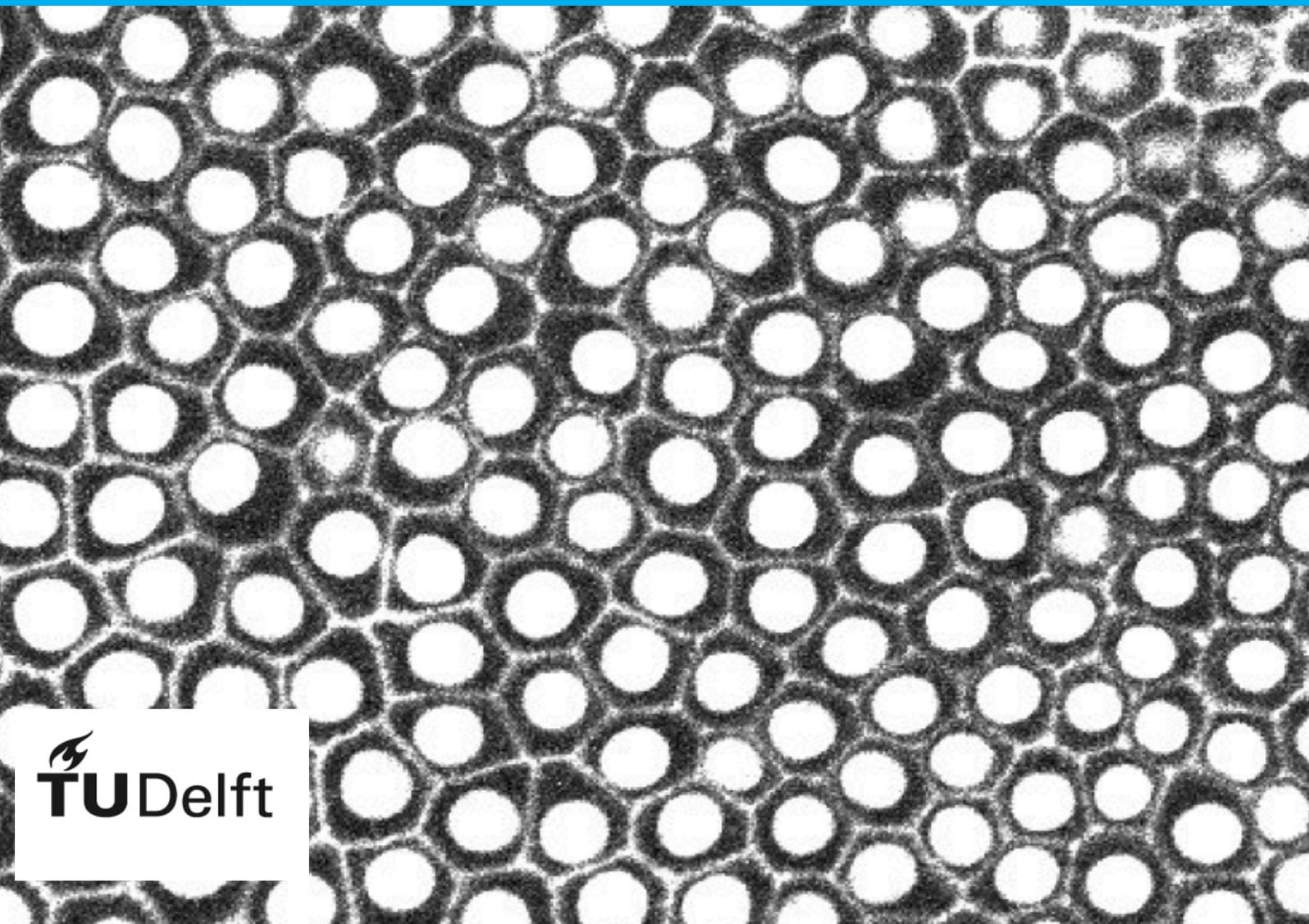


A graph theoretical approach to modelling growing cellular tissue

P. Kloet

Bachelor thesis



A graph theoretical approach to modelling growing cellular tissue

by

P. Kloet

to obtain the degree of Bachelor of Science
at the Delft University of Technology

Student number: 5172543
Project duration: February 13, 2023 – June 30, 2023
Supervisor: Dr. J. Komjáthy DIAM, Applied Probability
 Prof. Dr. T. Idema Bionanoscience, Idema group
Thesis committee: Prof. dr. J.M. Thijssen, Applied sciences, Thijssen group
 Dr. Ir. M. Keijzer, DIAM, Mathematical Physics

¹Source cover image: [1].

ABSTRACT. All organisms are built out of cellular tissue. Being able to recognise abnormalities in these tissues could be useful in recognizing cancerous cells. In this thesis we construct a mathematical model for cellular tissue based on its spatial structure. We consider cells as elements of the network. Touching cells are considered connected. Cells grow at different growth speeds. We determine the point in time when this network is fully connected, meaning there is a path between every pair of cells through touching cells. This point indicates the start of the last phase of the cellular growth, where friction restricts cell movement. We first use a Poisson point process to generate the locations of the cells. To make the model more similar to cellular tissue, we introduce determinantal point processes which have short-ranged repulsion, meaning points repel each other and thus spread. We compare the repulsion of Poisson and Determinantal point processes with real cellular tissue. We conclude that determinantal point processes have significantly higher repulsion than Poisson point processes. We also conclude that the repulsion in cellular tissue is higher than both point processes. Using simulations, we show that the model with determinantal point processes reaches connectivity significantly earlier than the Poisson model. We conclude that the analytically derived connectivity time point from the Poisson model can be used as an upper bound for the determinantal model.

CONTENTS

1. Introduction	3
1.1. Goal of this thesis	3
1.2. Representing networks mathematically	3
1.3. Graph theory	3
1.4. Spatial Boolean graphs	5
1.5. Point processes	5
1.6. Thesis overview	7
2. Mathematical model and main analytic results	8
3. Proofs	13
3.1. Conditional expectation of vertex degree	13
3.2. Probability of a vertex being isolated	15
3.3. Proof of properties of $g(w)$	15
3.4. Connectivity threshold	17
4. Simulations	22
4.1. Poisson simulations	22
4.2. Determinantal simulations	23
4.3. Physical model	26
4.4. Comparison Tribolium and spatial Boolean graphs	26
5. Results and discussion	29
5.1. Poisson point processes	29
5.2. Tribolium image analysis	30
5.3. Determinantal point processes	30
6. Conclusion	37
Appendix A. Proofs of claims	38
A.1. Proof of claim	38
A.2. Intuition for the average number of neighbours in a tiling of the 2d plane	38
A.3. Physical model	39
A.4. Confidence intervals	40
References	42

NON-EXPERT SUMMARY. In this thesis the goal is to construct a mathematical model for growing cellular tissue. We study the cellular tissue as a network. We take cells as elements of the network. A connection between two elements means those two cells are touching each other in the tissue. This together creates a big network. We determine the point in time this network becomes connected. Connected means that we can reach every cell from every other cell through touching cells. This connectedness indicates that the growth of the tissue is almost finished. Cells are already restricted by the cells around. This first version of the mathematical model, for which we have determined the connectedness, has the cells placed randomly. This causes some of them to be really close and some really far apart. In real cellular tissue the cells are spread quite evenly. Therefore we improve the model with better cell placement. We determine when this adapted model is connected and conclude that it is connected before the original model is. We compare both of the models to real cellular tissue and conclude that the cells in the real cellular tissue are still spread more evenly than both of our models.

1. INTRODUCTION

Cancer is the leading cause of deaths in the Netherlands [2]. Almost 30% of deaths in 2019 in the Netherlands were caused by cancer. The destructive property of cancerous cells is their rapid and uncontrolled cell growth and cell division. This makes the tumors, lumps of cancerous cells, increase in size and oppress the normal tissue. Non-cancerous cells only grow and divide when signaled by their surroundings to do so [3]. The difference between these growth characteristics results in quite different cell size distributions and therefore different cell structures. Intuitively smaller cells will have fewer neighbours and larger cells will have more neighbours. The different structural characteristics between healthy cells and cancerous cells might eventually be useful in recognizing cancer.

Cellular tissue can be modelled as a network. If we consider cells the elements of the network and neighbouring cells to be connected, we get a spatial network. In order to recognize the difference between different structural characteristics in cellular networks, it is useful to understand some properties of these networks first.

1.1. Goal of this thesis. In this thesis we introduce a mathematical model of a network. This model is already quite similar to the network of a growing cellular tissue. At time 0 the cells are point-like elements. They grow with different growth speeds. First, we analytically derive at which point in time this network is fully connected, meaning you can go from every cell to every other cell through touching cells. This point in time will depend on the parameters of the network. As this point in time indicates the transition between a disconnected and a connected network, it is a (mathematical) phase transition. This phase transition is useful in modelling cellular structures, because it indicates the start of the last phase of the cell growth. As cells are already touching other cells, friction prevents them from moving far. Second, we will make an adaptation to the model, which will improve its imitation of the spatial distribution of cells within a cellular tissue. Mathematically analysing the model with this adaptation is beyond this thesis. We compare simulations of both models. With these simulations we determine whether we can use the mathematically derived phase transitions in the adapted networks. Also we compare the original and adapted model to real cellular tissue to answer whether this adapted model is a good imitation of the real world.

1.2. Representing networks mathematically. Mathematically, we represent networks as graphs. Graph theory is a mathematical tool, which is useful for understanding networks. Examples of graphs are neural networks [4], or more generally brain networks [5]. Graphs are also used for representing protein-protein interaction networks [6]. In 1992, Godrèche, Kostov and Yekutieli made the connection between cellular tissues and graph theory, exploring the correlation between the number of neighbours of adjacent cells [7]. We now introduce some notions and literature on graphs used for modelling the networks of cells in a tissue. We also introduce the notion of a graph being connected.

1.3. Graph theory. Graph theory is the mathematical realm of studying networks. Graph theory is the study of network components and how they are connected. A mathematical graph consists of vertices and edges between those vertices. Vertices represent the network elements. Edges represent the connections between the elements in the network. We show an example of a graph in Figure 1. Graphs can represent a large range of things. Alongside the ones given above, examples range from friendships between people [8], food-webs [9], transportation networks [10], fungal networks [11] to topological knot theory [12].

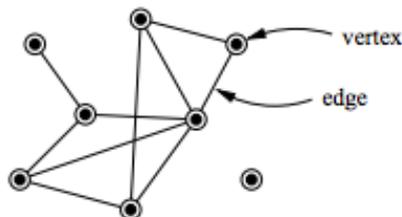


Figure 1. A simple example of a graph with 8 vertices and some edges between vertices. Source: [13].

1.3.1. *Directed graphs.* Connections in networks can represent a range of things. Therefore we can differentiate between different edges. One way of doing this is by stating whether an edge connecting vertex a to vertex b , also connects b to a . An edge going one way is called a directed edge. A directed edge from a to b does not connect b to a . A graph with directed edges is called a directed graph. Edges in an undirected graph or undirected network have no specific direction. They connect two vertices together. Directed graphs are used in modelling when one considers some interaction between the nodes, for example one cell, for example a neuron, sending some chemical signal to the other cell [14]. Here we are not concerned yet with the chemical interaction between cells. So for our purposes undirected graphs for modelling will suffice.

1.3.2. *Connected graphs.* An important property is the connectedness of graphs. First we need to define the notion of a path between vertices. A path between vertex a and vertex b is a sequence of edges which starts in a and ends up in b whilst going through other vertices. The existence of a path means there is a route between them. A graph is connected when between every pair of vertices a path exists. A graph can be not connected in multiple ways. An example could be Figure 1, which is not connected because there is one isolated vertex, a vertex with no connections. A graph could also be split into two larger connected parts which have no connections between them.

1.3.3. *Random graphs.* When a network is large, it may be impractical or infeasible to reconstruct it exactly. But we are interested in macroscopic properties of graphs. Connectedness is such a macroscopic property. There are way more microscopic interactions than macroscopic properties in the networks we are working with. So instead of reconstructing large networks, we can build models using randomness to approximate reality. In our case the randomness will be in the initialization of the locations of the cells and the growth speeds of the cells. In the construction of these graphs, we need to define some process which determines whether two vertices have an edge between them. This process could be random, like the Erdős-Rényi model. This random graph was introduced by Paul Erdős and Alfréd Rényi. It takes some probability p as input. For each pair of vertices, it constructs an edge between them with probability p . The construction of edges in these graphs is a pure probabilistic process. For Erdős Rényi graphs are proven to be connected when the probability p of vertices being connected grows faster than $c \log n/n$ [15]. In this equation n is the number of vertices in the graph and c is any constant, meaning p should grow faster than $\log n/n$ by more than just a constant factor. When p grows slower than $\log n/n$ by more than a constant, the graph is disconnected. Graphs can be extended with vertex properties. Using these vertex properties in the construction of graphs makes them appropriate for modelling real-life networks.

1.3.4. *Spatial graphs.* One possible vertex property is a location. This is useful when modelling cellular tissue, because that is a spatial network. When all vertices have some coordinate within a space, such a graph is called a spatial graph. The spatial graph usually implements a spatial dependency in the construction of edges. This usually results in vertices being more likely to connect when they close versus when they are far apart. An example of a spatial graph is a Gilbert disc model. Gilbert introduced this graph, which he called random plane networks, in 1961 [16]. This graph uses an infinite Poisson point process that generates vertex locations, see Section 1.5 below. Two vertices are connected by an edge if they are separated by a distance less than R , for some fixed $R > 0$. Connectivity for the Gilbert disc model was also proven in a finite box or torus, [17, 18, 19]. In this case the graph is connected with some probability close to one is $r \propto c\sqrt{\log n + \alpha}$.

1.3.5. *Weighted graphs.* Faloutsos, Faloutsos and Faloutsos determined that the topology of the internet follows a power-law, $\mathbb{P}(D \geq z) \propto z^{-(\tau-1)}$ where D is the number of connections of a certain internet page [20]. This means that there are hubs in the internet, which have a very large number of connections. This inhomogeneity can be modelled in a random graph with vertex weights. These hubs have a higher weight, which means they connect to more other vertices. Weight might also be called fitness. Weights represents the desire of a vertex to form edges with other vertices. Vertices with high weights are likely to have many more edges than vertices with a lower weight. These weights are oftentimes chosen from some probability distribution, either discrete or continuous. Other examples of weighted graphs are the Chung-Lu graph introduced in 2001, [21, 22] and the Norros Reitu graph introduced in 2006 [23]. In this thesis, the vertex weight will represent the speed of growth of the cell.

1.4. Spatial Boolean graphs. A cell structure can be transformed into a network. In this graph the vertices represent the cells. Edges between vertices will represent two vertices being neighbours, thus the edges are undirected. Vertices have both a coordinate and a growth speed (weight). Hence we have a time-dependent weighted spatial graph. Christian Hirsch introduced a graph with these structural properties in 2017 [24]. This graph was called a Poisson Boolean model or a scale-free Gilbert graph. This graph constructed a vertex set using a Poisson point process (PPP). The points in the PPP are the locations of the vertices. We show an example of a Poisson point process in Figure 3b. We introduce the Poisson point process mathematically in Section 1.5. He chooses the weights from a power-law distribution with parameter $\tau - 1 > 0$, see the section above. Two vertices are connected if one contains the other within its radius. We will use a generalised version of this model, which we will call a spatial Boolean graph. The generalisation of these graphs comes from switching the power-law distribution for generating growth speeds from some arbitrary probability distribution. Power-law speeds will result in power-law degree distributions, but in our data the degree distribution is definitely not heavy tailed, for this see Section 5.2.

A spatial Boolean graph requires five ingredients to be constructed:

- (1) A **graph size** n : an integer $n \in \mathbb{N}$ determining the size of the periodic space the graph exists in.
- (2) A **base space** \mathbb{T}_n^2 : a 2-dimensional space with size parameter n , constructed as a square $[0, \sqrt{n}]^2$, with a periodic norm. Note this square has volume n . This space is referred to as the torus, 2-dimensional torus or \mathbb{T}_n^2 .
- (3) A **vertex set** \mathcal{V} : a Poisson point process (PPP) Ξ with intensity parameter $\lambda = 1$ on the torus \mathbb{T}^2 .
- (4) A **speed distribution** W for vertices, w_v being the speed of growth of a vertex.
- (5) A **time point** t , the point in time we want to construct the graph for. We can calculate the radius of a vertex v with growth speed r_v by $r_v = w_v \cdot t$. We actually work with the square of the time, $\beta = t^2$. This simplifies the calculations and it is the mathematical convention for these graph types.

We need to define a connectivity function, which tells whether 2 vertices are connected, given their location. This is the following equation:

$$\mathbf{Prob}(\text{vertex } u \text{ is connected to vertex } v \text{ by an edge}) = \mathbb{1}\{t^2 [\max(r_v, r_w)]^2 \geq \|u - v\|^2\} \quad (1.1)$$

Conditionally on the vertex set and on the radii $W_v = w_v$, $W_u = w_u$, two vertices u and v are connected by an edge independently if the maximum of the radii is greater than the distance between the vertices. This is also illustrated in Figure 2. This connectivity function is often found in research on wireless networks [25]. In our case we would like to look at the sum of radii. The maximum function is a mathematical simplification for the sum. An argument for why this is reasonable is given in (2.2).

These ingredients together define a spatial Boolean graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where \mathcal{E} is the set of edges produced by the connectivity function. In the first part of this thesis we are going to look at connectivity for these graphs. Mathew Penrose proved that the absence of isolated vertices in these types of graphs implies connectivity [26]. Isolated vertices are defined as vertices which have no connections with other vertices. Larger isolated clusters like isolated pairs of cells, or three cells, disappear already at an earlier time, i.e. they are less likely to occur at a fixed time t . Larger isolated clusters would need to have no connections around their entire perimeter. This is much less likely than a single vertex being isolated.

In the second half of this thesis we will explore graphs where the vertex set is not a Poisson point process but a so-called determinantal point process. This may be a better choice for modelling cell tissue, because one can build in short-range repulsion between cells naturally in the point process. We show both a Poisson point process and a determinantal point process in Figure 3.

1.5. Point processes. An important difference between spatial Boolean graphs and real-world cells, is the fact that vertex locations in spatial Boolean graphs are fixed. Real cells exert forces on each other and respond to these forces by moving around. This difference means that the method for choosing the vertex locations in the spatial Boolean graphs is quite important. Mathematically a method of producing these vertex locations is called a point process. Usually spatial Boolean graphs use a Poisson point process.

1.5.1. Poisson point process. The Poisson point process (PPP) Ξ is a process that describes the location of a random point-set in \mathbb{R}^2 , or a subset $B \subset \mathbb{R}^2$. For any set $A \subset \mathbb{R}^2$, the number of points, i.e. $|\Xi \cap A|$, is a Poisson random variable with mean $\text{Vol}(A)$, the volume of A . If two sets A and A' are disjoint in

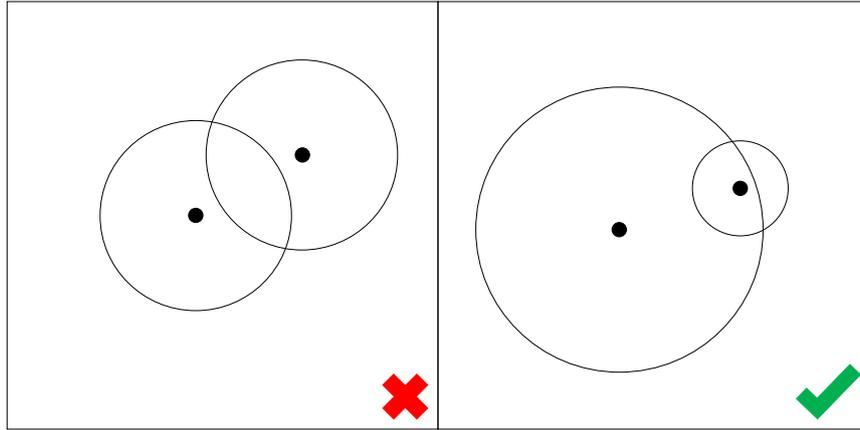
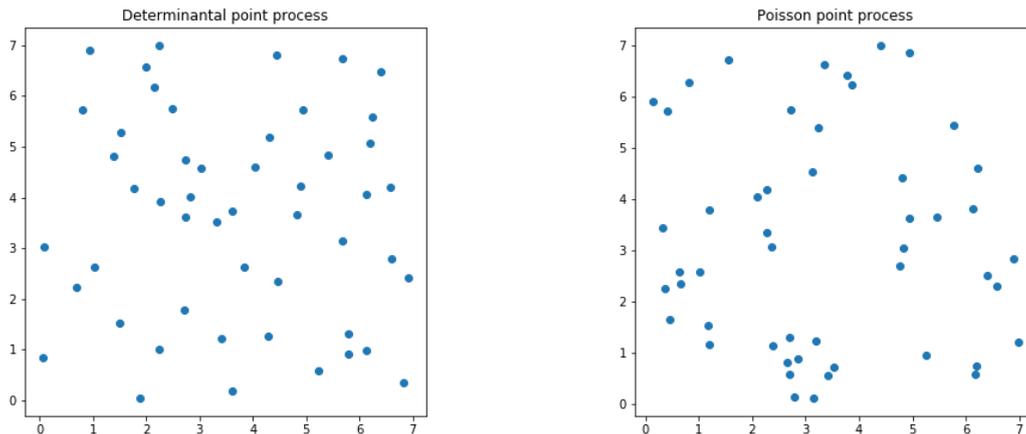


Figure 2. 2 examples of the behaviour of the connectivity function (1.1). In the left example, both vertices are not within the circle of the other vertex. Hence these vertices do not meet the connectivity condition. Therefore there will be no edges between these vertices. In the right example, the top-right vertex is within the circle of the other vertex. Therefore these vertices do meet the connectivity condition. Hence there will be an edge between them.



(a) A realisation of a determinantal point process with 50 points on a square of area 50.

(b) A realisation of a Poisson point process with intensity 1 on a square with area 50.

Figure 3. Two different point processes. (A) is a determinantal point process. (B) is a Poisson point process. Notice the points in (A) are much more evenly distributed.

\mathbb{R}^2 , $|\Xi \cap A|$ and $|\Xi \cap A'|$ are independent. One can prove that sampling a PPP in a torus or box is equivalent to first sampling a number $N \sim \text{Poi}(\text{Vol}(A))$, choosing the location of each of the N points independently and uniformly at random. Generating points this way leads to densely populated areas and areas where points are very scarce. In Figure 3b we show a realisation of a Poisson point process.

1.5.2. *Determinantal point process.* In cellular tissue, the centers of cells have a certain short-ranged repulsion. This is caused by cells exerting forces on other cells and responding to those forces. A Poisson point process has no internal repulsion. Points are chosen independently. Therefore we can improve the similarity of the model to reality by choosing points differently. A point process where internal short-ranged repulsion can be implemented is a Determinantal point process. This point process chooses points with a more even spread. In Figure 3a we show a realisation of a Poisson point process. We give a mathematical definition in Section 4.2 below.

1.6. Thesis overview. First, we will determine under which conditions spatial Boolean graphs are connected. More specifically we will determine how β , i.e. t , should depend on the graph size n and the growth speed distribution W . Apart from the interest of connectedness for cellular tissue as discussed in Section 1.1, this property is mathematically interesting in itself. It has not been investigated for these graphs.

Secondly as explained above, we will adapt the spatial Boolean graphs with the determinantal point process. We will measure internal repulsion for Poisson point processes, determinantal point processes and for a real cellular tissue and compare these. From this we conclude whether the determinantal point process has similar repulsion as cellular tissue. And finally, we explore and simulate both point processes and the according graphs. With these simulations we do two things. First, we confirm our results of the connectedness of spatial Boolean graphs. Second, we compare the phase transition for the Poisson and determinantal graphs. We answer the question whether the mathematical results found for Poisson graphs are useful and applicable in determinantal graphs.

In Section 2 we give the mathematical results. Section 3 contains the proofs of these results. Section 4 describes the simulation methods. This section also contains the theory about determinantal point processes. In Section 5, the results of the simulations are illustrated and discussed. Section 6 makes a conclusion about the comparison between the mathematical results and the simulations, both with a Poisson point process and with a internally repulsive determinantal point process. This thesis is the end result of my bachelor project which finishes the double bachelor Mathematics and Physics.

2. MATHEMATICAL MODEL AND MAIN ANALYTIC RESULTS

We start by defining the spatial random graphs with our specific connection function as given in (1.1). Note that we obtain a periodic interval by connecting end-points. Hence the periodic distance of $x, y \in [0, \sqrt{n}]$, becomes $|x - y|_p := \min\{|x - y|, 1 - |x - y|\}$. For the graphs in this thesis, we consider as ground space $\mathbb{T}_n^2 = \sqrt{n} \cdot \mathbb{R}^2 / \mathbb{Z}^2$, which is the 2-dimensional box $[0, \sqrt{n}]^2$, where opposite boundaries are connected. As distance function we use the L^2 -norm on \mathbb{T}^2 , i.e. for $x, y \in \mathbb{T}^2$ we define $\|x - y\| := \left[(x_1 - y_1)^2 + (x_2 - y_2)^2 \right]^{\frac{1}{2}}$. Now that we have a space with a norm, we can define the graph.

Definition 2.1 (Spatial Boolean graph). *Fix $n \in \mathbb{N}$, $\beta > 0$. Take a probability distribution W . Let Ξ_n denote a Poisson point process with intensity $\lambda = 1$ on \mathbb{T}^2 . For each vertex $v \in \Xi_n$ produce a weight w_v from R . We define the connectivity function as*

$$p((x_u, w_u), (x_v, w_v)) = \mathbb{1}\{\beta(\max(w_u, w_v))^2 \geq \|x_v - x_u\|^2\}. \quad (2.1)$$

Conditionally on Ξ_n , the random graph $\mathcal{G}_n = (\mathcal{V}(\mathcal{G}_n), \mathcal{E}(\mathcal{G}_n))$ is given by $\mathcal{V}(\mathcal{G}_n) = \Xi_n$ and for each pair $u = (x_u, w_u), v = (x_v, w_v) \in \Xi_n$ the edge $\{u, v\} \in \mathcal{E}(\mathcal{G}_n)$ with probability $p((x_u, w_u), (x_v, w_v))$; conditionally independently given Ξ_n .

Remark 2.2 (Time interpretation of β). In these graphs, β can be interpreted as time. But with $\beta = t^2$. The radius of a vertex v at time $t = \sqrt{\beta}$ would be $r_v = w_v \cdot t$. This interpretation connects a phase transition in the parameter β to a point in time.

Note that the probability distribution W producing the growth speeds for the vertices, has to produce non-negative values. In this paper, we will look at both discrete and continuous distributions. Some further restrictions are going to arise in order to obtain β threshold values.

Also note that this connection function is mathematically very similar to the case of overlapping radii. In our case one vertex needs to be in the radius of the other in order to connect them. But the following holds,

$$\max(w_v, w_u) \leq w_v + w_u \leq 2 \max(w_v, w_u). \quad (2.2)$$

This factor 2 is mathematically not significant. In some applications like wireless networks, the maximum would seem more reasonable. But in cell structures, the sum would seem a better choice. But mathematically the maximum is way easier to use. And the existence of different macroscopic regimes will not change. Just the exact boundary will vary slightly.

The goal is to say something about the existence of isolated vertices in these graphs, because like in [26] this might result in a statement about the connectivity of these geometric random graph types. Particularly we would like to know the existence of isolated points for very large amounts of points, thus we would like to investigate the existence as the size goes to infinity. Please note the density of points does not increase, only the size of the torus on which the graph lives. Even more specifically, what we would want is a value for β , which could depend on the size of the graph, above which there would be no isolated vertices and below which there would. In order to say something about the existence of isolated points, it would be interesting to know the probability of a single vertex v , with some radius r_v , being isolated. In order to get to this, we will first need to know the expected degree of such a vertex. In the discrete case there is a nice geometric intuition. For this intuition, 4 shows the geometry we need to have in mind. Suppose we have a vertex v with radius r_v . This implies all vertices radius r_v or smaller connect to v if they are within a radius of r_v to v . Vertices with the next larger radius after r_v , say $r_1 > r_v$ connect to v within a circle of radius r_1 . Vertices of radius $r_2 > r_1$ connect to v if they are within a circle of radius r_2 , and so on. This allows us to split the expected degree into the different discrete radii. For a graph like in Definition 2.1, we give the expected degree of a vertex in the following lemma.

Lemma 2.3 (Expected degree of a vertex conditioned on its radius). *Fix $n \in \mathbb{N}$ and $\beta > 0$. Take W a discrete or continuous probability distribution. With these parameters, construct a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ as given in Definition 2.1. Take a vertex $v \in \mathcal{V}$ arbitrary. There exists some $M \in \mathbb{N}$ such that if $n > M$, the expected degree of v conditioned on its radius r_v is,*

$$\mathbb{E}[D_v | w_v] = \pi\beta\mathbb{E}\left[\max(w_v, W)^2\right] \quad (2.3)$$

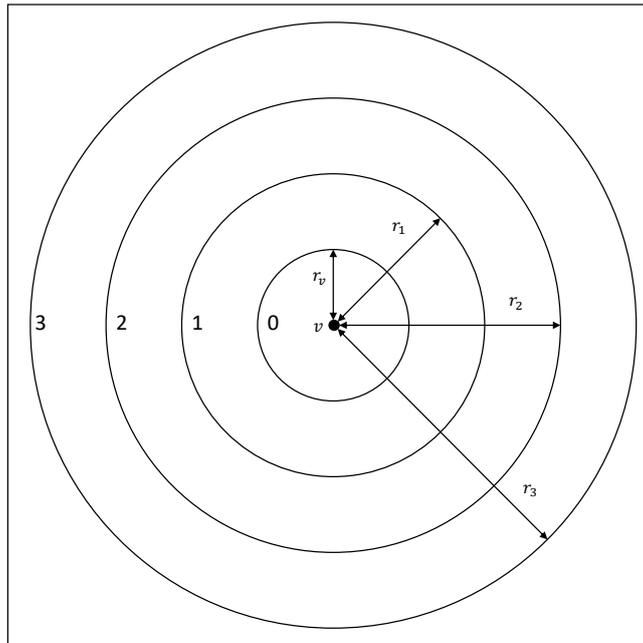


Figure 4. The geometry that is useful for the intuition on how to calculate the expected degree of a vertex. In this figure we are working with a discrete growth speed distribution and a certain point in time t . The figure shows a vertex v with radius r_v . $r_1 < r_2 < r_3$ are the first 3 radii larger than $r_v = w_v \cdot t$ in the distribution. All vertices with radius r_v or smaller connect to v if they are within circle 0. All vertices with radius r_1 connect to v within a radius r_1 . All vertices with radius r_2 connect to v if they are within a radius r_2 , and so on.

Remark 2.4 (Expected degree of a vertex increasing with conditioned radius). Please note that the RHS of (2.3) is increasing with w_v . Therefore $\mathbb{E}[D_v | w_v]$ is increasing as well.

Having a larger radius, means it is better at connecting to other vertices. Therefore Remark 2.4 follows quite intuitively. Next we give an example of this lemma to make it a bit more intuitive. In this example a continuous uniform distribution is used. Example 3.1 in Section 3 gives another example with a discrete distribution.

Example 2.5 (Conditional expectation of vertex degree for a continuous distribution). *Take Unif $([0, 1])$ as continuous probability distribution. Hence,*

$$f(w) = \begin{cases} 1 & \text{if } w \in [0, 1], \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

Let us calculate the expected degree of a vertex with weight w_v by using Lemma 2.3:

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \pi\beta \mathbb{E}[\max(w_v, W)^2] \\ &= \pi\beta \int_0^\infty f(w) [\max(w_v, w)]^2 dw = \pi\beta \int_0^1 [\max(w_v, w)]^2 dw \\ &= \pi\beta \left[\int_0^{w_v} w_v^2 dw + \int_{w_v}^1 w^2 dw \right] = \pi\beta \left[w_v^2 + \left[\frac{1}{3} w^3 \right]_{w_v}^1 \right] \\ &= \pi\beta \left[w_v^2 + \frac{1}{3} [1 - w_v^3] \right] = \frac{\pi\beta}{3} [1 + 2w_v^3] \end{aligned} \quad (2.5)$$

Hence we have calculated the expected degree of a vertex with some weight w_v . Let us plot this value and compare to some simulations of the model. Here we use $\beta = 0.01$ and $n = 1000$. For more on simulations see Section 4. The result is plotted in Figure 5.

Now we are prepared to calculate the probability of vertex v , with weight w_v , being isolated. There is a nice geometric intuition for this proof for discrete growth speed distributions. This intuition can be extended to continuous radius distribution. First assume a discrete weight distribution and assume

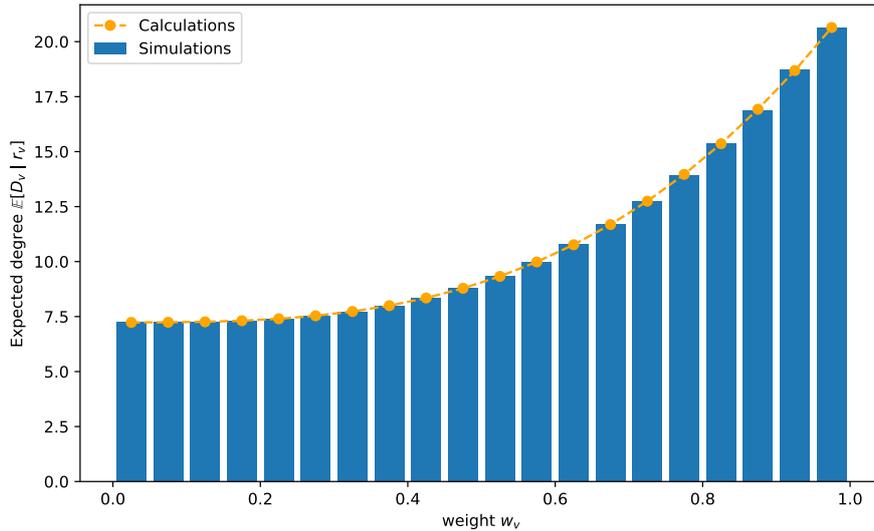


Figure 5. The expected degree of a vertex with weight w_v . The blue bar graph is based on simulations of the graph as given in Definition 2.1 with $n = 1000$, $\beta = \log n$ and a uniform radii distribution from 0 to 1. Note that the continuous spectrum of weights in the simulations are split in bins. The vertex degrees in each bin are averaged to approximate the expected degree. We calculate the orange graph in Example 2.5. The calculations are based on Lemma 2.3.

time t . Then $r = w \cdot t$. Vertex v should not have an edge with any of the other vertices. Knowing this, we can say a few things about where vertices can and cannot be and with which radius. First, there cannot be a vertex within a radius r_v of v . If there would be, this vertex would connect to v , and v would not be isolated. Second, suppose r^* is the next possible radius in the distribution. In the annulus $r_v < r \leq r^*$, vertices are allowed, but only with radius r_v or less. Because a vertex with radius larger than r_v would connect to v . The next annulus can contain vertices radius r^* or less, and so on. We show the intuition behind this proof in Figure 6, with a discrete distribution of 4 different weights, i.e. radii. The generalisation for the continuous case, would be turning these discrete annuli into a continuous range of annuli for all weights in the support of the distribution.

Lemma 2.6 (Probability of a vertex being isolated). *Consider a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ as given in Definition 2.1. And consider a vertex $v \in \mathcal{V}$ which has weight w_v . The probability of this vertex being isolated is,*

$$\mathbb{P}(D_v = 0 \mid R = w_v) = \exp(-\mathbb{E}[D_v \mid W = w_v]). \quad (2.6)$$

In order to form a conjecture about the growth of β on the existence of an isolated vertex, we will summarize the weight dependent part in (2.3) using the following definition:

Definition 2.7 (Radius dependent part of the expected degree of a vertex). *Take some weight distribution W , which is either discrete or continuous. Define $g(w)$ by,*

$$g(w) = \pi \mathbb{E} \left[\max(w, W)^2 \right]. \quad (2.7)$$

The following claim follows almost directly from Lemma 2.3. We give the proof of this claim in Section 3 on page 15.

Claim 2.8 (Connection between $g(w)$ and the expected degree of a vertex). *Construct a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ like in Lemma 2.3 and take an arbitrary vertex $v \in \mathcal{V}$. Take $g(w)$ as given in Definition 2.7. The following holds,*

$$\mathbb{E}[D_v \mid w_v] = \beta g(w_v). \quad (2.8)$$

Another property of $g(w)$ that will be necessary for the extension of proofs to the continuous case is the continuity of $g(w)$.

Claim 2.9 (Continuity of $g(w)$). *Take W a continuous weight distribution. Take $g(w)$ as defined in Definition 2.7. Then $g(w)$ is continuous.*

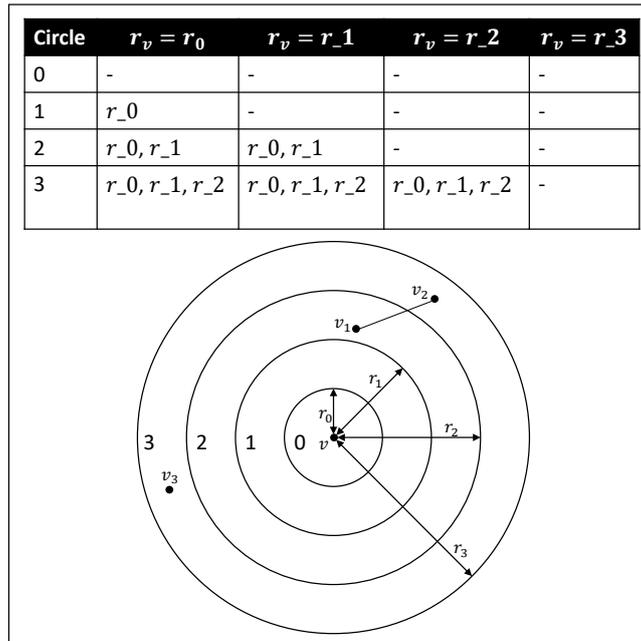


Figure 6. An intuition on the calculation of the probability of a vertex being isolated. This figure is only valid for a discrete weight distribution. It assumes only 4 possible weights, and thus 4 possible radii at any time: $r_0 < r_1 < r_2 < r_3$. Suppose we have a vertex v with radius r_v . In order for v to be isolated, there at least should not be any vertex within a radius r_v of v . Suppose $r_v = r_1$. Within a radius r_1 there should not be any other vertices. Vertices in the annulus $r_1 < r \leq r_2$ there can be vertices with radius r_0 or r_1 , because these radii are smaller than the distance between the annulus and v . In the annulus $r_2 < r \leq r_3$ vertices can exist with radius r_0, r_1 and r_2 . The table shows, given some r_v , whether there can be vertices in each of the annuli and which radii they can have. In the case where $r_v = r_0$ or $r_v = r_1$, vertex v is indeed isolated. In this case, vertex v_1 is allowed in annulus 2 because it has radius r_1 (not illustrated). Similarly vertices v_2 and v_3 are allowed in annulus 3, because they have radius r_2 and r_0 respectively. v_1 and v_2 have an edge between them. This does not influence v being isolated.

Next we need a definition for the smallest weight, w_0 , within a weight distribution. The smallest weight turns out to be important for the existence of isolated vertices. That has quite a clear intuition. A vertex with the smallest weight has the smallest radius at any time. Vertices with a smaller radius have a larger probability of being isolated, as follows from Remark 2.4. The expected degree is minimized for the smallest possible radius. Let us define w_0 .

Definition 2.10 (Minimal radius). *Take a radius distribution W . Define w_0 by:*

$$w_0 = \inf (\text{Supp } W) \quad (2.9)$$

Now let us try to get some intuition on the β value that is the threshold on the existence of isolated vertices. We have calculated some probability for a single vertex being isolated. Suppose this probability does not decrease with the graph size n . When increasing the graph size, there will be more vertices in the graph. Intuitively this will cause more attempts to be made on this fixed probability of a vertex being isolated. So when increasing the graph size enough, will lead to one of these attempts at this fixed probability being successful. Therefore we can conclude that the probability of a vertex being isolated, needs to decrease with n . Thus β needs to increase with n . To be more specific we would intuitively say that the threshold would be where the number of vertices, roughly n , multiplied with the probability of a vertex being isolated neither converges, nor diverges. For now write the $g(w)$ term in the probability of a vertex being isolated as a constant c , because this function does not depend on n . Keep in mind that this is not rigorous, but just to provide intuition on what we expect for the threshold β . We can write,

$$\mathbb{E}[\mathcal{Z}] \approx n\mathbb{P}(D_v = 0) = ne^{-\beta c} = e^{\log n - \beta c} \quad (2.10)$$

To capture the phase transition, we need $\log n - \beta c$ constant. Let us for now assume zero. We expect the following behaviour for the threshold β dependent on graph size n :

$$\beta = c^* \log n \quad (2.11)$$

We denote c^* as the threshold constant. From this non-rigorous reasoning, we would like to conjecture that β needs to be logarithmic with graph size at the threshold for existence of isolated an isolated vertex. The following theorem shows this conjecture is true for the convergence of the expected number of isolated vertices.

Theorem 2.11 (Convergence of the expected number of isolated vertices). *Take a discrete or continuous radius distribution W and $\beta_n = c \log n$. With these parameters, construct a sequence of graphs $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$. Take $c^* = \frac{1}{\pi \mathbb{E}[W^2]}$. For \mathcal{Z}_n being the number of isolated vertices in \mathcal{G}_n , as $n \rightarrow +\infty$ the following holds:*

- (i) if $c > c^*$, then $\mathbb{E}[\mathcal{Z}_n] \rightarrow 0$,
- (ii) if $c < c^*$, then $\mathbb{E}[\mathcal{Z}_n] \rightarrow +\infty$.

From this theorem we derive the following proposition about the existence of an isolated vertex in the limit. The proof of this proposition is just the application of Markov's inequality to Theorem 2.11(i).

Proposition 2.12 (Absence of isolated vertices). *Take a discrete or continuous weight distribution W and $\beta_n = c \log n$. With these parameters, construct a sequence of graphs $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$. Let $c^* = \frac{1}{\pi \mathbb{E}[W^2]}$. If $c > c^*$, the following holds:*

$$\mathbb{P}(\mathcal{Z}_n \geq 1) \rightarrow 0. \quad (2.12)$$

So we now know that when β grows faster than the threshold given in Theorem 2.11, the probability of an isolated vertex existing converges to zero with increasing graph size. Now we would like to show that the probability of an isolated vertex existing converges to 1 when β_n grows slower than the threshold $c^* \log n$. This would be the other side of the threshold, hence proving there is a, at least mathematical, phase transition. In order to prove this other side, the logical next step would be to attempt using Chebyshev's inequality. Let us set this up. We would like $\mathbb{P}(\mathcal{Z}_n \geq 1)$ to converge to 1 in this situation. Rewriting using the law of total probability, gives the following:

$$\mathbb{P}(\mathcal{Z}_n \geq 1) = 1 - \mathbb{P}(\mathcal{Z}_n = 0). \quad (2.13)$$

This means $\mathbb{P}(\mathcal{Z}_n = 0)$ needs to converge to 0. We approximate this value using Chebyshev's inequality, which results in the following:

$$\mathbb{P}(\mathcal{Z}_n = 0) = \mathbb{P}(\mathcal{Z}_n - \mathbb{E}[\mathcal{Z}_n] = -\mathbb{E}[\mathcal{Z}_n]) \leq \mathbb{P}(|\mathcal{Z}_n - \mathbb{E}[\mathcal{Z}_n]| \geq \mathbb{E}[\mathcal{Z}_n]) \leq \frac{\text{Var } \mathcal{Z}_n}{\mathbb{E}[\mathcal{Z}_n]^2}. \quad (2.14)$$

Therefore we need $\text{Var } \mathcal{Z}_n < \mathbb{E}[\mathcal{Z}_n]^2$. We did not succeed in getting the $\text{Var } \mathcal{Z}_n$ under $O(n)$. This means $\mathbb{E}[\mathcal{Z}_n]$ at least needs to grow with speed $O(\sqrt{n})$. But from the proof of Theorem 2.11 we can derive that when c is taken ϵ close to c^* , $\mathbb{E}[\mathcal{Z}_n]$ grows with n^ϵ . Hence this proof is not feasible.

Another way of proving the existence of an isolated point in the graph is by defining a rare event. This rare event would be the existence of an isolated vertex somewhere in the graph. When we can construct a lot of independent trials of this rare event, we eventually find an isolated vertex. This is only true, when the probability of this event happening decreases slower than the we can increase the number of trials. And this requires us to restrict the possible radius distributions to distributions with finite support. The following theorem States this existence of an isolated vertex. We give the proof at the end of Section 3.

Theorem 2.13 (Existence of isolated vertices). *Consider a weight distribution W with finite support. Denote $c^* = \frac{1}{\pi \mathbb{E}[W^2]}$. Consider a sequence of graphs $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$ all following Definition 2.1, taking $\beta(\mathcal{G}_n) = c \log n$ with $c < c^*$, and taking weight distribution W . Denote \mathcal{Z}_n the number of isolated vertices in graph n . This gives*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{Z}_n \geq 1) = 1. \quad (2.15)$$

This theorem, together with Proposition 2.12 show that, in the limit, there will be no isolated vertices when $\beta_n > c^* \log n$ and there will be an isolated vertex when $\beta < c^* \log n$. Hence proving $\beta = \frac{1}{\pi \mathbb{E}[W^2]} \log n$ is the threshold in the mathematical phase transition describing the existence of isolated vertices. In Section 5.1.2 this phase transition is confirmed using simulations.

3. PROOFS

This section contains the proofs to all the statements in Section 2. It follows the same order. Sometimes the proof is supported by a claim. The proof of 2.12 needs a little more structure. For this proof we give an intuition together with a structure. This structure consists of some Definitions and a Claim.

3.1. Conditional expectation of vertex degree. We will start with the proof for the expected degree of a vertex, given its weight.

Proof of Lemma 2.3. Notice we are working on the torus. Denote $w_{max} = \sup\{\text{Supp } W\}$. Take $M > 4\beta w_{max}^2$. This will make sure the box size is larger than the diameter around a circle. This assures vertices are not included twice inside the circumference of a vertex.

First we take the expectation of the Poisson point process conditioned expectation. After conditioning on the Poisson point process, we write the degree of a vertex as the sum of indicators over the vertex set,

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \mathbb{E}[\mathbb{E}[D_v | \Xi]] = \mathbb{E}\left[\mathbb{E}\left[\sum_{u \in \Xi} \mathbb{1}\{\{u, v\} \in \mathcal{E}(\mathcal{G})\} \mid \Xi\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{u \in \Xi} \mathbb{1}\{\beta [\max(w_u, w_v)]^2 \geq \|x_u - x_v\|^2\} \mid \Xi\right]\right] \\ &= \mathbb{E}\left[\sum_{u \in \Xi} \mathbb{E}\left[\mathbb{1}\{\beta [\max(w_u, w_v)]^2 \geq \|x_u - x_v\|^2\} \mid \Xi\right]\right]. \end{aligned} \quad (3.1)$$

Here we split the proof in a discrete and continuous part.

- (i) For the discrete weight distribution W , we assume a probability mass function $p_W(w)$. Denote the different weights possible in W as $w_1 < w_2 < \dots$. From (3.1), we calculate the inner expectation over the weight distribution. By dominated convergence, we can interchange this sum with the expectation over the Poisson point process. This gives,

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \mathbb{E}\left[\sum_{u \in \Xi} \sum_{i=1}^{\infty} p_W(w_i) \mathbb{1}\{\beta [\max(w_i, w_v)]^2 \geq \|x_u - x_v\|^2\}\right] \\ &= \sum_{i=1}^{\infty} p_W(w_i) \mathbb{E}\left[\sum_{u \in \Xi} \mathbb{1}\{\beta [\max(w_i, w_v)]^2 \geq \|x_u - x_v\|^2\}\right]. \end{aligned} \quad (3.2)$$

We can write the expectation over the Poisson point process as an integral by Mecke's formula [27]. The restriction on M we give in the beginning of this proof, makes sure vertices are not counted multiple times. The obtained integral simplifies to the area of a sphere, giving:

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \sum_{i=1}^{\infty} p_W(w_i) \int_{\mathbb{T}^2} \mathbb{1}\{\beta [\max(w_i, w_v)]^2 \geq \|x - x_v\|^2\} dx \\ &= \sum_{i=1}^{\infty} p_W(w_i) \pi \beta [\max(w_i, w_v)]^2 = \pi \beta \mathbb{E}[\max(w_v, W)^2]. \end{aligned} \quad (3.3)$$

This finishes the proof for a discrete distribution.

- (ii) Next we assume a continuous weight distribution with probability density function $f(w)$. From (3.1), we can take the expectation over the weights. This is because the choice of weight is independent of the Poisson point process. By dominated convergence we can take the integral out of the expectation. This gives,

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \mathbb{E}\left[\sum_{u \in \Xi} \int_0^{\infty} f(w) \mathbb{1}\{\beta [\max(w, w_v)]^2 \geq \|x_u - x_v\|^2\} dw\right] \\ &= \int_0^{\infty} f(w) \mathbb{E}\left[\sum_{u \in \Xi} \mathbb{1}\{\beta [\max(w, w_v)]^2 \geq \|x_u - x_v\|^2\}\right] dw. \end{aligned} \quad (3.4)$$

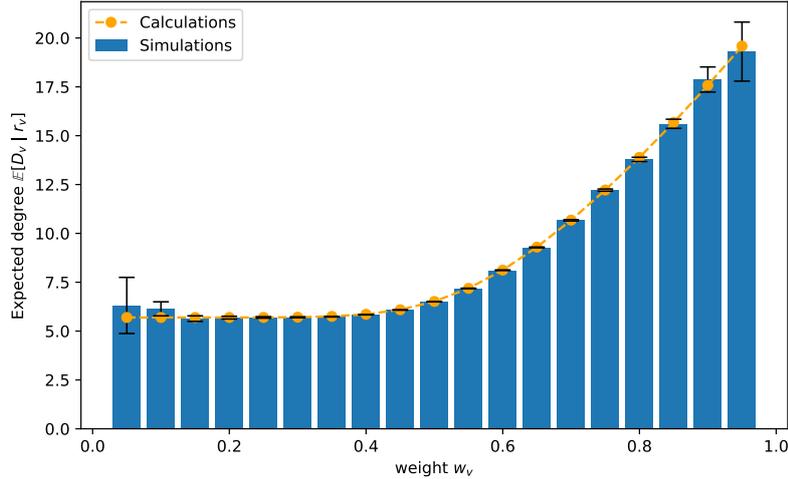


Figure 7. The expected degree of a vertex with weight w_v . The blue bar graph is based on simulations of the graph as given in Definition 2.1 with $n = 1000$, $\beta = \log n$ and a scaled down binomial as weight distribution with $n = 20$, $p = 0.5$ and scaling factor $\frac{1}{20}$. We add error bars, which show the confidence intervals of all vertices in 1000 runs with the given weight w_v . Notice weights w_v close to 0 and 1 are not likely in this probability distribution, hence there are less data points on them. This results in larger confidence intervals. The orange graph is calculated in Example 3.1. The calculations are based on Lemma 2.3.

We can write the expectation over the Poisson point process as an integral by Mecke's formula. The restriction on M we give in the beginning of this proof, makes sure vertices are not counted multiple times. The obtained integral simplifies to the area of a sphere, giving

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \int_0^\infty f(w) \int_{\mathbb{T}^2} \mathbb{1}\{\beta [\max(w, w_v)]^2 \geq \|x - x_v\|^2\} dx dw \\ &= \pi\beta \int_0^\infty f(w) [\max(w, w_v)]^2 dw = \pi\beta \mathbb{E}[\max(w, W)^2]. \end{aligned} \quad (3.5)$$

This finishes the proof. □

To give an intuition on this Lemma, we provide two examples. We show Example 2.5 with a uniform weight distribution in Section 2. The other one with a scaled binomial weight distribution we give here. The simulations supporting the example are explained in Section 4.1.2.

Example 3.1 (Conditional expectation of vertex degree for a discrete distribution). Take $\frac{1}{20}$ Binom(20, 0.5) as discrete weight distribution. Take $\beta = 1$. This gives,

$$p_W(k/20) = \binom{n}{k} p^k (1-p)^{n-k} \quad (3.6)$$

By Lemma 2.3 we have,

$$\begin{aligned} \mathbb{E}[D_v | w_v] &= \pi\beta \mathbb{E}[\max(w_v, W)^2] \\ &= \pi\beta \sum_{i=1}^\infty p_W(w_i) [\max(w_i, w_v)]^2 = \pi\beta \sum_{i=0}^{20} p_W(i/20) [\max(w_i, w_v)]^2 \\ &= \pi\beta \sum_{i=0}^{20} \binom{n}{i} p^i (1-p)^{n-i} [\max(w_i, w_v)]^2 \end{aligned} \quad (3.7)$$

This we can calculate for different w_v values, which is plotted and compared to simulations in Figure 7. In the simulations we construct graphs as in Definition 2.1. Here we use $\beta = 1$ and $n = 1000$. For more on simulations see Section 4.

3.2. Probability of a vertex being isolated. In this section we will prove a formula for the probability of a vertex being isolated.

Proof of Lemma 2.6. First translate the entire graph such that v is in the origin. Because of the periodic boundaries, this will not change any of its properties. Now $v = o$ and $w_v = w_o$. We partition the torus into squares $S_{t,s} := [t, t + dt] \times [s, s + dt]$. Define $X_{t,s} := \{v \in \mathcal{V}(\mathcal{G}_\infty) \text{ s.t. } x_v \in S_{t,s}\}$ as the set of vertices in a given square characterised by (s, t) . Now we split the probability into an intersection of separate squares as given above. Using the independence of the Poisson process, this yields

$$\begin{aligned} \mathbb{P}(D_o = 0 \mid W = w_o) &= \mathbb{P}\left(\bigcap_{S_{t,s} \in \mathbb{T}^2} \{\forall v \in X_{t,s} : \{v, o\} \notin \mathcal{E}(\mathcal{G}_\infty)\}\right) \\ &= \prod_{S_{t,s} \in \mathbb{T}^2} \mathbb{P}(\{\forall v \in X_{t,s} : \{v, o\} \notin \mathcal{E}(\mathcal{G}_\infty)\}). \end{aligned} \quad (3.8)$$

By taking a very small dt , we can assume there either being zero or one vertex in a given square $S_{t,s}$. This way we exclude the possibility of multiple vertices in a square. Using the law of total probability this gives the following:

$$\mathbb{P}(D_o = 0 \mid W = w_o) = \prod_{S_{t,s} \in \mathbb{T}^2} \left(\mathbb{P}(X_{t,s} = \emptyset) + \mathbb{P}(X_{t,s} \neq \emptyset, v \in X_{t,s}) \mathbb{P}(\{v, o\} \notin \mathcal{E}(\mathcal{G}_\infty)) \right). \quad (3.9)$$

By [28], $\mathbb{P}(X_{t,s} = \emptyset) = \exp(-dt^2)$ and therefore by the law of total probability, $\mathbb{P}(X_{t,s} \neq \emptyset) = 1 - \exp(-dt^2)$. This gives

$$\mathbb{P}(D_o = 0 \mid W = w_o) = \prod_{S_{t,s} \in \mathbb{T}^2} \left(e^{-dt^2} + (1 - e^{-dt^2}) (1 - \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty))) \right). \quad (3.10)$$

When taking small dt , the area dt^2 will be really small, hence we can use a Taylor approximation, $\exp(-dt^2) \approx 1 - dt^2$. Substituting this results in a cancellation. We rewrite the product as a sum in the exponent, giving

$$\begin{aligned} \mathbb{P}(D_o = 0 \mid W = w_o) &= \prod_{S_{t,s} \in \mathbb{T}^2} (1 - dt^2 \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty))) \\ &= \exp\left(\sum_{S_{t,s} \in \mathbb{T}^2} \ln(1 - dt^2 \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty)))\right). \end{aligned} \quad (3.11)$$

$1 - dt^2 \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty))$ is really close to one, because dt^2 is very small. Hence we Taylor approximate the logarithm,

$$\mathbb{P}(D_o = 0 \mid W = w_o) = \exp\left(-\sum_{S_{t,s} \in \mathbb{T}^2} dt^2 \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty))\right). \quad (3.12)$$

Taking $dt \rightarrow 0$ turns the sum into an integral, giving

$$\mathbb{P}(D_o = 0 \mid W = w_o) = \exp\left(-\int_{\mathbb{T}^2} \mathbb{P}(\{v, o\} \in \mathcal{E}(\mathcal{G}_\infty)) dt^2\right). \quad (3.13)$$

And here we recognize this integral, being the expectation of the degree of the origin given its weight w_o . Hence we get:

$$\mathbb{P}(D_o = 0 \mid W = w_o) = \exp(-\mathbb{E}[D_o \mid W = w_o]), \quad (3.14)$$

which finishes the proof. \square

3.3. Proof of properties of $g(w)$. This section contains the proofs for some properties of $g(w)$ defined in Definition 2.7.

Proof of Claim 2.8. By Lemma 2.3 we have

$$\mathbb{E}[D_v \mid w_v] = \pi\beta \mathbb{E}\left[\max(w_v, W)^2\right]. \quad (3.15)$$

Here we recognize $g(w)$ from Definition 2.7. Substituting $g(w_v)$ finishes the proof:

$$\mathbb{E}[D_v \mid w_v] = \beta g(w_v). \quad (3.16)$$

□

Proof of Claim 2.9. Take W a positive continuous weight distribution, characterised by its continuous probability density function $f(w)$. Take $w \in \text{Supp } W$ and $\epsilon > 0$ arbitrary. We split the proof in two cases.

- (1) Suppose $w > 0$. Take $x \geq 0$ such that $|w - x| < \min\{\frac{w}{2}, \frac{\epsilon}{4\pi f_{\max} w^2}\} = \delta$, with $f_{\max} = \sup\{f(w) : w \in \text{Supp } W\}$. This exists because $f(w)$ is continuous, positive and $\int f(w)dw = 1$. We simplify $|g(w) - g(x)|$ resulting in

$$|g(w) - g(x)| = \pi \left| \int_{\text{Supp } W} f(\tilde{w}) \max(x, \tilde{w})^2 d\tilde{w} - \int_{\text{Supp } W} f(\tilde{w}) \max(w, \tilde{w})^2 d\tilde{w} \right|. \quad (3.17)$$

From here we split the proof again in two parts. $w \geq x$ and $w < x$.

- (a) Suppose $w \geq x$. We split the integrals into 3 parts: $\tilde{w} < x$, $x < \tilde{w} < w$ and $w < \tilde{w}$. The third integral immediately cancels. This results in

$$\begin{aligned} |g(w) - g(x)| &\leq \pi \left| \int_0^x f(\tilde{w}) (w^2 - x^2) d\tilde{w} + \int_x^w f(\tilde{w}) (x^2 - \tilde{w}^2) d\tilde{w} \right| \\ &\leq \pi \left[\int_0^x f(\tilde{w}) |w^2 - x^2| d\tilde{w} + \int_x^w f(\tilde{w}) |w^2 - \tilde{w}^2| d\tilde{w} \right] \\ &\leq f_{\max} \pi \left[x(w - x)(w + x) + \int_x^w (w^2 - x^2) d\tilde{w} \right] \\ &< f_{\max} \pi [w(w - x)2w + (w - x)^2(w + x)] \\ &< \pi f_{\max} [2w^2\delta + \delta^2 2w] \\ &< \pi f_{\max} 2w^2\delta < \epsilon. \end{aligned} \quad (3.18)$$

Thus finishing the proof for $x \geq w$.

- (b) Now suppose $w < x$. We again split the integral into 3 parts of which the third cancels.

$$\begin{aligned} |g(w) - g(x)| &= \pi \left| \int_0^w f(\tilde{w}) (w^2 - x^2) d\tilde{w} + \int_w^x f(\tilde{w}) (\tilde{w}^2 - x^2) d\tilde{w} \right| \\ &\leq \pi \left[|w^2 - x^2| \int_0^w f(\tilde{w}) d\tilde{w} + \int_w^x f(\tilde{w}) |x^2 - w^2| d\tilde{w} \right] \\ &\leq \pi f_{\max} [w(w - x)(w + x) + (w + x)(x - w)^2]. \end{aligned} \quad (3.19)$$

We know $x - w < \frac{w}{2}$. Therefore $x < 2w$. Thus we can use the lowerbound $(w + x) < 3w$ to obtain

$$\begin{aligned} |g(w) - g(x)| &< \pi f_{\max} \delta [3w^2 + 3w\delta] \\ &\leq \pi f_{\max} \delta 3w^2 < \epsilon. \end{aligned} \quad (3.20)$$

This finishes the proof for the continuity of $g(w)$ at x when $x \neq 0$.

- (2) Now suppose $w = 0$. take $x \geq 0$ such that $|w - x| = |x| < \left(\frac{\epsilon}{\pi f_{\max}}\right)^{\frac{1}{3}} = \delta$. We simplify $|g(w) - g(x)|$ giving

$$\begin{aligned} |g(w) - g(x)| &= \pi \left| \int_0^x f(\tilde{w}) \max(0, \tilde{w})^2 d\tilde{w} - \int_0^x \max(x, \tilde{w})^2 d\tilde{w} \right| \\ &= \pi \int_0^x f(\tilde{w}) |\tilde{w}^2 - x^2| d\tilde{w} \leq \pi \int_0^x f(\tilde{w}) x^2 d\tilde{w} \\ &\leq \pi x^2 \int_0^x f_{\max} d\tilde{w} = \pi f_{\max} x^3 < \pi f_{\max} \delta^3 \\ &= \pi f_{\max} \frac{\epsilon}{\pi f_{\max}} = \epsilon. \end{aligned} \quad (3.21)$$

This finishes the proof for $w = 0$.

□

3.4. Connectivity threshold. In this section we will prove the phase transition in the expected number of isolated vertices in a graph. The next claim contains a property of increasing functions and is needed in the second part of the proof of this phase transition, Theorem 2.11. We give the proof of this Claim in the appendix on page 38.

Claim 3.2 (Property of increasing continuous functions). *Assume we have an increasing, non-negative, continuous function $g(x)$, and we have some $x_0 \geq 0$ and $c \in \mathbb{R}$. Given some δ_0 , if for all $\delta \in \mathbb{R}$ such that $0 < \delta < \delta_0$, we have $c < \frac{1}{g(x_0+\delta)}$, then $c < \frac{1}{g(x_0)}$.*

Proof of Theorem 2.11. This proof will be split into two parts. The first part is for discrete weight distributions. The second part is for continuous weight distributions.

- (i) For a discrete weight distribution W , we assume some probability mass function $p_W(w)$. We can write the number of isolated vertices as the sum of indicators, $\mathcal{Z} = \sum_{v \in \text{PPP}} \mathbb{1}_{\{D_v=0\}}$. Using this and conditioning on the different possible weights gives

$$\begin{aligned} \mathbb{E}[\mathcal{Z}] &= n \mathbb{E}[\mathbb{1}_{\{D_v=0\}}] = n \sum_{i=1}^{\infty} p_W(w_i) \mathbb{E}[\mathbb{1}_{\{D_v=0\}} \mid w_v = w_i] \\ &= n \sum_{i=1}^{\infty} p_W(w_i) \mathbb{P}(D_v = 0 \mid w_v = w_i). \end{aligned} \quad (3.22)$$

Take $g(w)$ from Definition 2.7. By 2.6 and using Claim 2.8, we get:

$$\begin{aligned} \mathbb{E}[\mathcal{Z}] &= n \sum_{i=1}^{\infty} p_W(w_i) \exp(-\mathbb{E}[D_v \mid w_v]) \\ &= n \sum_{i=1}^{\infty} p_W(w_i) \exp(-\beta_n g(w)) = n \sum_{i=1}^{\infty} p_W(w_i) \exp(-c \cdot g(w) \log n). \end{aligned} \quad (3.23)$$

We split the proof into $c > c^*$ and $c < c^*$.

- (a) By remark 2.4 and Claim 2.8, we know $g(w)$ is increasing with w . Thus $g(w) \geq g(w_0)$. This implies that, $\exp(-c \cdot g(w) \log n) \leq \exp(-c \cdot g(w_0) \log n)$. Therefore,

$$\mathbb{E}[\mathcal{Z}] \leq n \sum_{i=1}^{\infty} p_W(w_i) \exp(-c \cdot g(w_0) \log n) = n \exp(-c \cdot g(w_0) \log n) = n^{1-c \cdot g(w_0)}. \quad (3.24)$$

Hence if $1 - c \cdot g(w_0) < 0$, (3.24) converges to 0. Thus if $c > \frac{1}{g(w_0)} = \frac{1}{\pi \mathbb{E}[\max(w_0, W)^2]} = \frac{1}{\pi \mathbb{E}[W^2]} = c^*$, we have $\mathbb{E}[\mathcal{Z}] \rightarrow 0$.

- (b) By (3.23) we have,

$$\mathbb{E}[\mathcal{Z}] \geq n p_W(w_0) \exp(-c \cdot g(w_0) \log n) = p_W(w_0) n^{1-c \cdot g(w_0)}. \quad (3.25)$$

We know $p_W(w_0) > 0$. Hence if $1 - c \cdot g(w_0) > 0$, (3.25) diverges to $+\infty$. Thus if $c < \frac{1}{g(w_0)} = c^*$, $\mathbb{E}[\mathcal{Z}] \rightarrow +\infty$, finishing the proof for discrete radius distributions.

- (ii) For a continuous weight distribution W , we assume some probability density function $f(w)$. We can write $\mathcal{Z} = \sum_{v \in \text{PPP}} \mathbb{1}_{\{D_v=0\}}$. Using this and conditioning on the weight gives,

$$\begin{aligned} \mathbb{E}[\mathcal{Z}] &= n \mathbb{E}[\mathbb{1}_{\{D_v=0\}}] = n \int_{w_0}^{\infty} f(w) \mathbb{E}[\mathbb{1}_{\{D_v=0\}} \mid w_v = w] dw \\ &= n \int_{w_0}^{\infty} f(w) \mathbb{P}(D_v = 0 \mid w_v = w) dw. \end{aligned} \quad (3.26)$$

Take $g(w)$ from Definition 2.7. We get,

$$\mathbb{E}[\mathcal{Z}] = n \int_{w_0}^{\infty} f(w) e^{\beta_n g(w)} dw = n \int_{w_0}^{\infty} f(w) e^{-c \cdot g(w) \log n} dw. \quad (3.27)$$

We split the proof in two parts, $c > c^*$ and $c < c^*$.

- (a) By Remark 2.4 and Claim 2.8, we know $g(w)$ is increasing with w . Hence $\exp(-c \cdot g(w) \log n) \leq \exp(-c \cdot g(w_0) \log n)$. Therefore

$$\begin{aligned} n \int_{w_0}^{\infty} e^{-c \cdot g(w) \log n} f(w) dw &\leq n \int_{w_0}^{\infty} e^{-c \cdot g(w_0) \log n} f(w) dw = n e^{-c \cdot g(w_0) \log n} \int_{w_0}^{\infty} f(w) dw \\ &= n e^{-c \cdot g(w_0) \log n} = e^{\log n [1 - c \cdot g(w_0)]}. \end{aligned} \quad (3.28)$$

Hence if $1 - c \cdot g(w_0) < 0$, (3.28) converges to 0. Thus if $c > \frac{1}{g(w_0)} = \frac{1}{\pi \mathbb{E}[\max(w_0, W)^2]} = \frac{1}{\pi \mathbb{E}[W^2]} = c^*$, $\mathbb{E}[\mathcal{Z}] \rightarrow 0$.

- (b) By the fact that $w_0 = \inf\{w \geq 0 : f(w) > 0\}$, we can say that for all $\epsilon' > 0$, there exists an $\epsilon > 0$ such that $\epsilon < \epsilon'$ and $f(w) = f(w_0 + \epsilon) > 0$. By the fact that $f(w)$ is continuous, there exists an $\delta' > 0$ such that if $|w - (w_0 + \epsilon)| < \delta'$, $|f(w_0 + \epsilon) - f(w)| < \frac{f(w_0 + \epsilon)}{2}$. This implies $f(w) \geq \frac{f(w_0 + \epsilon)}{2}$ for all $w \in [w_0 + \epsilon, w_0 + \epsilon + \delta']$. Take $\delta \leq \delta'$. This gives,

$$\begin{aligned} \mathbb{E}[\mathcal{Z}] &= n \int_{w_0}^{\infty} f(w) e^{-c \cdot g(w) \log n} dw \\ &\geq n \int_{w_0 + \epsilon}^{w_0 + \epsilon + \delta} f(w) e^{-c \cdot g(w) \log n} dw \\ &\geq n \int_{w_0 + \epsilon}^{w_0 + \epsilon + \delta} \frac{f(w_0 + \epsilon)}{2} e^{-c \cdot g(w) \log n} dw. \end{aligned} \quad (3.29)$$

We also know in the current integration interval that: $\exp(-c \cdot g(w) \log n) \geq \exp(-c \cdot g(w_0 + \epsilon) \log n)$. Hence

$$\begin{aligned} \mathbb{E}[\mathcal{Z}] &\geq \frac{f(w_0 + \epsilon)}{2} n \int_{w_0 + \epsilon}^{w_0 + \epsilon + \delta} e^{-c \cdot g(w) \log n} dw \\ &\geq \frac{f(w_0 + \epsilon)}{2} n \int_{w_0 + \epsilon}^{w_0 + \epsilon + \delta} e^{-c \cdot g(w_0 + \epsilon) \log n} dw \\ &= \frac{f(w_0 + \epsilon)}{2} \delta e^{\log n [1 - c \cdot g(w_0 + \epsilon)]}. \end{aligned} \quad (3.30)$$

Therefore if $1 - c \cdot g(w_0 + \epsilon + \delta) > 0$, (3.30) diverges to $+\infty$. Both ϵ and δ can be chosen as small as needed. Hence we denote $\epsilon + \delta = \mu$. Hence for all small $\mu > 0$, if $c < \frac{1}{g(w_0 + \mu)}$, $\mathbb{E}[\mathcal{Z}] \rightarrow +\infty$. By Claim 3.2 this implies that if $c < \frac{1}{g(w_0)} = c^*$, $\mathbb{E}[\mathcal{Z}] \rightarrow \infty$, which finishes the proof. \square

Next we will prove Theorem 2.13. The plan for this proof is to create a rare event, where there is an isolated vertex in the middle of a box. By taking a weight distribution with a finite support, there exists a maximum weight and thus a maximum radius at any time t . For this rare event, we choose a box size such that the isolated vertex cannot connect to any vertices outside of the box, hence the event of such a box containing an isolated vertex is independent of everything outside of the box. Please note that this requires us to scale the box with β . The main proof idea is to increase the size of the space, which increases the number of boxes available. Therefore, even though the event is rare, it will occur with enough available boxes.

Assuming the maximum possible weight is w_m , we give a sketch of one of the boxes in Figure 8. In order for this box being a 'good box', we need exactly one vertex in the inner box, and it needs to be isolated. In order for it to be isolated with the highest probability, we require a small weight. For discrete weight distributions, this is the smallest possible radius. For continuous weight distributions, its weight is taken close to the smallest possible weight. We first need to define some structure.

Definition 3.3 (Embedding of disjoint boxes in the torus). *Fix $n \in \mathbb{N}$. And take $m(n) = \lfloor \frac{\sqrt{n}}{S} \rfloor^2$. Let $B_1, \dots, B_{m(n)}$ be disjoint boxes of side-length S embedded in \mathbb{T}_n^2 . For every $i \in \{1, 2, \dots, m(n)\}$ let B'_i be a smaller box with the same center as B_i and side-length $S^* < S$. S is called the outer side-length and S^* is called the inner side-length. We call this embedding $P(\mathbb{T}_n^2)$.*

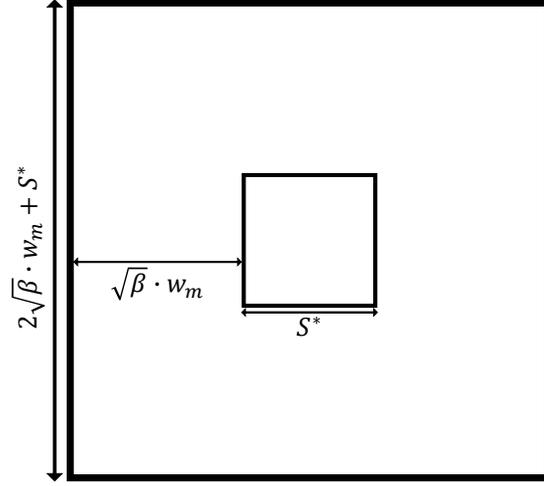


Figure 8. A box which is part of the embedding defined in Definition 3.3. The side-length of the inner box is S^* and the side length of the outer box is $2\sqrt{\beta}w_m + S^*$

Definition 3.4 (Correct box). Fix $n \in \mathbb{N}$, $\beta > 0$ and take a weight distribution W with finite support. Denote $w_m = \sup\{\text{Supp } W\}$. Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ be a graph from Definition 2.1 with these parameters. Fix $S^* > 0$. Consider the box embedding $P(\mathbb{T}_n^2)$ in Definition 3.3 with inner side-length S^* and outer side-length $2\sqrt{\beta}w_m + S^*$. A box $B_i \in P(\mathbb{T}_n^2)$ is correct if:

- (1) $\exists v \in \mathcal{V}$ s.t. $v \in B_i$,
- (2) v is isolated.

Claim 3.5 (Probability of a box being correct). Take an embedding $P(\mathbb{T}_n^2)$ as in Definition 3.3 and a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ as in Definition 3.4. Let W be the weight distribution with w_0 as defined in Definition 2.10 and $w_m = \sup\{\text{Supp } W\} < \infty$. Take a box $B \in P(\mathbb{T}_n^2)$ arbitrary. Take $g(w)$ as defined in Definition 2.7 and write $\beta = c \log n$. If W is discrete, the probability of B being correct is at least

$$\mathbb{P}(B \text{ correct}) \geq \mathbb{P}(W = w_0) S^{*2} e^{-S^{*2}} n^{-cg(w_0)}. \quad (3.31)$$

If W is continuous, take $\epsilon > 0$, the probability of B being correct is at least

$$\mathbb{P}(B \text{ correct}) \geq \mathbb{P}(W \leq w_0 + \epsilon) S^{*2} e^{-S^{*2}} n^{-cg(w_0 + \epsilon)}. \quad (3.32)$$

In both cases, the events of a boxes being correct are mutually independent.

We give the proof of this claim after the proof of Theorem 2.13.

Proof of Theorem 2.13 subject to Claim 3.5. Define $\beta_n = c^* \log n$. Take a sequence of embeddings $\{P_n\}_{n \in \mathbb{N}}$ for every graph $\mathcal{G}_n = \{\mathcal{V}_n, \mathcal{E}_n\}$ as in Definition 3.4. By Claim 3.5 the probability of a box $B \in P_n$ being correct is independent of the other boxes. Denote $q = \mathbb{P}(B \text{ correct})$. And denote the number of boxes in the embedding of graph n by m_n . We approximate the probability of there being an isolated vertex in a graph \mathcal{G}_n by the probability of a box in the embedding being correct:

$$\mathbb{P}(\mathcal{Z}_n \geq 1) = \mathbb{P}(\text{at least one box is correct}) \geq \mathbb{P}(\text{Bin}(m_n, q) \geq 1) = 1 - (1 - q)^{m_n}. \quad (3.33)$$

Now we use the fact that $e^{-q} \geq 1 - q$, which means that:

$$\mathbb{P}(\mathcal{Z}_n \geq 1) \geq 1 - (1 - q)^{m_n} \geq 1 - e^{-qm_n}. \quad (3.34)$$

This needs to converge to 1. Hence qm_n should converge to infinity. We split the proof into two cases. First, the case where W is discrete. And second, the case where W is continuous.

- (1) First take a discrete weight distribution. Using (3.31) to estimate q from below, let us expand q and m_n and put all n independent parts into a constant C .

$$\begin{aligned} qm_n &\geq \mathbb{P}(W = w_0) S^{*2} e^{-S^{*2}} n^{-cg(w_0)} \frac{n}{(2\sqrt{c \log n} w_m + S^*)^2} \\ &= C \frac{1}{(2\sqrt{c \log n} w_m + S^*)^2} n^{1-cg(w_0)}. \end{aligned} \quad (3.35)$$

Now taking the limit $n \rightarrow \infty$ will result in qm_n converging to infinity when $1 - cg(w_0) > 0$, because n^δ for $\delta > 0$ will outgrow the logarithm in the denominator. This follows from basic calculus. By assumption $g(w_0) = \pi \mathbb{E}[W^2] = \frac{1}{c^*}$. Hence $1 - \frac{c}{c^*} > 0$. Hence when assuming $c < c^*$, $qm_n \rightarrow \infty$. Together with (3.34), this means that:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{Z}_n \geq 1) = 1. \quad (3.36)$$

This finishes the proof for a discrete weight distribution.

- (2) Now let us assume we have a continuous weight distribution. We estimate qm_n below as in (3.32). Let us expand q and m_n and put all n independent parts into a constant C .

$$\begin{aligned} qm_n &= \mathbb{P}(W \leq w_0 + \epsilon) S^{*2} e^{-S^{*2}} n^{-cg(w_0 + \epsilon)} \frac{n}{(2\sqrt{c \log n w_m} + S^*)^2} \\ &= C \mathbb{P}(w \leq w_0 + \epsilon) \frac{1}{(2\sqrt{c \log n w_m} + S^*)^2} n^{1 - cg(w_0 + \epsilon)}. \end{aligned} \quad (3.37)$$

Observe that in this case there is an extra ϵ error in g and in $\mathbb{P}(W \leq w_0 + \epsilon)$. As in the discrete case taking the limit will result in qm_n converging to infinity if $1 - cg(w_0 + \epsilon) > 0$. This follows from basic calculus. Hence we need $c < \frac{1}{g(w_0 + \epsilon)}$. We know $c < c^* = \frac{1}{\pi \mathbb{E}[W^2]} = \frac{1}{g(w_0)}$. $g(w)$ is increasing and continuous in w , hence for a fixed $c < c^*$, we can take $\epsilon > 0$ small enough to force $c < \frac{1}{g(w_0 + \epsilon)} < c^*$. This means that qm_n converges to infinity. Just like in the discrete case, the proof of the theorem follows using (3.34). □

Proof of Claim 3.5. For these proof we define two events. Define A_1 as the event where there is exactly one vertex v in B' . And define A_2 as v being isolated.

$$A_1 = \{\text{PPP has one vertex } v \text{ in } B'\} \quad (3.38)$$

$$A_2 = \{v \text{ is isolated}\} \quad (3.39)$$

If both events A_1 and A_2 occur simultaneously, the box is correct. Hence we want to calculate

$$\mathbb{P}(B \text{ correct}) \geq \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1). \quad (3.40)$$

First the probability of A_1 occurring is $S^{*2} e^{-S^{*2}}$, [28]. The only thing we know about the surroundings of the vertex v , that exists when A_1 occurs is the fact that there are no other vertices in B' . This increases the probability of it being isolated compared to the setting of Lemma 2.6. In the proof of this lemma, the dt^2 squares that are inside B' have 0 vertices under the conditioning. In the discrete case this results in

$$\begin{aligned} \mathbb{P}(A_2 | A_1) &\geq \mathbb{P}(D_v = 0 | W_v = w_0 \text{ and } B' \cap \Xi = \emptyset) \geq \mathbb{P}(D_v = 0 | W_v = w_0) \\ &= \exp(-\mathbb{E}[D_v | w = w_0]). \end{aligned} \quad (3.41)$$

Therefore we can use the probability from this lemma as a lower bound on the probability of v being isolated. Note that the formula in this lemma is based upon knowing the weight of the vertex. We split the proof into two cases. The weight distribution being discrete or continuous.

- (1) First consider the case where the graphs have discrete weight distribution. The probability of v being isolated is lowest when we take weight w_0 , as defined in Definition 2.10. We fix that requirement. By 3.41,

$$\mathbb{P}(B \text{ correct}) \geq \mathbb{P}(A) \mathbb{P}(W = w_0) \mathbb{P}(D_v = 0 | W_v = w_0). \quad (3.42)$$

Define $g(w)$ as in Definition 2.7. Substituting (2.6) gives

$$\begin{aligned} \mathbb{P}(B \text{ correct}) &\geq \mathbb{P}(W = w_0) S^{*2} e^{-S^{*2}} e^{-\mathbb{E}[D_v | w = w_0]} \\ &= \mathbb{P}(W = w_0) S^{*2} e^{-S^{*2}} e^{-\beta g(w_0)} \\ &= \mathbb{P}(W = w_0) S^{*2} e^{-S^{*2}} n^{-cg(w_0)}. \end{aligned} \quad (3.43)$$

Thus finishing the proof for a discrete weight distribution.

- (2) Next, let us consider graphs with a continuous weight distribution. In this case we cannot choose weight w_0 , as defined in Definition 2.10. In this case we require this weight to be below $w_0 + \epsilon$ for some $\epsilon > 0$. This changes the probability of v being isolated as in Lemma 2.6. The following results from these adaptations:

$$\begin{aligned}
\mathbb{P}(B \text{ correct}) &\geq \mathbb{P}(W \leq w_0 + \epsilon) S^{*2} e^{-S^{*2}} e^{-\mathbb{E}[D_v|w=w_0+\epsilon]} \\
&= \mathbb{P}(W \leq w_0 + \epsilon) S^{*2} e^{-S^{*2}} e^{-\beta g(w_0+\epsilon)} \\
&= \mathbb{P}(W \leq w_0 + \epsilon) S^{*2} e^{-S^{*2}} n^{-cg(w_0+\epsilon)}.
\end{aligned} \tag{3.44}$$

All that is left, is the proof of the independence of correct boxes. By the definition of the Poisson point process, the existence and location of vertices within a box is independent of vertices in another box. The maximum connection distance for a vertex is fixed by $\sqrt{\beta}w_m$, with w_m the maximum possible weight generated by the weight distribution. This follows from (2.1). Therefore v cannot form edges with vertices outside this radius. This implies that the event $\{v \text{ is isolated}\}$ only depends on the Poisson point process in a ball around v of radius $\sqrt{\beta}w_m$. This ball is entirely contained in the box containing v . Therefore the event $\{v \text{ is isolated}\}$ is independent of the Poisson point process outside the box. Thus the event $\{B \text{ correct}\}$ is independent of events in other boxes. We can conclude that the events of boxes being correct are independent. This finishes the proof. \square

4. SIMULATIONS

In this Section we describe how we use simulations to confirm statements from Section 2 and to get the model from this section closer to real cellular tissue. In the first part we describe simulations of spatial Boolean graphs like in Definition 2.1. These graphs use Poisson point processes to construct the vertex set. This part will consist both of simulations for the expected degree of vertices, Lemma 2.3, and simulations for the phase transition, Theorem 2.11 and Proposition 2.12. This phase transition is the transition between the graphs being disconnected and the graphs being connected. The second part of this chapter will go into the theory behind determinantal point processes, which we use to improve the model from Section 2. We describe the methodology for the simulations of the threshold in spatial Boolean graphs with determinantal point processes. The third part goes into extracting the neighbour counts of Tribolium cells from images using image analysis and comparing the internal repulsion in both point processes and the real Tribolium cells. An example of these Tribolium cells is given in Figure 11a. We show the results of the simulations and image analysis described below in section 5. All of the code is available in the gitlab, [29].

4.1. Poisson simulations.

4.1.1. Graph construction. In order to form conjectures on these maximal spatial random graphs, simulations are run. We construct and execute these simulations using `c++`. These simulations construct realisations of mathematical graphs described in Definition 2.1. The growth speeds are chosen from different distributions, discrete and continuous. These graphs have some size, referred to as n in Definition 2.1. This size is size of the space we are working on. Remember that we choose the number of vertices in a Poisson point process from a Poisson random variable, hence there will not be exactly n vertices. We are working on a torus, n is the total area of the two dimensional space $[0, \sqrt{n}]^2$, with periodic boundaries. Lastly, we choose some value c . We choose these values of c around the analytically calculated transition values, where the Poisson spatial Boolean graphs should go from not being connected to being connected. We can calculate these analytical transition values using the statements from Section 2. Using a value of c , we can calculate $\beta = c \log n$, like in Theorem 2.11.

We can construct a realisation of a Poisson point process in two steps. First, we choose the number of points N in a space with size n using intensity $\lambda = 1$ from a Poisson random variable $N \sim \text{Poi}(\lambda n)$ [30]. Second, we choose N points uniformly at random. We give an example of such a Poisson point process in Figure 3b. After generating a realisation of a Poisson point process, we give all points a growth speed according to the growth speed distribution. For every set of vertices u and v , we calculate the presence of the edge $\{u, v\}$ using 2.1. After this the graph is fully constructed.

4.1.2. Expected vertex degree simulations. To get an intuition on the convergence of the vertex degree and in order to test Lemma 2.3 we construct 1000 spatial Boolean graphs. We fix the size at $n = 1000$ and we choose $\beta = 1$, thus we fix for time $t = 1$. We choose the growth speeds, which we call weights, from a uniform distribution, $\text{Unif}(0, 1)$, an exponential distribution $\text{Exp}(1)$ and a scaled binomial distribution $\frac{1}{20} \text{Binom}(20, 0.5)$. We construct 2500 graphs. We choose these distributions, because they include a discrete distribution, a finite continuous distribution and a light tailed continuous distribution. Therefore they represent a different characteristics for probability distributions. For all vertices in all graphs we collect the vertex weights (growth speeds), the number of connections and the vertex degree. All of these pairs of vertex degree and weights are bunched together. This is allowed, because all surroundings of all vertices are independent of the vertex location and fitness. Hence the environments of all vertices are equal, independent of which run it is in. All of these degree-weight pairs are sorted based on weight. For discrete weight distributions this is easy, because we can sort it in discrete piles. For continuous distributions, we sort them into weight-bins.

We calculate the mean and confidence interval of every bin. The means we compare to the calculation of expected degree as in the Lemmas. We give the calculation of confidence intervals in Section A.4. The discrete weight distributions might have some weights that are less common. And some bins of the continuous weight distributions might only have a few vertices included. Hence for these values the standard error might be larger. The confidence interval is plotted together with the data and the calculations for the exponential distribution in Section 3. We show results of the Uniform distribution in Figure 5 and for the binomial distribution in Figure 7.

4.1.3. *Threshold simulations.* In order to experimentally confirm Theorem 2.11, Proposition 2.12 and Theorem 2.13 we run simulations. Some of these simulations were performed before proving the statements. These early simulations were used to construct a conjecture about the n dependent phase transitional behaviour of β . Remember β can be interpreted as time with $\beta = t^2$. The Theorems and Proposition are all about the existence and expected number of isolated vertices in spatial Boolean graphs, Definition 2.1. We can get the number of isolated vertices in a graph from our simulations. We can approximate the expected number of isolated vertices by the average number of isolated vertices in the simulations. All simulations of the graphs can be seen as some realisation of a random variable. All simulations are independent, therefore these random variables are i.i.d.. By the central limit theorem, we can calculate confidence intervals for the average number of isolated vertices. We explain the construction of these confidence intervals in Appendix A.4. The intervals around the average number of isolated vertices is the 95% interval, meaning that the probability that the expected number of isolated vertices is within this interval, is 95%.

We use our simulation results to test Theorem 2.11. First, the right values for c need to be chosen. They need to be around the theoretically calculated value $c^* = \frac{1}{\pi \mathbb{E}[W^2]}$. We expect an upwards trend of the average number of isolated vertices for $c < c^*$. An upwards trend implies divergence in the limit. This implication can be made, because we do not expect the scaling of the graph to influence the surrounding of a specific vertex. Larger graphs will not lead to different local features. This means that, from a certain lower bound of vertices, we expect there not to be any fluctuating size effects. We expect the number of isolated vertices to scale monotonically with graph size. From this implication, we extrapolate an upwards trend in the average number of isolated vertices to the divergence of isolated vertices. Similarly a downwards trend can be extrapolated to implying a convergence of the number of isolated vertices. We give an example of an upwards and downwards trending number of isolated vertices in Figure 9. These are based on simulations with a uniform distribution between 0 and 1. On the horizontal axis are the graph sizes, which where 200, 450 and 800 in this case. On the vertical axis is the number of isolated vertices in the graph.

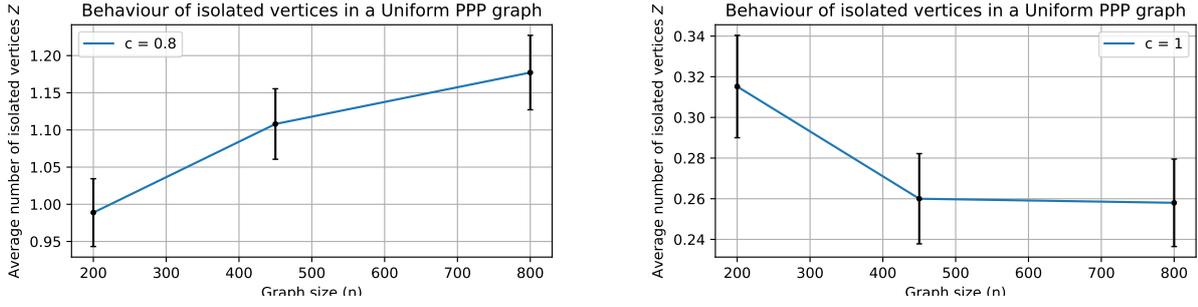
Proposition 2.12 follows from the convergence to 0 of the expected number of isolated vertices. Applying Markov's inequality is enough to see this. Intuitively this is reasonable as well. If the average number of isolated vertices converges to 0, the probability of an isolated vertex existing also converges to 0. Contrarily, if the probability of an isolated vertex existing would not converge to 0, the average number of isolated vertices would converge to some non-zero value. Our simulations actually cannot proof this, because convergence to zero versus some non-zero constant cannot be shown from these simulations. To show this, we would need simulations with infinite graph sizes. The simulations can still confirm the analytic statements and they can show that the convergence or divergence of isolated vertices is already visible for finite graph sizes.

To confirm this proposition together with Theorem 2.13 we will extract the number of runs where there exists at least one isolated vertex versus the number of runs where there does not. These values are extracted from the same 2500 runs per pair of graph size n and constant c as used for confirming Theorem 2.11.

4.2. Determinantal simulations.

4.2.1. *Theoretical determinantal point processes.* Poisson point processes have many advantages. The independence of the points makes it easy to do mathematical analysis on and makes them easy to construct. But when looking at cell structures, they are not really representative. Cells have an inherent repulsion, meaning their centers do not want to be close to each other. This is because they cannot exist at the same location, and their cell walls act like springs pushing other cells outwards. This spatial repulsion is not present in Poisson point processes. Therefore we would like to have another mathematical point process which has some repulsion between points. If simulating is our only goal, some ad hoc procedure can be designed to construct point sets. An example of this would be an acceptance/rejection algorithm. Doing mathematical analysis on these kinds of procedures is difficult. We need to know probabilistic properties of them. Those probabilistic properties are difficult to control with such a ad hoc algorithms, because of the large number of steps and decisions made during the construction.

That is where determinantal point processes enter [31, 32]. We show an example of a determinantal point process (DPP) with 50 points in Figure 3 together with a Poisson point process (PPP) with the same size. As visible in this figure, the points in the DPP are much more evenly distributed. However,



(a) The increasing trend indicates that the number of isolated vertices diverges as $n \rightarrow \infty$. Therefore for $c = 0.8$, $\beta = c \log n$ is below the threshold for isolated vertex convergence.

(b) The decreasing trend indicates that the number of isolated vertices converges to 0 as $n \rightarrow \infty$. Therefore for $c = 1$, $\beta = c \log n$ is above the threshold for isolated vertex convergence.

Figure 9. The expected number of isolated vertices in spatial Boolean graphs. The spatial Boolean graphs are defined in Definition 2.1. The spatial Boolean graphs is chosen with different sizes n and different values of c , where $\beta = c \log n$. For the growth speed a uniform distribution between 0 and 1 is chosen. In this graph the c values are 0.8 and 1.0. For each of these c values, we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. The data shown in these figures is the average number of isolated vertices in these graphs. The error bars give the 95% confidence intervals for these data points. The computed threshold for connectivity is $c^* = 3/\pi \approx 0.95$.

constructing a determinantal point process takes a lot of computational power, mainly if we would like a large set of points. Hence we edited the way to construct the point sets a bit compared to the standard determinantal point processes. First we will explain how determinantal point processes are theoretically constructed.

Determinantal point process are based on determinants, as the name suggests. Suppose we would like to create a DPP with n points. We take a square of area n , which makes sure we have density 1. We randomly choose a large set of points $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$ in the square from a uniform distribution. The goal is to get some subset of these points that has internal repulsion, meaning the points are spread out over the area. We want to define some function determining the likelihood two points end up in this final subset together. When points are close together, this likelihood needs to be low. Because we do not want points in the final subset to be close together. We want them to have short-ranged internal repulsion. When points are far apart, this likelihood is high. This will enforce the repulsion. The likelihood between two points is calculated using the Gaussian likelihood function:

$$L_{\vec{x}_i, \vec{x}_j} = \exp\left(-\frac{1}{2l}\|\vec{x}_i - \vec{x}_j\|^2\right), \quad (4.1)$$

where the norm is the periodic norm on \mathbb{T}_n^2 . The set of all likelihood values between all sets of two points creates the likelihood kernel L . From the large set of points X , a subset $\chi = \{\vec{\chi}_1, \vec{\chi}_2, \dots, \vec{\chi}_n\}$ of size n is randomly chosen. This subset χ has its own likelihood kernel L_χ , which is a sub matrix of the likelihood kernel L . The probability of choosing χ as point set is proportional to the determinant of the sub matrix: $p(\chi) \propto \det L_\chi$. Note, this requires all determinants of these sub matrices to be positive. For kernels with ones on the diagonal and values in $[0, 1]$ in the other spots, this is true. Hence the Gaussian likelihood kernel following (4.1) is allowed. It is given by:

$$L_\chi = \begin{bmatrix} L_{\vec{\chi}_1, \vec{\chi}_1} & L_{\vec{\chi}_1, \vec{\chi}_2} & \dots \\ \vdots & \ddots & \\ L_{\vec{\chi}_n, \vec{\chi}_1} & & L_{\vec{\chi}_n, \vec{\chi}_n} \end{bmatrix} \quad (4.2)$$

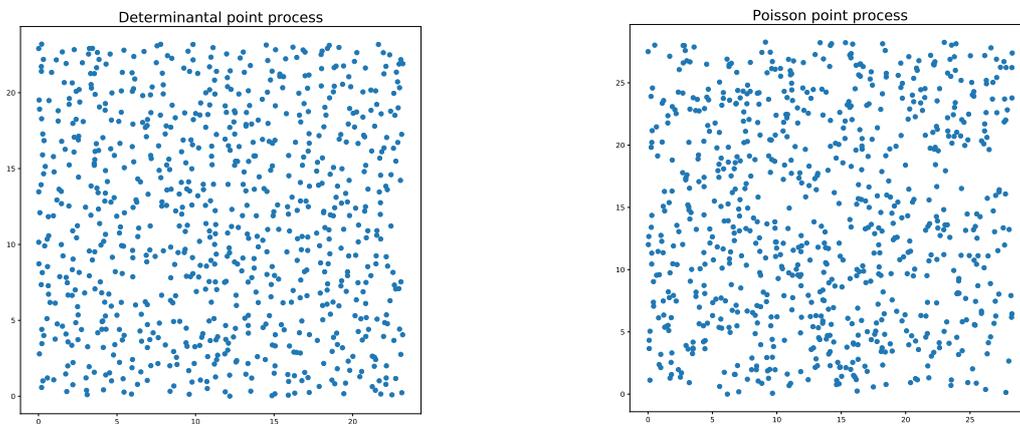
Taking $p(\chi) \propto \det L_\chi$ causes point sets with repulsion to have a higher probability of occurring. We will give some intuition on this. Remember the likelihood function is near 1 when points are close together and near 0 when points are far away. Hence when all points are really far apart, the likelihood kernel is going to look like the identity matrix with 1's on the diagonal and 0's everywhere else. The identity matrix has determinant 1. Hence point sets with points that are all far apart will have quite a high probability of occurring. Now suppose all points are really close together. All elements of the

likelihood kernel will be near 1. This likelihood kernel resembles a kernel with all 1's. The determinant of a matrix with only 1's is 0. Hence these point sets will have a small probability of occurring.

4.2.2. *Constructing Determinantal point processes.* Calculating all determinants of the likelihood kernels of all different subsets $\chi \subset X$ is only feasible for small n and small X . In practise, to generate point sets with repulsion using this mechanism we did the following. Define some limit d . Take $\chi \subset X$ randomly and calculate its determinant. If $\det L_\chi > d$, we take this point set. Otherwise take another $\chi \subset X$ randomly and try again with this subset. This allows us to generate point sets of about 50 points within a reasonable time, while taking d high enough that there is significant repulsion between the points. Figure 3a shows such a realisation of a determinantal point process. To compare the determinantal graph simulations with the Poisson simulations of Section 4.1.3, we require point sets of 800 points.

In order to generate these larger sets of points, we decide to use the realisations of 50 points. Multiple of these DPP realisations are arranged in a square grid. This allows us to create larger sets of repulsive points. The downside is that it is more difficult to analyse mathematically. It introduces another repulsive force, namely the fact that we evenly distribute points in these grid squares. No matter the distribution inside the grid squares, generating points this way is repulsive because we have distributed to some level already.

Another problem with this way of producing points is the clustering around the grid lines. The points from one square do not repel the points in the square next to it. Therefore at the grid lines, points can accumulate. We prevent this by the following three steps. First, instead of using a periodic norm for generating the point sets for the grid squares, we use a non periodic euclidean norm, as if it were a non-periodic square. This makes points at the edge of a grid square not repel points at the opposite edge, which results in more points at the edges of these squares. Second, we assemble these point sets into a grid, but this time between all of the squares we leave a gap of 0.5. Third, we scale the entire grid resulting in the original size and density 1. We show an example of a determinantal point process constructed with this process in Figure 10a. It includes 16 DPP realisations of 50 points.



(a) A point set that is a realisation of the adjusted determinantal point process described at the end of Section 4.2. It contains 16 sets of DPP realisations with 50 points each. Hence the figure contains 800 points. Repulsion between the points is clearly visible, because there are no large clusters of points and no low density areas.

(b) A realisation of a Poisson point process. This Poisson point process has intensity $\lambda = 1$. The area of this square is 800. Hence the expected number of points is also 800. This figure contains 790 points.

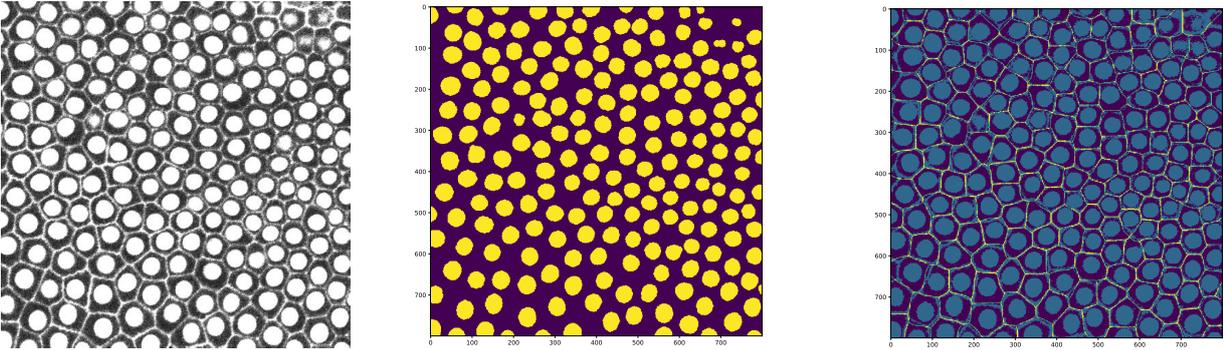
Figure 10. Two different point processes. (A) is a Determinantal point process of size 800. (B) is a Poisson point process of size 800. Notice the points in (A) are much more evenly distributed. There are regions in (B) with very few points and regions with very many points. This is much less visible in (A).

4.2.3. *Threshold simulations for determinantal point processes.* Analytically calculating the threshold of isolated vertex existence for spatial Boolean graphs with a determinantal point process goes beyond

this thesis. We compare the behaviour of spatial Boolean graphs with determinantal and Poisson point processes in simulations. We are going to do the same simulations as in Section 4.1.3 with a determinantal point process instead of a Poisson point process. Constructing a new determinantal point process for each run is computationally too expensive. We first constructed 100 realisations of determinantal point processes for graph sizes 200, 450 and 800. Their construction is described above. We choose sizes 200, 450 and 800 because they are $50 \cdot s^2$ for $s \in \{2, 3, 4\}$, which therefore can be constructed using the determinantal point processes blocks of 50 points. After constructing these DPP realisations, we choose a c . We run the simulation of the spatial Boolean graph 25 times for each DPP realisation. This is possible, because the weights are chosen at random from the weight distribution. Therefore we end up with 2500 runs while only constructing 100 DPP realisations. This does make the runs are not completely independent anymore. Therefore we should be more careful with the error bars. The simple version of the central limit theorem is theoretically not applicable here. We do plot the confidence intervals, but they should be approached with a bit more distrust.

4.3. Physical model. At first the plan was to create a more realistic physical model. In this model cells could push each other out of the way. After a while we decided to continue with just the mathematical model. This was decided because we were interested in the mathematical properties of the graphs and we were interested in making analytical statements about cellular structures. Therefore determinantal point processes were introduced as an alternative for the cell pushing mechanism in the physical model. The physical model is described in the appendix in Section A.3. Before deciding to continue with just the mathematical model, we were attempting to replicate the cellular structures with the physical model. How this was done and what the results were is also explained in Section A.3.

4.4. Comparison Tribolium and spatial Boolean graphs.



(a) Tribolium cells. Tribolium cells are cells of a Tribolium embryo. Tribolium is a species of beetle. This tissue are epithelial cells from Tribolium embryos. This figure was taken and shown in a paper on the mechanics of epithelial tissue formation, [1]. To make the image, both the nuclei and the actin in the embryo were labeled with GFP (green fluorescent protein). That is the white you see here.

(b) Nuclei of the Tribolium cells. This is a step in the image analysis of the Tribolium cells shown in Figure 11a.

(c) Voronoi cells resulting from the image analysis of Figure 11a. In the background we show the original image. This allows us to compare between the Voronoi boundaries and the original vague boundaries.

Figure 11. Steps in the image analysis of the Tribolium cells. The goal of this image analysis is to obtain the amount of neighbours for every cell. The width of these images is $150 \mu\text{m}$.

4.4.1. Tribolium cell image analysis. In the end it would be nice to be able to imitate real cell structures with the spatial Boolean graphs. One measure to compare these cell structures with the graphs is the degree distribution. The degree distribution is a distribution of the number of connections/neighbours of the vertices/cells. It is the frequency of a vertex/cell having a certain number of connections/neighbours. We can visualize this as a histogram with on the horizontal axis the number of neighbours and on the

vertical axis the number of cells with that number of neighbours. Before deciding to continue with the mathematical model, we wanted to compare cellular structures to our physical model, described in Section A.3. This comparison would use a degree distribution. The goal in this section is to retrieve the neighbour counts/degree distribution of the Tribolium cells that we give in Figure 11a, [1]. Tribolium is a species of flour beetle. The cells are epithelial cells in Tribolium embryos. We use these images as a real-world example of cellular tissue. Two cells are denoted as neighbours in this tissue if they share a cell wall. The image analysis in this section is done in python using the skimage morphology library [33]. The boundaries of cells are not defined clearly, as visible in Figure 11a. That makes it more reasonable to use the cell nuclei. Hence first we do a threshold, which turns the image into an array of Booleans. And we use the morphology opening operation with disks, which removes all small objects and keeps the nuclei. In Figure 11b we show the resulting image.

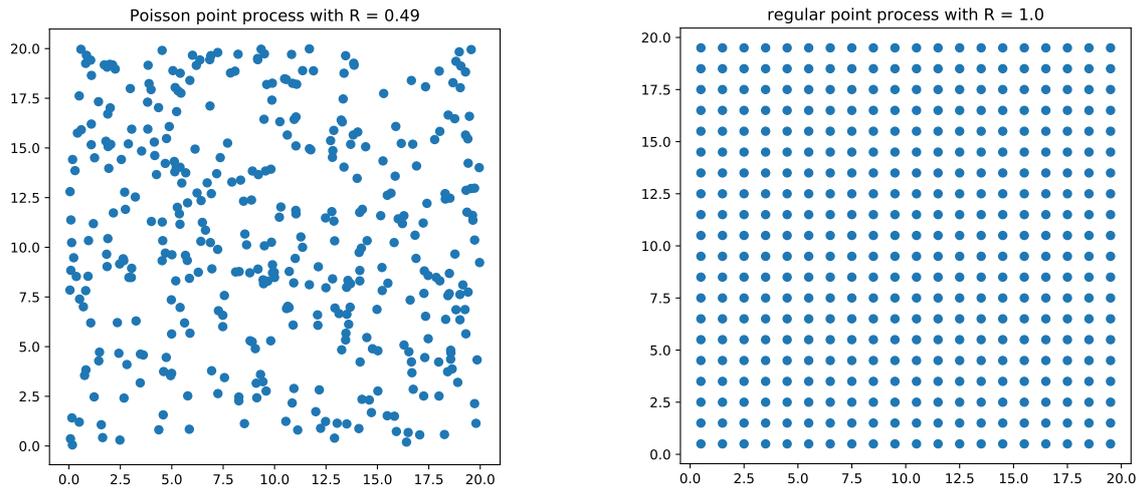
Next the Voronoi cells are constructed out of these nuclei. This might change the boundaries a bit compared to the boundaries in the original image. This slight change is unavoidable when using the nuclei, but the alternative of determining the neighbour counts from the cell boundary we give in Figure 11a, is way more difficult and therefore probably less precise. The Voronoi cells, [34], are constructed by growing each of the nuclei until they cannot grow anymore. Note that all cells need to grow at the same speed. We show the resulting Voronoi cells in Figure 11c. We show the original image in the background. This allows us to compare the Voronoi cell boundaries to the original boundaries.

After constructing the Voronoi cells, we need to extract the number of neighbours for each cell. Here the assumption is made that at a single point, at most 3 cells meet. This assumption is argued in A.2 and illustrated in Figure 25. This assumption allows us to look at a 3x3 square around each pixel, and say two cells are neighbours if they both appear in such a square. Lastly, we need to take into account the boundaries. The cells at the boundaries have fewer neighbours, because they are outside of this image. Therefore all the neighbour counts of the cells that touch the boundary are not taken into account. Finally this analysis results in the number of neighbours per cell, for all cells not at the boundary. In Section 5.2 we show the results.

4.4.2. *Quantifying repulsion.* In addition to the degree distributions, we would like to quantify the repulsion in the point processes. A nice approach for this is to calculate the mean distance to the nearest neighbour. This was derived by Hertz in 1909 [35] and independent from him by Clark & Evans in 1954 [36]. The repulsion of a set of n points is,

$$R = \sqrt{\lambda} \frac{1}{n} \sum_{i=1}^n \min_{j \neq i} \{\|x_i - x_j\|\}. \quad (4.3)$$

Note that it is normalized with $\sqrt{\lambda}$ and thus independent of density. A completely independent point process should have $R = 0.5$. If the points are located on the grid points of a square grid, $R = 1$ [37]. The average repulsion will be calculated for 100 Poisson point processes and 100 Determinantal point processes. It will also be calculated for 7 images of Tribolium similar to Figure 11a, [1]. There are more images, but some are very difficult to analyse. Very vague nuclei definition or a wide range in brightness of nuclei make the image analysis difficult. We retrieve the cell nuclei for the easier images. This process is described in Section 4.4.1 above. The centers of these nuclei are calculated using the skimage measure regionprops method, which can calculate the centers of mass of these disk like shapes. We calculate the repulsion for these centers. We show the results in Section 5.3.1.



(a) A realisation of a Poisson point process on a 20×20 square. It has 395 points. This number is chosen from $\text{Poi}(400)$. The points are independently chosen from a uniform distribution.

(b) Points located on the intersections of a square grid. It contains 400 points in the 20×20 square. It is a very regular repulsive point set. It has repulsion $R = 1$.

Figure 12. A point set from an independent point process and a very regular point set. For both point sets their repulsion has been calculated using (4.3).

5. RESULTS AND DISCUSSION

This section will contain results from the simulations described in Section 4. The first part of the results are simulations based on a model defined in Definition 2.1, called spatial Boolean graphs. These work with a Poisson point process. The goal of these is to confirm the theorems stated in Section 2. Second, this section also contains results on spatial Boolean graphs with a determinantal point process instead of a Poisson point process. This part will contain results on the comparison between both point processes and real cell structures. Also we give results on the threshold for isolated vertex existence for these determinantal spatial Boolean graphs, similar to Theorem 2.11.

5.1. Poisson point processes.

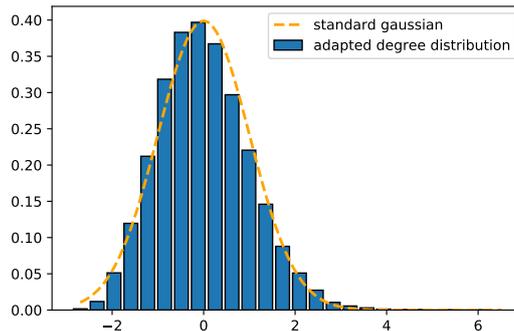


Figure 13. A standard Gaussian plotted with simulation results from simulations described in Section 4.1.2. We constructed 1000 graphs as explained in Section 4.1.2. We use a uniform distribution like in Example 2.5. Degrees of all vertices with weights (growth speeds) between 0.15 and 0.20 were taken. This set of degrees can be interpreted as a set of realisations of i.i.d. random variables. Therefore we can apply the central limit theorem. This is explained in Section A.4. To check this assumption, we shift the values by the mean μ to the right and divide by $\frac{\sigma}{\sqrt{n}}$ with n the number of samples. That is what is shown as the blue graph. According to (A.12), this should be distributed as a standard Gaussian distribution. We plot a standard Gaussian distribution in orange in this figure.

5.1.1. *Expected vertex degree simulations.* We check the assumption that we can apply the central limit theorem to calculating the confidence intervals for the expected degree of a vertex. We show this confirmation in Figure 13. We transform simulations from one bin of weights according to the central limit theorem, see section A.4. Then we plot both a standard Gaussian and the simulation results. They agree quite well, which tells us the application of the central limit theorem is justified. We show the result for the exponential distribution with $\lambda = 1$ in Figure 14. As we can see, the higher weights are less common, which causes them to have a larger confidence interval. The blue histogram is the result of simulations and the orange line is the result of the calculations following Lemma 2.3. The last bin contains only one value. Therefore we cannot calculate a confidence interval. In example 3.1 we show simulation results for the scaled binomial distribution. In Example 2.5 we give the result for the uniform distribution. All three the simulations agree with the calculations really well. This confirms the calculations.

5.1.2. *Threshold simulations for Poisson point processes.* In this section we give the results regarding the simulations described in Section 4.1.3. We show simulations for a uniform distribution, $\text{Unif}(0, 1)$, and an exponential distribution with $\lambda = 1$, $\text{Exp}(1)$. In both cases we calculate the theoretically derived threshold constants c^* , Theorem 2.11. We choose values of c around c^* . We normalize the plots of the expected number of isolated vertices with respect to the expected number at graph size 200. We do this for all values of c . It allows us to show and compare all of them in one figure. Note that this changes the confidence intervals. We show the original confidence intervals.

First for the uniform distribution we can calculate c^* as follows,

$$c^* = \frac{1}{\pi \mathbb{E}[W^2]} = \frac{1}{\pi \frac{1}{3}} = \frac{3}{\pi} \approx 0.95. \quad (5.1)$$

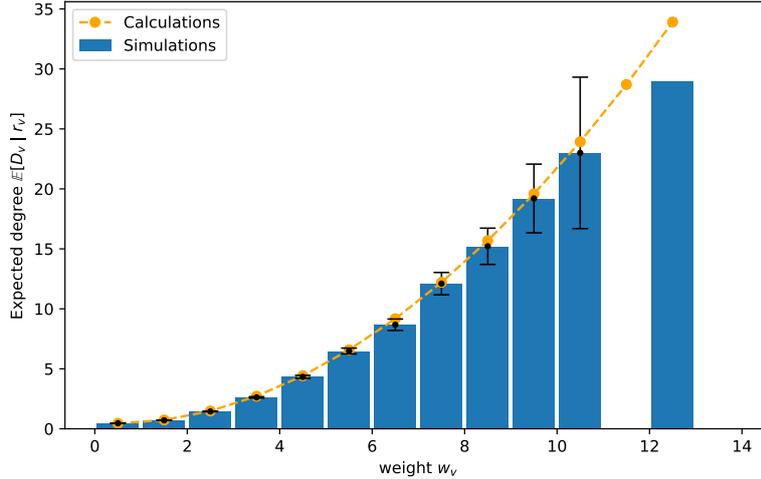


Figure 14. The expected degree of a vertex with weight w_v . The blue bar graph is based on simulations of the graph we give in Definition 2.1 with $n = 1000$, $\beta = 0.01 \log n$ and an exponential weight distribution with parameter $\lambda = 1$. Note that we split the continuous spectrum of weights in the simulations in bins. This means we can take an average of the expected degree for weights in this bin. The calculations are based on Lemma 2.3. The confidence intervals increase with weight w_v , because the probability that these weights occur is increasingly small. Less data points means less confidence. That is also why we miss data for weights between 11 and 12. There are no vertices with weights in this range. There is one vertex with a weight between 12 and 13. This means that we cannot calculate a confidence interval.

As c values we choose $\{0.6, 0.8, 0.9, 1.0\}$. We give the plot for the expected number of isolated vertices in Figure 15. From this figure, we conclude that for $c = 0.6$ and $c = 0.8$ the expected degrees clearly diverge as we had expected. We had expected $c = 0.9$ to diverge as well, but this plot is inconclusive. The errors are larger than the deviation. For $c = 1.0$ it converge to zero, which we also expected. We give the plot for the probability of an isolated vertex existing in Figure 16. For $c = 0.6$ and $c = 0.8$ we see an increasing trend. We expect this from Theorem 2.13. $c = 0.9$ is indecisive. The probability of an isolated vertex existing for $c = 1.0$ decreases, which confirms Proposition 2.12. For the exponential distribution with $\lambda = 1$, c^* is calculated as follows,

$$c^* = \frac{1}{\pi \mathbb{E}[W^2]} = \frac{1}{\pi} \frac{\lambda^2}{2} = \frac{1}{2\pi} \approx 0.16. \quad (5.2)$$

As c values we have chosen $\{0.10, 0.14, 0.18, 0.22\}$. We give the plot of the expected number of isolated vertices in Figure 17. From this figure we conclude that for $c = 0.10$ and $c = 0.14$ the expected degree diverges. For $c = 0.14$ this result is just about significant, as the confidence interval for $n = 800$ just excludes 1. For $c = 0.18$ and $c = 0.22$ the expected degrees converge to 0. Both of these are what we had expected from Theorem 2.11. We give the plot for the probability of an isolated vertex existing in Figure 18. We can see increasing plots for $c = 0.10$ and $c = 0.14$ and decreasing plots for $c = 0.18$ and $c = 0.22$, which is what we expected. Although $c = 0.18$ and $c = 0.22$ do not show very significant results. To conclude all isolated vertex convergence or divergence in Poisson spatial random graphs either confirm the proven statements or are indecisive.

5.2. Tribolium image analysis. The Tribolium image analysis was described in Section 4.4.1. This results in a degree distribution for all cells not at the boundary. We show the histogram in Figure 19. We compare these histograms to the results of the physical model in Section A.3. The model with a medium amount of relaxation time imitates the Tribolium cells best. Quantifying the amount of relaxation precisely is beyond this thesis.

5.3. Determinantal point processes.

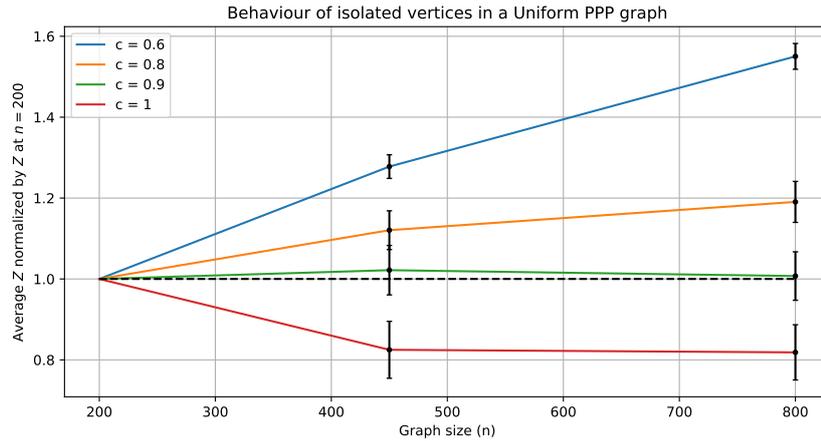


Figure 15. The expected number of isolated vertices in spatial Boolean graphs normalized by the number of isolated vertices at graph size $n = 200$. We define the spatial Boolean graphs in Definition 2.1. We choose the spatial Boolean graphs with different sizes n and different values of c , where $\beta = c \log n$. For the weights we choose from a uniform distribution between 0 and 1. In this figure the c values plotted are $\{0.6, 0.8, 0.9, 1.0\}$. For each of these c values, we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. We show the original 95% confidence intervals derived before normalization.

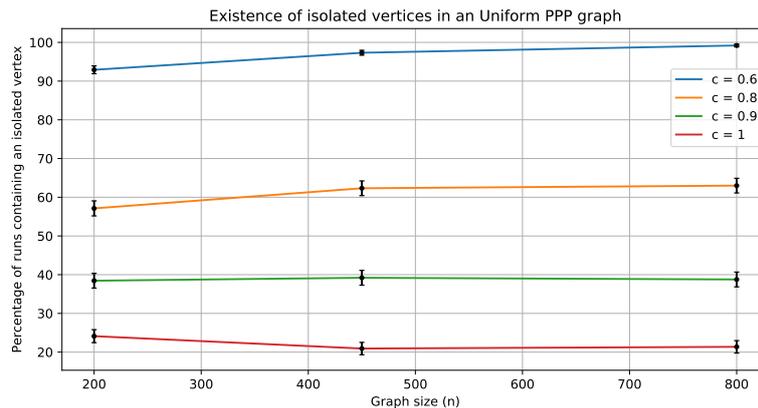


Figure 16. The percentage of runs of a spatial Boolean graph that contain an isolated vertex. This is for some graph size n and constant c . The graph size n is on the horizontal axis. For these graphs $\beta = c \log n$. The weights for these graphs are chosen from a uniform distribution between 0 and 1. The c values plotted in this figure are $\{0.6, 0.8, 0.9, 1.0\}$. For each of the values of c , we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. Plotted are the percentages of these 2500 graphs that contain an isolated vertex. The error bars are the 95% confidence intervals.

5.3.1. *Comparison of Poisson and Determinantal point processes with Tribolium.* The repulsion in determinantal point processes is calculated according to (4.3). This is done for 100 realisations of a Poisson point process and 100 realisations of a determinantal point process, constructed as described in Section 4.2. The repulsion values are plotted in Figure 20. The central limit theorem tells us that such a set of repulsion values can be described by a Gaussian distribution. We also show the Gaussian fit in this figure. In Table 1 we give the sample mean and sample variance.

We show an example Tribolium nuclei in Figure 11b. This results from image analysis on Figure 11a. In Table 1 we give the results of repulsion calculation for 7 Tribolium images. We see that the Tribolium has significantly larger repulsion than both point processes.

5.3.2. *Threshold simulations for Determinantal point processes.* In this section we give the results regarding the simulations described in Section 4.2.3. We run these simulations for both a uniform

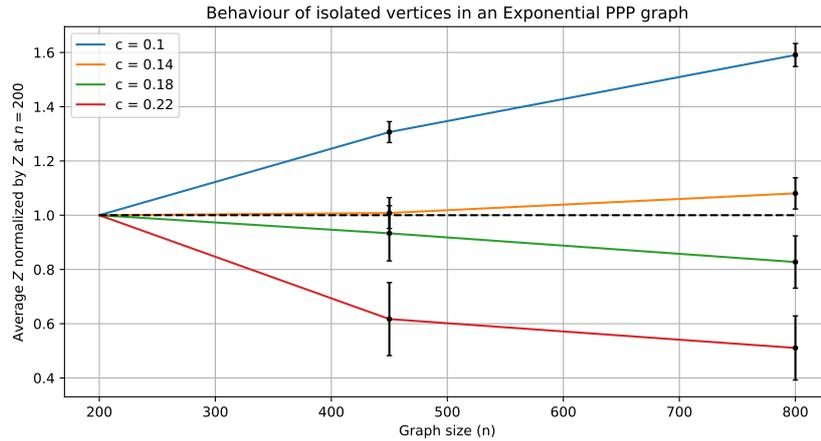


Figure 17. The expected number of isolated vertices in spatial Boolean graphs normalized by the number of isolated vertices at graph size $n = 200$. The spatial Boolean graphs are defined in Definition 2.1. The spatial Boolean graphs is chosen with different sizes n and different values of c , where $\beta = c \log n$. We choose the weights from an exponential distribution with $\lambda = 1$. In this figure the c values plotted are $\{0.10, 0.14, 0.18, 0.22\}$. For each of these c values, we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. We show the original 95% confidence intervals derived before normalization.

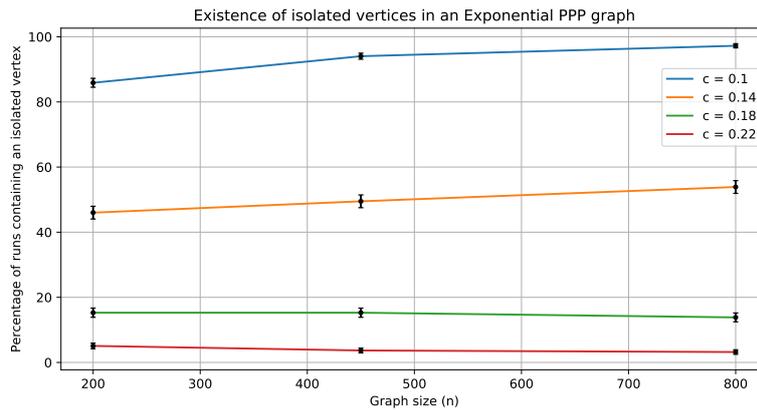


Figure 18. The percentage of runs of a spatial Boolean graph that contain an isolated vertex. This is for some graph size n and constant c . The graph size n is on the horizontal axis. For these graphs $\beta = c \log n$. The weight distribution for this graph is an exponential distribution with $\lambda = 1$. The c values plotted in this figure are $\{0.10, 0.14, 0.18, 0.22\}$. For each of the values of c , we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. Plotted are the percentages of these 2500 graphs that contain an isolated vertex. The error bars are the 95% confidence intervals.

Table 1. Mean and Variance of the repulsion of 100 Poisson point processes, 100 determinantal point processes and 7 Tribolium images. And example of such a Tribolium image is given in Figure 11a.

Point process	Mean	Variance
Poisson	0.50	0.01
Determinantal	0.58	0.01
Tribolium	0.88	0.02

distribution, $\text{Unif}(0, 1)$, and an exponential distribution with $\lambda = 1$, $\text{Exp}(1)$, as weight distributions. In both cases we calculate the theoretically derived threshold constants c^* for Poisson point processes, Theorem 2.11. We choose values of c around c^* . We expect the threshold of isolated vertex existence

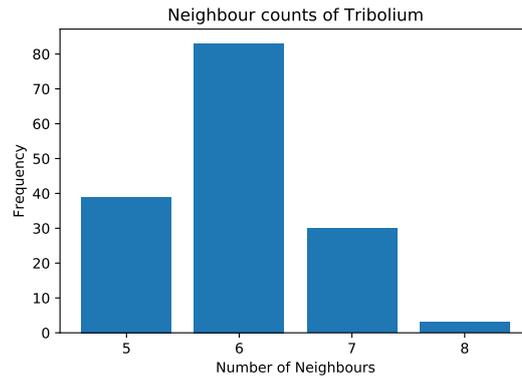


Figure 19. A histogram of the number of neighbours per cell for the Tribolium cells in Figure 11a. This data is generated using image analysis. The total number of cells in the image of the Tribolium cells is 201. Therefore the total frequency should add up to 201. We only did this analysis for one image.

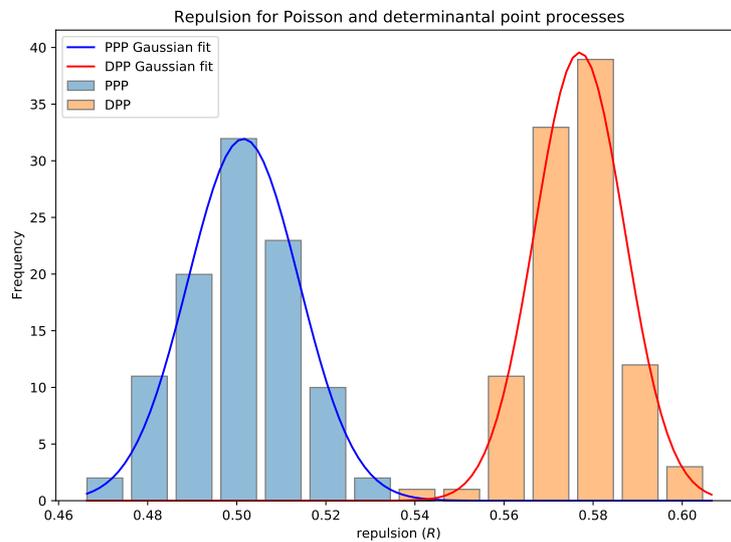


Figure 20. The repulsion of both Poisson and determinantal point processes. The repulsion of 100 determinantal point processes and 100 Poisson point processes was calculated using (4.3). The determinantal point processes consisted of 450 points and the Poisson point processes consisted of around 450 points, Poisson distributed. The Poisson point process is an independent point process and should on average obtain a repulsion of 0.5. This is clearly visible in the blue histogram and also in the Gaussian approximation of this histogram. The mean and sample variance was calculated and with these the Gaussian approximation was constructed. This distribution can be approximated using the Gaussian distribution because of the central limit theorem. The orange histogram and Gaussian are for the determinantal point process. As visible, the determinantal point process is clearly more repulsive.

in determinantal graphs to be lower than the analytical thresholds for Poisson graphs. The repulsion between the points makes their spread more regular, which decreases the probability that a vertex is isolated. Therefore we would require a lower β and thus a lower c . Therefore the simulations for values of c smaller than c^* are most important. The plots of the expected number of isolated vertices are normalized with respect to the expected number at graph size 200, just like in Section 5.1.2. This allows us to show and compare all of them in one figure.

First for the uniform distribution we calculated $c^* \approx 0.95$. As c values we have chosen $\{0.6, 0.8, 1.0, 1.2\}$. We give the plot for the expected number of isolated vertices in Figure 21. From this figure, we conclude that for $c = 0.6$ the expected degrees clearly diverge. In the Poisson case we expect $c = 0.8$ to diverge as well, but in this determinantal case, this does not diverge. It actually converges. Both $c = 1.0$ and $c = 1.2$ converge as well. These we expect to converge, because the threshold c decreases compared to

the Poisson case. This decrease should only make the convergence of these c -values stronger. We give the plot for the probability of an isolated vertex existing in Figure 22. For $c = 0.6$ this again increases, following the Poisson case. $c = 0.8$ decreases, but is quite indecisive. And the probability of an isolated vertex existing for $c = 1.0$ and $c = 1.2$ decreases, which we expect.

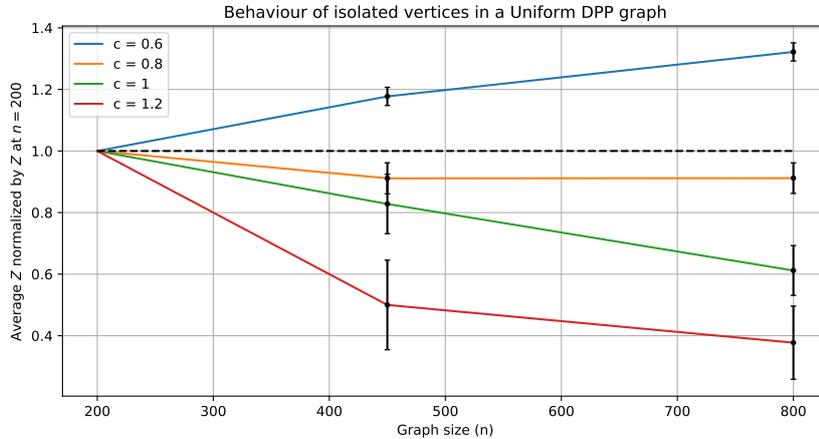


Figure 21. The expected number of isolated vertices in determinantal spatial Boolean graphs normalized by the expected number of isolated vertices for graph size $n = 200$. The spatial Boolean graphs are defined in Definition 2.1, but these are using a determinantal point process instead of a Poisson point process. Determinantal point processes are described in Section 4.2. The spatial Boolean graphs is chosen with different sizes n and different values of c , where $\beta = c \log n$. The weights are chosen from a Uniform distribution between 0 and 1. In this figure the c values plotted are $\{0.6, 0.8, 1.0, 1.2\}$. For each of these c values, we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. The data shown in the figure is the average number of isolated vertices in these graphs. The error bars are the 95% confidence intervals. These should not be trusted completely, because the determinantal point processes used are not completely independent. Hence the basic central limit theorem is not applicable here. Also they are the original confidence intervals before normalization.

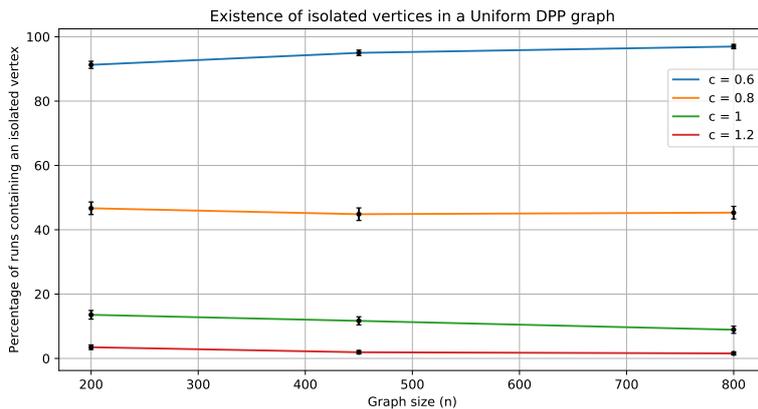


Figure 22. The percentage of runs of a determinantal spatial Boolean graph that contain an isolated vertex. This is for some graph size n and constant c . The graph size n is on the horizontal axis. For these graphs $\beta = c \log n$. The weight distribution for this graph is a uniform distribution between 0 and 1. The c values plotted in this figure are $\{0.6, 0.8, 0.9, 1.0\}$. For each of the values of c , we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. Plotted are the percentages of these 2500 graphs that contain an isolated vertex. The error bars are the 95% confidence intervals. These should not be trusted completely, because the determinantal point processes used are not completely independent. Hence the basic central limit theorem is not applicable here.

For the exponential distribution with $\lambda = 1$, $c^* \approx 0.16$. As c values we have chosen $\{0.10, 0.14, 0.15, 0.18, 0.22\}$. We give the plot of the expected number of isolated vertices in Figure 23. From this figure we conclude that for $c = 0.10$ the expected degree diverges. For $c = 0.14$ and $c = 0.15$ the value already converges. In the Poisson case we expect this to diverge. However, the case $c = 0.14$ is not so convincing. For $c = 0.18$ and $c = 0.22$ the expected degrees converge to 0. We expect this from Theorem 2.11 for the Poisson case and the fact that we expect the Determinantal threshold to be lower. We give the plot for the probability of an isolated vertex existing in Figure 24. Here as well we see an increasing trend for $c = 0.10$, which we expect. $c = 0.14$ and $c = 0.15$ are indecisive. For $c = 0.18$ and $c = 0.22$ the trend is decreasing, which we expect.

Both for the uniform and the exponential distribution, the threshold has decreased. Lower values of c give converging numbers of isolated vertices. In the uniform case the Poisson threshold is $c^* \approx 0.95$. From the simulations we conclude that the determinantal threshold c^* is likely between 0.6 and 0.8. In the exponential case the Poisson threshold is $c^* \approx 0.16$. From the simulations we conclude that the determinantal threshold c^* is likely below 0.14. It is definitely above 0.1. We conclude that in both cases the threshold dropped significantly. A decreasing threshold is likely, because in a more evenly spread point set, the probability that a vertex is not connected is smaller. This is because the probability of a vertex being in an area with a very low vertex density is small. We conclude that the Poisson threshold can be used as an upper bound to the connectivity threshold for determinantal spatial Boolean graphs. Therefore, when c is above the Poisson threshold for determinantal spatial Boolean graphs, the graph is connected in the limiting case of infinite graph size.

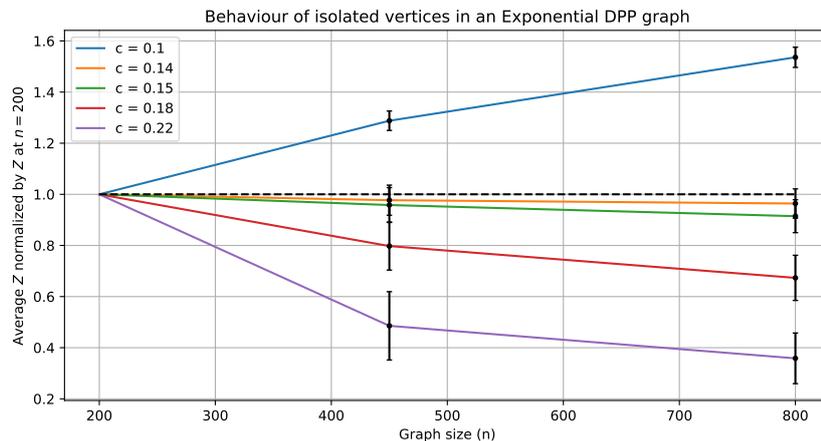


Figure 23. The expected number of isolated vertices in determinantal spatial Boolean graphs normalized by the expected number of isolated vertices for graph size $n = 200$. We define the spatial Boolean graphs in Definition 2.1, but in this figure we use a determinantal point process instead of a Poisson point process. We describe determinantal point processes in Section 4.2. We choose the spatial Boolean graphs with different graph sizes n and different values of c , where $\beta = c \log n$. For the weight distribution we use an exponential distribution with $\lambda = 1$. In this figure the c values plotted are $\{0.10, 0.14, 0.18, 0.22\}$. For each of these c values, we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. In the figure we show is the average number of isolated vertices in these graphs. The error bars are the 95% confidence intervals. These should not be trusted completely, because the determinantal point processes used are not completely independent. Hence the basic central limit theorem is not applicable here. Also they are the original confidence intervals before normalization.

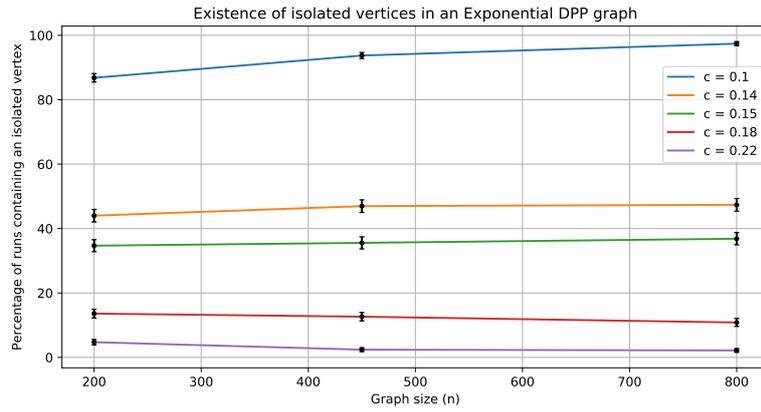


Figure 24. The percentage of runs of a determinantal spatial Boolean graph that contain an isolated vertex. This is for some graph size n and constant c . The graph size n is on the horizontal axis. For these graphs $\beta = c \log n$. The weight distribution for this graph is an exponential distribution with $\lambda = 1$. The c values plotted in this figure are $\{0.10, 0.14, 0.18, 0.22\}$. For each of the values of c , we run 2500 simulations of spatial Boolean graphs for graph size 200, 450 and 800. We plot the percentages of these 2500 graphs that contain an isolated vertex. The error bars are the 95% confidence intervals. These should not be trusted completely, because the determinantal point processes used are not completely independent. Hence the basic central limit theorem is not applicable here.

6. CONCLUSION

In this thesis we investigated the property of connectivity of certain graphs. We argued why these graphs could represent cellular structures with some small adaptation. Finally we saw how this adaptation influences the statement about connectivity made about the original graph. The connectivity of similar graphs has been linked to the existence of isolated vertices, vertices with no connections, by Matthew Penrose [26]. We showed the existence of a phase transition in the number of isolated vertices in a spatial Boolean graph as the graph size tends to infinity. We also showed that the threshold is $\beta = c^* \log n$. The threshold β can be interpreted as time, using $\beta = t^2$. We proved that c^* can be calculated using the second moment of the growth speed distribution, $c^* = \frac{1}{\pi \mathbb{E}[W^2]}$. For $\beta > c^* \log n$, the probability of an isolated vertex existing tends to zero. And for $\beta < c^* \log n$, the probability tends to one. If we translate β to t , the transition is at $t = \sqrt{c^* \log n}$. We confirmed this transition for a uniform and exponential growth speed distribution with simulations that are given in Section 5.1.2.

We introduced determinantal point processes in Section 4.2 and a repulsion measure in Section 4.4.2. With this measure we showed that the repulsion within our determinantal point sets ($R = 0.58 \pm 0.01$) is significantly greater than in the Poisson point processes (0.50 ± 0.01). We also concluded that the real world example of a cellular structure, Tribolium, had a significantly higher repulsion ($R = 0.88 \pm 0.02$), than both point processes.

We ran simulations for determining the threshold for isolated vertex existence of spatial Boolean graphs with determinantal point processes. We gave results of these simulations for the uniform and exponential growth speed distribution in Section 5.3.2. We compared these with the Poisson graph simulations which confirmed the mathematical results. We concluded that the threshold for the determinantal graphs significantly differs from that of Poisson point processes. The Poisson threshold can be used as an upper bound for the determinantal graphs. More mathematical analysis into spatial Boolean graphs with determinantal point processes might result in an analytical proof of an isolated vertex existence phase transition. More research into determinantal point processes can lead to a faster way of constructing point processes with a certain repulsion. These would both be beneficial for the simulation of cellular structures. With which we might be able to eventually recognize abnormal and endangering cell structures.

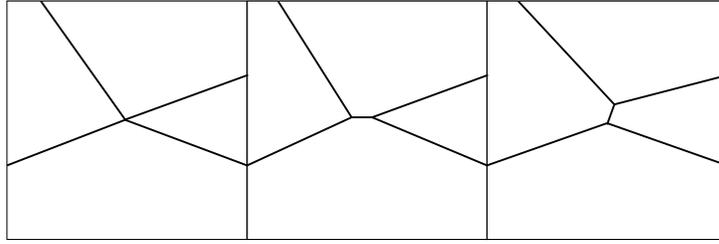


Figure 25. The unlikelihood of there being a point in the 2d plane where more than 3 regions meet. The left image shows such a point. There is no boundary between the top and bottom region and no boundary between the left and right region, because the 4 regions meet exactly at one single point. The middle image shows the top and bottom region sharing a boundary and the right and left regions sharing a boundary. These both result in the point where 4 regions meet, splitting into two points where 3 regions meet.

APPENDIX A. PROOFS OF CLAIMS

A.1. Proof of claim.

Proof of Claim 3.2. Take a δ such that $0 < \delta < \delta_0$. Then $c < \frac{1}{g(x_0 + \delta)}$. We know $g(x_0 + \delta) \geq g(x_0) + \epsilon$ for some $\epsilon > 0$. Hence,

$$c < \frac{1}{g(x_0 + \delta)} \leq \frac{1}{g(x_0) + \epsilon} \leq \frac{1}{g(x_0)}. \quad (\text{A.1})$$

This completes the proof. \square

A.2. Intuition for the average number of neighbours in a tiling of the 2d plane. Suppose we have a tiling of the 2d plane, meaning the 2d plane is divided into connected regions with no space left. Also suppose there are no points in the 2d plane where more than 3 regions meet. This is a logical assumption, because for this to occur coincidentally, it would be really unlikely. There would always be some small boundary between two opposite areas, like in figure 25. This figure shows that a point where 4 regions meet, only happens when 2 points of 3 meeting regions overlap. This overlapping is theoretically possible, but would in practice be not likely. An example of such a tiling is given in Figure 26. These assumptions result in the average number of neighbours of a region in such a tiling being 6. We can give an intuition on the proof. First, we need to go from a tiling to a graph. In order to do this, we will turn regions into vertices. Two regions with a shared boundary get an edge between them. The tiling in Figure 26 is transformed into such a graph in Figure 27. The points where 3 regions meet in the tiling will become the faces of the graph. Hence these faces all have exactly 3 edges surrounding them. We also know each edge belongs to two faces. Let us denote the number of edges by e , the number of faces by f and the number of vertices by v . We discovered that $2e = 3f$, because each face has 3 associated edges and each edge has 2 associated faces. Euler's formula tells us $v - e + f = 2$. Substituting one into the other results in:

$$2e = 3f = 3(2 + e - v) = 6 + 3e - 3v. \quad (\text{A.2})$$

Hence we get the following:

$$3v - e = 6 \quad (\text{A.3})$$

$$2e = 6v - 12 \quad (\text{A.4})$$

$$\frac{2e}{v} = 6 - \frac{12}{v} \quad (\text{A.5})$$

Know for a real tiling the number of vertices, edges and faces would be infinite. So these formula's would not make sense. Therefore this is just an intuition. But in the number of vertices v being infinite, would mean we can remove the $\frac{12}{v}$ term, which results in $\frac{2e}{v} = 6$. Hence the average number of edges per vertex is 6. Translating back to the tiling, this would mean the average number of neighbours is 6.

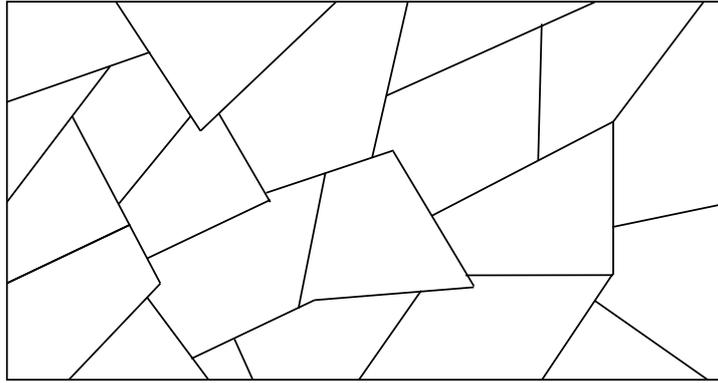


Figure 26. An example of a tiling of the 2d plane. This tiling contains no points where 4 or more regions meet. In Figure 27 this graph is transformed into a graph. Note that this is not a full infinite tiling, but just a section.

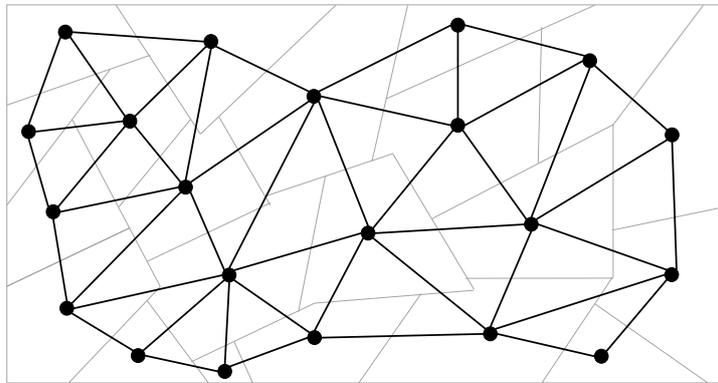


Figure 27. The transformation into a graph of the tiling given in Figure 26. In this tiling the regions are transformed into vertices and two regions with a shared boundary will have an edge between them. Note that this is not a full infinite tiling, but just a section. The edges to the regions outside of the visible section have not been drawn.

A.3. Physical model. In the first part of this project we started working on a physical model which could be imagined somewhere between the graph theoretical model and the experimental data shown in Figure 11a. It is constructed in c++ and consisted of cells starting of as points initialized by a Poisson point process. They live on a periodic box, \mathbb{T}_n^2 , with n the number of cells. Every cell will also have an initial growth speed. This growth speed is chosen from some distribution just like the growth speed from the spatial Boolean graphs. In this model the cells start growing as circles with increasing radius. This growth is done in time steps. Every step consists of a growth phase and a relaxation phase. In the growth phase the cells grow with $dt * v$, with dt the time step and v the growth speed of the cell. In the relaxation phase the forces get a change to act. The cells push each other away according to Hook's law. Hence the force from one cell onto the other is proportional to the overlap of the radii. The environment was assumed very viscous, meaning we include the viscous drag force into the force and get

$$F = \xi v - kd. \quad (\text{A.6})$$

With ξ some drag constant, k the spring constant, d the overlap between the two cells/circles and v the speed of the cell. Applying newton to this force gives

$$F = \xi v - kd = ma. \quad (\text{A.7})$$

Because of high viscosity, high resistance to movement, in the environment the ma term can be neglected in this equation. ξv and kd are both large compared to ma . We get the following equation

$$\frac{dx}{dt} = \frac{kd}{\xi}. \quad (\text{A.8})$$

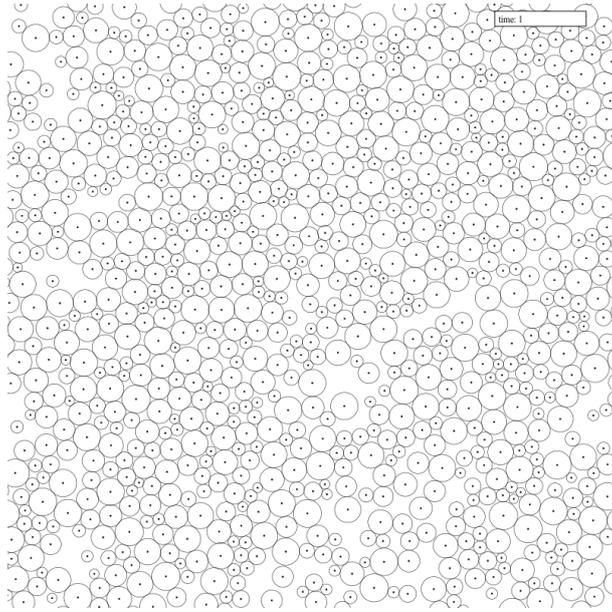


Figure 28. The physical model as described in Section A.3. This model lives on a periodic square box. It contains 1000 cells or vertices. The cells grow and push each other out of the way with a spring force. The growth speed is chosen from a uniform random distribution between 0.5 and 1.5. The cells grow when they are under a certain pressure threshold. This image shows a step which is already quite close to the end of the growth. The points start out with zero radius. In this time step there is at least one cell which has no touching cells, so we could say this structure is not connected yet.

In the simulations we will be using a non-zero time step dt . The displacement of the cell for this time step can be calculated,

$$dx = \frac{kd}{\xi} dt. \quad (\text{A.9})$$

The cells stop growing when they accede a certain pressure. The pressure is calculated as the sum of the absolute value of the forces that neighbouring cells apply to a cell. The amount of relaxation after each growth step has a significant influence on the final structure. We give an example of this physical model during some time step in its growth in Figure 28.

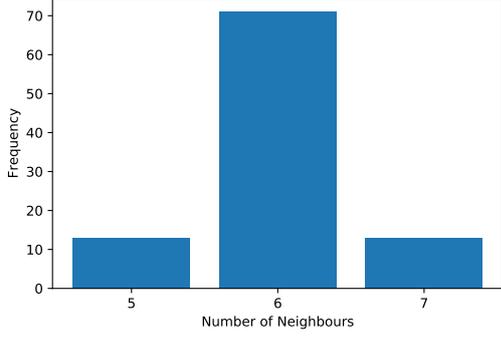
The structure is measured using the degree distribution/ neighbour distribution, just as calculated for the Tribolium cells. Two cells are defined as neighbours if their radii overlap. A lot of relaxation will result in the cells being able to settle completely after each step. This causes the cells to be more crystallized. This crystallization will result in a very narrow degree distribution, with most cells having either 5, 6 or 7 neighbours. In A.2 we give an intuition for why this number is 6 on average. With a small amount of relaxation after each growth step, the cells have less opportunity to crystallize. This will result in a wider distribution. We show results of this in Figure 29.

This growth implies that the threshold occurs at some point in time, which can be derived from $\beta = t^2$. The difference between this model and the spatial Boolean graphs is that cell push each other and therefore move around. This causes a repulsion effect between the cells. The cells will to be more evenly distributed. This does not happen in the spatial Boolean graphs. Therefore we started looking at more repulsive initialization of the points. More on this in Section 4.2.

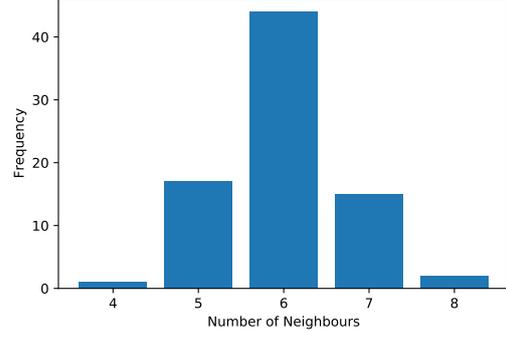
A.4. Confidence intervals. In this thesis we often run a lot of simulations and extract some value from all of the simulations. These simulations are, in most cases, independent. Extracting this value from a simulation where random processes are involved is like taking a realisation of a some random variable, albeit a quite complex random variable. Therefore we choose this list of values from a set of independent identical distributions. These are exactly the conditions for applying the central limit theorem. With the central limit theorem we calculate confidence intervals.

Suppose we have a sequence of i.i.d. random variables $\{X_1, X_2, \dots, X_n\}$. Define \bar{X} as

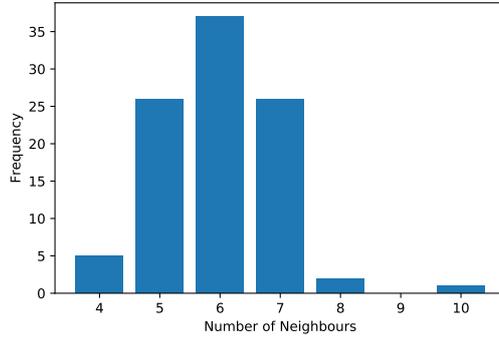
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (\text{A.10})$$



(a) The neighbourcount of the Voronoi cells resulting from the model with plenty of relaxation steps and relaxation time after reaching the growth-stop threshold pressure.



(b) The neighbourcount of the Voronoi cells resulting from the model with medium relaxation steps and relaxation time after reaching the growth-stop threshold pressure.



(c) The neighbourcount of the Voronoi cells resulting from the model with 1 relaxation step and almost no relaxation time after reaching the growth-stop threshold pressure.

Figure 29. The neighbour count distribution for the physical model described in Section A.3. The vertical axis gives the frequency of a cell having the corresponding number of neighbours. The three figures all have different amounts of relaxation between growth steps. With low relaxation we expect it to crystallize less. Therefore we expect a wider distribution. For more relaxation we expect more crystallization and thus a narrower distribution.

Take $\{x_1, x_2, \dots, x_n\}$ a set of realisations of the random variables $\{X_1, X_2, \dots, X_n\}$. Define $\mu = \frac{1}{n} \sum_{i=1}^n x_n$ as their average. And $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$ as their standard deviation. By the central limit theorem

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right). \quad (\text{A.11})$$

Rearranging this will give the following,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1). \quad (\text{A.12})$$

For the standard Gaussian distribution we know the confidence intervals. The above tells us that,

$$\begin{aligned} \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in \left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)\right) &\cong \mathbb{P}\left(\mathcal{N}(0, 1) \in \left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)\right) \\ &= \mathbb{P}\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) \\ &= \mathbb{P}\left(\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right). \end{aligned} \quad (\text{A.13})$$

For a 95% confidence interval of a standard Gaussian we have $\alpha = 5$. Thus we need $z_{2.5} = 1.96$. Hence the confidence intervals are constructed:

$$\left(\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right). \quad (\text{A.14})$$

REFERENCES

- [1] Ruben van Drongelen, Tania Vazquez-Faci, Teun APM Huijben, Maurijn van der Zee, and Timon Idema. Mechanics of epithelial tissue formation. *Journal of Theoretical Biology*, 454:182–189, 2018.
- [2] Centraal bureau voor de statistiek. 54 procent sterfgevallen in 2019 door kanker of hart- en vaatziekten. <https://www.cbs.nl/nl-nl/nieuws/2020/27/54-procent-sterfgevallen-in-2019-door-kanker-of-hart-en-vaatziekten#:~:text=Longkanker%20blijft%20het%20meest%20voorkomende,duizend%20mensen%20stierven%20aan%,20dikkedarmkanker.>, 2020.
- [3] National Cancer Institute. What is cancer? <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>, 2021.
- [4] Edmund Rolls and Alessandro Treves. Neural networks and brain function. 1997.
- [5] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.
- [6] Krishnan Balasubramanian and Satya P Gupta. Quantum molecular dynamics, topological, group theoretical and graph theoretical studies of protein-protein interactions. *Current topics in medicinal chemistry*, 19(6):426–443, 2019.
- [7] C Godreche, I Kostov, and I Yekutieli. Topological correlations in cellular structures and planar graph theory. *Physical review letters*, 69(18):2674, 1992.
- [8] Salvatore Catanese, Pasquale De Meo, Emilio Ferrara, and Giacomo Fiumara. Analyzing the facebook friendship graph. *arXiv preprint arXiv:1011.5168*, 2010.
- [9] Jennifer A Dunne. The network structure of food webs. *Ecological networks: linking structure to dynamics in food webs*, pages 27–86, 2006.
- [10] Sambor Guze. Graph theory approach to the vulnerability of transportation networks. *Algorithms*, 12(12):270, 2019.
- [11] Tancredi Caruso, Matthias C Rillig, and Diego Garlaschelli. On the application of network theory to arbuscular mycorrhizal fungi–plant interactions: the importance of basic assumptions. *New phytologist*, 194(4):891–894, 2012.
- [12] Kunio Murasugi. On invariants of graphs with applications to knot theory. *Transactions of the American Mathematical Society*, 314(1):1–49, 1989.
- [13] Wikipedia. Network theory — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Network%20theory&oldid=1160501078>, 2023. [Online; accessed 20-June-2023].
- [14] Lei Shi, Yifan Zhang, Jian Cheng, and Hanqing Lu. Skeleton-based action recognition with directed graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7912–7921, 2019.
- [15] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 43. Cambridge university press, 2016.
- [16] Edward N Gilbert. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961.
- [17] Paul Balister, Amites Sarkar, and Béla Bollobás. Percolation, connectivity, coverage and colouring of random geometric graphs. *Handbook of large-scale random networks*, pages 117–142, 2008.
- [18] Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.
- [19] Mathew D Penrose. On the spread-out limit for bond and continuum percolation. *The Annals of Applied Probability*, 3(1):253–276, 1993.
- [20] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262, 1999.
- [21] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for power law graphs. *Experimental mathematics*, 10(1):53–66, 2001.
- [22] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002.
- [23] Ilkka Norros and Hannu Reittu. On a conditionally poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.
- [24] Christian Hirsch. From heavy-tailed boolean models to scale-free gilbert graphs. 2017.
- [25] Piyush Gupta and Panganamala R Kumar. Critical power for asymptotic connectivity in wireless networks, 1999.
- [26] Mathew D Penrose. Connectivity of soft random geometric graphs. 2016.
- [27] Wikipedia. Poisson point process — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Poisson%20point%20process&oldid=1151448460>, 2023. [Online; accessed 21-June-2023].
- [28] Giovanni Peccati and Matthias Reitzner. *Stochastic analysis for Poisson point processes: Malliavin calculus, Wiener-Itô chaos expansions and stochastic geometry*, volume 7. Springer, 2016.
- [29] Pim Kloet. Bep pim kloet. <https://gitlab.tudelft.nl/ig-projects/bep-pim-kloet>, 2023.
- [30] H Paul Keeler. Notes on the poisson point process. *University of Melbourne, Tech. Rep*, 2015.
- [31] Simon Barthélémy. Determinantal point processes: a quick introduction. https://barthesi.gricad-pages.univ-grenoble-alpes.fr/personal-website/dpps/2018-26-11-dpps_intro/, 2018.
- [32] Alexei Borodin. Determinantal point processes. *arXiv preprint arXiv:0911.1153*, 2009.

- [33] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.
- [34] Franz Aurenhammer and Rolf Klein. Voronoi diagrams. *Handbook of computational geometry*, 5(10):201–290, 2000.
- [35] Paul Hertz. Über den gegenseitigen durchschnittlichen abstand von punkten, die mit bekannter mittlerer dichte im raume angeordnet sind. *Mathematische Annalen*, 67(3):387–398, 1909.
- [36] Philip J Clark and Francis C Evans. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4):445–453, 1954.
- [37] Matt Dry, Kym Preiss, and Johan Wagemans. Clustering, randomness and regularity: Spatial distributions and human performance on the traveling salesperson problem and minimum spanning tree problem. *The Journal of Problem Solving*, 4(1):1–17, 2012.