

Topology and geometry of clusters in the Vicsek model across scales

Ashiqur Rahman ^{*}

Department of Mechanical Engineering, Virginia Tech, Blacksburg, Virginia, USA

Michael Napoli ^{*} and Ofek Lauber Bonomo 

*Department of Mechanical and Aerospace Engineering and Center for Urban Science and Progress,
Tandon School of Engineering, New York University, Brooklyn, New York, USA*

Maurizio Porfiri [†]

*Department of Mechanical and Aerospace Engineering, Center for Urban Science and Progress,
Department of Civil, Urban, and Environmental Engineering, and Department of Biomedical Engineering,
Tandon School of Engineering, New York University, Brooklyn, New York, USA*

Nicole Abaid [‡]

Department of Mathematics and the Center for the Mathematics of Biosystems, Virginia Tech, Blacksburg, Virginia, USA



(Received 1 December 2025; accepted 21 May 2026; published 10 July 2026)

Collective motion in active matter can arise from interaction rules that generate group-level order. A seminal example is the Vicsek model, which describes flocking using a simple local alignment rule. Despite three decades of research, the network topology induced by the particles' alignment remains understudied. Here, we explore clusters—defined as connected subgraphs of the system-wide interaction network—through topological and geometric measures at the critical noise and across various speeds and densities. We find that the average degree increases polylogarithmically with cluster mass while the variations from the average degree remain fixed across scales. The average path length scaling reveals a two-dimensional network structure, and the average clustering coefficient converges to a fixed value. Geometric analyses reveal that cluster area grows linearly with mass, with the perimeter growing more slowly. We discuss the findings using the degree distribution and positional roles of particles within a cluster. Our results indicate that clusters in the Vicsek model have uniform density with a self-similar network topology across scales, despite the lack of an explicit mechanism to form such interaction networks.

DOI: [10.1103/hkrb-wpdr](https://doi.org/10.1103/hkrb-wpdr)

I. INTRODUCTION

Collective behavior offers advantages to individuals in a biological group through efficient decision-making, resource sharing, and enhanced protection against external threats [1–5]. The coordinated movement patterns in schools of fish, flocks of birds, and swarms of bacteria showcase how collective behavior maintains group structure, a phenomenon whose mechanisms have inspired a large body of scientific research [6–10]. Mathematical models offer a valuable lens for examining collective dynamics, as individual interactions and their corresponding social networks can be directly defined or measured within these models [11–14].

Understanding collective dynamics in flocking models requires analyzing the underlying network, which facilitates interactions and the transfer of information among individuals. The topology of these networks can take many forms. For example, Ballerini *et al.* demonstrated that birds in a flock interact, on average, with a fixed number of neighbors rather than with all the individuals within a fixed metric distance [15]. They also found that this so-called topological interaction model leads to significantly higher group cohesion, as agents are capable of forming long-range interactions in the network. Similarly, individual ants respond to the rate and pattern of local interactions with other colony members, generating a self-organizing network that regulates activity across the entire colony [16–18].

The Vicsek model is perhaps the simplest and most studied model of collective dynamics. In this model, self-propelled particles with constant speed adjust their heading in response to the direction of their neighboring particles and under the effect of uniform additive noise [19]. If the noise is sufficiently small, the particles follow a common heading direction, displaying long-range order despite the locality of their interactions [12]. If noise is high, particles act

^{*}These authors contributed equally to this work.

[†]Contact author: mporfiri@nyu.edu

[‡]Contact author: nabaid@vt.edu

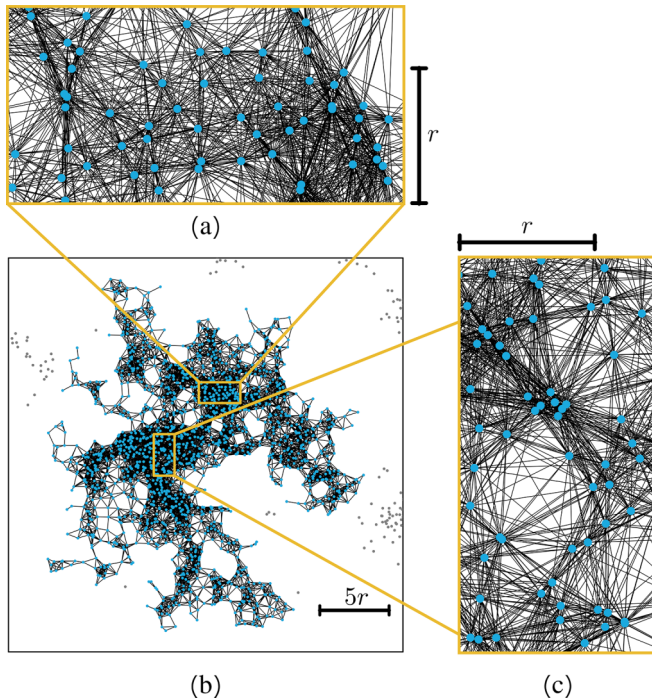


FIG. 1. Substructures with uniform density emerge in large clusters in the Vicsek model. Regions emerge where particles maintain near-uniform spacing and comparable local connectivity as seen in panel (a), a subregion of a cluster in the Vicsek model [panel (b)]. This organization coexists with regions where spatial regularity relaxes [panel (c)], yet most node degrees remain moderate rather than extreme because no organizing pattern concentrates. Overall, particles in the Vicsek model create clusters with a uniform spatial density. In this example, the Vicsek model is simulated with $N = 32,768$ total particles, speed $s = 0.1$, density $\rho = 0.25$, and the interaction radius is $r = 1$.

independently, as their heading is dominated by random perturbations [14]. The presence of an order-disorder transition at a critical noise value has been studied across a range of particle densities and speeds [19–22]. Chaté *et al.* demonstrated that, close to the transition, the dynamics of the ordered phase are governed by high-density, high-order solitary structures [21]. Away from the transition, the ordered phase becomes statistically homogeneous, exhibiting superdiffusion and intermittent dynamics arising from interactions between distinct clusters [21,23]. As speed increases, the phase transition of the Vicsek model approaches that of the vectorial network model, where connections become uncorrelated in time due to the rapid coverage of the space [24,25].

Near the order-disorder transition, particles group in space, leading to locally high alignment (Fig. 1). These clusters are transient in nature, colliding and fragmenting repeatedly, akin to the behavior seen in reacting polymers [26,27]. In the most extreme cases, stable high-density moving regions of local alignment—referred to as bands in two dimensions or sheets in three dimensions—may emerge as a function of density and speed [21]. As noise is decreased in the Vicsek model, clusters coalesce into a single, rigid group, referred to as a “macrocluster,” which contains a large fraction of the system particles [28]. In this article, we formally define clusters as

connected components in the interaction network built by connecting neighboring particles within a fixed radius.

Changing the definition of neighbors can have a profound effect on the interaction network in the Vicsek model and its subsequent clusters. For instance, adding a restricted vision cone—limiting the spatial region from which each agent can detect and interact with neighbors—changes the connections from undirected to directed [29–31]. This directionality can be reversed depending on whether interactions between agents are considered to occur based on visual or auditory cues [31]. Introducing a hierarchy for interactions creates multiscale networks where particles can respond to neighbors at different organizational levels, allowing for the emergence of leader-follower relationships and complex spatial arrangements [32].

While the Vicsek model is well studied from the view of the whole system, the topology of the interaction network coevolving with the particle motion remains elusive [20,33]. To characterize the connected components of this network, Huepe and Aldana defined the cluster mass distribution (CMD; synonymous with the cluster “size” distribution reported in some references) as the probability that a particle is a member of a cluster of a given “mass” or cardinality [23]. Their analysis demonstrated that particle aggregation into clusters is a key factor in intermittent dynamics. They further showed that events involving the gain or loss of a small number of particles are the most common, and that clusters of all masses have an equal probability of gaining or losing a given number of particles. Peruani and Bär, and Barberis demonstrated that the CMD undergoes a transition from a monotonically decreasing function to one exhibiting a pronounced peak at large cluster masses as the noise decreases [27,28,34]. To the best of our knowledge, only Baglietto *et al.* have directly addressed the network topology of particle clusters in the Vicsek model [35]. They found that flocks are “highly clustered, assortative, and nonhierarchical networks with short-tailed degree distributions” and that they have “an effective network dimension higher than that of the space where the actual dynamics actually occur.” However, the spatial structure of the clusters and their dependence on system parameters have yet to be illuminated.

How does the system speed and density shape the CMD? What is the topology of cluster networks? How do geometric characteristics of the clusters relate to the network topology? To answer these questions, we take a three-pronged approach, where we (1) study the effects of speed and density on the trends displayed in the CMD and cluster attributes, (2) analyze the trends of classical network features (degree, path length, and clustering coefficient) as a function of the cluster mass, and (3) describe the geometric shape of clusters and explore their dependency on mass. We focus our analysis near the critical noise, where the Vicsek model yields clusters spanning a broad range of masses, and compare our findings with those of Baglietto *et al.* [35].

Similar to canonical examples of spatially constrained networks [36]—such as random geometric graphs (RGGs) and regular lattices [36,37]—we find that clusters in the Vicsek model maintain a consistent network structure across orders of magnitude in mass. This finding is coherent with our analyses of cluster geometry in the Vicsek model, which reveal that clusters possess a nearly uniform density and consist of

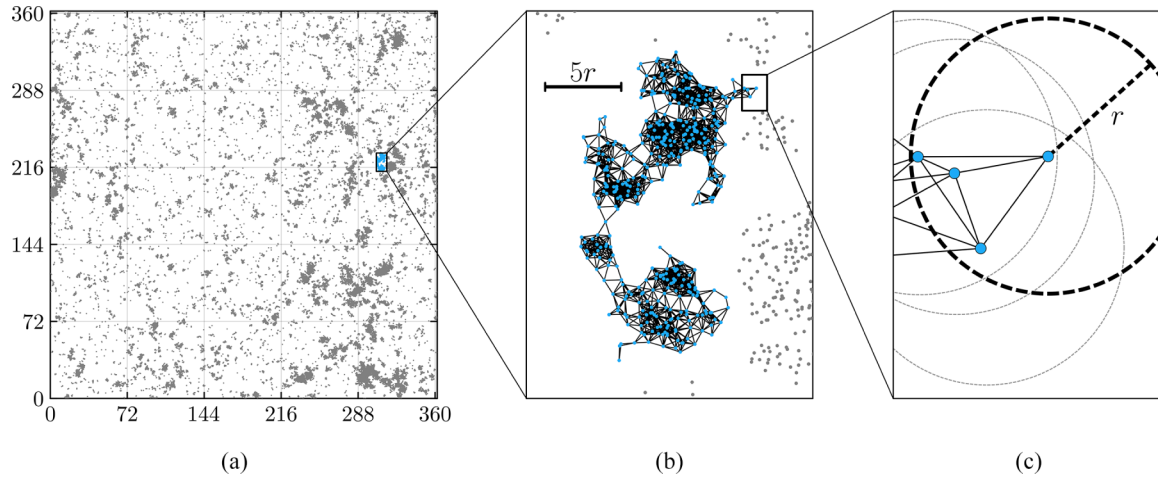


FIG. 2. Snapshot of the interaction network of the Vicsek model. We simulate a Vicsek model for $N = 32\,768$ particles moving with speed $s = 0.1$, density $\rho = 0.25$, and critical noise $\eta_c = 0.13$. (a) Interacting particles form dispersed clusters, with axis labels showing coordinates in arbitrary units equal to the interaction radius r . (b) Clusters can range in mass $m \in \{1, 2, \dots, N\}$, with larger clusters characterized by sparsely connected subgroups. (c) Within each cluster, particles are connected if they are within $r = 1$ units of each other, and use Eq. (2) to align their direction of motion over these connections.

regions with minimal variation in spatial characteristics (see Fig. 1). We study the dependence of these clusters' characteristics on the two remaining parameters in the model: speed and density, with the critical noise calculated as a function of each speed-density pair.

II. THE VICSEK MODEL

We consider N particles in a square domain of length L and density $\rho = N/L^2$, with periodic boundary conditions. The j th particle has position $\mathbf{x}_j \in \mathbb{C}$, velocity $\mathbf{v}_j \in \mathbb{C}$, and heading direction $\theta_j \in \mathbb{R}$. All particles move with the same constant speed $s = |\mathbf{v}_j(t)|$ for all time steps t , so that

$$\mathbf{x}_j(t + \Delta t) = \mathbf{x}_j(t) + \mathbf{v}_j(t)\Delta t, \quad (1)$$

where $\mathbf{v}_j(t) = s e^{i\theta_j(t)}$, with i being the complex unit and Δt the length of the time step. We use a backward update scheme, as in the original Vicsek model, in which particle headings are updated first, and the resulting velocities are then used to update particle positions [19].

The particle's heading direction is adjusted at each point in time to align with neighbors within an interaction radius $r \in \mathbb{R}^+$, with alignment hampered by independent realizations of a uniformly distributed random variable $\xi_j \in [-\pi, \pi]$, scaled by the noise amplitude $\eta \in [0, 1]$. These mechanisms lead to the following update rule for the heading:

$$\theta_j(t + \Delta t) = \text{Arg} \left[\sum_{u \in \mathcal{N}_j(t)} \mathbf{v}_u(t) \right] + \eta \xi_j(t), \quad (2)$$

where $\mathcal{N}_j(t)$ is the time-varying set of particles inside the interaction radius r for particle j , including itself. Under the analogy of a flock of birds, each particle adopts the mean heading of its nearest neighbors in an attempt to align with the local "flock" of particles [38]. Noise is added after computing the average heading in an attempt to capture the inability to follow a perfect consensus among conspecifics. The periodic boundary conditions imply that particles that leave the

boundary from one side reenter on the opposite side. The alignment function couples the dynamics of neighboring particles.

Interactions between particles induce an undirected, time-varying network, where the (u, j) th edge exists if particles u and j are interacting. In this context, particles separated by distances less than r reside in the same connected component of the interaction network (see Fig. 2). Clusters are defined as connected subgraphs in the interaction network and are characterized by the cardinality of their node set $m \in \{1, 2, 3, \dots, N\}$, which we call the cluster mass.

When $\eta \rightarrow 0$, the Vicsek model tends toward a perfectly aligned state with a single, dominating macrocluster [28]. Particle alignment can be quantified via the order parameter

$$\varphi(t) = \frac{1}{Ns} \left| \sum_{j=1}^N \mathbf{v}_j(t) \right|, \quad (3)$$

where $\varphi \approx 1$ in perfectly aligned systems. As $\eta \rightarrow 1$, the direction of motion for an individual particle is independent of its coupling to its neighbors, so $\varphi \rightarrow 0$ (up to numerical fluctuations) and any clusters that form are small and temporally fleeting [28].

In the intermediate range of η , the Vicsek model exhibits a phase transition from the ordered to the disordered phase, described by critical noise η_c . Near this critical noise threshold, the model produces groups of particles with coupled dynamics and clusters of a wide range of masses [28]. The critical noise η_c is a function of both the system speed and density [20]. Due to the increase in the richness of the cluster dynamics, we choose to simulate the model at the critical noise for all combinations of speed and density (see Sec. V A for a detailed method).

For low speed and low density, $\eta_c \sim s\sqrt{\rho}$, but this relationship does not hold at higher speeds or densities [20,22]. As speed or density increases, the clusters collide more frequently, resulting in more alignment and a higher value of η_c [21]. When $s \rightarrow \infty$, the phase transition in the Vicsek

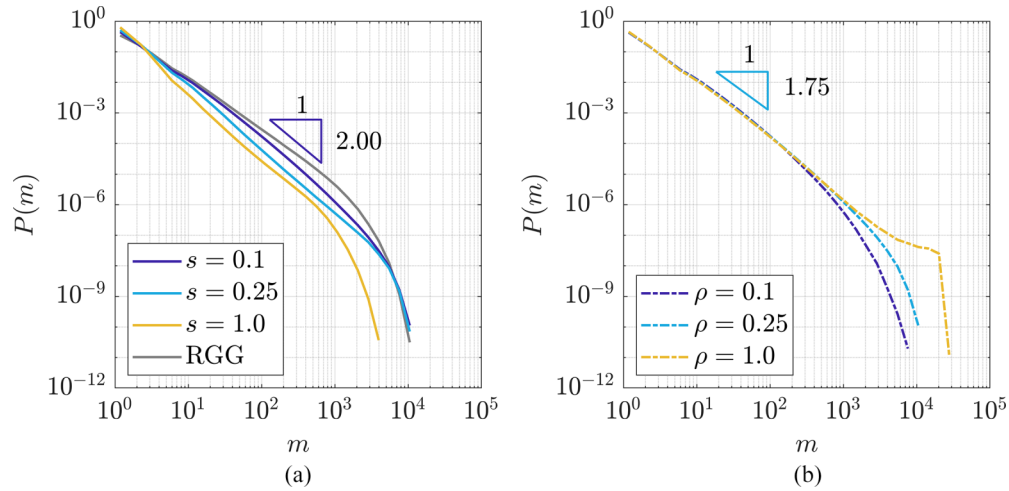


FIG. 3. Cluster mass distributions at critical noise for varying density ρ and speed s . The CMD is computed over each bin as an average over a total of 6030 snapshots, comprising 201 independent samples from 30 replicates (see Sec. VB for computational details). Shading around each curve shows the standard deviation across the 6030 snapshots, although it is too narrow to be visible. (a) For increasing s at fixed density $\rho = 0.25$, the scaling of the cluster mass distributions remains approximately constant for $m \in \{10, 11, \dots, 10^3\}$, with a downward translation in the trend. The gray line corresponds to CMD created by 6030 replicates of RGGs at $\rho_{\text{RGG}} = 1.30$ (see Sec. A of the Supplemental Material [49] for alternative densities). At the chosen density, the CMD in the RGG has a heavier tail but a similar frequency of occurrence over the entire range of cluster sizes, compared to the CMDs of the Vicsek model at $\rho = 0.25$. (b) For increasing ρ at fixed speed $s = 0.1$, the distribution for clusters of mass $m < 10^2$ is constant. However, large clusters ($m > 10^3$) become more probable with increasing density and are capable of exhibiting traveling bands spanning the length of the domain. In each figure, the triangle represents the slope for $s = 0.1$ and $\rho = 0.25$.

model is similar to that of the vectorial network model, where connections are effectively random across time steps [24,25,39]. When observed instantaneously, snapshots of the interaction network coupling Vicsek particles moving at high speeds resemble that of an RGG, defined by a uniform spatial distribution and fixed interaction radius [40]. We thus identify the emergent properties of networks in the Vicsek model by comparing with this effective “null” model, where interactions between particles do not influence the position of the particles.

The nature of the order-disorder transition can vary based on the particle speed, density, and how noise is implemented in the model. Nagy *et al.* found that at low speeds the transition is second order (the first derivative of the order parameter curve is continuous), but it is first order at high speeds (a discontinuity in the first derivative appears at the transition point) [41]. Pimentel *et al.* showed that, for so-called extrinsic noise—where errors are introduced when estimating the interactions as opposed to after the averaging rule is applied as in the original Vicsek model—the phase transition is first order [42]. Mwaffo *et al.*, and Porfiri and Ariel explored alternative distributions of additive noise to better capture individual dynamics of biological systems and identify salient statistical properties shaping the phase transition [25,43].

Applying kinetic theory in hydrodynamic descriptions of the Vicsek model has been useful to extract analytical solutions for macroscopic behavior as defined by the order parameter [44,45]. The flocking transition in the Vicsek model takes the form of a liquid-gas transition in stochastic hydrodynamic equations [33]. Peshkov *et al.* derived continuous equations describing large-scale behavior for the Vicsek model at low density [46]. Ihle studied the phase transition

using kinetic equations and the stability of homogeneous states near critical noise [47,48].

III. RESULTS

A. Cluster mass distribution

The CMD represents the probability $P(m)$ that a particle exists in a cluster of mass m . In agreement with the literature, the CMD at critical noise scales with a power-law relationship to the cluster mass, $P(m) \propto m^\gamma$, where γ is the scaling exponent (see Secs. VA and VB for computational details). Note that, while the distribution is said to follow a power law, the scaling exponent is sensitive to the choice of cluster mass range considered for fitting due to the finite size effects at the boundary of the system size. We choose ranges from the literature to allow for direct comparison to previous results [23,35].

Figure 3(a) illustrates variations in the CMD for changing speed, s , while the density is fixed at $\rho = 0.25$. For higher speed, the proportion of isolated particles becomes more prevalent, as identified by the increasing sharpness of the initial slope of the CMD. More specifically, in the range of small clusters ($m < 10$) the scaling exponents are $\gamma = -1.74 \pm 0.09$, -1.96 ± 0.14 , and -2.46 ± 0.23 for speeds $s \in \{0.1, 0.25, 1.0\}$, respectively, where the bounds describe one standard deviation around the mean. However, we find that increasing speed does not affect the scaling of the CMD for clusters in the range $m \in \{10, 11, \dots, 10^3\}$, yielding exponents $\gamma = -2.003 \pm 0.049$, -2.120 ± 0.019 , and -2.108 ± 0.062 for the same three speed values. Instead of changing in steepness, the overall curve translates downward as the speed increases due to the variation in the initial slopes

of the distribution. This scaling exponent matches the results of Baglietto *et al.*, who reported $\gamma \in [-2.6, -2.0]$ for $\rho = 0.25$ and $s = 0.1$ [35].

For increasing density at fixed speed $s = 0.1$, we observe a negligible change in the scaling of the CMD for clusters in the range $m \in \{1, 2, \dots, 10^2\}$ with scaling exponents $\gamma = -1.740 \pm 0.047$, -1.754 ± 0.035 , and -1.768 ± 0.033 for $\rho \in \{0.1, 0.25, 1.0\}$, respectively [see Fig. 3(b)]. This scaling is comparable to the values reported by Huepe and Aldana ($\gamma \in \{-1.9, -1.5\}$ for $N \in \{500, 5000\}$, respectively) [23]. For $m > 10^3$, the slope of the CMD—in the log-log domain—varies nonlinearly with the cluster mass. Generally, as the density is increased, the exponent describing the CMD decreases with the cluster mass, with stronger decrease associated with lower density. This is, in part, due to the existence of traveling bands that span the length of the domain, wrapping around the torus created by the periodic boundary condition and increasing the probability of observing large clusters. At higher densities, the mean free path of particles is reduced, which facilitates the formation of such bands [20]. This mechanism is further promoted by higher particle speed, even though higher density and speed raise the critical noise threshold. Because traveling bands are regions of exceptionally high local density, snapshots containing them are more likely to exhibit very large clusters than those in which no bands are present. The role of traveling bands in producing dense regions has been highlighted in prior studies of the ordered phase near criticality [50].

To compare cluster networks in the Vicsek model with a spatially constrained null model, we present an RGG as an alternative for cluster formation. RGGs are parametrized by the system-wide particle density, ρ_{RGG} , which sets the typical interparticle spacing and controls cluster formation [40] (see Sec. V C for details). Consequently, any comparison between RGG and Vicsek clusters is sensitive to this choice of density. The particles in both the Vicsek model and RGGs are capable of exhibiting similar CMDs. However, the CMD created in RGGs by simply considering the system-wide density in the Vicsek model, $\rho_{\text{RGG}} = \rho$, is vastly different, with large clusters becoming significantly less probable. In the absence of motion, the Vicsek model will maintain the same CMD as that measured in RGGs with the same system-wide density.

With this in mind, we parametrize the RGG by its density, and choose the value where the CMDs obtained from RGG and Vicsek are qualitatively comparable (see Sec. A of the Supplemental Material [49] for alternative choices of density). This similarity in cluster frequency occurs when $\rho_{\text{RGG}} \approx 1.30$, as illustrated in Fig. 3(a), where similarities in the frequencies of clusters of a given size allow for a fair comparison across cluster metrics. The notably large ρ_{RGG} suggests that clusters in the Vicsek model restrain themselves to a higher “effective” system-wide density that is approximately fivefold greater than the true density. Finally, we note that, for increasing speed in the Vicsek model, large clusters occur less frequently than for their RGG counterpart.

B. Cluster topology

The average degree, $\langle k \rangle$, is a common metric for quantifying the connectedness of a network and can be computed by

considering the average number of connections at each node in the network. Similarly, the cluster-specific average degree, $\langle k \rangle$, calculated as the mean average degree of the clusters with mass in the same mass bin, tells us the intercluster average and highlights the connectedness of the cluster relative to the expected connectedness of the system. We also consider the path length, $\langle \ell \rangle$, and clustering coefficient, $\langle c \rangle$, averaged over the cluster networks [37,51]. The average path length is a measure of how distant nodes are topologically, while the average clustering coefficient reflects the relationship between a node’s neighbors, and its neighbor’s neighbors, helping to quantify the “cliquishness” of a network by measuring its substructures [52]. Here, we apply each of these metrics to clusters formed in the Vicsek model, with computation described in Sec. V D.

In Fig. 4, we plot the system-wide degree distribution, the cluster-specific average degree for variations in cluster size, and an example of the normalized cluster-specific degree distribution. Each of these metrics is compared against its RGG counterpart so as to highlight the direction of the trends for growing system speed. We find that the probability of a given degree follows an exponential distribution such that, $P(k) \propto e^{-\lambda k}$, where $\lambda = 0.087 \pm 0.001$, 0.113 ± 0.003 , and 0.346 ± 0.008 , for $s \in \{0.1, 0.25, 1\}$, respectively. In the case of RGGs, the exponent is $\lambda = 0.805 \pm 0.155$, which is significantly higher than any of the exponents measured in the Vicsek model. Borrowing terminology from classifications of small-world networks, we argue that the system-wide degree distribution is indicative of a single-scale network [53]. A single-scale network is one in which the degree distribution decays faster than a power law (including exponential and Gaussian distributions).

Interestingly, we find that the cluster-specific average degree [plotted in Fig. 4(b)] does not trend with a power law of the cluster mass, in contrast to observations from Baglietto *et al.* [35]. Instead, the degree follows a polylogarithmic trend of the form $\langle k \rangle \propto (\ln m)^\beta$, where β is a scaling exponent that decreases with an increase in the speed s . The exponents are $\beta = 1.315 \pm 0.017$, 1.106 ± 0.029 , and 0.799 ± 0.046 for $s \in \{0.1, 0.25, 1.0\}$, respectively. While the average degree increases monotonically, the polylogarithmic trend of $\langle k \rangle$ implies that the growth becomes progressively slower with increasing m and approaches the trend in the RGG. It is difficult to say if $\langle k \rangle$ is truly converging to a single value, especially at lower speeds, where $\langle k \rangle$ becomes large. Although one might attempt to fit asymptotic functions (such as a decaying power law) to estimate $\langle k \rangle$ as $N \rightarrow \infty$, computational capacity limits our ability to measure system sizes exceeding $N = 32\,768$ reported in this study in a reasonable time frame. Nevertheless, simulating systems for at any finite N may not necessarily remove finite size effects since cluster size is bounded above by N and thus cannot be taken to infinity independently (see Sec. C of the Supplemental Material [49] for simulations with varying N).

The sharp peak of the normalized degree distribution [illustrated in Fig. 4(c)] supports our argument that particle degrees within clusters are concentrated around the average degree $\langle k \rangle$, whereby the variance from the average degree is consistent for variations in speed, density, and cluster size (see Sec. B of the Supplemental Material [49] for

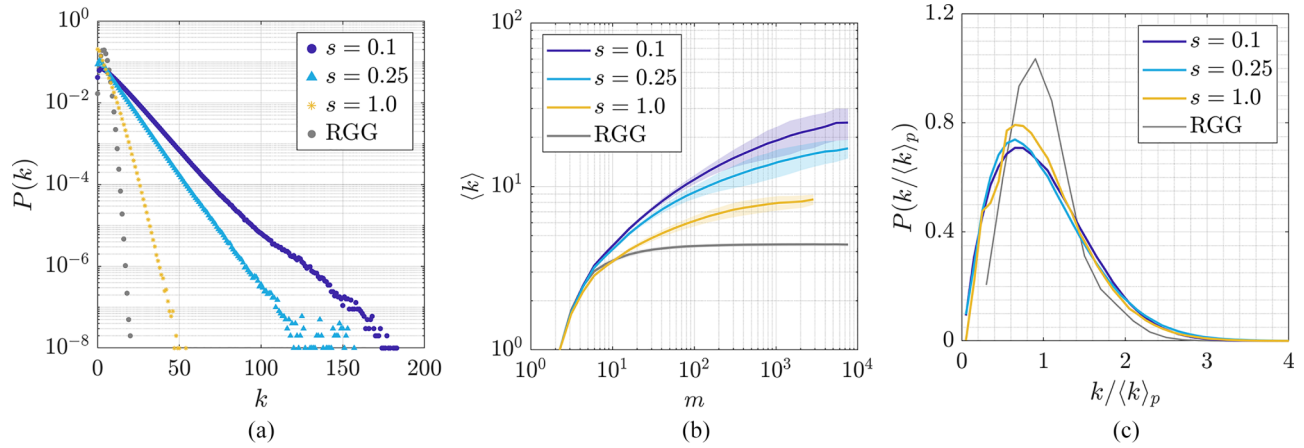


FIG. 4. Global and cluster-specific degree distributions in the Vicsek model. We plot the system-wide degree distribution $P(k)$, cluster-specific average degree $\langle k \rangle$, and normalized degree distribution $P(k/\langle k \rangle_p)$, with fixed density $\rho = 0.25$. Curves are averages over 6030 snapshots, with shading showing the standard deviation, and each trend is accompanied by its RGG counterpart (compiled from 6030 replicates). (a) The global degree distribution follows an exponential decay, $P(k) \propto e^{-\lambda k}$, indicating the resilience of a single-scale network topology over the entire system, which is significantly less steep than clusters formed in RGGs. (b) The average degree $\langle k \rangle$ (see Sec. VD for definition) follows a trend of the form $\langle k \rangle \propto (\ln m)^\beta$, which has a continuously decreasing slope as m is increased. For increasing speed, the average degree trend approaches that measured in clusters of RGGs, which reaches an asymptote of $\langle k \rangle_{\text{RGG}} \approx 4.50$. (c) Distribution of the normalized degree over the particles of the clusters of mass $919 \leq m \leq 1271$. We normalize each degree by the average degree of all the particles in the same cluster mass range, $\langle k \rangle_p$, and show that particles in clusters of comparable mass have similar degrees. This result is insensitive to cluster mass m , system speed s , and density ρ (see Sec. B of the Supplemental Material [49]). This sharp peak also indicates that the local density is approximately uniform within each cluster. In the RGG, the normalized degree is comparably sharp around the mean value, but with less overall variance as indicated by the increased peak height.

comparisons across speed, density, and mass). This regularity is illustrated by the limited growth of the average degree, which suggests their connectedness is comparable both within clusters of the same mass and across clusters with varying mass. In line with results from Baglietto *et al.* [35], we find that the normalized degree distribution—computed by taking the ratio between each particle’s degree and the average degree of their corresponding cluster ($\langle k \rangle_p$)—maintains a sharp peak around $k/\langle k \rangle_p = 1$, with an exponential-like tail. A Kolmogorov-Smirnov comparison test between the normalized degree distributions for the Vicsek model at $s = 1.0$ and the RGG shows that there is no quantitative difference in the distributions.

In Fig. 5, we plot the path length, and clustering coefficient at a fixed density $\rho = 0.25$, and varying speed $s \in \{0.1, 0.25, 1.0\}$. The average path length scales with a power law of the cluster mass, $\langle \ell \rangle \propto m^\gamma$, with exponents $\gamma = 0.428 \pm 0.009$, 0.483 ± 0.009 , and 0.554 ± 0.011 , respectively, for $\rho = 0.25$ and $s \in \{0.1, 0.25, 1.0\}$. With an increase in speed, the exponent describing the average path length also increases.

For $s = 0.1$ and $s = 0.25$, the average clustering coefficient $\langle c \rangle$, shown in Fig. 5(b), exhibits maxima with values of approximately 0.67 and 0.65, respectively, at the cluster mass $m = 8$. The peak in $\langle c \rangle$ is the result of the ability of slow-moving, small clusters to form more triangular connections, as opposed to fast-moving and large clusters, where the ratio of existing-to-potential triplets becomes less pronounced. The trend then gradually decreases with an asymptotic approach to ≈ 0.61 . In contrast, the high-speed case with $s = 1.0$ monotonically increases toward the same value of $\langle c \rangle \approx 0.61$. This trend seems to highlight that, while variations in speed

have a large effect on the average degree and path length, the cliquishness of particles within Vicsek clusters remains consistent in very large clusters. However, smaller clusters ($m \approx 10$) exhibit noticeably lower clustering coefficient when speed is increased. We hypothesize that for larger system sizes, where clusters larger than $m = 10^4$ may emerge, the average clustering coefficients will converge to similar $\langle c \rangle$ with increasing m for a wide range of speeds, including those selected for this study. Although $\langle c \rangle$ varies for small clusters ($m < 10^2$), its convergence toward a constant highlights a persistent level of local interconnectedness in large clusters.

As speed increases, we observe a clear increase in the similarity between the network metrics measured in the Vicsek model and those measured in RGGs [Figs. 4(b) and 5]. For large clusters, the average degree in RGGs is mass invariant with an asymptote at approximately $\langle k_{\text{RGG}} \rangle = \pi \rho_{\text{RGG}}$, a result of the average spacing between particles at ρ_{RGG} . The scaling of the average path length increases from $\gamma \approx 0.428$ to $\gamma \approx 0.554$ as the speed increases from 0.1 to 1.0, approaching the scaling of $\gamma_{\text{RGG}} \approx 0.650$ measured in RGG clusters. Interestingly, the clustering coefficient is the only metric that does not trend toward its RGG counterpart, but generally has a comparable magnitude in large clusters (≈ 0.610 and ≈ 0.542 for the Vicsek model and RGG, respectively). These results suggest that the repeated local interactions of particles in the Vicsek model create clusters with a regular network organization across scales, suppressing the emergence of highly interconnected nodes.

By identifying Vicsek clusters as single-scale networks, we accentuate (1) the slow growth of the average degree with cluster mass, and (2) the collapse of the degree

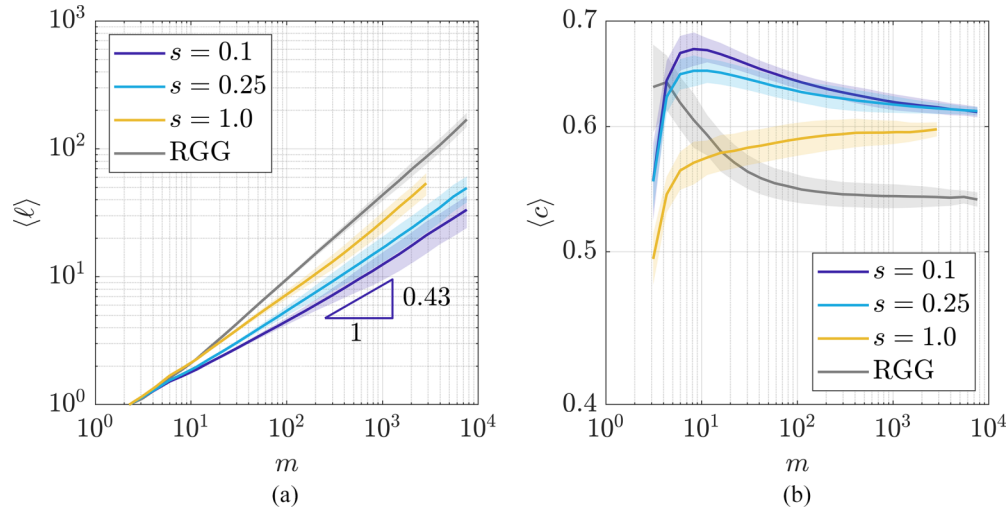


FIG. 5. Network topology metrics of clusters in the Vicsek model. We plot cluster-specific average path length $\langle \ell \rangle$ and average clustering coefficient $\langle c \rangle$ as functions of the cluster mass m with fixed density $\rho = 0.25$ (same snapshots used in Fig. 4). Curves are averages over 6030 snapshots, with shading showing the standard deviation, and each trend is accompanied by its RGG counterpart (compiled from 6030 replicates). (a) The average path length $\langle \ell \rangle$ grows with a sublinear power law of the cluster mass m , with a higher exponent for higher speeds, suggesting that elongated clusters emerge for systems with fast-moving particles. The triangle represents the slope for $\rho = 0.25$ and $s = 0.1$, with exponents for other parameter sets given in Table I. As speed increases, the exponent approaches that of clusters in RGGs. (b) The clustering coefficient exhibits a peak at $m = 8$ in both of the low-speed cases, before approaching a constant value. For high speed, the peak value does not emerge, but $\langle c \rangle$ approaches the same asymptotic value for large cluster masses. While the trend measured in the Vicsek model does not approach that of the RGGs, the asymptotes measured in both cases are comparable.

distribution around said average degree. This differs from scale-free networks or networks with preferential attachment, where the degree distribution lacks a characteristic scale and one observes the increased prevalence of high-degree nodes [51,54]. Measurements on flocks of birds by Ballerini *et al.* demonstrate that the individuals in the flock maintain an approximately constant degree, pointing away from “hubs,” or nodes with a comparatively high connectivity that are common in scale-free networks and act as leaders for the group [15,51]. The slow, saturating growth of the average degree

thus indicates that clusters retain a regular local structure even as they become very large.

On the other hand, the growing differences between Vicsek clusters and their RGG counterparts for decreasing speed suggest that the network features measured in the Vicsek model are an emergent property of the dynamics. The increased connectedness in large clusters arise from correlated motion and repeated local alignment, while maintaining the single-scale topology indicative of spatially constrained networks like RGGs and lattices.

TABLE I. Critical noise η_c and trends on topological and spatial metrics for changing speed s and density ρ . For each parameter set, we report the average degree $\langle k \rangle$, average path length $\langle \ell \rangle$, cluster area A , and the number of interior n_i , boundary n_b , and exterior n_e nodes. With the exception of the average degree, which was fitted with a polylogarithmic trend $\langle k \rangle \propto (\ln m)^\beta$, each metric was fit using a power law of the form $\propto m^\nu$, with upper and lower bounds denoting the 95% confidence interval. The cluster mass range considered for fitting the curves is reported at the bottom of the table. The R^2 values for all fits were greater than 0.985 and so are not reported individually.

Parameters			Scaling exponents describing trends of fit for metrics					
s	ρ	η_c	$\langle k \rangle$	$\langle \ell \rangle$	A	n_i	n_b	n_e
0.1	0.1	0.08	1.39 ± 0.02	0.40 ± 0.01	1.03 ± 0.03	1.13 ± 0.03	0.77 ± 0.02	0.62 ± 0.04
0.1	0.25	0.13	1.32 ± 0.02	0.43 ± 0.01	1.03 ± 0.02	1.12 ± 0.03	0.80 ± 0.01	0.66 ± 0.02
0.1	1.0	0.25	1.20 ± 0.01	0.46 ± 0.01	1.04 ± 0.02	1.14 ± 0.02	0.83 ± 0.00	0.69 ± 0.01
0.25	0.1	0.11	1.17 ± 0.03	0.47 ± 0.01	1.06 ± 0.02	1.14 ± 0.03	0.86 ± 0.01	0.71 ± 0.02
0.25	0.25	0.17	1.11 ± 0.03	0.48 ± 0.01	1.06 ± 0.02	1.14 ± 0.02	0.87 ± 0.01	0.73 ± 0.01
0.25	1.0	0.31	1.06 ± 0.02	0.47 ± 0.01	1.07 ± 0.01	1.19 ± 0.01	0.87 ± 0.02	0.71 ± 0.02
1.0	0.1	0.18	0.88 ± 0.04	0.53 ± 0.01	1.09 ± 0.02	1.16 ± 0.05	0.95 ± 0.00	0.81 ± 0.04
1.0	0.25	0.26	0.80 ± 0.05	0.55 ± 0.01	1.08 ± 0.01	1.15 ± 0.04	0.97 ± 0.00	0.83 ± 0.02
1.0	1.0	0.39	0.75 ± 0.03	0.52 ± 0.01	1.08 ± 0.01	1.19 ± 0.03	0.96 ± 0.01	0.80 ± 0.01
$\rho_{\text{RGG}} = 1.30$				0.65 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	0.99 ± 0.01	0.98 ± 0.00
Mass range			$m > 2$	$m > 2$	$m > 10$	$m > 100$	$m > 100$	$m > 100$

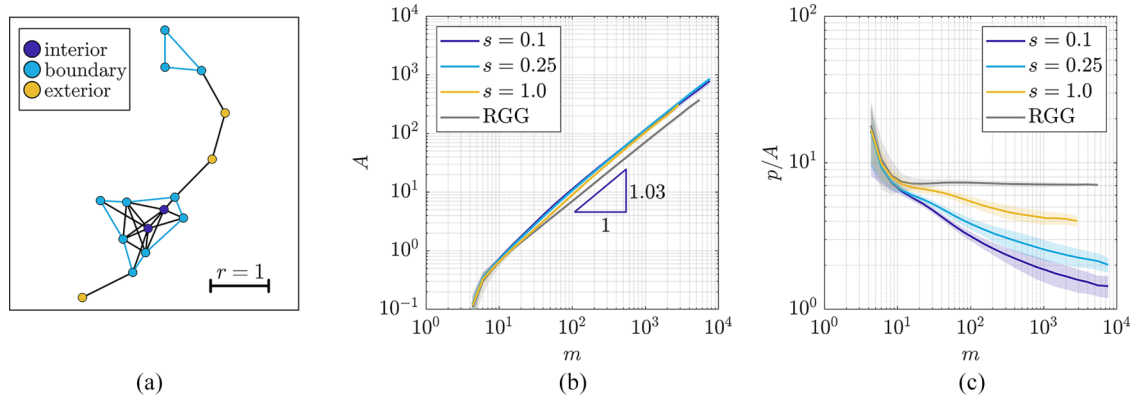


FIG. 6. Node labeling and the variation in the area coverage A and perimeter-to-area ratio p/A of clusters as a function of the cluster mass m with fixed density $\rho = 0.25$. Curves are averages over 6030 snapshots, with shading showing the standard deviation. Each trend is accompanied by its RGG counterpart. (a) Interior nodes are those contained within the boundary created by the perimeter (blue) of the cluster, while boundary nodes are those on the perimeter. Exterior nodes are those that are not contained within, nor contribute to, the area of the cluster. (b) The area of a cluster increases linearly with its mass, implying that the particle density is fixed across cluster sizes. The triangle indicates the scaling for $\rho = 0.25$ and $s = 0.1$ while the other parameter sets are reported in Table I. The trend is comparable to that measured in RGG clusters for each measured value of speed. (c) When compared to the perimeter of the cluster p , however, the perimeter-to-area ratio decreases with cluster mass—illustrating an elongation as the cluster grows. For increasing speed, p/A decreases more slowly due to a change in composition of the node types. For increasing speed, the perimeter-to-area ratio approaches that measured in clusters of RGGs.

C. Cluster geometry

As clusters in the Vicsek model are complex, spatially defined objects, their topology is closely related to the spatial distribution of their corresponding particles. With this in mind, we study trends in area and perimeter for changing cluster mass. To compute the area A and perimeter p , we used alpha shapes [55], which are well suited for capturing the geometry of irregular point sets (particles in a cluster). An alpha shape is a scale-dependent geometric envelope of a point set consisting of all particles whose circumscribed radius is at most α , which in our case is equal to half the interaction radius, $r/2$, and allows for the appreciation of clusters with concave geometric features. This method partitions particles within a cluster into three groups [illustrated in Fig. 6(a)]: (1) interior nodes that lie inside the envelope; (2) boundary nodes that are on the envelope; and (3) exterior nodes that are connected to the network but are not enclosed by the envelope and therefore do not contribute to the area of the cluster. The area of a cluster is given by the area enclosed by the envelope and the perimeter is defined by the edges connecting the boundary nodes. We note that for exterior nodes, removing any incident edge causes the cluster to fragment into smaller clusters. These methods are described in detail in Sec. V E. Classifying the individual particles in a cluster using the perspective of alpha shapes allows for justifying observed trends in the cluster network topology.

Figure 6(b) shows that the cluster area scales with an approximately linear relationship with its mass, $A \propto m^\gamma$, where $\gamma = 1.033 \pm 0.022$, 1.059 ± 0.018 , and 1.084 ± 0.013 for $s \in \{0.1, 0.25, 1.0\}$, respectively, and $\rho = 0.25$. These exponents imply that, for increasing cluster mass, the network area grows proportionally, suggesting that incoming particles join the periphery of the cluster, and ultimately contribute an equal quantity of incremental area, independently of cluster mass. The linear scaling of the area suggests that the local

density of a particle within a cluster is consistent across scales, and that the emergent spatial organization of the cluster is a persistent feature of the group. These results point toward a uniform spatial configuration, whereby the spatial distribution of particles is reminiscent of their RGG counterparts as $s \rightarrow \infty$.

Next, we use the perimeter-to-area ratio as a metric to understand if clusters with more mass are more elongated. Viewed from the shape created by the node positions, clusters are naturally nonconvex entities, with spatially heterogeneous distributions of edges that result in some subsets of nodes being highly connected, and others less so. Figure 6(c) illustrates the comparison between the perimeter and the area, where p/A decreases with increasing cluster mass, with a similar but less pronounced trend for increasing speed that converges toward the area-to-perimeter ratio measured for RGG clusters. In other words, while the cluster area scales linearly, the ratio between perimeter and area decreases with mass due to the slow growth of p . This feature does not detract from the central claim that clusters are uniform, as the lower perimeter-to-area ratio indicates a more regular shape for large clusters. Higher system speed increases the influence of noise on the boundary nodes and creates nonconvex parts in the cluster boundary, increasing p/A in the process. For variations in speed, however, the area maintains its linear relationship with the cluster mass, indicating that the uniform spatial configuration is speed invariant.

To elucidate this trend, Fig. 7 demonstrates the population of each subgroup of nodes—interior (n_i), boundary (n_b), and exterior (n_e)—in all clusters for changing mass and speed. We note that $n_i + n_b + n_e = m$. When averaged across snapshots for $s = 0.1$ and $\rho = 0.25$, we find that $n_e < n_i$ for $m > 15$. Figures 7(a) and 7(b) demonstrate the number of interior and boundary nodes in a cluster of mass m , respectively. We observe that, for clusters of mass $m < 180$, the number of

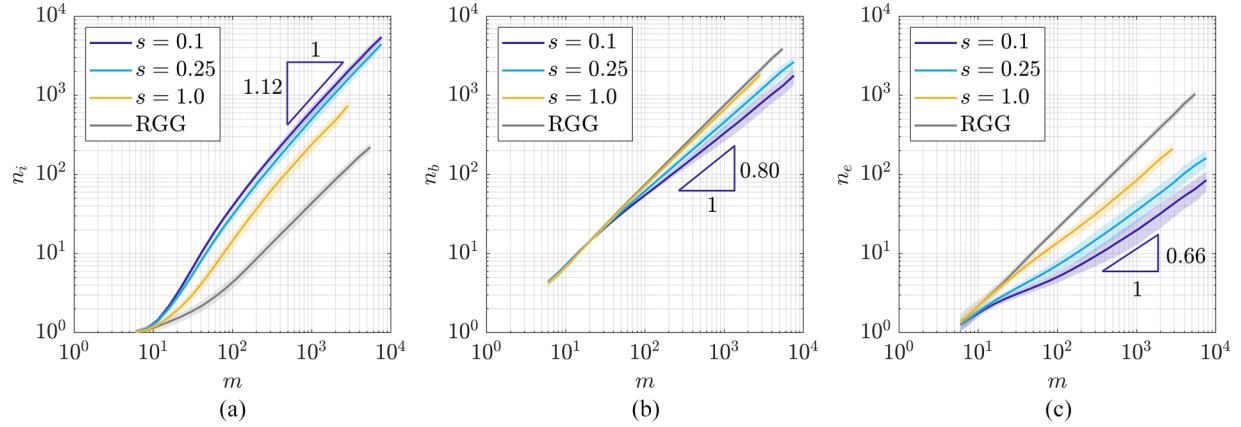


FIG. 7. Variation in the number $n_{\langle \rangle}$ of boundary b , interior i , and exterior e nodes as a function of the cluster mass m for fixed density $\rho = 0.25$. Curves are averages over 6030 snapshots, with shading showing the standard deviation, and are compared to their RGG counterparts. For clusters of mass $m > 10^2$, the number of interior nodes [panel (a)] increases superlinearly, while the numbers of boundary [panel (b)] and exterior [panel (c)] nodes increase sublinearly. The triangles represent the slopes for $\rho = 0.25$ and $s = 0.1$, with exponents for other parameter sets given in Table I. As s increases, clusters of similar mass have less interior nodes, with more nodes existing on the periphery (corresponding to sparser particle connections). Consequently, all trends approach those measured in clusters generated by RGGs, in which each metric scales linearly with cluster mass, indicating that the cluster composition is scale invariant.

boundary nodes composes the largest subgroup, while, for larger clusters, interior nodes become more prevalent. When speed is increased, the presence of boundary and exterior nodes—shown in Fig. 7(c)—likewise increases, with exterior nodes composing up to $\approx 1/5$ of the cluster mass when $s = 1$.

The number of interior nodes increases with a superlinear relationship to the cluster mass, $n_i \propto m^\gamma$, where $\gamma = 1.124 \pm 0.026$, 1.140 ± 0.023 , and 1.150 ± 0.037 for $s \in \{0.1, 0.25, 1.0\}$, respectively, and $\rho = 0.25$. Alternatively, both the boundary and exterior nodes increase with sub-linear trends of the cluster mass: $n_b \propto m^\gamma$, where $\gamma = 0.802 \pm 0.008$, 0.874 ± 0.008 , and 0.965 ± 0.004 for $s \in \{0.1, 0.25, 1.0\}$, respectively, and $\rho = 0.25$; and $n_e \propto m^\gamma$, where $\gamma = 0.659 \pm 0.020$, 0.732 ± 0.014 , and 0.826 ± 0.021 under the same parameter conditions. The exterior and boundary nodes typically contribute less to the total number of connections than the interior nodes. In small clusters, the number of exterior and boundary nodes outweighs the interior, yielding a lower average degree due to a large proportion of the nodes existing on the periphery. When computing area, however, the interior and boundary nodes contribute similarly while exterior nodes are neglected. Thus, the increase in average degree with cluster mass as shown in Fig. 4(b) does not maintain a strict relationship with the area computed via alpha shapes. As clusters increase in size, the relative influence of exterior nodes on the area of the cluster network diminishes and the balanced contribution of interior and boundary nodes creates a linear trend between area and cluster mass [Fig. 6(b)].

As speed increases, the trends approach that of their RGG counterparts, each of which scales with a linear relationship of the cluster mass. This result suggests that, in the limit of $s \rightarrow \infty$, each of the node subgroups grows proportionally, and clusters in the Vicsek model should maintain a fixed ratio between the interior, boundary, and exterior nodes. In RGG clusters, this relationship tells us that interior nodes

will eventually dominate the network topology, and that while the Vicsek model may create local regions with nonuniform spatial distributions of particles, the “expected” cluster in fast systems has a uniform density and behaves similarly to that of the RGG.

IV. DISCUSSION

This paper seeks to fill an important gap in the Vicsek model literature, whereby taking a network-rigorous approach to studying clusters enables us to ascertain a clear picture of their shape and structure. Our evidence suggests that the network topology of Vicsek clusters resembles a two-dimensional single-scale network structure with uniform density, which becomes more well defined in massive clusters—illustrated in Fig. 1. More specifically, clusters in the Vicsek model are characterized by (1) narrowly peaked degree distributions, which is insensitive to changes in cluster mass, (2) clustering coefficient reaching an asymptote, (3) power-law scaling of average path length, and (4) linear growth of the spatial coverage of a cluster’s network. The cluster-specific average degree and average clustering coefficient vary slowly with increasing mass, indicating the presence of large clusters composed of homogeneous substructures. This is further demonstrated by the linear increase in the area covered by the cluster network, suggesting that particles added to the periphery do not considerably compress or expand the existing set of particles in a spatial sense. We find that the trends exhibited for $N = 32\,768$ particles are equivalent to those of smaller systems (see examples for $N = 512$ and 4096 in Sec. C of the Supplemental Material [49]).

We examine how network organization is reflected in cluster geometry, while accounting for the contributions of different node types. Following the results summarized in

Table I, the linear scaling of the cluster area, $A \propto m^\gamma$ with $\gamma \approx 1$, tells us that the cluster-specific density does not change for increasing mass. The decay of the perimeter-to-area ratio, p/A , tells us that the perimeter grows slower than the area [Fig. 6(c)]. While not a precise power law, this trend lies in agreement with expectations for the “effective surface area” of reacting clusters (or the perimeter in $\mathbb{R}^2$), where Ziff *et al.*, and Peruani and Bär argue that the scaling exponent should lie in the interval $[0.5, 1]$ [26,27]. The lower limit, 0.5, reflects the surface area of a compact solid in $\mathbb{R}^2$, and 1 corresponds to the emergence of topologies reminiscent of lattices [26]. In RGG clusters, the perimeter-to-area ratio illustrates a linear relationship between cluster mass and the perimeter, demonstrating that RGG clusters exhibit properties similar to lattices. Likewise, as we increase the system speed, Vicsek clusters more closely resemble interacting lattices in chemical kinetics.

Considering a more expansive definition of cluster area than that of the alpha shape, which includes the union of all interaction areas of its particles, we find $\gamma < 1$ with decreasing magnitude as speed increases (see Sec. D of the Supplemental Material [49]). This arises due to the increased number of exterior nodes in small or fast-moving clusters, where sparsely connected particles contribute more unique area than those that are more dense.

As cluster mass m increases, the interconnectedness of the nodes grows with diminishing returns, following a polylogarithmic relationship between the average degree and cluster mass, $\langle k \rangle \propto (\ln m)^\beta$. For increasing speed, this trend persists, but the exponent β decreases, indicating a further reduced growth rate. Many features of the Vicsek model contribute to this observation. First, the persistence length—the expected distance over which two particles maintain correlated directions—is proportional to particle speed, $\ell_p \sim s/\eta^2$, thereby increasing the probability that a particle loses a connection between time steps in faster systems [20]. The persistence length argument also supports the increase in the number of exterior nodes for increasing system speed. While exterior nodes may exhibit at most five neighbors due to the geometric limitation of the interaction radius in two dimensions, the exterior node-specific average degree does not exceed three for any m for any of the nine parameter sets we considered (see Sec. B of the Supplemental Material [49]). Thus, as speed increases, exterior nodes are more prevalent and $\langle k \rangle$ decreases. Interestingly, the deviation of the degree of a node from the cluster-specific average degree remains invariant with respect to system speed, density, and cluster mass. This is shown in Sec. B of the Supplemental Material [49], where we construct the normalized degree distribution $P(k/\langle k \rangle_p)$ and observe that the probability mass function peaks for $k/\langle k \rangle_p \approx 1$, before decaying toward zero. This trend collapse tells us that, while $\langle k \rangle$ varies as a function of the cluster mass, the topology within the cluster is a simple function of its average degree.

The average path length scales approximately with a power law whose exponent is in the range $\gamma \in [0.40, 0.55]$. Expectations from the literature tell us that the average path length in a complex network typically scales $\langle \ell \rangle \propto m^{1/d}$, where d is the network’s effective dimension, which represents the hypercube whose vertices describe the furthest possible dis-

tance traveled in the network [37]. The scaling exponents extracted from simulations then imply that clusters evolve with $d \in [1.81, 2.50]$ in the Vicsek model—in contrast with the results of Baglietto *et al.*, whose results predict a much higher network dimension $d = 4$ [35]. For high speed, the scaling exponent observed in the Vicsek model approaches that of its RGG counterpart (≈ 0.650). For decreasing speed, the exponent likewise decreases, suggesting that the alignment mechanism helps in creating a higher-dimensional network topology. Finally, we note that, while the network dimension is not exactly equal to the spatial dimension ($d = 2$), its close proximity at moderate speed is reminiscent of a network constrained by the space where the dynamics occur. In particular, the upper limit is given by the slow-speed case and the lower limit is represented by the RGG, whose network properties depend explicitly on the embedding dimension [56].

At this point, we acknowledge a number of key differences between the results shown here and the network metrics computed in Baglietto *et al.* [35]. In their work, they identified $\langle k \rangle \propto m^{1/2}$ and $\langle \ell \rangle \propto m^{1/4}$, with similar values for the mass distribution and average clustering coefficient to those in this work. We speculate that the mismatch in trends is due to differing analytical choices, for example, how temporal independence between snapshots is ensured or whether networks are integrated over time. In the case of a time-integrated network, the network is capable of exhibiting more connections in clusters than the single time step snapshots analyzed here. Supporting this potential cause of disparity, the highest cluster-specific average degree we observe (for parameters $s = 0.1$ and $\rho = 0.25$) has value $\langle k \rangle < 30$, contrary to the $\langle k \rangle > 200$ observed in Ref. [35]. On a similar note, the maximum average path length we measure is $\langle \ell \rangle \approx 50$, whereas Baglietto *et al.* reported $\langle \ell \rangle \approx 10$. We choose not to pursue a time-integrated definition of the interaction network for two reasons: (1) we consider our approach to be more biologically salient, where we assume that animals are more likely to preferentially respond to instantaneous, rather than past, connections, and (2) results from a time-integrated network are largely dependent on the time window over which the analysis is performed.

For low-speed systems, the average clustering coefficient exhibits a peak at clusters with mass $m \approx 8$. Similar to Baglietto *et al.*, we find that the average clustering coefficient has only a weak dependence on the mass of the cluster, with large masses having an asymptoticlike trend toward $\langle c \rangle \approx 0.61$. This differs from many real-world networks, where the average clustering coefficient typically decays with a power-law-like function of the network size [57]. In fast-moving systems, small clusters do not exhibit the peak we see when $s = 0.1$ or 0.25 , due to the decreased likelihood of forming triangular connections.

In systems with high speed or density, the existence of traveling bands plays a role in determining the metrics computed over the range of cluster masses [21]. These bands wrap around the torus created by the periodic boundary conditions, partitioning the domain into regions of low and high order, creating an increase in the probability for larger clusters to emerge. For higher speed and density, traveling bands become

more prominent, increasing the critical noise required to break them apart and achieve a time-averaged order parameter of $\varphi \approx 1/2$ [39]. While phenomenologically interesting, traveling bands are considered an artifact of the finite size effect and careful consideration should be taken when computing spatially correlated metrics [21]. To avoid these issues, we do not consider the upper limits of the CMD when computing trends. That said, the metrics computed over the cluster networks are assumed to be independent of the local order in the environment.

Results from our analysis suggest that particles within clusters in the Vicsek model have a narrow degree distribution with an approximately fixed local spatial density. Cluster networks provide key insights into persistent topologies in the absence of mechanisms for regular structure formation—such as attraction or repulsion—and are consistent with measurements taken on silverside fish schools [6] and flocks of starling birds [15,58]. Alignment in the Vicsek model governs only local interactions among particles, reflecting the limited awareness of conspecifics in mobile animal groups. In particular, Ballerini *et al.* demonstrated that bird flocks interact through a topological distance rule, forming networks with approximately uniform degree, consistent with a decentralized, hub-free form of self-organization [15]. Our results suggest that the Vicsek model may be more capable of representing animal behavior than is currently appreciated, where the microscopic dynamics of individual particles recreate the macroscopic uniform spacing of natural flocks using a simple local alignment mechanism.

Clusters in the Vicsek model consistently exhibit uniform substructures; however, the present work is limited to instantaneous snapshots. While not explicitly studied here, as speed increases further, the spatial organization of a cluster is expected to restructure much faster with time due to (1) frequent collisions with other clusters and (2) the rapid shedding of individual particles [34]. Note that while giant number fluctuations characterize anomalously large density fluctuations at the system scale, the uniform organization we report emerges at the scale of the cluster and reflects local spatial regularities with suppressed density fluctuations, similar to those observed in “hyperuniform” states of matter [59,60]. A recent study on the Vicsek model adjusted to use active nematic alignment has shown that these seemingly opposite phenomena can appear when considering large and intermediate length scales [61].

Our analysis of the spatial structure and network topology of the Vicsek model clusters opens several promising avenues for future research. Here, we limit our analysis to static images of the cluster network. A critical next step may involve studying the behavior of cluster-cluster interactions in space and how they coalesce into the features measured in this paper. A compelling approach is taken by Peruani and Bär, where the authors utilize a coagulation-fragmentation framework to describe critical features in the Vicsek dynamics [27]. Future efforts may be focused on adapting the authors’ approach to elucidate the network structure exposed here. While our analysis of geometric features and network properties provides compelling evidence of a single-scale system-wide network with approximately uniform substructures, a rigorous mathematical proof of this

phenomenon remains an important open question. Future theoretical work can focus on developing analytic frameworks to establish how these arrangements emerge from local particle dynamics. Another important avenue is understanding the effect of different interaction rules on cluster formation, by revisiting Vicsek model variants through the lens of our approach.

V. METHODS

A. Simulation parameters and critical noise

We considered $N = 2^{15}$ particles for this study, and set the interaction radius $r = 1$ as the unit length, with time step length $\Delta t = 1$. The density of the periodic domain was varied $\rho \in \{0.1, 0.25, 1.0\}$ and the speed of the particles was varied $s \in \{0.1, 0.25, 1.0\}$. Accordingly, we had nine parameter sets of density and speed. The simulations and numerical procedures were performed using C++17 and MATLAB (version 2024b).

We calculated the critical noise for each parameter set by simulating the model and computing the order parameter. Starting from initial conditions of uniformly distributed particle positions and headings at $t = 0$, we ran 30 independent replicates for 10^6 time steps. For each replicate, the mean order parameter of the last 10^5 time steps was calculated. It is then confirmed as being in steady state by checking that the coefficient of variation of the mean order parameter for ten nonoverlapping consecutive windows of 10^4 time steps was $<10\%$. Incrementally adjusting η by steps of 0.01, we identified the critical noise η_c as the maximum noise that had a steady-state order parameter value ≥ 0.5 when averaged across replicates. We computed the critical noise up to two decimal places.

Figure 8(a) shows a sample phase transition plot of the steady-state order parameter varying η and Fig. 8(b) describes how the critical noise η_c depends on the model parameters. We notice that the particle speed s and density ρ both have strong effects on the critical noise value, with increase in either parameter resulting in a higher value of critical noise. This translates to maintaining order for higher values in the noise, rendering the system more robust to noise [21]. The values for η_c are reported in Table I and used for the simulation study.

For the 30 replicates at critical noise, we explored clustering in the interaction network. To consider systems in steady state, we simulated the dynamics for an additional 2×10^6 time steps, from which we sampled snapshots for our analysis.

B. Clusters in the interaction network

To analyze the clusters formed at steady state, we considered 201 snapshots starting from $t = 10^6$ and sampled every 10^4 time steps for all replicates. This sampling rate was determined by noticing that the mean autocorrelation of the order parameter time series was below e^{-1} for a 7×10^3 time step lag when averaged across all replicates [62]. As a result, the snapshots could be considered independent. Combining all replicates, $N_s = 6030$ snapshots were considered for each of the nine parameter sets. For each snapshot, edges were created between particles separated by less than the interaction radius. We note that this procedure can also be accomplished using

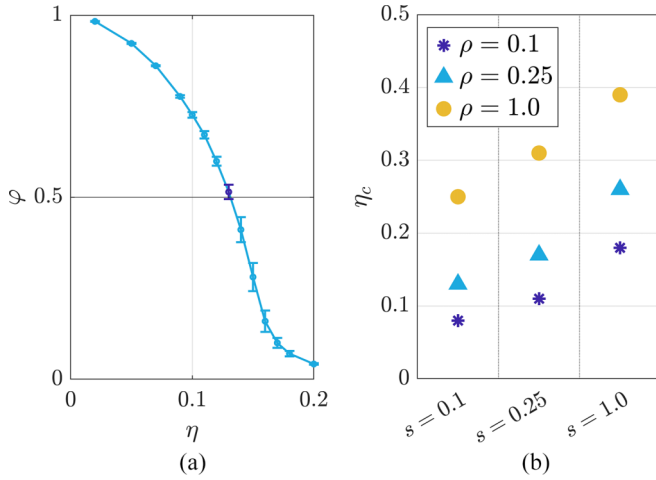


FIG. 8. The high-to-low order transition in the Vicsek model. We computed the critical noise η_c for changing speed s and density ρ . (a) The steady-state order parameter ϕ was computed for 14 values of $\eta \in (0, 0.2]$ for $\rho = 0.25$ and $s = 0.1$. These simulations are repeated for 30 replicates, and the critical noise η_c is determined to be the maximal η where $\phi \geq 0.5$. The dark blue marker represents the critical noise value and corresponding order parameter. (b) We then repeated this process for all combinations of $(\rho, s) \in \{0.1, 0.25, 1.0\}^2$, solving for the critical noise η_c in each case. We found that η_c increases with both ρ and s .

density-based spatial clustering of applications with noise (DBSCAN) [63]. The network construction is designed to produce time-varying networks, rather than time-integrated networks whose links accumulate, making the results independent of the selected number of snapshots or, analogously, simulation time.

We divided the possible cluster mass range (1 to $N = 32\,768$) in $B = 32$ logarithmically spaced bins. The probability of finding a cluster in bin b can be written as

$$P_B(b) = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{n_b(i)}{\sum_{j=1}^B n_j(i)}, \quad (4)$$

where $n_j(i)$ indicates the number of clusters in bin j in snapshot i . As we did not differentiate between different-sized clusters in the same bin, the CMD can be expressed as

$$P(m_b) = \frac{1}{w_b} P_B(b), \quad (5)$$

where m_b is the geometric mean of the edges of bin b and w_b is the number of integer values in the range of bin b . When plotting trends as a function of m_b in Sec. III, we omit subscript b for conciseness.

With this clear definition of clusters, we computed their geometric and topological features as functions of their mass. For these features, we balanced the relative influence of each snapshot by first averaging feature values for each bin within each snapshot, and then across snapshots. Bins were excluded if clusters were found in less than ten snapshots.

To avoid the disproportionate complexity of calculating geometric or topological properties for the clusters crossing one or more boundaries of the physical domain, clusters were excluded if any of their nodes corresponded to particles whose

position was $\leq r/2$ from the boundary. Nevertheless, a negligible fraction of clusters were ignored in this process.

C. Generating random geometric graphs

To compare clusters in the Vicsek model with a spatially constrained null model, we generated RGGs by uniformly distributing $N = 32\,768$ particles on a torus with length parametrized by the system-wide density, ρ . As RGGs do not possess any dynamics, density was the only parameter to consider when comparing the CMD to those computed over Vicsek model clusters. The comparison density, denoted by ρ_{RGG} , is selected by computing the CMD for densities in the range $\rho \in [0.1, 1.8]$ and by minimizing the mean logarithmic root-mean-square deviation between the RGG and Vicsek cluster mass distributions over their common mass range, considering only cases with sufficiently large maximum cluster mass for reliable overlap (see Sec. A of the Supplemental Material [49] for CMD for alternative densities). In this way, clusters of a given size emerge with similar frequency in both systems, allowing for a direct comparison between their associated trends.

D. Computing topological features of a cluster

The degree k_i of node i is the number of edges incident at the node. Mathematically, the average degree of a cluster of mass m can be defined as [51]

$$\langle k \rangle_c = \frac{1}{m} \sum_{i=1}^m k_i. \quad (6)$$

A high value of $\langle k \rangle_c$ indicates the nodes in the cluster are relatively well connected with each other. For a fully connected cluster, $\langle k \rangle_c \propto m^1$, and for sparsely connected clusters, which may have more pathlike topologies, $\langle k \rangle_c \propto m^0$. When performing analysis, we bin the cluster-specific average degrees by the mass ranges used to construct the CMD in Sec. VB. The mean of the average degrees in each bin, $\langle k \rangle$, is then used to extract trends in the inset of Fig. 4(b). To normalize the degrees for a given bin in Fig. 4(b), we consider the average of the degrees of all the particles in the clusters of said bin, $\langle k \rangle_p$.

The minimum number of edges between two different nodes i and j of the same cluster is the minimum path length ℓ_{ij} between them. For a cluster of mass m , there are $C(m, 2)$ distinct node pairs, where $C(\cdot, \cdot)$ is the binomial coefficient. Mathematically, the average path length of a cluster of mass m can be defined as [51]

$$\langle \ell \rangle_c = \frac{1}{C(m, 2)} \sum_{1 \leq i < j \leq m} \ell_{ij}. \quad (7)$$

A cluster with high $\langle \ell \rangle_c$ indicates a sparse or elongated network. Similar to the average degree, the cluster-specific average path length is binned using the mass ranges from the CMD. The mean of each is used to compute trends in Fig. 5(a).

The clustering coefficient of node i measures the density of connections among its neighbors. For a node of degree $k_i \geq 2$, the local clustering coefficient is defined as the ratio between the number of existing edges e_i among its k_i neighbors and

the maximum possible number of such edges. For leaf nodes, this value is zero. For a cluster of mass $m \geq 3$, the average clustering coefficient is [52]

$$\langle c \rangle_c = \frac{1}{m} \sum_{i=1}^m \frac{2e_i}{k_i(k_i - 1)}. \quad (8)$$

A cluster with higher value of $\langle c \rangle_c$ indicates a prevalence of triangular connections among nodes, while a zero value indicates a treelike network structure with no cycles among nodes. The average clustering coefficients are binned for each mass range in the CMD and used to compute trends in Fig. 5(b).

E. Computing geometric features of a cluster

The area and perimeter of a cluster, illustrated in Fig. 6(a), were calculated by creating an alpha shape from the nodes of the cluster with $\alpha = r/2$ [55]. The interior area A created by the boundary nodes of the alpha shape corresponds to the total area enclosed by the boundary edges. Nodes outside the alpha shape are either leaf nodes (nodes with a single edge) or articulation points (i.e., nodes with edges whose removal would create more connected components). Note that these exterior nodes do not contribute to either the perimeter or area calculations, as they represent structural extensions rather than enclosed regions. When articulation points are present within a cluster, multiple distinct alpha shapes are created. In

such case, we computed the cluster's total area as the sum of all individual alpha shape areas. The perimeter calculation follows a similar approach, where the physical length of the edges along the boundary nodes constitutes the total perimeter.

ACKNOWLEDGMENTS

This work was supported in part by funds from the US National Science Foundation (NSF: EF Grant No. 2222418). The work of M. Napoli was supported in part by a Mitsui USA Foundation scholarship. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources.

M.P. and N.A. conceptualized the work, acquired the funding, and supervised the project. A.R. created the software. A.R., M.N., O.L.B., M.P., and N.A. conducted the investigation and performed the formal analysis. A.R., M.N., and N.A. prepared the original draft. A.R. and M.N. carried out the visualization. A.R., M.N., O.L.B., M.P., and N.A. reviewed the manuscript.

DATA AVAILABILITY

The data that support the findings of this article are openly available [64].

-
- [1] I. Giardina, Collective behavior in animal groups: Theoretical models and empirical studies, *HFSP J.* **2**, 205 (2008).
 - [2] J. K. Parrish and L. Edelstein-Keshet, Complexity, pattern, and evolutionary trade-offs in animal aggregation, *Science* **284**, 99 (1999).
 - [3] D. J. T. Sumpter, *Collective Animal Behavior* (Princeton University Press, Princeton, NJ, 2010).
 - [4] L. A. Dugatkin, *Cooperation Among Animals: An Evolutionary Perspective* (Oxford University Press, New York, 1997).
 - [5] J. Krause and G. Ruxton, *Living in Groups, Oxford Series in Ecology and Evolution* (Oxford University Press, Oxford, New York, 2002).
 - [6] B. L. Partridge, The structure and function of fish schools, *Sci. Am.* **246**, 114 (1982).
 - [7] A. Cavagna, A. Cimorelli, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, R. Santagati, and F. Stefanini, New statistical tools for analyzing the structure of animal groups, *Math. Biosci.* **214**, 32 (2008).
 - [8] A. Cavagna and I. Giardina, Bird flocks as condensed matter, *Annu. Rev. Condens. Matter Phys.* **5**, 183 (2014).
 - [9] N. C. Darnton, L. Turner, S. Rojevsky, and H. C. Berg, Dynamics of bacterial swarming, *Biophys. J.* **98**, 2082 (2010).
 - [10] A. Be'er and G. Ariel, A statistical physics view of swarming bacteria, *Mov. Ecol.* **7**, 9 (2019).
 - [11] C. W. Reynolds, Flocks, herds and schools: A distributed behavioral model, in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (ACM Press, New York, NY, 1987), pp. 25–34.
 - [12] J. Toner and Y. Tu, Long-range order in a two-dimensional dynamical XY model: How birds fly together, *Phys. Rev. Lett.* **75**, 4326 (1995).
 - [13] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, Collective memory and spatial sorting in animal groups, *J. Theor. Biol.* **218**, 1 (2002).
 - [14] G. Gregoire and H. Chate, Onset of collective and cohesive motion, *Phys. Rev. Lett.* **92**, 025702 (2004).
 - [15] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, *Proc. Natl. Acad. Sci. USA* **105**, 1232 (2008).
 - [16] D. M. Gordon, *Ant Encounters: Interaction Networks and Colony Behavior* (Princeton University Press, Princeton, NJ, 2010).
 - [17] M. Porfiri, N. Abaid, and S. Garnier, Socially driven negative feedback regulates activity and energy use in ant colonies, *PLoS Comput. Biol.* **20**, e1012623 (2024).
 - [18] M. Napoli, R. Sipahi, and M. Porfiri, The role of colony size on activity rhythms of ants, *J. Theor. Biol.* **617**, 112286 (2026).
 - [19] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel type of phase transition in a system of self-driven particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
 - [20] F. Ginelli, The physics of the Vicsek model, *Eur. Phys. J.: Spec. Top.* **225**, 2099 (2016).

- [21] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, *Phys. Rev. E* **77**, 046113 (2008).
- [22] M. L. Rubio Puzzo, A. De Virgiliis, and T. S. Grigera, Self-propelled Vicsek particles at low speed and low density, *Phys. Rev. E* **99**, 052602 (2019).
- [23] C. Huepe and M. Aldana, Intermittency and clustering in a system of self-driven particles, *Phys. Rev. Lett.* **92**, 168701 (2004).
- [24] M. Aldana and C. Huepe, Phase transitions in self-driven many-particle systems and related non-equilibrium models: A network approach, *J. Stat. Phys.* **112**, 135 (2003).
- [25] M. Porfiri and G. Ariel, On effective temperature in network models of collective behavior, *Chaos* **26**, 043109 (2016).
- [26] R. M. Ziff, E. M. Hendriks, and M. H. Ernst, Critical properties for gelation: A kinetic approach, *Phys. Rev. Lett.* **49**, 593 (1982).
- [27] F. Peruani and M. Bär, A kinetic model and scaling properties of non-equilibrium clustering of self-propelled particles, *New J. Phys.* **15**, 065009 (2013).
- [28] L. Barberis, Emergence of a single cluster in Vicsek's model at very low noise, *Phys. Rev. E* **98**, 032607 (2018).
- [29] B.-M. Tian, H.-X. Yang, W. Li, W.-X. Wang, B.-H. Wang, and T. Zhou, Optimal view angle in collective dynamics of self-propelled agents, *Phys. Rev. E* **79**, 052102 (2009).
- [30] P. T. Nguyen, S.-H. Lee, and V. T. Ngo, Effect of vision angle on the phase transition in a flocking behavior of animal groups, *Phys. Rev. E* **92**, 032716 (2015).
- [31] S. Roy, M. J. Shirazi, B. Jantzen, and N. Abaid, Effect of visual and auditory sensing cues on collective behavior in Vicsek models, *Phys. Rev. E* **100**, 062415 (2019).
- [32] Y. Jia and T. Vicsek, Modelling hierarchical flocking, *New J. Phys.* **21**, 093048 (2019).
- [33] A. P. Solon, H. Chaté, and J. Tailleur, From phase to microphase separation in flocking models: The essential role of nonequilibrium fluctuations, *Phys. Rev. Lett.* **114**, 068101 (2015).
- [34] F. Peruani, L. Schimansky-Geier, and M. Bär, Cluster dynamics and cluster size distributions in systems of self-propelled particles, *Eur. Phys. J.: Spec. Top.* **191**, 173 (2010).
- [35] G. Baglietto, E. V. Albano, and J. Candia, Complex network structure of flocks in the standard Vicsek model, *J. Stat. Phys.* **153**, 270 (2013).
- [36] M. Barthélemy, Spatial networks, *Phys. Rep.* **499**, 1 (2011).
- [37] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2016).
- [38] J. T. Emlen, Flocking behavior in birds, *The Auk* **69**, 160 (1952).
- [39] M. Aldana, V. Dossetti, C. Huepe, V. M. Kenkre, and H. Larralde, Phase transitions in systems of self-propelled agents and related network models, *Phys. Rev. Lett.* **98**, 095702 (2007).
- [40] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003).
- [41] M. Nagy, I. Daruka, and T. Vicsek, New aspects of the continuous phase transition in the scalar noise model (SNM) of collective motion, *Phys. A* **373**, 445 (2007).
- [42] J. A. Pimentel, M. Aldana, C. Huepe, and H. Larralde, Intrinsic and extrinsic noise effects on phase transitions of network models with applications to swarming systems, *Phys. Rev. E* **77**, 061138 (2008).
- [43] V. Mwaffo, R. P. Anderson, and M. Porfiri, Collective dynamics in the Vicsek and vectorial network models beyond uniform additive noise, *J. Nonlin. Sci.* **25**, 1053 (2015).
- [44] P. Degond, J.-G. Liu, S. Motsch, and V. Panferov, Hydrodynamic models of self-organized dynamics: Derivation and existence theory, *Meth. Appl. Anal.* **20**, 89 (2013).
- [45] P. Degond and S. Motsch, Continuum limit of self-driven particles with orientation interaction, *Math. Models Meth. Appl. Sci.* **18**, 1193 (2008).
- [46] A. Peshkov, E. Bertin, F. Ginelli, and H. Chaté, Boltzmann-Ginzburg-Landau approach for continuous descriptions of generic Vicsek-like models, *Eur. Phys. J.: Spec. Top.* **223**, 1315 (2014).
- [47] T. Ihle, Kinetic theory of flocking: Derivation of hydrodynamic equations, *Phys. Rev. E* **83**, 030901(R) (2011).
- [48] T. Ihle, Invasion-wave induced first-order phase transition in systems of active particles, *Phys. Rev. E* **88**, 040303(R) (2013).
- [49] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/hkrb-wpdr> for additional numerical results.
- [50] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani, and F. Raynaud, Modeling collective motion: Variations on the Vicsek model, *Eur. Phys. J. B* **64**, 451 (2008).
- [51] A.-L. Barabási and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).
- [52] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature (London)* **393**, 440 (1998).
- [53] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, Classes of small-world networks, *Proc. Natl. Acad. Sci. USA* **97**, 11149 (2000).
- [54] S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, *Adv. Phys.* **51**, 1079 (2002).
- [55] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, On the shape of a set of points in the plane, *IEEE Trans. Inf. Theory* **29**, 551 (1983).
- [56] J. Dall and M. Christensen, Random geometric graphs, *Phys. Rev. E* **66**, 016121 (2002).
- [57] M. E. J. Newman, The structure and function of complex networks, *SIAM Rev.* **45**, 167 (2003).
- [58] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Empirical investigation of starling flocks: A benchmark study in collective animal behaviour, *Anim. Behav.* **76**, 201 (2008).
- [59] S. Dey, D. Das, and R. Rajesh, Spatial structures and giant number fluctuations in models of active matter, *Phys. Rev. Lett.* **108**, 238001 (2012).
- [60] S. Torquato, Hyperuniform states of matter, *Phys. Rep.* **745**, 1 (2018).
- [61] S. D. Cengio, R. Mari, and E. Bertin, Giant density fluctuations in locally hyperuniform states, *Phys. Rev. E* **112**, L042101 (2025).
- [62] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, San Francisco, CA, 1970).
- [63] H. Miyahara, H. Yoneki, T. Mizohata, and V. Roychowdhury, Vicsek model meets DBSCAN: Cluster phases in the Vicsek model, *J. Phys. Soc. Jpn.* **94**, 084002 (2025).
- [64] A. Rahman (2026). https://github.com/sqrrahman/Vicsek_interaction_network_critical_noise_geometry_and_topology.