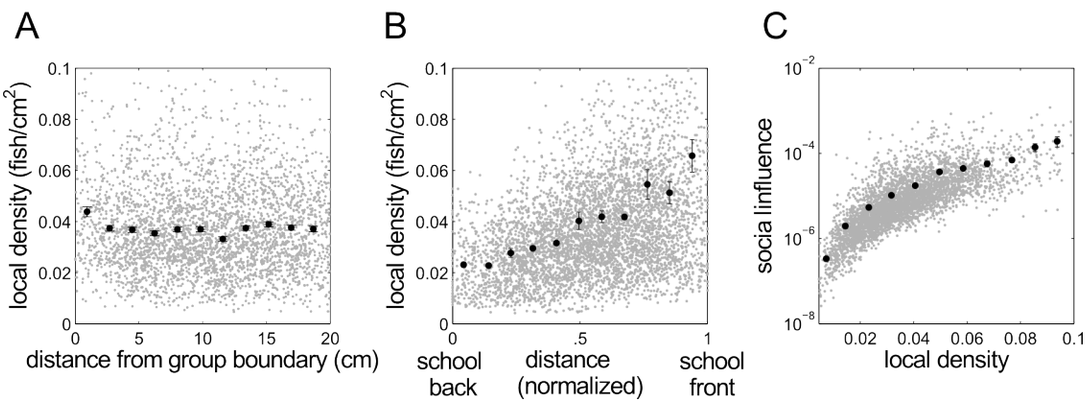


Fig. S15. Spatial structure of local density and social influence. (**A**)Local density is not strongly correlated with distance from boundary, but (**B**) density does tend to be higher near the group front. (**C**) social influence (approximated by local clustering coefficient) varies positively with local density.

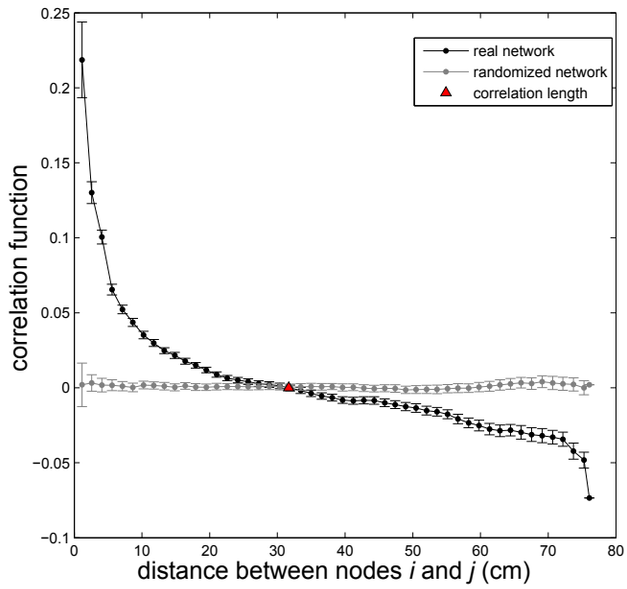


Fig. S16. Correlation function between clustering coefficients of pairs of nodes, as a function of the distance between them. Nodes which are close together tend to have more similar values of local clustering coefficients than nodes which are far apart.

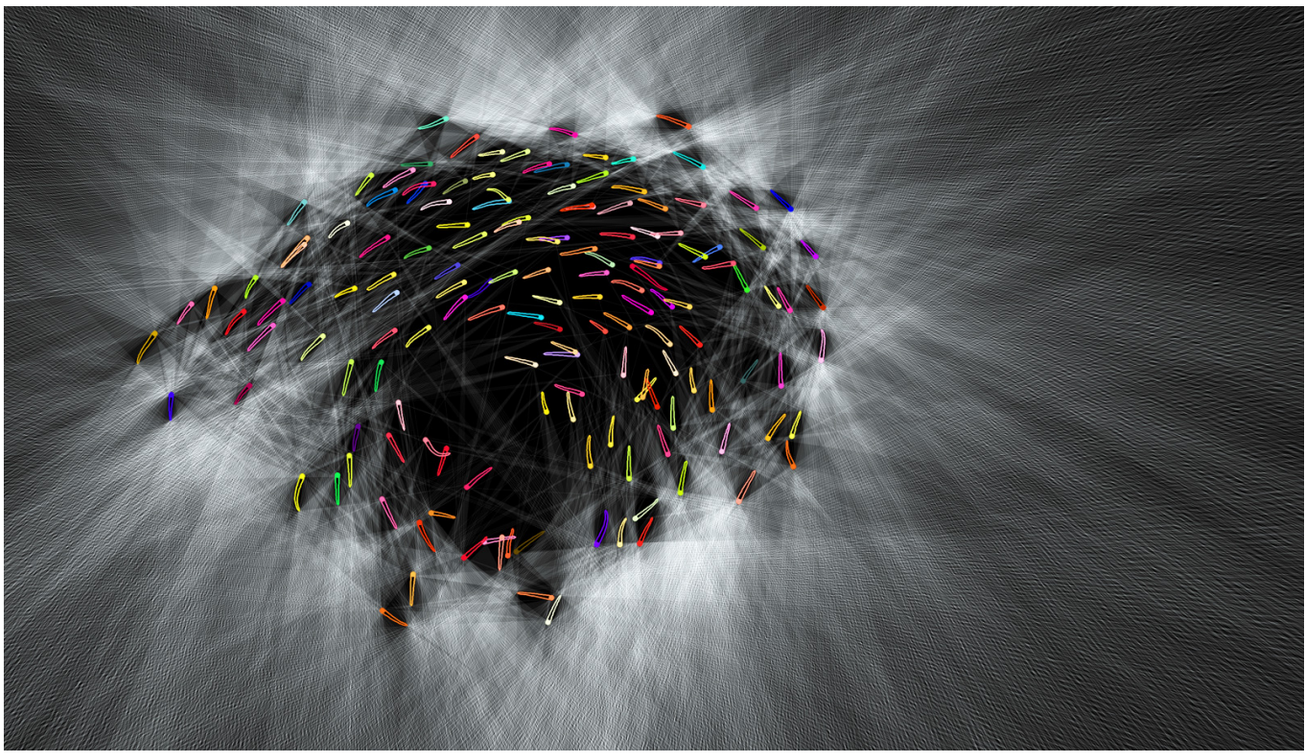


Fig. S17. Visualization of external field of view; rays (white lines) unobstructed by other fish are drawn from eye positions to corresponding positions on the tank wall. Fish on the group boundary can see much more outside the group than those in the group interior.

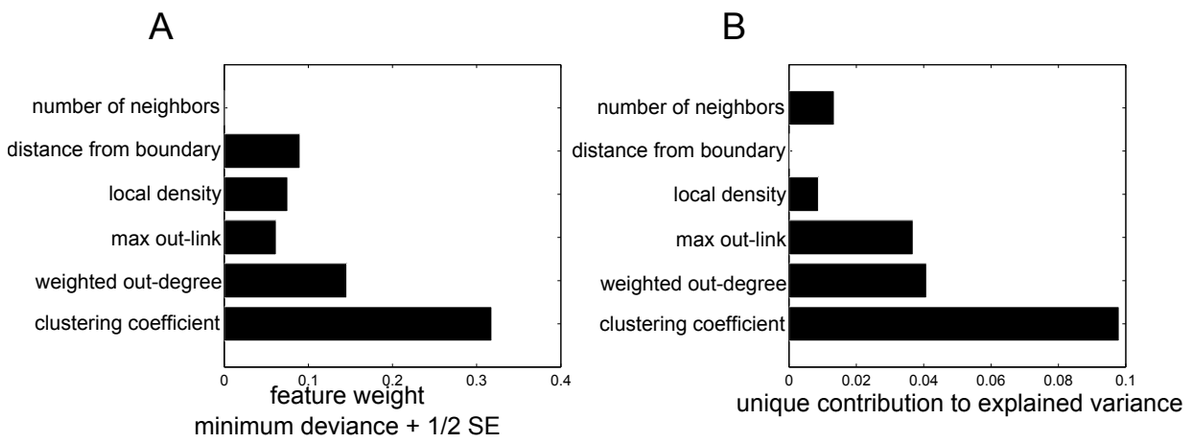


Fig. S18. Relative contribution of each variable to the model's fit to the cascade size distribution. (**A**) relative feature weight from L_1 regularization. (**B**) Each variable's unique contribution to total variance explained by the model. The local clustering coefficient has the highest contribution to the model in both cases.

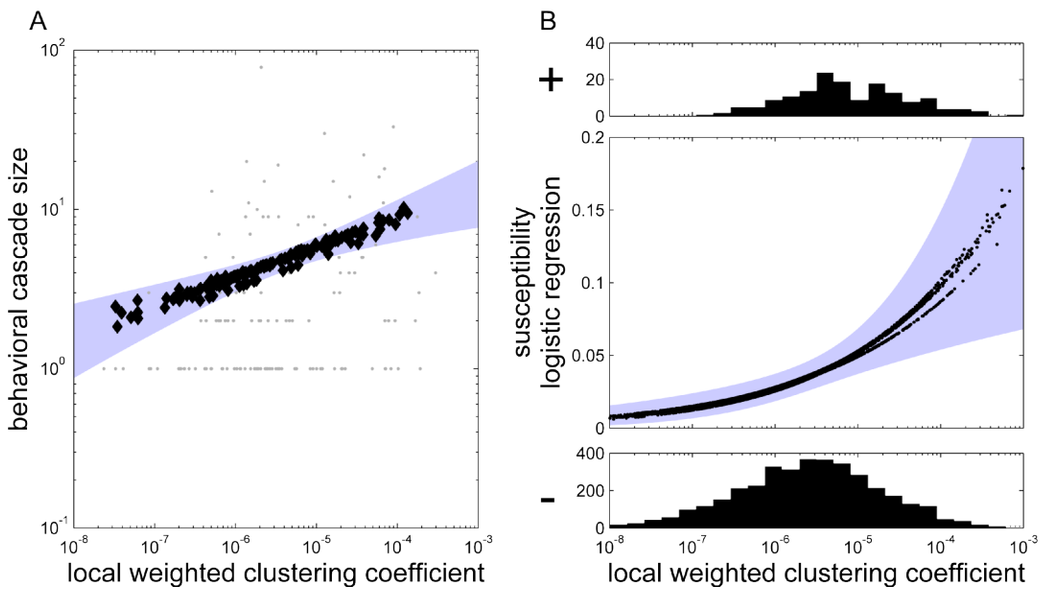


Fig. S19. Cross validating the model. The data was repeatedly divided into 10 subsets, trained on 9 of them, and tested on the remaining one (plotted in black diamonds). 95% confidence interval from model fits are shown in the blue shaded areas. (**A**) Cross-validation for the relationship between cascade size and local clustering coefficient. (**B**) Cross-validation for relationship between susceptibility to influence and local clustering coefficient. Histograms show the frequencies of positive responses (above) and no responses (below) as functions of clustering coefficient.

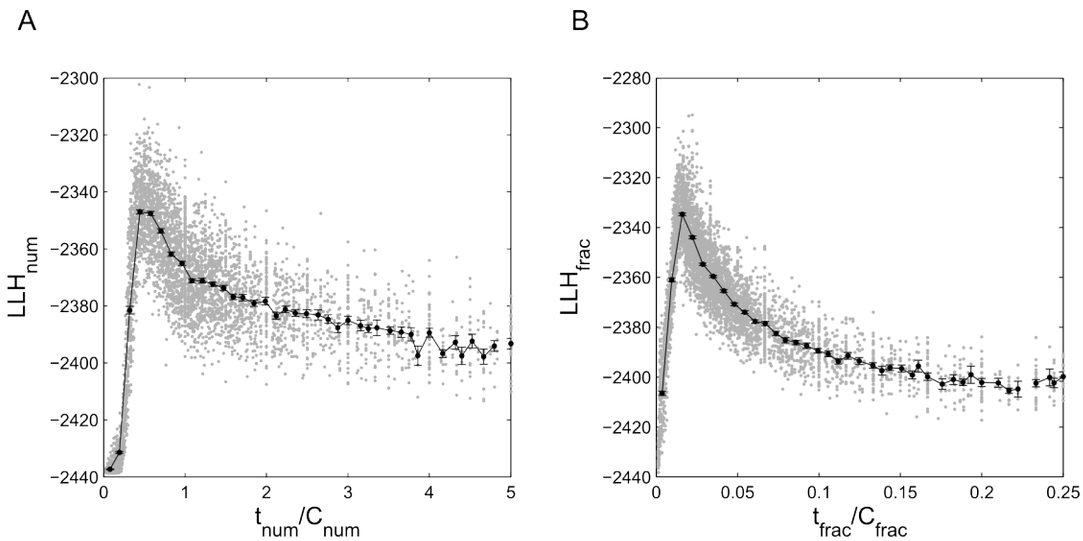


Fig. S20. Log likelihood for (**A**) numeric and (**B**) fractional models are maximized at a ratio t/C of 0.5, and 0.5/30, respectively.

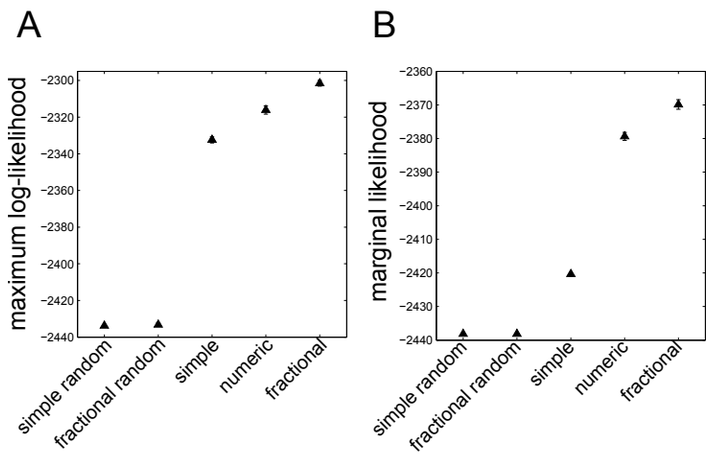


Fig. S21. (**A**) Results from the simple, complex numeric, and complex fractional contagion model fits. The models are evaluated by how well they predict the identities of the responded individuals, seeded by the initiators from the data. We show the mean values of the maximum likelihoods, averaged over 10 iterations of 500 sample parameter values. Simple random and fractional random are the results from randomized versions of the data, and represent the baseline maximum log-likelihood. Both complex contagion models (numeric and fractional) fit our data better than the simple contagion model. (**B**) Marginal likelihoods for candidate models (alternative to maximum likelihood (**A**)). Error bars indicate standard error (note that some are smaller than the marker size). Marginal likelihoods were calculated by averaging the log-likelihoods from 10 trials using 500 parameters in each trial, fit to the responder identity data.

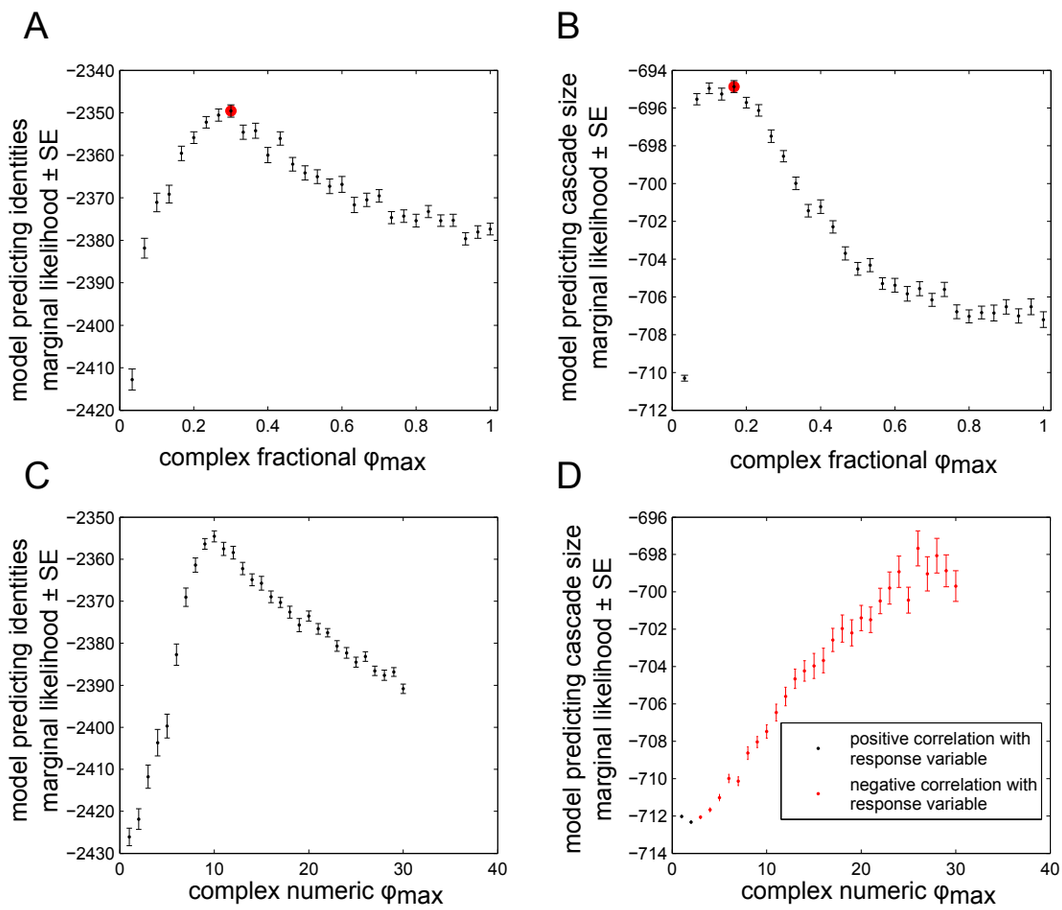


Fig. S22. Optimal threshold values (ϕ), organized by row for complex fractional (**A**, **B**) and complex numeric (**C**, **D**) contagion models, and by column for identity (**A**, **C**) and cascade size (**B**, **D**) model scoring metrics.

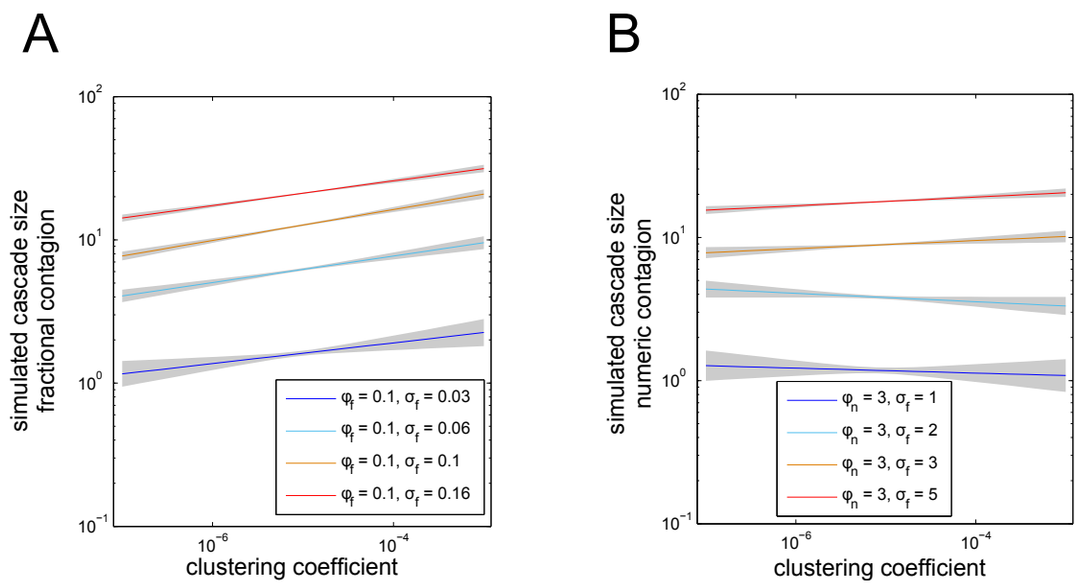


Fig. S23. Simulated cascade size plotted against clustering coefficient of initiator, for (**A**) fractional contagion, (**B**) numeric contagion. Four different levels of standard deviation in threshold response distributions are tested, while the mean threshold response is held constant. Solid lines represent best fit of cascade size to clustering coefficient, after controlling for node out-degree and local density, while shaded regions represent the 95% confidence interval around the line of best fit. Cascade size varies more strongly with clustering coefficient when using fractional contagion (A) than when using numeric contagion (B).

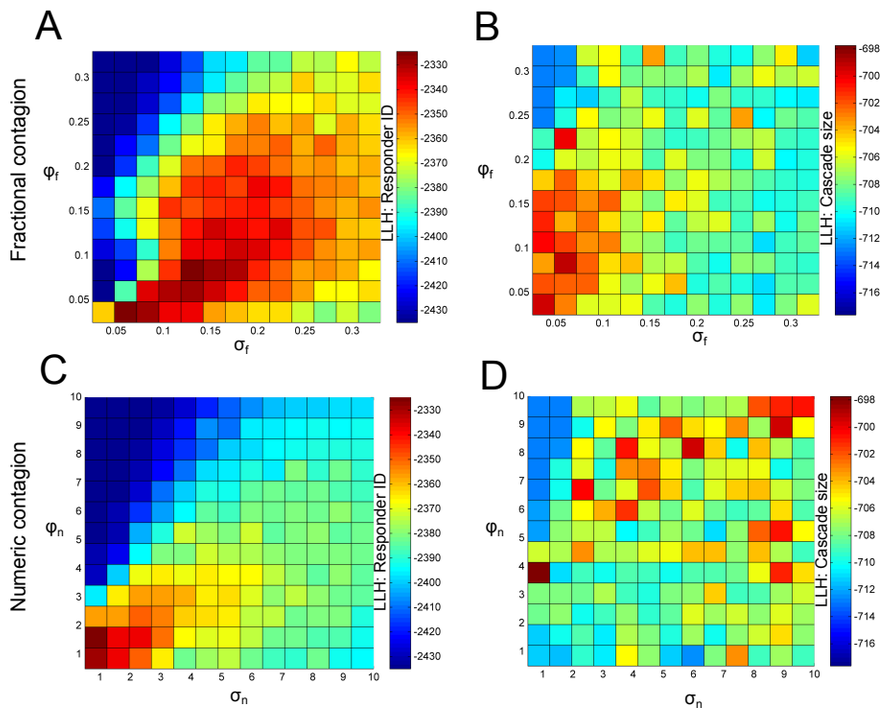


Fig. S24. Gaussian threshold distribution: best fit parameters. Heatmaps show the average log-likelihood value at each value of ϕ and σ (5000 total parameter values). This figure is a 2-dimensional version of Figure S22. (**A**) Fractional contagion, model predicting responder identities. (**B**) Fractional contagion, model predicting cascade size. (**C**) Numeric contagion, model predicting responder identities. (**D**) Numeric contagion, model predicting cascade size. Simulated cascade size is negatively correlated with experimental cascade size, for numeric contagion, which explains why there is no clear optimal parameter region here.

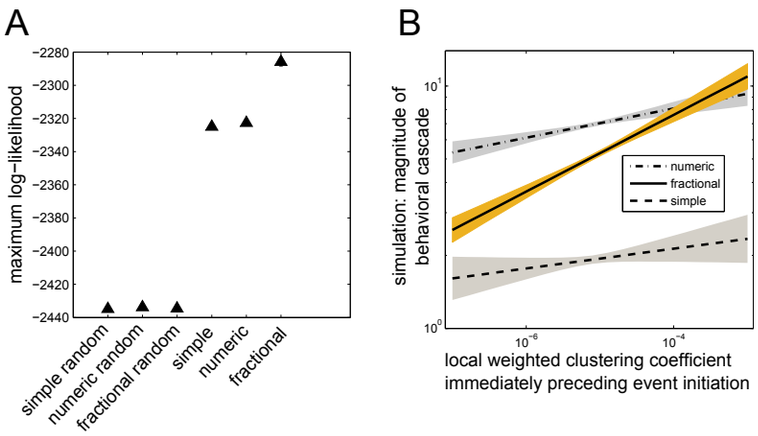
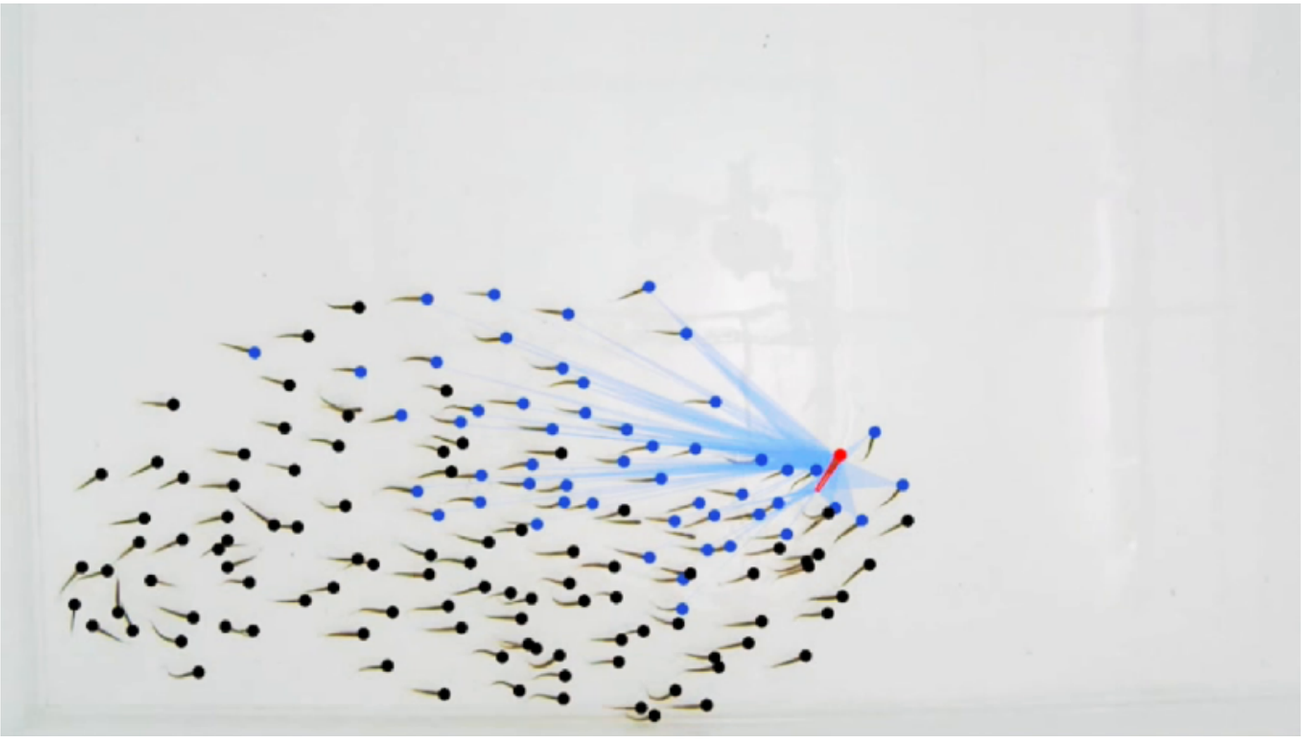
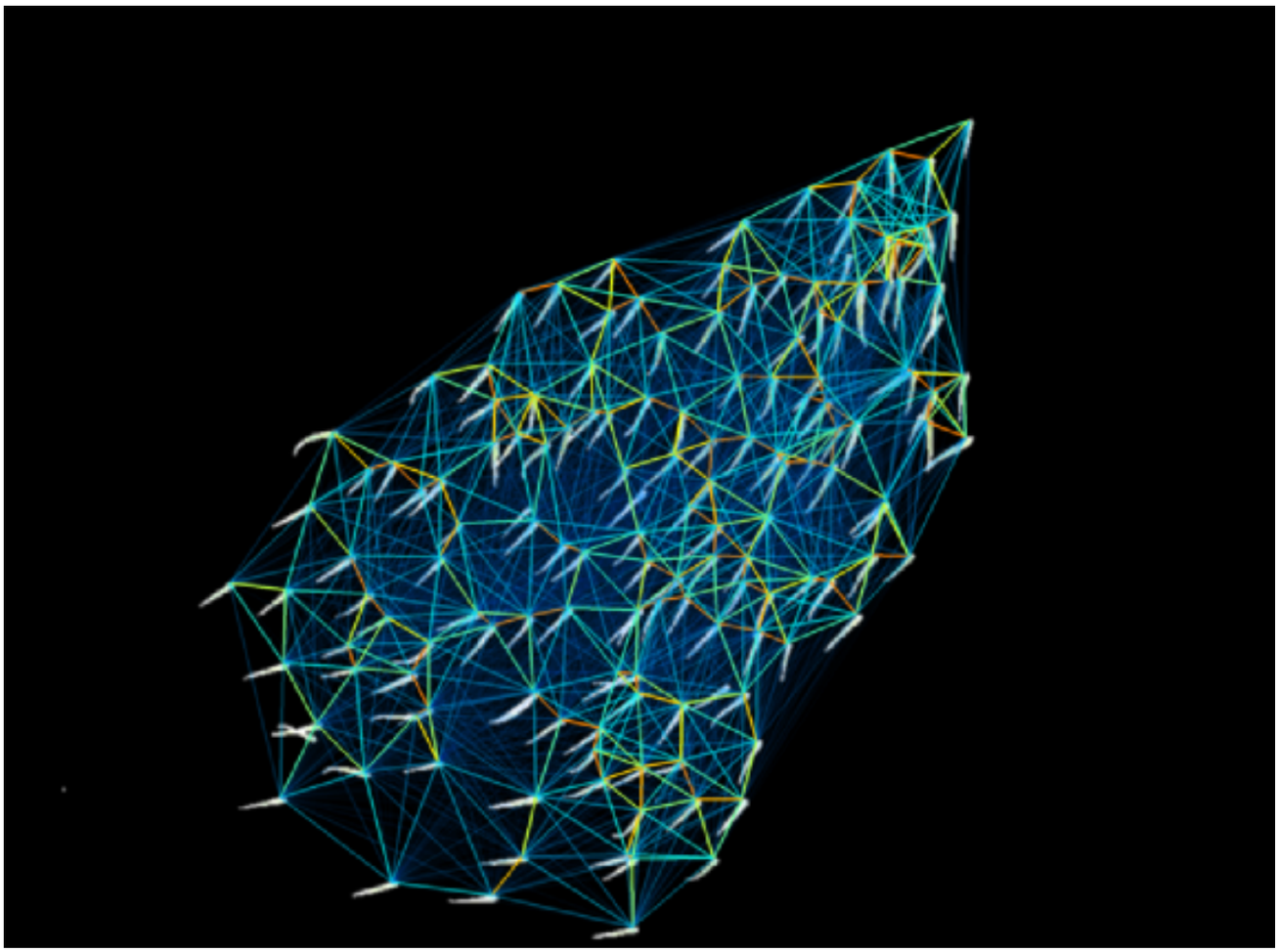


Fig. S25. Gaussian threshold distribution results (**A**) Maximum log-likelihood for models predicting responder identities in experimental cascades. Mean values of the maximum likelihoods are shown, with standard errors averaged over 10 iterations of 500 sample parameter values (note that error bars are smaller than marker size). Simple random, numeric random, and fractional random are the results from randomized versions of the data, and represent the baseline maximum log-likelihood. Similar to Figure S21, fractional contagion is the best model. (**B**) Relationship between simulated cascade size and local weighted clustering coefficient, for all three contagion models, using Gaussian distributions. Solid and dotted lines are regression fits to the simulated data, while shaded regions are 95% confidence intervals.



Movie S1. See attached movie. Here we show the dynamic visual connections between one selected focal individual and its neighbors in a group of freely swimming fish. The red dot is placed over the head of the focal fish, while the red outline is drawn around the focal fish's body. Blue dots are placed over the head of every fish that has a direct line of sight to the focal individual. Light blue triangles show the angular area occupied by the focal individual on the eye of each observer.



Movie S2. See attached movie. The dynamic complex network is shown for five selected cascade events, at half speed, 15 frames per second. Lines are drawn between every fish pair that has visual contact, with more transparency for weaker connections. Strong connections are plotted in brighter colors, with red being the strongest.