

# A random graph model of density thresholds in swarming cells

Siddhartha G. Jena \*

Harvard College, Cambridge, MA, USA

Received: October 3, 2015; Accepted: October 30, 2015

## Abstract

Swarming behaviour is a type of bacterial motility that has been found to be dependent on reaching a local density threshold of cells. With this in mind, the process through which cell-to-cell interactions develop and how an assembly of cells reaches collective motility becomes increasingly important to understand. Additionally, populations of cells and organisms have been modelled through graphs to draw insightful conclusions about population dynamics on a spatial level. In the present study, we make use of analogous random graph structures to model the formation of large chain subgraphs, representing interactions between multiple cells, as a random graph Markov process. Using numerical simulations and analytical results on how quickly paths of certain lengths are reached in a random graph process, metrics for intercellular interaction dynamics at the swarm layer that may be experimentally evaluated are proposed.

**Keywords:** aquaporin-1 • membrane composition • water transport • cholesterol

## Introduction

Swarming behaviour refers to collective movement in a population of organisms and has been found to occur in individual cells, herds of cattle and flocks of birds [1, 2]; it is best described as a group of organisms moving purely through individual directives. Swarming can occur on two or three dimensions; however, for the purpose of this study, we will focus on populations of single cells, moving through intercellular interactions. A textbook example of this cell-cell interaction is swarming in bacterial colonies [3]. This process, in comparison to tissue formation and bird flocking, occurs largely on the two-dimensional plane of the media on which the cells grow, and much data have been collected on swarming and continue to be collected, as bacterial motility and its governing forces are of great interest in the laboratory and clinic. *Proteus mirabilis*, a gram-negative bacterium, displays swarming behaviour and has been demonstrated to form multicellular rafts of elongated and hyperflagellated swarmer cells [4]. This differentiation into the swarmer phenotype has been found to precede the swarming motility of the entire population of cells. Due partly to a large consensus that an understanding of swarming in bacteria could hold the key to understanding collective biological behaviour in larger cellular ensembles, a great number of mathematical models of swarming have arisen in recent years. However, since swarming comprises so many individual cellular and population-level behaviours, these mathematical models differ significantly in how effectively they are able to explain certain beha-

viours. For example, the stepwise process of swarming has been modelled as a series of differential equations [5, 6]. This model faithfully recreates the concentric 'swarm rings' that are the trademark of swarming behaviour, but tells us very little about the underlying dynamics occurring at each swarm front. However, this is not due to a lack of adequate modelling strategies; surprisingly little is known about these dynamics on the experimental level. Intercellular interactions are incredibly important and are required for a wide range of biological phenomena to occur. In some cases, particularly in bacterial swarming motility, these phenomena are dependent on local cell density [7]; the compounded effect of numerous identical interactions occurring within a subset of the population. In this study, we offer an explanation of bacterial swarming using graph theoretical methods.

If we represent the bacterial cells as discrete 'vertices' and the interactions between them as 'edges' or lines connecting these vertices, we have a simple framework for quantifying interaction-related characteristics of a population. Using this *graph theoretical* framework, which we will rigorously define later in this article, we can not only quantify the number of interactions in a given subset of a population but also measure how these quantities change as the graph evolves over time.

To better motivate the current investigation, let us take a brief detour into physics, turning to the tools that are used to understand phase changes in condensed matter systems. Specifically, we refer to a model in condensed matter physics called the random cluster model, which makes extensive use of a mathematical object called a *random graph*. This is a graph defined as a set of vertices, with an additional parameter  $p$  incorporated which represents the probability

\*Correspondence to: Siddhartha G. JENA  
E-mail: sjena@college.harvard.edu

that there exists an edge between any two vertices. This model is used extensively to represent interactions between particles on a dimensional lattice as a graph. Using random cluster models, materials such as metals, liquids and other ensembles of atoms or particles can be analysed to a surprising level of accuracy. This model can be applied to systems which are variations on the Ising model, in which particles are coupled on a lattice. Phase changes signify changes in states of matter and therefore in the underlying arrangement of particles in matter. In nature, these can be as simple and easy to observe as watching water freeze or ice melt; indeed, the phase diagrams for the more common materials have been in existence for quite some time. However, the atomic level interactions that lead to phase changes are a bit more subtle and hard to understand using conventional approaches. Yet, when placed under the lens of the relatively simple random cluster model, we find a great deal of connections between random graphs and measurable phenomena in physics. For example, many phase changes occur at a critical edge parameter of the random graph; in other words, when a certain level of connectivity  $0 < p_c \leq 1$  is achieved, the system as a whole will undergo a phase change [8]. For generalized systems, this critical point has been proven to be  $p_c = 0.5$ ; however, in our model, we will derive a specific result for our bacterial system.

We apply the above theory to our model of threshold cell densities, as a direct parallel can be drawn from assemblies of particles in an amorphous solid or other highly correlated condensed state to assemblies of cells which have a threshold density. A unique property of these graphs is that they maintain an understanding of the interactions between individual particles, while yielding important measures of critical phenomena such as phase changes in the material. In the present study, we use random graph models, often used to give analytical results of phase changes in condensed matter systems, to demonstrate the dynamics through which local density thresholds are reached. We treat a given cellular density threshold as a ‘connected component’ of a fixed size and work through an analytical result for the time taken to reach such a connected component in a random graph process.

## Mathematical formalism

### Graphs

*Definition:* A graph is described as a collection of vertices  $V = \{v_1, v_2, v_3, \dots, v_i\}$  and edges  $E = \{e_1, e_2, e_3, \dots, e_j\}$  where  $j \leq i$ , where an edge is simply a line drawn between two vertices. The presence of edges in a graph can be represented as a matrix with dimensions  $i \times i$ , where the existence of an edge between vertices  $v_n$  and  $v_m$  is represented by a positive non-zero value at  $(m, n)$ , which signifies the weight of the edge [9].

*Definition:* A graph can be *weighted* in that some of the non-zero values can be larger or smaller than others, which is useful for representing, for example, signalling networks where certain signals are stronger or more robust than others.

There are many measurable graph properties, but only several key measurable traits of a given graph  $G\{V, E\}$  that we will study in this

article are discussed here. They are *graph isomorphisms*, *supergraphs*, *subgraphs* and *connectedness*.

*Definition:* Two graphs  $G = \{V, E\}$  and  $G' = \{V', E'\}$  are *isomorphic* if there exist a pair of functions  $f: V \rightarrow V'$  and  $g: E \rightarrow E'$  such that  $f$  associates each element in  $V$  with exactly one element in  $V'$  and vice versa;  $g$  associates each element in  $E$  with exactly one element in  $E'$  and vice versa, and for each  $v$  in  $V$ , and each  $e$  in  $E$ , if  $v$  is an endpoint of the edge  $e$ , then  $f(v)$  is an endpoint of the edge  $g(e)$  [10].

*Definition:* A graph  $H\{V'', E''\}$  is a *subgraph* of the graph  $G = \{V, E\}$  if  $V''$  is a subset of  $V$  and  $E''$  is a subset of  $E$ .  $G$  is, therefore, the *supergraph* of  $H$ .

*Definition:* Two nodes  $v_1, v_2$  are *connected* if there exists an edge between them.

*Definition:* A graph  $G = \{V, E\}$  is *connected* if for any two  $v_i, v_j$  in  $V$ ,  $E$  is such that there exists a path between  $v_i$  and  $v_j$ .

There are various different types of graphs, but as we are not dwelling on theory but rather exploring the applications of graphs to a topic in biology [11], it will be more enlightening if we cover these graphs as they naturally arise.

### Markov chains

*Definition:* A *Markov chain* is a sequence of random variables  $X_1, X_2, \dots, X_n$ , that follow the property that the probability  $P(X_n = x_n | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$ . The countable set  $S = \{X_1, X_2, X_3, \dots, X_j\}$  is called the *state space* of the Markov chain [12].

In a set  $S$  containing a certain number of states, called Markov states, we can imagine a Markov chain as ‘jumping’ from state to state, while following pre-described probabilities for the transition between different states. What this means is that the process is non-deterministic. The matrix corresponding to the Markov chain is, as we would expect, a *stochastic* matrix, which means that for each state that the chain is in, the sum of the probabilities of getting to any of the other states is 1. Now, even though these structures are referred to as ‘chains’ in much of the literature, this simply refers to a 1-to-1 Markov process. Any process, however, that describes moving between states *via* a stochastic matrix in which the probabilities are non-deterministic can be described as a Markov process. Individual Markov states need not be quantities; rather, in this application, we are treating each state in the Markov process as a graph.

### Creating the model

Let us treat the time propagation of an assembly of cells as a random graph process, with individual cells represented as discrete vertices on the graph. Our objective is to find the time for the consolidation layer to form, given a certain interaction threshold  $m$ .

A *graph process* [11, 12] is a sequence of graphs on the same set of vertices,  $V$ . In our model, this means that the spatial arrangement of the vertices stays the same. We define the graph process with the definition from Bollobas [13]:

- Each  $G_t$  is a graph on  $V$  the set of vertices
- $G_t$  has  $t$  edges for  $t = 0, 1, \dots, N$ .
- $G_0$  is a subset of  $G_1$  which is a subset of  $G_2 \dots$  is a subset of  $G_{t_{\max}}$ .

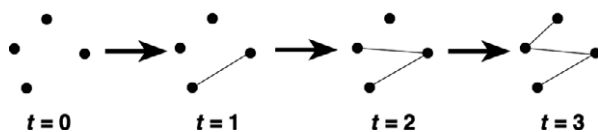
In most biological phenomena, the system is under a certain amount of ‘pressure’ from its surroundings to respond to a given stimuli within a certain timescale. Therefore, in this particular application of this model, we examine graphs with a certain  $t_{\max}$ , so that each random graph  $G(n, \alpha)$  can be defined by two indices:  $N$  the number of vertices in the graph and  $\alpha$  the probability that any two randomly chosen vertices in the graph are connected at  $t_{\max}$ . We next define the *hitting time*  $\tau$  of the given monotone property  $Q$ , a term of our own devising which is the time it takes to reach a certain property:  $T = \min\{t \geq 0 \text{ such that } G_t \text{ has } Q\}$ . These random graph processes can be treated as Markov chains, with the corresponding state space displayed in Figure 1 [15, 16].

It has been found that swarming bacteria of multiple species undergo a phenotypical differentiation during their swarming phase that results in swarming cells being hyperelongated in addition to being grouped in ‘raft’ structures [17, 18, 19]. These rafts consist largely of cells lying side by side, so that a rafting group of cells can be approximated as a network of cells that forms a long chain, with each cell connected to a maximum of two adjacent cells [20, 21]. This simplifies our search for a connected component, since the limitation placed upon the degree of connectivity of each cell refines our search to a connected ‘chain’ subgraph. We can show, using the definition of random graphs, that this property is indeed monotone; because  $G_0$  is a subset of  $G_1$  which is a subset of  $G_2 \dots$  is a subset of  $G_{t_{\max}}$ , the adjacencies that are present in  $G_t$  will carry over to  $G_{t+1}$  and so at any timestep the length of a given string of cells will either stay the same or will increase but will not decrease. We can utilize this particular property of random graph processes by assuming that the interactions leading up to the formation of a cell–cell raft component are stochastic but that once the component is formed it is permanent.

Having established the general parameters for our model and the biological assumptions that are being made in our calculations, we can approach the problem of reaching the cellular density threshold in several ways. These are described and explored here.

### Model 1

In our first, and most basic model, we consider the graph  $G_1(V, E)$ , abbreviated as  $G_1$ , where the set of vertices  $V = v_1, v_2, \dots, v_N$ . The graph process is then computed as follows:



**Fig. 1** Random graph process on 4 vertices. Each transition has a probability of 1, so achieving a certain result, *i.e.* a path of connected cells, is controlled by the randomness of the graph.

- 1 At each timestep, two random numbers  $1 \leq i \leq N$  and  $1 \leq j \leq N$  are generated from a uniform distribution.
- 2 If  $i = j$ , nothing happens.
- 3 If an edge already exists between  $v_i$  and  $v_j$  nothing happens.
- 4 If an edge does not already exist between  $v_i$  and  $v_j$ , an edge is created.
- 5 If either  $v_i$  or  $v_j$  is a node of degree 2 (the maximum degree of any node on our graph), nothing happens.

This graph process can be compared to an assembly of cells faithfully forming a new interaction at almost every timestep but assumes that a maximum of only one interaction is formed at each timestep. It also excludes the biologically meaningless outcome of a cell connecting to itself. Regardless of its limitations, it provides a good starting model of how we imagine the swarming process to occur. The matrix representing this graph is an  $N \times N$  square matrix, which we denote as the *adjacency matrix*  $M_1$ , with  $M_1(i, j) = 1$  when  $v_i$  and  $v_j$  are interacting and  $M_1(i, j) = 0$  when  $i = j$  or if  $v_i$  and  $v_j$  do not interact. If we numerically simulate this process and measure the average degree of all of the nodes  $v_1, v_2, \dots, v_N$  over time for a graph with  $N = 1000$ , we have the figure in (Fig. 2):

As we can see, as  $t \rightarrow \infty$ , the average degree of  $G_1(t)$  eventually becomes 2, our limit, as expected. The limit in the figure is taken over half of the adjacency matrix, which is symmetric over the diagonal. We can also measure the size of the largest connected component within our graph. We find that akin to descriptions of random graph processes, the giant component, once formed, ‘engulfs’ smaller connected components as time progresses, explaining the sudden increases in size that we see in our numerical simulation. Of course, in Model 1 if the simulation is allowed to continue, the size of the giant component approaches  $N$  (Fig. 3).

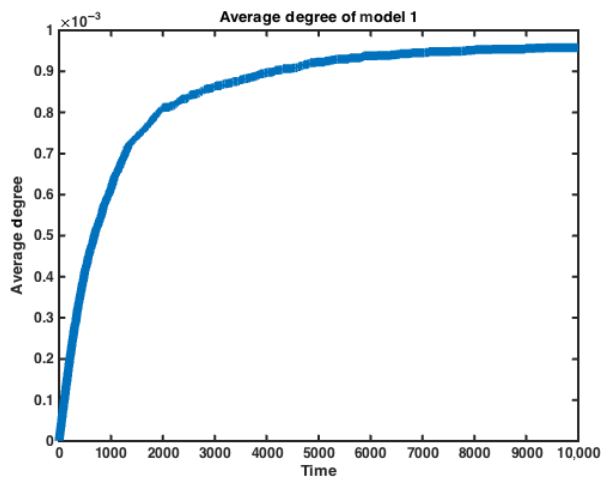
### Model 2

For our second model, we introduce a truly probabilistic element into our random graph process. Our algorithm for generating the graph process on  $G_2$  is similar to that for  $G_1$ , with an added step. This new algorithm also allows for a realistic development of the swarming colony; in experimental conditions, the colony is not limited to one inter-cellular interaction per timestep. At each timestep, the following process loops over all pairs of vertices  $v_i, v_j$  in  $V$ .

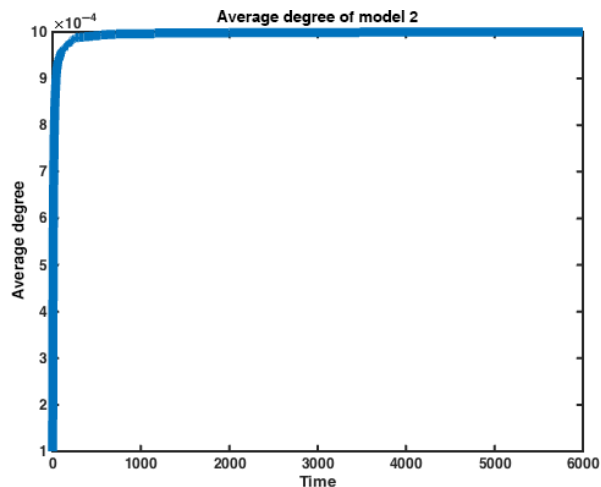
- 1 If  $i = j$ , nothing happens.
- 2 If an edge already exists between  $v_i$  and  $v_j$ , nothing happens.
- 3 If either  $v_i$  or  $v_j$  is a node of degree 2 (the maximum degree of any node on our graph), nothing happens.
- 4 If an edge does not already exist between  $v_i$  and  $v_j$ , an edge is created with probability  $P$ .

Again, we examine the average degree of Model 2 for convergence. All numerical simulations were run with  $N = 1000$  and with a probability of interaction  $P = 0.0001$  (Fig. 4).

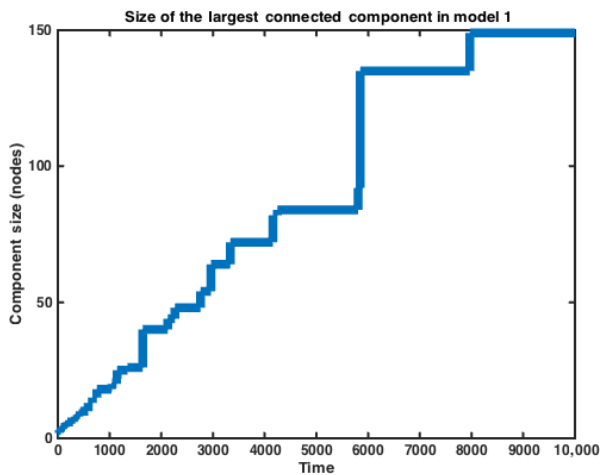
As we did in Model 1, we track the size of the largest connected component as well. Model 2 converges on the giant component equalling the entire vertex space much more quickly than does Model 1. We next examine the role of the interaction probability  $P$  on the rate of this convergence. As we may expect, the rate of convergence



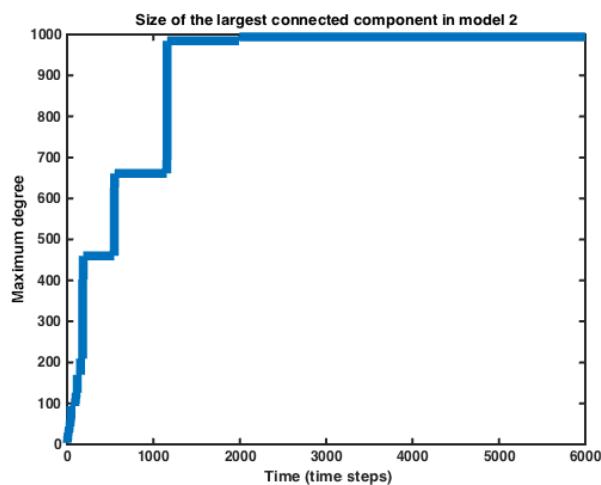
**Fig. 2** Average degree of nodes vs. time in Model 1. We can see that the average degree levels off as time approaches infinity. The average degree is measured over half of the adjacency matrix  $M_1$ . Data represent average of  $n = 10$  simulations.



**Fig. 4** Average degree of nodes vs. time in Model 2. We can see that the average degree levels off as time approaches infinity. The average degree is measured over half of the adjacency matrix  $M_2$ . Data represent average of  $n = 10$  simulations.



**Fig. 3** Size of the largest connected (giant) component vs. time in Model 1. Observations of random graph processes describe the 'engulfing' of smaller components by the giant component as time progresses, explaining the sudden increases in size of the giant component at  $t \approx 1700$  and  $6000$ . Data represent average of  $n = 10$  simulations.



**Fig. 5** Size of the giant component vs. time in Model 2. Data represent average of  $n = 10$  simulations.

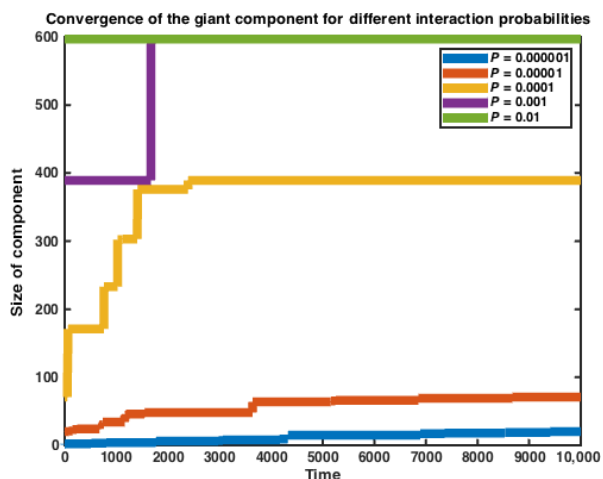
decreases as the probability of any two cells interacting decreases (Fig. 5).

Finally, we look at the distribution of connected component sizes at a fixed time. Mathematically, we can always extend our number of timesteps so that the size of the giant component approaches  $N$ . However, for biological systems, it is likely that there is some ideal time  $t_s \ll \infty$  after which the giant cell cluster is large enough that the population of cells begins to swarm. We, therefore, calculated the distribution of connected component sizes at time  $t = 500$  for a

population of size  $N = 1000$ . We can see that the distribution of component sizes is largely disparate when  $P \approx 0$  and the giant component gets larger as  $P \rightarrow 1$  (Figs 6 and 7).

### Model 3

A characteristic element of cellular simulations is a birth–death process. This is an additional biological assumption that is often brought into use when considering populations of cells. However, the significance of the birth–death process changes when we are looking at interactions between cells in addition to scalar quantities representing



**Fig. 6** Size of the giant component vs. time in Model 2, for different values of the interaction probability  $P$ . We can see that as  $P$  approaches 1, convergence occurs at faster rates. Data represent average of  $n = 10$  simulations.

connected components. Specifically, if a single cell dies, its connections to other cells are lost. We denote the probability of cell death to be  $p_D$  and construct a random graph process  $G_3(t)$  with the following governing algorithm, which loops over all pairs of vertices  $v_i$  and  $v_j$  in  $V_3$ .

- 1 If  $i = j$ , nothing happens.
- 2 If an edge already exists between  $v_i$  and  $v_j$ , nothing happens.
- 3 If either  $v_i$  or  $v_j$  is a node of degree 2 (the maximum degree of any node on our graph), nothing happens.
- 4 If an edge does not already exist between  $v_i$  and  $v_j$ , an edge is created with probability  $P$ .
- 5 Again looping over all cells in  $V_3$ , a cell  $v_k$  in  $V_3$  has a probability  $p_D$  of ‘dying’, which is represented by  $M_3(i, k) = M_3(k, j) = 0$  for all  $i$  where  $1 \leq i \leq N, i \neq k$ .

It should be noted here that:

- There are no explicit ‘birth’ processes in this algorithm. Because our system is completely defined by the vertices and edges present and because the number of vertices remains constant, a birth process would simply be an addition of edges with a certain probability, which would essentially be indiscernible from our stepwise edge addition.
- The above algorithm is not strictly that of a random graph process in the mathematical sense. We will recall that the requirements for a random graph process are that each graph is a subgraph of the one after it and as we are incorporating a certain probability of taking edges away in the present model, this is purely a stochastic simulation and the death process will not be considered in our analytical model in the following section.

We run the above algorithm for an assembly of  $N = 1000$  cells and can see that the imposed death rate prohibits

convergence of the giant component at  $t = 10,000$ . These results are very interesting in light of our results from Models 1 and 2, and contextualizing them with experimental results suggests that death rates in bacterial colonies are not necessarily prohibitive to convergence of the giant connected component, although the development of the component itself is due to intercellular interaction. In our ‘Experimental Directions’ section, we discuss the implications of this finding, both for the understanding of the underlying population structure of the swarming population and for our understanding of the nature of cellular interactions during swarming (Fig. 8).

## Analytical results

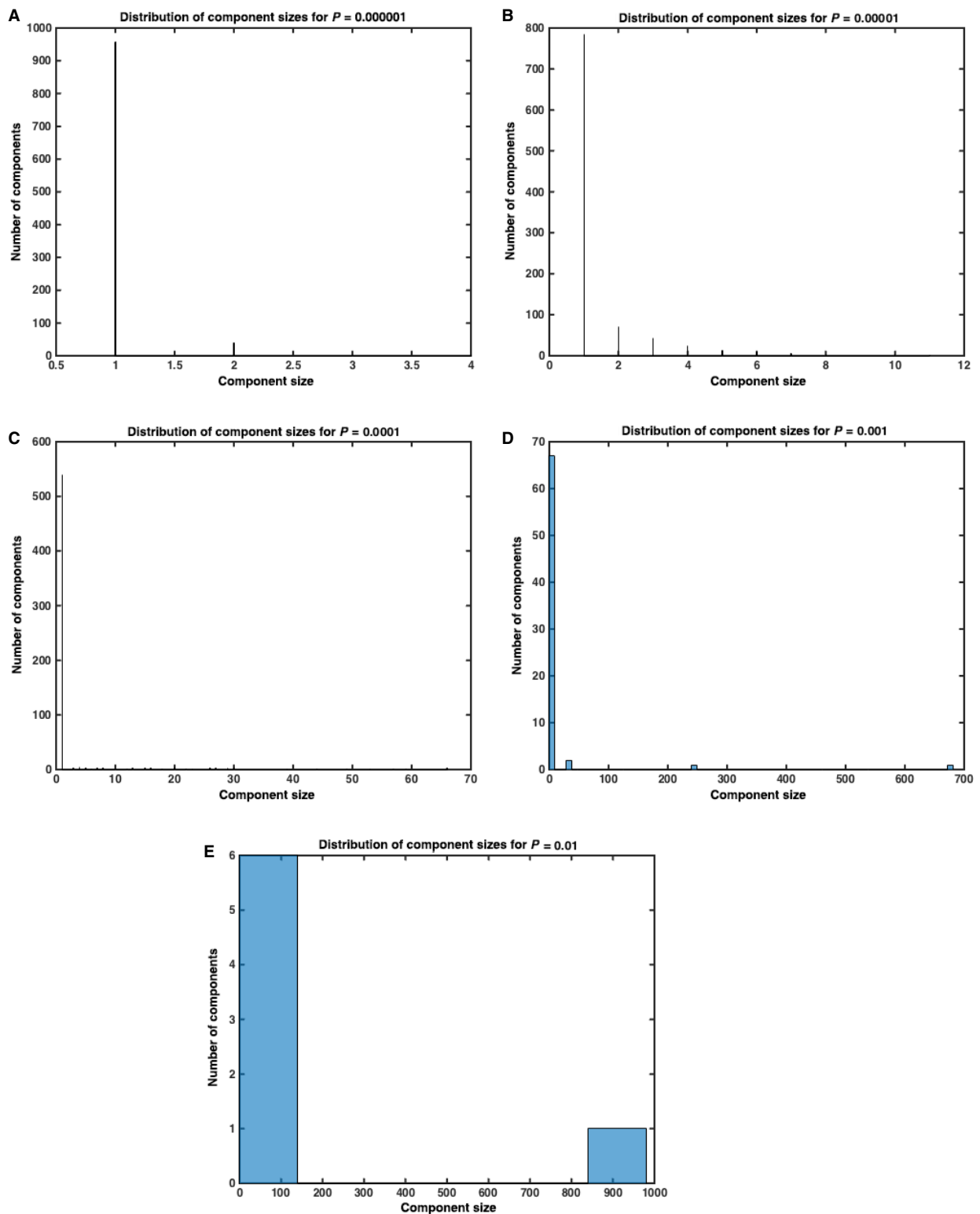
Although the above treatment is good as a general model for understanding the dynamics of swarming and possible parameters affecting this process, it benefits us to extend this model to some measurable characteristic of the bacterial population so that it may be tested. In this section, we give analytical results for the probability of reaching a connected component of a certain size and of maximum degree 2 in a population of cells, given a certain interaction probability.

Experimentally, it is observed that swarming bacteria form both tendrill-like, uneven ring and concentric ring patterns. We hypothesize that tendrill-like and uneven ring shapes in swarming colonies can be attributed to localized clusters of cells reaching their threshold connected component sizes at different times, whereas concentric circles are formed through synchronicity in achievement of the connected component. In other words, the distribution in radii measured around the colony is related to the interaction probability of the cells. Let us denote the cellular density threshold  $m$ , where  $1 \leq m \leq n$ . Bollobas [13] found that for the graph on the set of vertices  $V$ , almost every Erdos–Renyi random graph of this size contains a path of minimum length  $(1 - \alpha(c))^n$ , where  $\alpha(c)$  is an exponentially decreasing function of  $c$ , a measure of the edge probability of the random graph. We can change some of the parameters for the model to make it fit our application, but it requires that we think about the assembly somewhat abstractly. We take the total number of cells in the entire assembly to be  $N$ , and represent it as a sum of clumps of cells, which we can take to be equal. Let us choose the number of elements in each clump to be such that within the context of the structure of the ring, they are ‘dense clumps’ of vertices, so that for a given clump, a connection that makes biological sense (*e.g.* an interaction through cell–cell adhesion proteins) can be made between any two vertices in the clump.

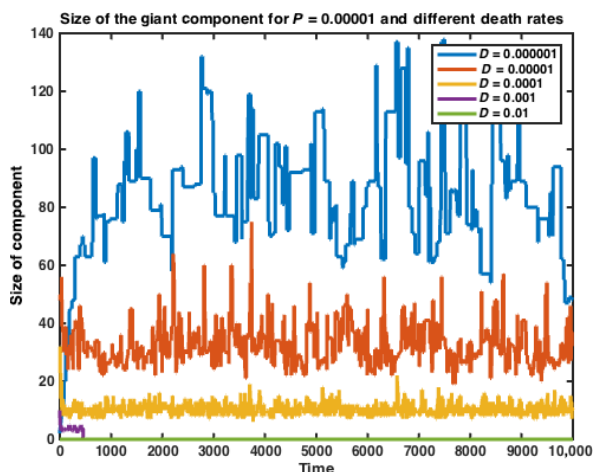
Let us say that within a given one of these clumps, the edges are chosen with probability  $c/n_i$ . If we have  $n_1 = n_2 = n_3 = n_i$ , we can simplify this to  $c/n$ . If we take into account the threshold  $m$ , we have

$$m \geq (1 - e^{-c})n \tag{1}$$

Because each edge has an equal chance of appearing, we will also assume that there are equal weights attributed to travelling along an edge, and once you reach the threshold probability, it is only a



**Fig. 7** Distribution of component sizes for  $P = 0.000001, 0.00001, 0.0001, 0.001, 0.01$ , respectively, after 500 steps of numerical simulation under Model 2. As expected, the components are largely disparate when  $P \approx 0$  and the giant component gets larger as  $P$  approaches 1. Data represent average of  $n = 10$  simulations.



**Fig. 8** Size of the giant component vs. time in Model 3, for different values of the death rate  $D$ . Data represent average of  $n = 10$  simulations. We can see that convergence does not occur within 10,000 timesteps and that for high death rates, the giant component does not have a chance to grow at all.

question of a Markov chain reaching time to reach your result. In this case, each step in the Markov chain is a stage in the graph process, and so we have this time being proportional to the inverse of the probability of reaching this step:

$$P = \left(\frac{c}{n}\right)^i \left(1 - \frac{c}{n}\right)^{v-i} \quad (2)$$

$$T \propto \left(\left(\frac{c}{n}\right)^i \left(1 - \frac{c}{n}\right)^{v-i}\right)^{-1} \quad (3)$$

Alon and Chung [22] demonstrate the following: For every  $\varepsilon > 0$  and every integer  $m \geq 1$ , there is a graph  $G$ , which can be explicitly constructed, with  $O\left(\frac{m}{\varepsilon}\right)$  vertices and maximum degree  $O\left(\frac{1}{\varepsilon^2}\right)$  such that after deleting all but  $\varepsilon$  portion of its vertices or edges, the remaining graph still contains a path of length  $m$ .

The graph described above (after  $0 < x < \varepsilon$  of the vertices have been removed) is essentially a random graph of the Erdos–Renyi model. To investigate the connections between these two graphs, we can draw a parallel between the  $\frac{\varepsilon}{n}$  in Eq. 2 and Eq. 3 and the  $\varepsilon$  in the second. From this, we have  $\frac{\varepsilon}{n} = \varepsilon$  and so we know that the graph in question has  $\frac{m}{\varepsilon}$  vertices and maximum degree  $\frac{n^2}{\varepsilon^2}$ . From Eq. 1, we have that

$$\frac{m}{n} \geq 1 - e^{-c} \quad (4)$$

$$c \geq \ln\left(1 - \frac{m}{n}\right) \quad (5)$$

The primary biological and mathematical question that is being asked in this study is that of randomness and probabilities within the biological network. We create a framework for understanding the

architecture of an interacting assembly of cells here. For a given cell, the probability that it will participate in an interaction that falls within the threshold  $m$  that was defined earlier is

$$P(t) \propto \binom{t}{j} \left(\frac{c}{n}\right)^j \left(1 - \frac{c}{n}\right)^{t-j} \quad (6)$$

where to make Eq. 6 physically realistic,  $c \geq |\ln(1 - \frac{m}{n})|$ . We have  $i = 1$  since after a chain of cells (the giant component) has formed we are done. We simply want the probability of the Markov chain hitting the intended result within a certain threshold time  $t$  which we can take to be measured in discrete steps.

We set  $c$  to be minimal, i.e.  $c = |\ln(1 - \frac{m}{n})|$ , so that we have now:

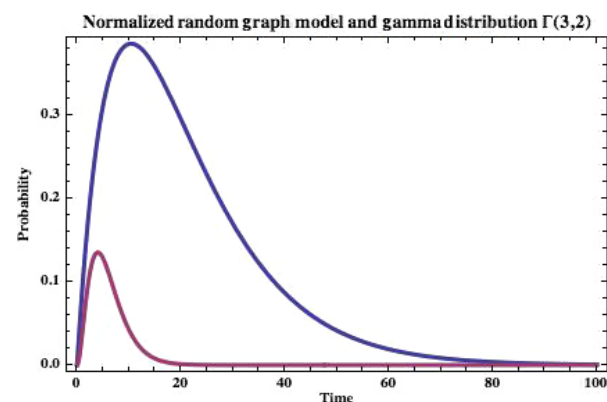
$$P(t) \propto t \left(\frac{|\ln(1 - \frac{m}{n})|}{n}\right) \left(1 - \frac{|\ln(1 - \frac{m}{n})|}{n}\right)^{t-1} \quad (7)$$

which is the lower bound on  $P$  for a given  $m$ ,  $n$  and  $t$ . Let us graph this result for  $1 < t < 100$ ,  $m = 6$  and  $n = 10$ . We can see that it assumes the general form of a gamma ( $\Gamma$ ) distribution (Fig. 9).

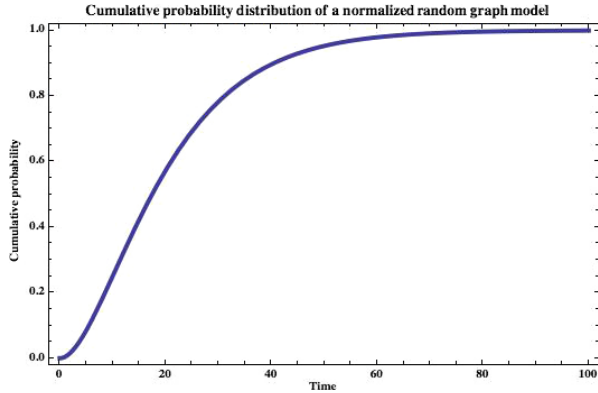
To observe asymptotic behaviour, we can represent our probability distribution function as a cumulative function (Fig. 10).

As expected, we can see that as the time is increased from  $t = 0$ , the probability of hitting the rafting threshold within that time increases as well, to a maximum. We can analytically find out where this maximum lies by taking the first derivative, which is

$$P'(t) \propto \frac{|\ln(1 - \frac{m}{n})| \left(\frac{n - |\ln(1 - \frac{m}{n})|}{n}\right)^{t-1}}{n} + \frac{t}{n} |\ln(1 - \frac{m}{n})| \ln\left(\frac{n - |\ln(1 - \frac{m}{n})|}{n}\right) \left(\frac{n - |\ln(1 - \frac{m}{n})|}{n}\right)^{t-1} \quad (8)$$



**Fig. 9** Our probability function for hitting a raft at a given time  $t$ . Graph shown is for  $m = 6$  and  $n = 10$ . The smaller graph shown is the gamma distribution  $\Gamma(3,2)$  for comparison.



**Fig. 10** The cumulative probability distribution function for hitting a raft at a given time  $t$ . Graph shown is for  $m = 6$  and  $n = 10$ .

and then setting this equal to 0, which yields

$$t_{\max} = \frac{-1}{\ln\left(1 - \frac{|\ln(1 - \frac{m}{n})|}{n}\right)} \quad (9)$$

As a coherence check, having a  $n$  value that is too large makes this value approach infinity, as we might expect. The function in Figure 9 resembles a gamma distribution, also shown. An interesting idea to present here is that we have not yet ruled out the possibility of our threshold being a function of  $t$  we can still see if this is the case. For some proportionality constant  $A$ , we represent our model in the form

$$A \frac{t^{k-1} e^{-t/\theta}}{\theta^k (k-1)!} = t(C)(1-C)^{t-1} \quad (10)$$

where

$$C = \frac{|\ln(1 - \frac{m}{n})|}{n} \quad (11)$$

Performing a series expansion of the two sides of the equation and matching the first-order terms (terms that are approximately proportional to  $t$ ), we have

$$-\frac{AC}{C-1} = \frac{\theta^{-k-2}}{2(k-1)!} \quad (12)$$

We solve for  $C$  to obtain:

$$C \approx \frac{t^k}{2A(k-1)!\theta^{k+2} + t^k} \quad (13)$$

and using our definition for  $C$ , we have:

$$\frac{-\ln(1 - \frac{m}{n})}{n} \approx \frac{t^k}{2A(k-1)!\theta^{k+2} + t^k} \quad (14)$$

$$m(t) = n \left( e^{\frac{m^k}{2A(k-1)!\theta^{k+2} + t^k}} - 1 \right) \quad (15)$$

When  $k \ll 1$ , we approximate  $t^k \approx 1$ :

$$m(t) = n \left( e^{\frac{n}{2A(k-1)!\theta^{k+2} + 1}} - 1 \right) \quad (16)$$

As we can see,  $m$  is in fact approximately a constant independent of  $t$  for small  $k$  and can be related to the values generated for gamma fits of experimental data by the above equation.

## Experimental directions

Having formulated a graph theoretical model for the dynamics of swarming bacteria, we turn our sights to the experimental domain. There are several ways in which to expand upon this model, which are briefly outlined in this section.

## Normalizing our analytical model

Although this model is enlightening in its explanation of trends in various swarming phenomena, it lacks very much practical application without at least a general understanding of the values that  $m$  and  $n$  can take, which can in turn help us understand our scaling factor  $A$ . With reference to papers seeking to find the density threshold of *P. mirabilis* swarming [19, 20], we can see that the ratio of swimmers to swimmers in a consolidation layer of *P. mirabilis* strain PM23 that is ready to swarm is roughly 1:1.5, which corresponds to  $\frac{m}{n} = 0.6$ . We plug in  $m = 60$ ,  $n = 100$  and our experimentally derived values for  $k$ ,  $\theta$  to get  $A \approx -3.06 \times 10^{-6}$ .

It is important to note that this is only an example of a value that  $A$  can take. Swarming behaviour is different from strain to strain of *P. mirabilis*, and so the same experiments on density would have to be performed to properly normalize the model for different *P. mirabilis* strains.

## Discussion

In the current study, we explore the use of random graph processes to find analytical times for attaining a threshold giant 'chain' component within a spatial cellular network, in which edges signify interactions between cells. We demonstrate a possible use for this model in interpreting density-dependent biological data and show that under a small gamma distribution parameter  $k$ , the interaction threshold for a certain population is independent of time and can be calculated from experimental data.

The inspiration for this model, as discussed previously, stems from the phase change threshold of connectivity within a random graph as described by random cluster theory in condensed matter physics. In the aforementioned model, an energetic consideration can be made, since we can simply assign an energy to different



types of interactions and compute the partition function of the entire graph to obtain a free energy measurement for the whole graph. The free energy measurement of massive particles such as cells in media cannot be generalized in this way to a high degree of accuracy; however, the emergence of certain density-dependent behaviours over time can be used to collect data to test the accuracy of the model.

It should also be noted that the phenomenon studied is similar to the emergence of the 'giant component' in a random graph process; however, this is a specific example of the evolution of a giant component in the graph. The evolution of a giant component in a random graph process occurs through the existence of a variable distribution of degrees of connectivity throughout the vertices in the connected component. Here, we attempt to apply this formalism to biological systems, in which cells and organisms have an upper bound on the number of connections they can make; a good example of this is through cell–cell adhesion in the cells of the stomach

lining, which occurs on the basolateral sides of cells rather than on the apical side facing the lumen of the stomach. Swarming is also an example of this process in that there is a definite upper limit on the degree of any vertex in a spatial population graph of bacterial cells. For this reason, the current study examines 'one-to-one' chains of interacting cells.

## Funding

This work was supported by a summer grant from the Origins of Life Initiative awarded to SGJ.

## Conflicts of Interest

The author has no competing interests to declare.

## References

1. **Bialek W, Cavagna A, Giardina I, et al.** Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci USA*. 2012; 109: 4786–91.
2. **Bull L.** On the evolution of multicellularity and eusociality. *Artif Life*. 1999; 5, 1–15.
3. **Coetsee J.** Transduction of swarming in *Proteus mirabilis*. *Microbiology*. 1963; 33: 1–7.
4. **Jones B.** Role of swarming in the formation of crystalline *Proteus mirabilis* biofilms on urinary catheters. *J Med Microbiol*. 2005; 54: 807–13.
5. **Ayali B.** A structured-population model of *Proteus mirabilis* swarm-colony development. *J Math Biol*. 2006; 52: 93–114.
6. **Ayali B.** Modeling the role of the cell cycle in regulating *Proteus mirabilis* swarm-colony development. *Applied Mathematics Letters*. 2007; 20: 913–8.
7. **Matsuyama T, Takagi Y, Nakagawa Y, et al.** Dynamic aspects of the structured cell population in a swarming colony of *Proteus mirabilis*. *J Bacteriol*. 2000; 182: 385–93.
8. **Grimmett G.** *The random-cluster model*. Berlin: Springer; 2006.
9. **Harary F.** *Graph theory*. Reading, MA: Addison-Wesley Pub; 1969.
10. **Franc A.** Metapopulation dynamics as a contact process on a graph. *Ecol Complex*. 2004; 1: 49–63.
11. **Lieberman E, Hauert C, Nowak MA.** Evolutionary dynamics on graphs. *Nature*. 2005; 433: 312–6.
12. **Norris J.** *Markov chains*. Cambridge, UK: Cambridge University Press; 1998.
13. **Bollobas Bela.** Long paths in sparse random graphs. *Combinatorica*. 1982; 2: 223–8.
14. **Bollobas Bela.** The evolution of random graphs—the giant component. *Random Graphs*. 184: 130–59.
15. **Luczak T.** Component behavior near the critical point of the random graph process. *Random Struct Algor*. 1990; 1: 287–310.
16. **Sood V, Redner S, Ben-Avraham D.** First-passage properties of the Erdos-Renyi random graph. *J Phys A-Math Gen*. 2005; 38: 109–23.
17. **Kearns D.** A field guide to bacterial swarming motility. *Nat Reviews Microbiol*. 2010; 8: 634–44.
18. **Darnton N, Turner L, Rojevsky S, et al.** Dynamics of bacterial swarming. *Biophys J*. 2010; 98: 2082–90.
19. **Dick H, Murray GE, Walmsley S.** Swarmer cell differentiation of *Proteus mirabilis* in fluid media. *Canadian J Microbiol*. 1985; 31: 1041–50.
20. **Copeland M, Weibel D.** Bacterial swarming: a model system for studying dynamic self-assembly. *Soft Matter*. 2009; 5: 1174–87.
21. **Esipov S, Shapiro J.** Kinetic model of *Proteus mirabilis* swarm colony development. *J Math Biol*. 1998; 36: 249–68.
22. **Alon N, Chung FRK.** Explicit construction of linear sized tolerant networks. *Discrete Math*. 1988; 72: 15–9.
23. **Rauprich O, Matsushita M, Weijer CJ, et al.** Periodic phenomena in *Proteus mirabilis* swarm colony development. *J Bacteriol*. 1996; 178: 6525–38.
24. **Steager E, Kim C, Kim M.** Dynamics of pattern formation in bacterial swarms. *Phys Fluids*. 2008; 20: 073601.