Isolation and Connectivity in Networks of Autonomous Entities

Michael Wilsher

University of Bristol
EPSRC Centre for Doctoral Training in Communications

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Engineering

March 2021

Abstract

With the rapid growth of the number of wireless communications devices in the last two decades, device-to-device networks are becoming increasingly important. One sub-group of these are networks of autonomous entities where there is no centralised controller and users actions are based solely on local interactions. The question of connectivity is of paramount importance and is often seen as a pre-requisite of such systems. This thesis will concentrate on this question for two such networks: Vehicular Ad-hoc NETworks (VANETs) and Swarm Robotics.

Connectivity is crucial in VANETs due to the safety critical nature of the information being broadcast. The current state of the art is to model them using a spatial stochastic network model called a one-dimensional hard Random Geometric Graph (RGG). Using a Poisson point process to represent the vehicles in the network, pairs of points are connected if their mutual distance is less than a specified radius. We analyse an extension of this model which incorporates the randomness of the wireless medium by now using a distance dependent, probabilistic connection function to determine the existence of edges: a 1-D *soft RGG*.

We derive bounds on the probability that this graph is fully connected for a large class of connection functions by analysing two key barriers to full connectivity: isolated nodes and uncrossed gaps. Firstly, analytic expressions are derived for the mean and variance of the number of isolated nodes, and a sharp threshold established for their occurrence. At this threshold we show that the number of isolated nodes can be well approximated by a Poisson random variable. Bounds are then derived for uncrossed gaps, and it is shown analytically that uncrossed gaps have negligible probability in the scaling at which isolated nodes appear. The critical scaling for the occurrence of uncrossed gaps is also investigated.

Connectivity is also also a key feature in robotic swarms since it enables them to complete tasks as a group which would be impossible for individual robots. We derive an algorithm, implemented on virtual robots, to mimic the behaviour of a group of spiders whose social interactions determine their disposition to travel to a dangerous, unknown area. Simulations of the proposed algorithm demonstrate that this social behaviour improves the swarms ability to explore an unknown environment whilst maintaining suitable levels of connectivity.

List of Publications

Published

- Michael Wilsher, Carl P Dettmann, and Ayalvadi Ganesh. "Connectivity in one-dimensional soft random geometric graphs".*Physical Review E* vol. 102, no. 6, (2020), p. 062312.

- Edmund R Hunt, George Jenkinson, Michael Wilsher, Carl P Dettmann,and Sabine Hauert. "SPIDER: a bioinspired swarm algorithm for adaptive risk-taking".*Artificial Life Conference Proceedings*. MIT Press. 2020, pp. 44–51.

In Preparation

- Michael Wilsher, Carl P Dettmann, and Ayalvadi Ganesh. "The Distribution of the Number of Isolated Nodes in the 1-Dimensional Soft Random Geometric Graph"

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

To Nanny Joan and Nanny Eileen.

Acknowledgements

I would like to start by thanking Carl Dettmann and Ayalvadi Ganesh for both their academic as well as pastoral support. They have been an invaluable source of knowledge and guidance as well as supporting me through some very difficult times. I would also like to thank all of the other members of my cohort in the Communications CDT for making the taught year very enjoyable, and especially for their constant help when it came to learning the myriad of acronyms used within Electrical Engineering. The organisers of the CDT, and especially Suzanne, were also always on hand to help with any queries I had, no matter how mundane. Within the Mathematics department there are also too many people to thank individually but who have all made the last four years incredibly enjoyable and inspiring. The University of Bristol Volleyball club has also been an integral part of my University life and allowed me to take out many of my frustrations as well as giving me friends for life.

Finally, I would like to thank my family for always being there for me and allowing me to bore them by attempting to explain my research. In particular to my dad for helping me with the fourth line on page 4, to my mum for years of doing my maths homework with me, and to Matthew and Claire for not always being annoying.

Contents

List of Figures

List of Tables

List of Notation

$\mathbb{N}$	The set of integers
$\mathbb{N}_0$	The set of non-negative integers
$\mathbb{R}$	The set of real numbers
$\mathbb{P}(A)$	Probability of the event A
$\mathbb{E}[A]$	Expectation of the event A
$\mathcal{B}$	Borel set
c_V^2	Coefficient of variation
δ_x	Dirac measure at x
d_{TV}	Total variational distance
η	Path loss exponent
$\mathcal{G}(\Phi, r_0)$	Hard random geometric graph with vertex set Φ and connection range r_0
$\mathcal{G}_H(\Phi)$	Soft random geometric graph with vertex set Φ and connection function H
$H(r)$	Connection function: Probability of two nodes a distance r apart being connected
$H^L(\cdot)$	Scaled connection function
$\mathcal{L}_\Phi$	Laplace functional of the point process Φ
Λ	Intensity measure (of a point process)
P_x	Palm measure conditional on the Poisson process having a point at x
Φ	Point process (described as a random set)
$\mathcal{P}$	Poisson point process
$\mathcal{R}$	Erdős–Rényi random graph
$\rho(x, y)$	The circular (or toroidal) distance between x and y
$\rho_\infty(x, y)$	The Euclidean distance between x and y

Acronyms

ASTM	American Society for Testing and Materials
CSMA/CA	Carrier-Sense Multiple Access with Collision Avoidance
DSRC	Dedicated Short Range Communication
EA	Evolutionary Algorithm
GPS	Global Positioning System
GPSR	Greedy Perimeter Stateless Routing
MAC	Medium Access Control
MGF	Moment Generating Function
PBCs	Periodic Boundary Conditions
PDF	Probability Distribution Function
PLP	Poisson Line Process
PPP	Poisson Point Process
QoS	Quality of Service
RGG	Random Geometric Graph
RSU	Road Side Unit
RWP	Random Waypoint
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
TVWS	TV White Space
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VANET	Vehicular Ad-Hoc Networks
WAVE	Wireless Access in Vehicular Environments

Asymptotic Notation

Within this thesis, we will be using the standard asymptotic notation. These are defined as follows:

$$f(x) = O(g(x)) \text{ at } a \quad \text{if } \limsup_{x \to a} \left| \frac{f(x)}{g(x)} \right| < \infty,$$

where a is often clear from context and thus omitted (with a typically taking the value of 0 or ∞). Similarly,

$$f(x) = o(g(x)) \text{ at } a \quad \text{if } \limsup_{x \to a} \left| \frac{f(x)}{g(x)} \right| = 0.$$

Furthermore,

$$\begin{aligned} f(x) &= \omega(g(x)) && \text{if } g(x) = o(f(x)), \\ f(x) &= \Omega(g(x)) && \text{if } f(x) = O(g(x)) \text{ and } g(x) = O(f(x)), \\ f(x) &\sim_x g(x) && \text{if } \lim_{x \to \infty} \frac{f(x)}{g(x)} = 1. \end{aligned}$$

Chapter 1

Introduction

Connectedness has fast become an indispensable part of modern life. Whether this means simply connecting to the public Wi-Fi at your local coffee shop, or it means that a sensor in a pipeline is able to communicate a critical failure back to a central computer. Although these have differing levels of importance, they both rely on the same underlying technologies and fundamental framework. Communication networks have allowed for the rapid geographical expansion of our everyday lives as well as advancements in technologies that improve our safety, help us to stay connected, and even enable us to watch endless reruns of our favourite TV shows from the comfort of our bedrooms.

Wireless communication networks give a way for data to be transferred between machines or users without the need for expensive cabling to every end user in the network. The traditional cellular network set up consists of base stations (BSs) and end users or devices. To transfer data between end users, one must firstly transfer that data to the nearest BS where it is then routed via the backhaul network to the nearest BS to the destination user where it can be sent wirelessly to the intended target. This is the same idea as the Wi-Fi network one may have at home in which devices wirelessly connect to a central router. This router is then connected to the backhaul network via some form of Ethernet connection. In the traditional set up, one BS will service many users over a large geographical area. Maintaining a connected network requires an incredibly robust and well designed network architecture. The large number of users connected to one BS means that there is a large amount of strain on this infrastructure due to factors such as interference in which the data the users wish to transfer collide, leading to lost data. To address this issue, more sophisticated network protocols can be implemented at the Media Access Control

(MAC) layer to only permit certain subsets of users to attempt to communicate with the BS at the same time or one could simply build more BSs. However, the MAC layer changes will unlikely be able to cope with exponential growth in network users and building more BSs is an incredibly costly process as well as introducing the issue of interference between BSs [Pra19]. Due to the large physical distances involved in these networks, the ability for the network to transfer at large data rates can be a challenge. In the design of 5G networks for example, the use of pico and femto cells (which act as mini BSs) have been utilised to improve capacity at a more local level.

These types of traditional wireless networks are of crucial importance to everyday life, whether that be socially, in business, or even in use by emergency services. However, due to their reliance on fixed infrastructure, they are not always the best choice of network topology. To alleviate the issues involved with the set up of these networks, other system designs have been implemented.

We will now introduce the types of network which will be considered in this thesis before introducing the mathematical framework which we will use to model these networks and analyse their performance.

1.1 Wireless Ad-hoc Networks

The class of wireless communication networks considered in this thesis are so-called *ad-hoc networks*. Here, rather than users being connected to some central router, they are connected to each other directly. They have been utilised in a wide variety of applications such as in search and rescue operations, disaster recovery, wireless sensor networks [HMD+04], along with the two main application topics within this thesis of swarm robotics [LLZ+08], and vehicular networks [TML+08]. With the birth of millimetre wave communications, short-range device to device communication is likely to play a major role in the deployment of 5G networks. Alongside the rise in number of Internet of Things (IoT) devices and the large scale deployment of wireless sensor networks, the use of ad-hoc networks in every day life will only grow. The three main benefits they have over centralised communication networks are robustness, scalability, and improved mobility and flexibility:

Robustness: this is the ability of the network to withstand node failures and perturbations, thus meaning it is still able to operate even if nodes are removed from the network or there is a change in the environment. The absence of a central point of failure as there is in a centralised network (e.g. a base station or Wi-Fi router)

improves the robustness of these networks.

Scalability: this is the ability of the network to be able to handle an increased number of users being added into the system. This improvement is facilitated by the fact that new users can join anywhere in the network and hence the issues of overloading associated with all new users connecting to the central router are alleviated.

Mobility and Flexibility: these networks can be set up without any infrastructure already being in place and are able to more easily adapt to topology changes allowing for greater mobility: the users don't have to remain connected to a single centralised machine, they simply have to remain connected to any other member of the network. The decentralising of the network means that devices can simply react to their local environment without having to use up vital resources in finding out information about areas of the network that don't affect them.

However, these networks are not without their downfalls. Although they scale well in terms of computational cost, the increased topological complexity leads to additional routing costs i.e. finding the best route for a message to be sent through the network. Furthermore, providing Quality of Service (QoS) is very difficult due to the rapid changes in network configurations and topology, and the potential for large amounts of interference between nodes. Included in the QoS metrics is latency (the time taken for data or a request to go from source to destination) which is of vital importance to safety critical applications such as vehicular networks as will be discussed in more detail in Section 1.1.1. Ad-hoc networks are also susceptible to security attacks due to their low computational power and the fact that it is very difficult to maintain secure connections due to their dynamic nature. Finally, due to their use of battery power to allow for mobility, power control becomes a very important issue. If one of the nodes in the network loses power, this could fragment the network into disconnected clusters, leading to a disconnected network [HA14].

1.1.1 Vehicular Ad-Hoc Networks

One of the main applications of these networks is in the area of vehicular ad-hoc networks (VANETs) which has rapidly become a very important area of research due to the idea of self-driving vehicles becoming more of a reality. Here, the users or devices in the network are the vehicles. They will be equipped with some form of on-board radio enabling them to transmit messages either between each other (V2V communication), with Road Side Units (RSUs), or with other infrastructure (V2I

communication). These networks can be used for entertainment purposes: streaming videos, sending emails etc. whilst on the move, as well as for safety purposes: accident information, intersection management, and traffic flow for example [TML+08]. Ensuring the best possible design of the network infrastructure is thus crucial due to the necessity for low latency and high connectivity and has many challenges. For example the high level of mobility of the nodes in the system can easily lead to a more sparse network, increasing the chances of it becoming disconnected. The fact that drivers are constrained to the roads for travel introduces predictability into users movements which can make routing easier since node locations are more predictable but it also reduces the possible path redundancy. Alongside these issues, security is a crucial factor to consider due to the high risk nature of these networks. Thus, large areas of the literature focus on this issue, for example see [YOW08; MBH14; EBP+14; HSB+17] for more in depth discussion.

Whilst there are many challenges involved with this new technology, one of the main challenges is designing a network in which these vehicles will be able to effectively communicate. This communication is paramount for these vehicles to be safe, and hence incorporating this into their design is of utmost importance. Therefore, the full connectivity of these networks is where the research in this thesis is concentrated. Having a fully connected network means that it is possible for every vehicle in the network to communicate with every other vehicle either via a single- or multi-hop path. In a vehicular network, this means that it is possible to relay any information (safety or otherwise) to every vehicle that is on the road. This kind of communication is not only for the future application of driverless cars, but also for improving safety and efficiency in current vehicular networks. For example, it will be possible for cars to receive information about crashes ahead, causing the car to automatically slow down, and it will be possible to get a better understanding of areas of congestion and update driver's routes accordingly. It will also be possible to use tools such as platooning where groups of vehicles that are nearby will accelerate and brake at the same time. This will mean that they can drive closer together and remain safe since the factor of human reaction time is taken out of the equation [ADC+15]. This will both decrease route times in areas of high congestion and reduce the number of crashes.

With the question of connectivity in mind, it is useful to model these networks mathematically to develop an understanding of their behaviour without the need to physically build and test them, or run simulations which require a high number of parameters. The level of tuning needed to build accurate simulations can lead

to a lack of reproducibility as results become nearly impossible to compare with each other. Mathematically, the current state of the art is to model them using a spatial stochastic network model called a *one-dimensional hard Random Geometric Graph (RGG)*[HM06]. In this set-up, nodes are placed on a line segment and an edge placed between nodes that are within a certain distance of each other. Here, the nodes represent the vehicles in the network, the line segment represents the road on which they are driving, and an edge between two nodes signifies that the vehicles are able to communicate with each other. This model, and its associated results, will be discussed in more detail in Section 1.2.

1.1.2 Swarm Robotics

A second important application of ad-hoc networks is to *Swarm Robotics*, an area that has been of interest to researchers for many years. In the early 1980s there was a lot of interest in the topic of cellular automota [Wol02]. Here, seemingly simple initial rules can lead to highly complex patterns being formed. The idea of "cellular robots" [FN88; Ben88] was then introduced in the late 1980s as a way of describing systems of robots which could work co-operatively to complete a task and to self organise, an extension of the idea of cellular automata in which the units of the automaton interact dynamically [Ben04]. The inspiration for these groups of robots was the observation of social insects such as ants, termites, wasps, and bees [Şah04]. In these biological systems, the insects are able to complete tasks as a group which they would not be able to as individuals. Although these types of networks can be used to simulate biological systems to understand them better [HJW+20], they also have real-world applications such as exploration, surveillance, and search and rescue [BFB+13; BGN17].

To emulate the biological systems, there are three key properties of which these multi robot swarms must have. These are almost identical to those which are desirable for ad-hoc networks [Şah04]:

Robustness: the individual robots should have low complexity and be dispensable via redundancy with the large number of users meaning that there is multiplicity of sensing leading to increased Signal-to-Noise Ratio (SNR). They should also co-ordinate in a decentralised fashion thus removing the central point of failure.

Flexibility: they should have the flexibility to respond to changes in the environment and offer solutions to the tasks at hand by utilising different co-ordination strategies.

Scalability: they should be able to complete these tasks and operate under different system sizes which are generally of the order of around $10^2 - 10^{23}$ [Ben04]. The co-ordination mechanisms used should be undisturbed by changes in the system size since co-ordination should be based on local interactions and self-organisation.

One of the key ideas to note here is the importance of the underlying co-ordination system to allow for these properties to be possible. The use of local interactions in the design means that the robots can be low powered and low complexity, but most importantly, low cost. In this scenario, the robots are the users and they communicate with each other via short range communication channels (sometimes of the order of centimetres). Therefore, for information to be distributed throughout the network, there is the necessity for almost constant contact with the rest of the swarm and hence swarm cohesion comes from minimising the time that the individual robots are disconnected from the rest of the swarm [HWZ+08].

The reduction in cost in terms of producing these robots has been one of the main drivers to move away from small groups and allowing for much larger swarms to now be possible. One example of these low cost, scalable robots is the Kilobot [RAN12]. They cost just \$14 per robot and take around 5 minutes to assemble. They are relatively small, with a diameter of 33mm and height of 34mm, and have two vibrating motors instead of wheels, reducing the cost of production but also leading to noisy trajectories. They have a communication range of up to 10cm via an isotropic infrared LED transmitter and an isototropic infrared photodiode receiver. Since all the robots will be communicating over the same infrared channel, a standard carrier sense multiple access with collision avoidance (CSMA/CA) method is used to mitigate this interference. They also have limited sensing and localisation capabilities. Their low cost and ease of set up allow for large scale simulations using actual robots and are thus the motivation behind the use of the Kilobox [JSH+18] simulator for the work done in this thesis which has been shown to give fast and accurate simulations of Kilobots [RAN12].

As with VANETs, connectivity in a swarm network is highly important: it is what allows the swarm to complete tasks which would be impossible for an individual robot and hence must be integrated into their design. One way to aid with this design is to model them mathematically, with the most popular approach being to model them as a random graph. In this set up, the robots are represented by nodes on the graph and an edge between two nodes means that those two robots are able to communicate with each other and therefore send information across this communication channel. One way to generate these graphs is by constructing an

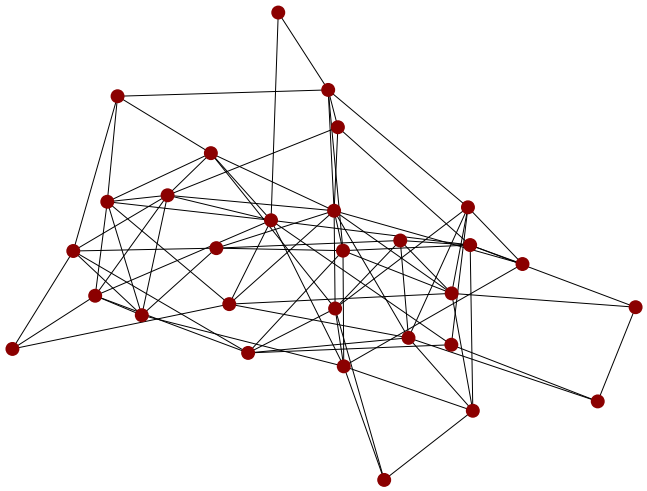


Figure 1.1: An Erdős–Rényi random graph with $n = 30$ nodes and edge probability $p = 0.2$. This is denoted by $\mathcal{R}(30, 0.2)$.

Erdős–Rényi graph. To construct this graph, one has n nodes and an edge is placed between any pair of nodes with a fixed probability, p. (p is often a function of the system size and so written as $p = p(n)$.) We will denote this graph by $\mathcal{R}(n, p)$. As an example, Figure 1.1 shows the graph $\mathcal{R}(30, 0.2)$. These have been used to study the stability of swarm beliefs in which half of the agents have one belief and half have another [BKB+13; BSB15]. The edge probability, p, and thus the level of intercommunication between the agents, is then tweaked to see how this effects the stability of the final outcome. Other random graph models have also been used to investigate the optimal population topologies for various tasks [KM02]. Here, the key aspects of the networks to understand are the numbers of neighbours that each agent has, how much clustering (when an agents neighbours are neighbours of each other) occurs in the network, and finally, the average shortest hop count between any two nodes in the graph. This work has been extended to real world communications applications [DBS12], and also to look at geometric settings [LC08; FCH+20]. In these final two examples, the underlying graph has a geometric structure meaning that it is a more accurate representation of the real world since communication between agents depends on the distance between them. This type of random graph is called a Random Geometric Graph (RGG) and will be the focus of the following section.

1.2 Random Geometric Graphs

When investigating these varying types of ad-hoc communication networks and attempting to understand them, it is important to be able to mathematically model them in such a way that they are simple enough to achieve tractable mathematical results but also realistic enough that these results help us to understand the real-world networks and not just some theoretical object. One such way of modelling an ad-hoc communication network is by using Random Geometric Graphs (RGGs). A RGG is created by firstly defining a space in which the graph will be defined (thus adding additional geometric structure). This could be a unit square, or the infinite real line for example. Points, or nodes, are then generated within this space according to some probability distribution. An edge is then placed between any pairs of points which are within a certain distance of each other. An example of a RGG is shown in Figure 1.2. RGGs have been used to model a plethora of different real world situations. For example, they have been used to model disease spread in social networks [EGK+04], the climate [DZM+09], infrastructure [RBJ+14], and neuronal networks [NVS+13].

One area of particular interest in recent years has been wireless networks [Hae12] in which the nodes represent users in a communications network and an edge between two nodes means that they are able to communicate with each other. This is not a new idea and has been around since the early 1960s [ER60; Gil61]. In Erdős–Rényi random graphs, all nodes within a network have the same probability of being connected to each other, independent of the distance between them and hence there is no geometry present in the model. In a RGG, a geometric layer is added into the model by implementing the rule that the existence of an edge between two nodes is based on the distance between them. The application of this to the study of wireless networks can be seen immediately, and was hence the main focus of Gilbert's 1961 paper on *random plane networks* where he studied the idea of *network criticality*. Here, nodes are distributed on the 2-dimensional infinite plane uniformly at random with a fixed density per unit area and an edge placed between two nodes if their mutual distance is less than a certain radius, termed here as the *connection radius*. Within this setting, the paper develops an understanding of when the network transitions from being comprised of connected clusters of finite size into a network in which there is almost surely an infinite size (giant) component [Pra19]; a phenomenon known as *percolation*. This work finds an upper bound for the critical radius by using a hexagonal tiling of the plane (a technique widely used in the mathematical modelling of

communication networks) and uses results from bond percolation.

However, especially when studying communication networks, a more appropriate question to ask may be whether the network is *fully connected*. I.e. when is there a single- or multi-hop path between any pair of nodes in the network. In the context of Erdős–Rényi random graphs, it is well known that the question of full connectivity is equivalent to asking whether there are isolated nodes in the network [ER60]. In this context, in the limit as the number of nodes in the graph grows to infinity, there is a sharp transition from having a system in which an isolated node is present almost surely to one in which there are no isolated nodes almost surely. This occurs when the mean degree (average number of edges of each node) is close to the natural logarithm of the number of nodes in the graph. An analogous result has also been shown to be true for RGGs [Pen97]. However, for random plane network as described in Gilbert's work, a *connectivity transition* does not occur since, with probability 1, an isolated node will exist somewhere in the network. In Penrose's work, the nodes are distributed uniformly at random random in a unit square and this model is studied in the regime in which the density tends to infinity with the connection radius decreasing as a function of the density. Once again, the limiting factor with regards to connectivity is whether an isolated node exists and it is shown that a mean degree close to the natural logarithm of the number of nodes in the graph defines a critical threshold for the almost sure existence (non-existence) of isolated nodes and hence for the network (not) being fully connected almost surely. This result was also extended to higher dimensional models and a more general collection of point distributions [Pen03], however the result that isolated nodes are the only barrier to full connectivity is one that doesn't hold true in dimension one.

In the 1-dimensional version of the RGG, the graph can split into separate, disconnected clusters due to a large gap occurring between two nodes somewhere in the graph. Defining an *uncrossed gap* to be the event that there is a gap between two nodes on the line which is greater than the connection radius of the RGG, we note that an isolated node cannot occur without the presence of an uncrossed gap. Therefore, in the case of the 1-D RGG, the critical event for full connectivity becomes uncrossed gaps rather than isolated nodes. This 1-D version has been a well-studied problem for a long time once again with origins in continuum percolation [Sha81; Dom89; MR96; Dro97]. With regards to the question of connectivity, much of the original work in this area was based around understanding the largest spacing between any pair of nodes in the network: if the largest spacing between any pair of nodes in the graph is less than the connection radius, then the graph is fully con-

nected. One of the foremost works in this area being that of Devroye [Dev81] in which it was shown that the largest gap scales roughly with the natural logarithm of the number of nodes in the graph. This idea has been used in most studies of the application of this work to 1-D ad-hoc networks such as 1-D MANETs [FL04], and to (quasi-1-D) wireless sensor networks [BBS+07]. However, the most prominent application of 1-D RGGs has been to VANETs. Here, the vehicles are represented by the vertices of the graph and an edge represents a reliable communication channel. It is generally assumed that nodes are distributed uniformly at random along the line, hence the underlying vertex set constitutes a Poisson point process (PPP). The PPP has been shown to be a good approximation for traffic in its free-flow state [RPM04], but is also used due to its mathematical tractability which, in general, is thought to outweigh its possible inaccuracies for real-world networks [HBS+21]. Under this model, it has been shown that a sharp threshold for the network to be fully connected occurs when the connection radius is such that the mean degree of the graph is close to the natural logarithm of the number of nodes in the graph [HM06; HM07; HM08]. In Han and Makowski's work, they build on the work done by Devroye but also derive another important result: the distribution of the number of so-called *breakpoint nodes* (i.e. nodes that signify an uncrossed gap) can be well approximated by a Poisson distribution. This fact can then be used to reconstruct the critical threshold discussed.

However, this model doesn't take into account the random nature of the wireless medium. One way to include this randomness in the model is by using soft RGGs [Pen16]. The construction of a soft RGG begins in the same way as the original RGG by defining a space and generating the points in this space according to some probability distribution. However, once the points have been generated, an edge is now placed between two points with a probability that is a function of their mutual distance. This probability is called the *connection function* and is denoted by $H(r)$, where r is the distance between the two nodes in the graph. The connection function itself is a very general object and can take on many forms (discussed in more detail in Section 2.1.3), but it is generally assumed to be monotonically decreasing (especially in the context of wireless communications) and integrable (to avoid triviality). These graphs have been variously termed a random connection model [MR96; MA13], a soft RGG [Pen16; DG16], or a Waxman graph [Wax88]. For this work we will call them soft RGGs and we will call the original RGG model a hard RGG. (The hard RGG is a special case of the soft RGG where the connection function is the indicator that the distance between two nodes is less than the connection radius.) They have been

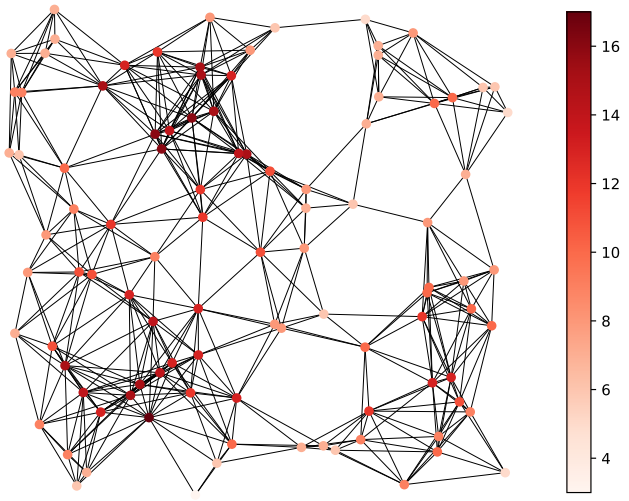


Figure 1.2: A random geometric graph generated on the unit square with nodes placed uniformly at random and an edge placed between any two nodes within a distance of $r = 0.2$ of each other. The colours of the nodes represent the mean degree (average number of edges) of the node.

widely used to model the randomness of wireless communication systems in 2- and 3-dimensional models, for example in [MA13; MZA10; MA11a; MA11b; GDC14; DG16]. The main discovery of these works was that the dominant contribution to lack of connectivity in confined geometries is that of isolated nodes (a node that is not connected to any other within the network) and hence the probability of full connectivity (the probability of the whole network being connected) can be well approximated by the probability of not finding an isolated node. These results will be discussed in more detail in Chapter 2.

No work as of yet, however, has looked at the connectivity of the *one-dimensional* version of the soft RGG and the properties of this graph. Other properties of these graphs have been analysed such as their diameter [DF14], and with respect to VANETS, the "average connectivity distance" (i.e. the expected length of the connected path from a node) and "average platoon size" (i.e. expected number of nodes in a cluster) were calculated in [CA12] using a queuing theoretic approach. Furthermore, the number of nodes (vehicles) which have two-hop connectivity (i.e. are either connected directly or can connect via a single relay node) to a fixed RSU is analysed in [KKL+20] with its expectation being calculated and a Normal approximation derived. Although interesting results, none tackle the problem of calculating the probability that the graph is fully connected. The connectivity of this model has

been mentioned in other work, for example [Han07], but wherever mentioned, either the author remarks on the difficulty of the problem and hence doesn't attempt to understand it in more detail, or the approach given in the work is not satisfactory. With this in mind, the work done on understanding the connectivity of the 1-D soft RGG is, as far as the author is aware, all novel and is where the main body of research within this thesis is concentrated. In depth analysis of this model is given in Chapters 3, 4, and 5.

1.3 About this thesis

The main idea that all of the applications described thus far have in common is the importance of maintaining a fully connected network. In other words, a network in which every user in the network is connected to every other via a single- or multi-hop path. This allows for the dissemination of information throughout the whole network even if users aren't connected to each other directly. In VANETs, this means that if there is an accident, all other vehicles will know about the incident causing nearby vehicles to slow down to avoid collision, and rerouting vehicles further away in the network to avoid congestion and reduce danger to those users. In a swarm of robots, full connectivity ensures that the swarm is able to act as one coherent group, enabling them to more efficiently search an area or to complete a foraging task. The idea of full connectivity is what underpins this thesis .

In Chapters 3, 4, and 5, we investigate the 1-dimensional soft RGG. Within these three chapters, we firstly state the conjecture that there are only two main causes of disconnection in this model: isolated nodes and uncrossed gaps (as defined in Definitions 3.2.1 and 3.2.2). This claim is backed up by simulation data. Under fairly loose constraints on our connection function, we then calculate, for the first time, a threshold for the existence of isolated nodes in this graph by calculating their expectation and variance. Placing the condition on the connection function that it is monotonically decreasing, we are able to prove that at this critical threshold the number of isolated nodes converges in total variational distance to a Poisson random variable as the system size tends to infinity: a novel proof for the 1-D soft RGG. We are then able to show that in an infinite sized system that uncrossed gaps appear with negligible probability in the scaling at which isolated nodes appear. This means that in an infinite sized system, the isolated nodes are the main barrier to full connectivity: a new and very interesting result since in the 1-D hard RGG, it

has been shown that it is the uncrossed gaps that are more important. Reasons as to why this is the case, along with an investigation into the critical scaling for the appearance of uncrossed gaps, are also given.

In Chapter 6, a novel decentralised algorithm is proposed to model the behaviour of social spiders using Kilobots (a type of cheap, simple robot). The algorithm uses inter-node connectivity (i.e. social interactions between robots/spiders) as a way of increasing an individuals propensity to explore an unknown environment. This algorithm was tested in software to analyse its performance against less complex algorithms for different swarm sizes and the results of these simulations are discussed in this chapter. Furthermore, the simulations investigated how well the swarm adapted to sharp drops in population size thus examining its robustness. This is the first time that this natural social behaviour has been tested in the context of swarm robotics and in a decentralised control system. The results give a good indication that this algorithm could be used in real-world field trials in hazardous environments where there is a high risk of loss or destruction of robots.

Each chapter is summarised below:

Chapter 2: An overview of the mathematical tools used in this thesis are given, along with some important results associated with this work.

Chapter 3: The 1-dimensional soft RGG model is introduced. Analytic expressions for the mean and expectation of the number of isolated nodes in this graph are derived before calculating a threshold for the existence of isolated nodes.

Chapter 4: The behaviour of the isolated nodes is analysed at the critical threshold. A theorem stating that, at this threshold, the number of isolated nodes converges in total variational distance to a Poisson random variable is proven.

Chapter 5: Uncrossed gaps are investigated in more detail and it is shown that they appear with negligible probability in the scaling at which isolated nodes appear. Reasons for this, along with an investigation into the critical scaling for the appearance of uncrossed gaps, are also given.

Chapter 6: An algorithm used to model social spiders is introduced. A scoring mechanism is then defined to analyse how well this algorithm performs in the context of exploring an unknown environment.

Chapter 7: The conclusions of the work discussed in this thesis are given along with potential future questions that have come out of this project.

Chapter 2

Mathematical Preliminaries

In this chapter, an overview of the more general mathematical topics in this thesis will be given. The aim being to give the reader an introduction to the ideas used along with making the thesis as self-contained as possible. As was discussed in the previous chapter, many real-world systems can be modelled by spatial networks, with one of the main applications being to communication networks. In the study of these networks, mathematical tools enable us to study their properties without the costs associated with physically building and testing them. For example, one of the main concerns in their design is the Signal-to-Noise-Ratio (SNR) or more importantly the Signal-to-Interference-plus-Noise-Ratio (SINR) which is seen as the performance limiting metric in the design of communication networks [Hae12]. In fact since Claude Shannon's work around the middle of the 20th century [Sha48], the SNR has been the main quantity of interest to communication engineers. However in wireless communication networks, this doesn't give the full picture when analysing the network since concurrent transmissions from users within the network complicate the problem, and SINR becomes the key metric. To study this, we need to be able to model the random spatial locations of users and objects within the network which will lead to shadowing (fluctuations in received signal power due to obstructions) and fading (attenuation of signals). To understand how they interact with each other and thus enable us to answer questions, we look to Stochastic Geometry as a tool for us to model and study these networks.

2.1 Stochastic Geometry

Stochastic geometry is the study of random spatial patterns, with random point patterns, or point processes, being the most basic object studied. Initially used as a way to understand fields such as material science, astronomy, and biology [Sos00; TSZ08; BB10], more recently stochastic geometry (and in particular point processes) has also been used to study areas such as computational neuroscience [Joh96], ecology [LIB+09], crime [MSB+11], disease spread [MH+14], and even machine learning [SD19]. One of the main areas of application is to the world of wireless networks (see e.g. [NBK07; BB10; Hae12; EHH13] to name a few) and this is where our research lies.

2.1.1 Point Processes

We begin this section by formally defining a Point process.

Definition 2.1.1 (Point process, [Hae12] Defn 2.1)**.** *A point process is a countable random collection of points that reside in some measure space, usually the Euclidean space $\mathbb{R}^d$. The associated σ-algebra consists of the Borel sets $\mathcal{B}^d$, and the measure is the Lebesgue measure.*

Borel sets are the sets that can be constructed from open (or closed) sets and repeatedly taking the countable unions and intersections of these sets. The collection of all such sets forms the Borel σ-algebra, and all singletons are contained within this. All sets and functions referred to in this thesis are assumed to be Borel measurable, even if not explicitly stated.

It is important to note that a point process can be described in two different formalisms: a random set formalism and a random measure formalism. Since we will use both descriptions in this thesis, it is important to gain an understanding of each of them. Once again taking the descriptions from [Hae12], we describe these two ideas as follows.

Random set formalism: the point process is regarded as a *countable random set* $\Phi = \{x_1, x_2, ...\} \subset \mathbb{R}^d$ consisting of random variables $x_i \in \mathbb{R}^d$ as its elements.

Random measure formalism: the point process is characterised by counting the number of points falling in sets $B \subset \mathbb{R}^d$. Denoting by $N(B)$ the number of points in B, $N(B)$ is a random variable that assumes values from the non-negative integers, $\mathbb{N}_0$. N is called a (random) *counting measure*.

This duality gives two different ways of talking about point processes as it enables one to define them as random sets or as random counting measures. There is a one-to-one mapping between Φ and N. If Φ is given, N is obtained from

$$N(B) = \#\left\{\Phi \cap B\right\},$$

where $\#\{\cdot\}$ gives the number of elements in the set. In reverse, Φ can be retrieved if N is known from

$$\Phi = \left\{x \in \mathbb{R}^d : N(\{x\}) = 1\right\}.$$

In this thesis, we will make the assumption that all point processes are *simple* i.e. only one point can exist at a specific location. More formally,

$$N(\{x\}) \in \{0, 1\} \quad \text{a.s.} \quad x \in \mathbb{R}^d.$$

For *simple* point processes both formalisms are identical and hence allow us to write the following equivalence statement,

$$\Phi(B) = N(B), \quad \forall B \in \mathcal{B}^d.$$

For simplicity, within this thesis we will use the notation $\Phi(B)$ to denote the number of points in the point process, Φ, lying within some (Borel-measurable) set, B. Along with this, we also make the assumption that our point processes are locally finite i.e. for any finite area, there are only finitely many points. More formally,

$$|B| < \infty \implies \Phi(B) < \infty.$$

Finally, this brings us to the definition of the intensity measure, i.e. the expected number of points within a set.

Definition 2.1.2 (Intensity Measure)**.** *The intensity measure* Λ *is defined as*

$$\Lambda(B) := \mathbb{E}[N(B)], \quad \forall B \in \mathcal{B}^d,$$

or equivalently

$$\Lambda(B) := \mathbb{E}[\Phi(B)], \quad \forall B \in \mathcal{B}^d.$$

A key result from point process theory is Campbell's theorem for sums which tells us that the sum of a measurable function f over the point process is a random

variable [CSK+13] by relating the expectation of a function summed over a point process to an integral of the function involving the mean measure of the point process. This allows us to take a spatial average over all possible node locations.

Theorem 2.1.1 (Campbell's Theorem for Sums, [CSK+13] Thm. 4.1)**.** *The sum of a measurable function $f : \mathbb{R}^d \mapsto \mathbb{R}$ over the point process Φ is a random variable with mean*

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x)\right] = \int_{\mathbb{R}^d} f(x)\Lambda(dx),$$

provided that the right hand side is finite. If Φ has an intensity function λ, then we can rewrite this as

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x)\right] = \int_{\mathbb{R}^d} f(x)\lambda(x)dx.$$

So far we have only described point processes as very general objects, but we wish to be able to compare them. The main way to do this is via the Laplace functional of a point process.

Definition 2.1.3 (Laplace Functional, [LP17] Def. 2.8, [Hae12] Def. 4.5)**.** *For a point process Φ on $\mathbb{R}^d$, and letting $\mathcal{F}$ be the set of all bounded non-negative measurable functions f of bounded support, the Laplace functional of Φ is given by*

$$\mathcal{L}_\Phi(f) = \exp\left(-\int_{\mathbb{R}^d}\left(1-e^{-f(x)}\right)\Lambda(dx)\right), \quad \textit{for } f\in\mathcal{F}. \tag{2.1}$$

An interesting result is that the Laplace functional of a point process determines its distribution. This is expressed in Proposition 2.10 of [LP17] in which the following equivalence is stated: For point processes Φ and Φ', the following are equivalent

$$\begin{aligned} &\text{(a)} \quad \Phi \overset{d}{=} \Phi', \\ &\text{(b)} \quad \mathcal{L}_\Phi(f) = \mathcal{L}_{\Phi'}(f), \end{aligned}$$

where $\overset{d}{=}$ denotes that two random objects are equal in distribution. If our Point process is simple, then we also have two more ways in which to compare the distributions of point process, namely the void probabilities and finite-dimensional (fidi) distributions.

Definition 2.1.4 (Void Probability)**.** *The void probability, V, of a simple point process Φ is given by*

$$V(K) := \mathbb{P}(\Phi(K) = 0),$$

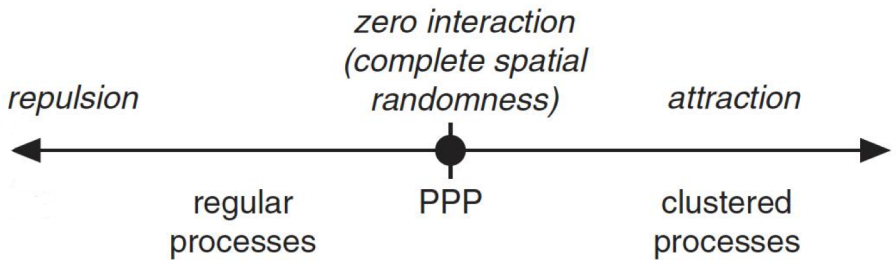


Figure 2.1: Figure taken from [Hae12] (with some edits) showing the different types of point process based on the level of interaction between the points with the Poisson PP having zero interaction between points. As we move to the left, the repulsion between points increases forming more regular point processes such as hard-core PPs. To the right, we see point processes in which the points are attracted to each other and hence form clustered processes such as Cox processes.

for K compact.

Definition 2.1.5 (Finite-Dimensional (Fidi) Distributions, [Hae12], Def. 2.21)**.** *The fidi distributions of a point process, Φ are the joint probability distributions of*

$$(\Phi(K_1), \Phi(K_2), ..., \Phi(K_m))$$

for all finite integers m and all $K_1, K_2, ..., K_m$ compact.

This leads us to the following theorem.

Theorem 2.1.2 (Fidi distributions and void probabilities, [Hae12], Thm. 2.23)**.**
(a) If the fidi distributions of two point processes are identical, then they have the same distribution.
(b) If the void probabilities of two simple point processes are identical, then they have the same distribution.

Point processes are able to describe phenomena in which objects can be mapped to some space and hence are very general objects. This has incredibly wide reaching applications [BBS07] such as statistical ecology [LR88], astrostatistics [BF96], neuroscience [BZG18], and epidemiology [Dig00; PSS+12]. For a wider review of their applications see for example [CSK+13], and for an overview on the use of spatial networks to model real-world systems see [Bar11]. Using a useful pictorial description of point processes (Figure 2.1) from [Hae12], we are able to see them on a spectrum in terms of how much interaction there is between the points within the model.

For this thesis, we will focus our attention on a small subsection of them, with special attention being given to the Poisson point process (PPP). We will also introduce other point processes for comparison and discussion of applications, namely the Binomial, Cox, and Matérn PPs.

Poisson Point Processes

One of the most widely used point processes is the Poisson point process (PPP) which we define formally.

Definition 2.1.6 (Poisson Point Process)**.**
Let $\lambda : \mathbb{R}^d \to \mathbb{R}_+$ be a function whose integrals on bounded subsets of $\mathbb{R}^d$ are finite. A Poisson point process (PPP) $\mathcal{P}$ on $\mathbb{R}^d$ with intensity $\lambda(\cdot)$ is a random set of points such that $\mathcal{P}(B)$, the number of points in $B \subset \mathbb{R}^d$, has a Poisson distribution with mean $\Lambda(B) = \int_B \lambda(x)dx$, while the numbers of points in a (finite) collection of disjoint subsets are mutually independent random variables. If $\lambda(x)$ is identically equal to a constant λ, we say that the PPP is homogenous with intensity λ.

Its wide use in the literature is in part due to the fact that its spatial independence property simplifies the analysis of many problems, but also due to its ability to accurately model many real world systems. Most importantly for this work, a simple PPP has been shown to be a reasonably good approximation for the locations of base stations in dense urban environments [LD15] as well as being realistic for traffic in its free flow state [RPM04], and hence for VANETs. An important idea when designing communication networks is that of a "typical user" which is formalised mathematically in Palm probability, or the Palm measure of a point process, which defines the probability of an event given that a point exists at some location. We denote this by

$$\mathrm{P}_x(Y) := \mathbb{P}(\mathcal{P} \in Y | x \in \mathcal{P}),$$

where $\mathcal{P} \in Y$ is the event that the point process, $\mathcal{P}$, has some property, Y. In addition, we denote by $\mathrm{P}_x^!$ the reduced Palm measure where the point on which we have conditioned is not included in the distribution. More formally,

$$\mathrm{P}_x^!(Y) := \mathbb{P}(\mathcal{P} \backslash \{x\} \in Y | x \in \mathcal{P}).$$

Due to its spatial independence property, the PPP has many very nice properties. One such example being that if we condition on there being a point at some location

$\{x\}$ within our space, this does not change the distribution of the rest of the process. This is formalised in Slivnyak's theorem.

Theorem 2.1.3 (Slivnyak's Theorem, [Hae12], Thm. 8.10)**.** *For the Poisson point process with intensity function λ, $\mathcal{P}_\lambda$,*

$$\mathbb{P}\left(\mathcal{P}_\lambda \in Y | x \in \mathcal{P}_\lambda\right) = \mathbb{P}\left(\mathcal{P}_\lambda \cup \{x\} \in Y\right), \quad \forall x \in \left\{y \in \mathbb{R}^d : \lambda(y) > 0\right\}.$$

More compactly,

$$\mathrm{P}_x^! = \mathrm{P}.$$

Writing this in terms of (reduced) Palm measures tells us that if we condition on there being a point at some location, and then remove this point, the distribution of the point process is not affected. We leave out the the details of the proof of this theorem, but the general idea follows from the fact that (as described in Theorem 2.1.2 of Section 2.1.1) if the void probabilities of two simple point processes are the same, then they have the same distribution.

Using Def. 2.1.4, we can define the void probability for the PPP on some Borel measurable space $S \in \mathbb{R}^d$.

Definition 2.1.7 (Void Probability of a PPP)**.** *For a PPP on $\mathbb{R}^d$ with intensity function λ, $\mathcal{P}_\lambda$, the void probability is given by*

$$\mathbb{P}\left(\mathcal{P}_\lambda(S) = 0\right) = \exp\left(-\int_S \lambda(\boldsymbol{x}) d\boldsymbol{x}\right).$$

This idea, along with Campbell's Theorem (Thm. 2.1.1) is used in Chapters 3 and 4 when analysing node isolation within our graph. Although this thesis solely focuses on RGGs constructed on Poisson point processes, within the wireless communications literature there are some further examples of point processes used to model these systems. These will also be useful to understand when discussing the future directions in which this work could be taken in Chapter 7.

Binomial Point Processes

The first of these is the Binomial point process (BPPs). This is where a fixed number of points, say n, are distributed uniformly at random within a compact (Borel-measurable) set $V \subset \mathbb{R}^d$. A BPP can also be constructed from a Poisson point process by conditioning on the number of points within the point process. This

conditioning results in a loss of complete spatial randomness since if the number of points within a subset of the space is known, then number of points outside of this set is also known, and hence not independent. However, the BPP still produces simple formulas for the void probabilities and nearest neighbour distances and results for the BPP are very similar to those of the PPP when the number of points is large [DGP18].

Cox Point Processes

A Cox point process, also known as a *doubly stochastic Poisson process*, allows for the intensity measure itself to also be random, where the intensity measure is the realisation of non-negative, locally finite random measure [Hae12]. One such example is the *mixed Poisson point process*. Let L be a non-negative random variable, then given $L = \lambda$, the mixed Poisson point process, $\mathcal{P}^{\mathrm{M}}$, is a homogeneous PPP of intensity λ. For a single realisation of this process, it is impossible to distinguish it from the PPP. The Cox point process is used in Chapter 5 when analysing the number of nodes which have an edge across the origin to determine the scaling for the existence of uncrossed gaps.

Another important example is the Poisson line Cox process [CD18]. A Poisson line process (PLP) is constructed by firstly generating a Poisson point process in the *representation space*, $[0, \infty) \times [0, 2\pi]$. These $x-$ and $y-$co-ordinates can then be mapped back to Euclidean space and give the perpendicular distance of the line from the origin and its angle respectively [Che20]. A 1-D PPP is then generated on each line. The point process generated from this construction is itself a Cox process. This has been used by Chetlur and Dhillon as a way of modelling VANETs in which the PLP resembles the road and the points on these lines represent either a vehicle or a RSU.

Matérn (hard-core) Point Processes

Hard-core processes can be used to model networks in which points are not allowed to be close together. A Matérn (hard-core) point process [Mat13] of type I and II are defined as [Hae12]:

Definition 2.1.8 (Matérn hard-core process of type I)**.** *Starting with a homogeneous PPP $\mathcal{P}_b$ of rate λ_b, then remove all points which have a neighbour within a distance, r_0.*

Definition 2.1.9 (Matérn hard-core process of type II)**.** *Starting with a homogeneous PPP $\mathcal{P}_b$ of rate λ_b, add a mark to each point which is a value uniformly picked from $[0, 1]$. Then remove all points that are within a distance r_0 of a point with a larger mark than themselves.*

These types of processes can be used to model communication networks in which users have to be separated by a set distance. A modified version of the Matérn type II in which the marks are drawn uniformly from $\{0, 1, ..., W\}$, where W is typically taken as 15, has been used to model the use of 802.11p for VANETs [TLH+16]. The mark here is used to represent the random back-off counter employed by the standard (i.e. if a channel is idle, a message is sent, otherwise the user waits a random time before trying again).

Now that we have a general understanding of Point processes of different kinds, we can begin to construct network models using these processes as the basis.

2.1.2 Hard Random Geometric Graphs

In his 1961 paper [Gil61], Gilbert introduced the idea of a *random plane network* in which points are picked from the infinite plane according to a Poisson process with some fixed density per unit area and edges are then placed between any two points which are separated by a distance of less than some connection radius. The work used this construction to model a communication network and to study how information flowed through the system. More recently, these types of networks are more generally known as Random Geometric Graphs (RGG), or Gilbert graphs, and in this thesis will be termed *hard* RGGs to avoid confusion with their *soft* counterpart. More formally [Pen03], a hard RGG is constructed by firstly defining a space (generally $\mathbb{R}^d$), then defining a norm $\|\cdot\|$ (e.g. the Euclidean norm) on this space. Then for the positive parameter r_0, the so-called connection radius, and a (point) set $\Phi \subset \mathbb{R}^d$, we denote by $\mathcal{G}(\Phi, r_0)$ the undirected graph with vertex set Φ and (undirected) edge set containing all pairs $\{x, y\}$ if $\|y - x\| \leq r_0$. An excellent monograph on this subject is that of Penrose [Pen03] in which a much deeper study of these objects is given than is feasibly possible here, however we will discuss some key results from the field.

One such result is that of connectivity. Denoting by $\mathcal{P}_n^2$ the points of a unit rate Poisson process on $\mathbb{R}^2$ that lie within a square of area n (i.e. the expected number of points of $\mathcal{P}_n^2$ is n), and connect any points within distance $r(n)$ of each other. Then the associated hard RGG, $\mathcal{G}(\mathcal{P}_n^2, r(n))$, has the following property [Pen97]:

Theorem 2.1.4. *If* $n\pi r(n)^2 = \ln n + \alpha$, *then*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}(\mathcal{P}_n^2, r(n)) \text{ is connected}) = e^{-e^{-\alpha}}, \quad \alpha \in \mathbb{R}.$$

In particular, this probability tends to 1 if $\alpha = \alpha(n) \to \infty$ *in this limit.*

A similar result was also proved independently by Gupta and Kumar [GK99] a few years later. This result has an exact analogue in the theory of classical random graphs (see [Bol01; BSB08]) and the main idea for both of these is that in the limit as n grows to infinity, the only obstruction to connectivity is *isolated nodes*. To arrive at Theorem 1, one has to prove that isolated nodes are the only barrier to connectivity and then also gain an understanding of the behaviour of these isolated nodes. This understanding is gained by showing that the isolated nodes themselves form a Poisson process of rate $e^{-\alpha}$ via a technique called the Chen-Stein method for Poisson approximation [Che75; AGG+89; BHJ92]. (This will be discussed in much more detail in Section 2.2.2.) The probability that this Poisson process is empty (i.e. that there are no isolated nodes) is then given by $e^{-e^{-\alpha}}$.

However, the idea that isolated nodes are the main barrier to connectivity does not carry into the 1-dimensional case since, in this special case, there is also the potential for the graph to split into two or more large components. This low dimensional model has been widely studied, originally in the context of 1-D continuum percolation [Sha81; Dom89; MR96; Dro97] but more recently it has become of keen interest to the wireless communications community due to its use as a way of modelling networks over "linear highways" and in particular to VANETs (see e.g. [HM07; ANB11; KKG+16]). When studying the low-dimensional version of this model, it is key to notice that studying the connectivity of the network is the identical to understanding the size of the largest gap between any two points distributed on the line. In the context of a uniform distribution of points, Devroye's paper [Dev81] is one of the key works studying this problem. In this paper, the main result is given as follows (Eqn. (1.1) in [Dev81]):

Lemma 2.1.1. *For the sequence* $X_1, ..., X_{n-1}$ *of independent, uniformly distributed random variables on* $[0, 1]$ *and their associated order statistics,* $X_{(1)} < X_{(2)} < ... < X_{(n-1)}$, *we define the maximal gap as*

$$M_n = \max_{i\in\{1,...,n\}} S_i,$$

where $S_1 = X_{(1)}$, $S_i = X_{(i)} - X_{(i-1)}$, *and* $S_n = 1 - X_{(n-1)}$. *Under this definition,* M_n

has the following property,

$$\limsup \left(nM_n - \ln n\right)/2\ln\ln n = 1, \quad \textit{almost surely.}$$

In other words, the largest gap in the network grows as $\frac{\ln n}{n}$ (with variations of order $\frac{\ln\ln n}{n}$). This result was a tightening of earlier results from Sukhatme [Suk37], Lévy [Lév39], Darling [Dar53] and others from which the following Lemma was derived (Lemmas 2.4 and 2.5 in [Dev81]):

Lemma 2.1.2. *For n points distributed uniformly at random on the line segment $[0,1]$,*

$$\lim_{n\to\infty} \mathbb{P}(nM_n < \ln n + \alpha) = e^{-e^{-\alpha}}, \quad \alpha \in \mathbb{R},$$

and

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{nM_n}{\ln n} - 1\right| \leq \delta\right) = 1, \textit{ for any fixed } \delta > 0.$$

If we denote by $\mathcal{P}_n$ the points of a unit rate Poisson process falling within the line segment $[0,n]$, these previous two Lemmas are summarised in the following Theorem [GJ96]

Theorem 2.1.5. *If $nr(n) = \ln n + \alpha + o(1)$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{G}(\mathcal{P}_n, r(n)) \textit{ is connected }\right) = e^{-e^{-\alpha}}.$$

We remark here that Theorems 2.1.4 and 2.1.5 are the 1- and 2-D analogues of each other, however the reasons for which they are both true are different. For the 2-D version (as was discussed) the cause of this lack of connectivity is isolated nodes, whereas in the 1-D model, it is due to a splitting of the network into separate (disconnected) clusters. This occurs when there is a large physical gap between two consecutive nodes and hence in this thesis is termed an *uncrossed gap*. Defining an uncrossed gap in this way, we note that an isolated node cannot occur without the presence of an uncrossed gap since, in the bulk of the system, an isolated node is equivalent to two consecutive uncrossed gaps.

The main application of this 1-D model has been to 1-D mobile ad-hoc networks (MANETs) [FL04] and to VANETs. The nodes in the 1-D hard RGG represent vehicles (or users) in the network, the ability of two users to communicate is represented by an edge between the two nodes, and the road which the vehicles are on is represented by a line segment. It is important to clarify the assumptions made through

the use of this model (and specifically in this thesis) with respect to a real-world VANET:

Snapshot in time: we assume that we are looking at the model at a single moment and hence no mobility is considered. Mobility has been considered by many authors including [ANB11; AOE13], and will also be discussed in Chapter 7.

Free-flow traffic: We assume that the vehicles are in a free-flow state and hence the distribution of their positions along the road are independent and can be well approximated by a Poisson distribution [RPM04].

"Narrow" roads: Finally, we make the assumption that the communication range of the vehicles is much wider than the road and hence we are able to ignore the width of the road in our analysis, reducing the problem to a 1-D model.

With this application in mind the work of Han and Makowski [HM08] looks at this critical scaling for connectivity and reproves the result of Theorem 2.1.5 in two different ways. The first of which once again uses the idea of maximal spacings, but the second looks at so-called *breakpoint nodes*. A breakpoint node is defined as a node which is not the leftmost node in the network and there is no node within communication range of the node to its left hand side. In other words, it is the right hand node of an uncrossed gap. They are then able to show (via the Chen-Stein method) that the number of breakpoint nodes can be well approximated by a Poisson distribution and once again return to the result that the critical scaling of the connection radius for an almost surely fully connected network occurs when the connection radius is such that the mean degree of the graph is close to the natural logarithm of the number of nodes. Breakpoint nodes don't make sense as an object in higher dimensions in which instead isolated nodes are the object of investigation.

With the application of wireless networks in mind, it is important to note that the hard RGG is not able to adequately capture the randomness associated with the wireless communication medium, hence we must turn to a different way of modelling the communication channel between nodes in our network.

2.1.3 Soft Random Geometric Graphs

One way to include this randomness is to instead model the wireless network using a soft RGG in which a probabilistic function is used to determine the existence of an edge between two nodes in our graph, rather than the deterministic function used in the construction of a hard RGG. To do this, we introduce the idea of a *connection function*, $H : \mathbb{R}^+ \to [0, 1]$. The connection function maps the distance between two

nodes to a probability, i.e. the connection function tells us the probability that two nodes are connected. The hard RGG is therefore a special case of the soft RGG in which H is defined as

$$H(r) = \begin{cases} 1, & \text{if } r \leq r_0, \\ 0, & \text{otherwise,} \end{cases}$$

where r denotes the (Euclidean) distance between the two nodes. Although many different forms of this connection function have been used in the literature (see [DG16] for a summary of those used in the wireless communications literature), the results in this thesis hold for a very general class of connection functions. In Chapter 3, the only condition placed on the connection function is that it is integrable (and non-zero), i.e.

$$0 < \int_0^\infty H(r)dr < \infty.$$

This is generally considered as a necessary condition to make the problem non-trivial since this quantity is related to the mean degree of the system. In Chapters 4 and 5, we introduce the additional constraints that the connection function is monotonically decreasing. This condition is widely applied to the connection function when this model is applied to wireless networks since it makes sense that if two users are further apart, they are less likely to be able to communicate with each other. Also with this context in mind, in the simulations in Chapter 3 and in the analysis at the end of Chapter 5, we pay particular attention to the *generalised Rayleigh* connection function. This is calculated as the probability that the signal-to-interference-plus-noise ratio (SINR) of a received signal from a node a distance r away is greater than some threshold, q, when the channel gain follows a Gaussian distribution. We denote by G_{T_j} and G_R the antenna gains of transmitter j and receiver respectively, s denotes the channel gain, and $\mathcal{N}$ the noise power. The path-loss is given by $r^{-\eta}$ where η is the path-loss exponent, a quantity that depends on the environment in which the network exists. $\eta = 2$ is used for free space path loss (i.e. line-of-sight communication) but η can take values in the range from 1 (no absorption) to around 6 (very dense urban environments such as Manhatten) [DG16]. The signal-to-interference-plus-noise ratio (SINR) is thus defined as

$$SINR = \frac{G_{T_i} G_R |s_i|^2 r^{-\eta}}{\mathcal{N} + \gamma \sum_{k \neq i} G_{T_k} G_R |s_k|^2 r_k^{-\eta}}.$$

The parameter $\gamma \in [0,1]$ is used to apply a random thinning to the set of interfering signals, as would be present in an ALOHA transmission scheme (devices are active with probability γ) [DGP18]. We have assumed that the channel gain follows a Gaussian distribution, hence its magnitude is Rayleigh distributed and thus $|s|^2$ follows an Exponential distribution. We also assume that there is no interference within the network (i.e. $\gamma = 0$), hence we return to the signal-to-noise ratio (SNR). Therefore, we can derive the generalised Rayleigh connection function as

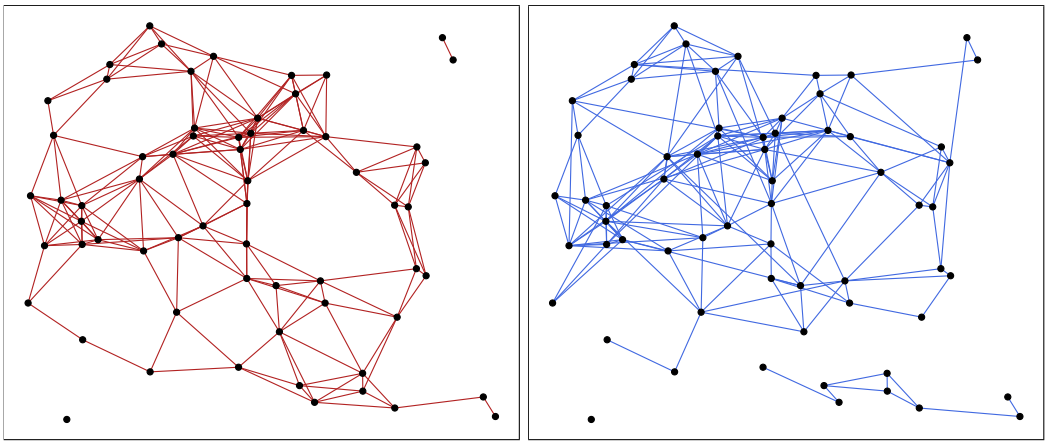
$$\begin{aligned} H(r) &= \mathbb{P}(\text{SNR} > q) \\ &= \mathbb{P}\left(\frac{G_T G_R |s|^2 r^{-\eta}}{\mathcal{N}} > q\right) \\ &= \mathbb{P}\left(|s|^2 > \frac{q\mathcal{N}}{G_T G_R} r^{\eta}\right) \\ &= \mathbb{P}\left(|s|^2 > \left(\frac{r}{r_c}\right)^{\eta}\right) \\ &= \exp\left(-\left(\frac{r}{r_c}\right)^{\eta}\right), \end{aligned} \tag{2.2}$$

where $r_c = (G_T G_R / q\mathcal{N})^{1/\eta}$. Therefore, we have that the probability of an edge between two nodes a distance r apart is given by

$$H(r) = \exp\left(-\left(\frac{r}{r_c}\right)^{\eta}\right). \tag{2.3}$$

An illustration of edge probabilities as a function of distance for different connection functions of this class is given in Figure 3.1. To note the topological differences between the soft and hard RGGs, we show an example (Figure 2.2) of both edge sets created on the same point process. We also remark that the edge set in the hard RGG is deterministic once the points have been generated whereas this is just one realisation of the edge set for the soft RGG. In these figures we can see that some of the shorter edges present in the hard RGG are not there in the soft RGG (e.g. bottom right hand corner) however some longer edges appear in the soft RGG that aren't present in the hard RGG (e.g. top right hand corner) leading to two quite different network topologies even on the same underlying point process. From a wireless network design perspective, this demonstrates the importance of the use of the soft RGG.

Eqn.(2.3) has an infinite connection range, which may seem to be an unrealistic assumption when discussing communication networks since they will live in some



(a) Hard RGG (b) Soft RGG

Figure 2.2: A side by side comparison of a soft and hard RGG both defined on the same point process. For (a), the connection function is given by $H(r) = \mathbb{1}\{r \leq 0.2\}$, and for (b), the connection function is given by $H(r) = \exp(-(r/0.2)^2)$.

bounded space, however, it possibly allows for better insights into the behaviour of such networks. The main reason being that it removes the need to place a predefined cut-off on the connection range. This is very difficult to quantify since it will depend in some complex manner on the local topography of the network and hence assumptions will have to be made in defining this. To avoid this, we make the assumption of infinite range whilst being conscious of the fact that at large distances, the probability of connectivity becomes arbitrarily close to zero.

The connectivity of soft RGGs in dimensions 2 and higher has been studied by many different authors. In the work of Coon, Dettmann, and Georgiou [CDG12], isolated nodes are analysed via a cluster expansion approach as a way to average over all node and edge configurations, thus allowing the authors to understand how the boundaries affect the connectivity of the network. This work gives a better understanding of finite sized systems by taking these boundary conditions into consideration. In Mao and Anderson's work [MA11b; MA12] they investigate the connectivity of soft RGGs by looking at the points of a PPP with density ρ which fall in the unit square and connect nodes with a probability that is inversely proportional to the density, i.e. a scaled connection function. The critical scaling for the threshold at which isolated nodes exist (or don't exist) almost surely was shown to be $H^\rho(r) = H(r/r_\rho)$, where $r_\rho = \sqrt{(\ln \rho + b)/(\|H\|_1 \rho)}$ for some $b > 0$, and $\|H\|_1 = \int_0^\infty H(z)dz$. Therefore, the scaling is insensitive to the connection function and depends only on its integral. In this work, they showed that if the connection

function satisfies $H(r) = o(1/(x^2 \ln x))$, then in the limit as $\rho \to \infty$ that (via the Chen-Stein method) the number of isolated nodes in the network follows a Poisson distribution with mean e^{-b} and that asymptotically almost surely there is no fixed and finite order strictly greater than one (i.e. either a node is part of the giant component or it is an isolated node). Therefore, a sufficient and necessary condition for the graph to be fully connected is that there are no isolated nodes which occurs if $b = b(\rho) \to \infty$ as $\rho \to \infty$. Penrose [Pen16] then continued this investigation by looking at a large class of exponentially decaying connection functions and proving this result for dimensions greater than or equal to 2 along with a *de-Poissonised* version of the result for any locally finite (point) set $\mathcal{X} \subset \mathbb{R}^d$. Denoting by $\mathcal{K}$ the class of connected graphs, and $N_{\text{iso}}(\mathcal{G})$ the number of isolated nodes in the graph $\mathcal{G}$, then for a large number of nodes in the graph, the main result from these papers is neatly summarised in Penrose's paper as

$$\mathbb{P}[\mathcal{G} \in \mathcal{K}] \approx \mathbb{P}[N_{\text{iso}}(\mathcal{G}) = 0] \approx \exp(-\mathbb{E}[N_{\text{iso}}(\mathcal{G})]).$$

However, as with the hard RGG, when discussing the 1-D soft RGG there is still the disparity between the two models in understanding where the main barrier to connectivity comes from since we are once again in a position where it is not obvious that this obstruction will be isolated nodes. As before, this comes from the fact that there is the potential for the graph to split into two large clusters. No work as of yet has been done on investigating the connectivity of the 1-D model and any work where this problem has been mentioned has either remarked on its difficulty, or has not been satisfactory. Therefore, any results on the 1-D soft RGG given in this thesis are novel and hence this model is the main focus of Chapters 3, 4, and 5.

2.1.4 A Brief Note on the Domain

Before continuing, it is important to make a quick note on the domain within which we will be working in this thesis. With respect to the results within this thesis for one-dimensional soft RGGs, we are generally interested in a line segment of length L, i.e. $[0, L] \cap \mathbb{R}$. However, the results given in Chapters 3 and 4 on *isolated nodes* will instead look at a torus of length L, i.e. $\mathbb{R}/L\mathbb{Z}$ (or a line segment of length L for which the end points are identified). The use of the torus is to remove the issue of dealing with the boundary conditions which add a large amount of technical complexity without much insight, especially in the regime considered here in which L tends to

infinity. In Chapter 5, we will instead once again consider the line segment $[0, L]$. The reason here being that when discussing *uncrossed gaps*, the existence of a single uncrossed gap does not necessarily fragment the graph. To ensure that the network is not fully connected requires two uncrossed gaps rather than one. Therefore it makes sense to consider the line segment and analyse the event of the existence of at least one uncrossed gap.

2.2 Probabilistic Tools

In addition to the construction of the graphs as described in the previous sections, we must also introduce some tools to understand the mathematics involved in analysing them. We begin by introducing the Markov and Chebyshev inequalities before moving on to the Chen-Stein method for Poisson approximation.

2.2.1 Concentration Inequalities

Two of the most widely concentration inequalities when analysing random variables are the Markov and Chebyshev inequalities. Although they only give very weak bounds on the behaviour of these random quantities, since they only depend on their first two moments, they are able to give good insight in a relatively simple way.

Theorem 2.2.1 (Markov's Inequality)**.** *Let X be a non-negative random variable with finite mean $\mathbb{E}[X]$. Then for all $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{2.4}$$

In particular, if X is a non-negative integer valued random variable,

$$\mathbb{P}(X = 0) \geq 1 - \mathbb{E}[X]. \tag{2.5}$$

Chebyshev's inequality puts a bound on the probability that the random variable X is far from its mean and is given by:

Theorem 2.2.2 (Chebyshev's Inequality)**.** *Let Y be a random variable with finite mean, $\mathbb{E}[Y]$, and finite variance, $\mathrm{Var}(Y)$. Then for all $a > 0$,*

$$\mathbb{P}\left(|Y - \mathbb{E}[Y]| \geq a\right) \leq \frac{Var(Y)}{a^2}. \tag{2.6}$$

In particular,

$$\mathbb{P}(Y = 0) \leq \frac{Var(Y)}{(\mathbb{E}[Y])^2}. \tag{2.7}$$

A corollary of Markov's inequality is Chernoff's inequality (also known as the Bernstein inequality) which is given by:

Theorem 2.2.3 (Chernoff's Inequality)**.** *For a non-negative random variable X and any $t > 0$,*

$$\mathbb{P}\left(e^{tX} \geq e^{ta}\right) \leq e^{-ta}\mathbb{E}[e^{tX}].$$

2.2.2 Poisson Approximations and the Chen-Stein method

When analysing RGGs, our aim is to understand the underlying behaviour of these random objects. One way to do this is to try to work out whether their properties can be approximated by a known probability distribution. As discussed above, one approach is to see whether such a property can be approximated by a Poisson distribution; for example the number of isolated nodes in the network. The most widely used tool when making this approximation is the Chen-Stein method for Poisson approximation (see e.g. [AGG+89; BHJ92]). These references give a very in depth study of the method, but here we will simply aim to give the reader a brief overview of the method to enable them to follow the work done in this thesis. This method places a bound on the total variational distance, d_{TV}, between a sum of (almost) independent indicator functions and a Poisson random variable (or more formally between their distributions). Therefore, since total variation is a distance metric and hence non-negative, the general method is to find an upper bound on d_{TV} and show that this tends to 0 in some limit. This therefore proves that the distribution of the sum of indicator functions converges in total variational distance to that of a Poisson random variable. We also note that convergence in total variational distance is stronger than weak convergence and hence implies convergence in distribution.

The total variational distance is a measure of the distance between two probability distributions and is defined as:

Definition 2.2.1. *(Total variational distance) Let $\mathcal{B}$ denote the set of all Borel sets. For two integer valued probability distributions, P and Q, defined on the same probability space,*

$$d_{TV}(P, Q) = \sup_{A \in \mathcal{B}} |P(A) - Q(A)|,$$

i.e. the maximum possible difference that the two probability distributions can assign to the same event.

We begin with a the graph $\mathcal{G} = (V, E)$, i.e. a graph with vertex set V, and edge set E. For $i \in V$, denote by I_i a Bernoulli random variable with $p_i := \mathbb{E}[I_i] = \mathbb{P}(I_i = 1)$ and $p_{ij} = \mathbb{E}[I_i I_j] = \mathbb{P}(I_i = 1, I_j = 1)$. Let $W = \sum_{i \in V} I_i$ and $\zeta = \mathbb{E}[W] \in (0, \infty)$. The indicator I_i defines whether the vertex $i \in V$ has some particular property. For example, within the context of this work, $I_i = 1$ may denote the fact that vertex i is isolated. By this notation, W then denotes the total number of vertices within our vertex set, V, which have this property. Continuing the previous example, W would therefore denote the total number of vertices within V which are isolated, i.e. how many isolated nodes are there in the graph $\mathcal{G} = (V, E)$.

For each $i \in V$, we define a *neighbourhood of dependence*, B_i, such that I_i is (nearly) independent of all the I_j for j outside of B_i. Define

$$\begin{aligned} b_1 &:= \sum_{i \in V} \sum_{j \in B_i} p_i p_j, \\ b_2 &:= \sum_{i \in V} \sum_{i \neq j \in B_i} p_{ij}, \\ b_3 &:= \sum_{i \in V} \mathbb{E}\left[\left| \mathbb{E}\left[I_i | (I_j : j \in V \backslash B_i) \right] - p_i \right| \right]. \end{aligned}$$

As described in [AGG+89], b_1 measures the neighbourhood size, b_2 measures the expected number of neighbours of a given occurrence, and b_3 measures the dependence between an event and the number of occurrences of that event outside of its neighbourhood. Denoting by $Po(\lambda)$, a Poisson random variable of rate λ, we state the main result of this section ([BHJ92, Th. 1.A], and [MA12, Th. 12]):

Theorem 2.2.4. *For b_i, $i = 1, 2, 3$ and W defined as above,*

$$d_{TV}\left(W, Po\left(\mathbb{E}[W]\right)\right) \leq \min\left(1, \frac{1}{\mathbb{E}[W]}\right)(b_1 + b_2 + b_3).$$

A derivation of this result can be found in many books and papers (see e.g. [Che75; BHJ92] or Chapter 2 of [Pen03]) but will be skipped in this thesis for the sake of brevity. Theorem 2.2.4 therefore places a bound on the total variational distance between the (distribution of the) sum of (almost) independent indicator functions, W, and a Poisson random variable with mean $\mathbb{E}[W]$ in terms of the b_i as defined above. Thus, to prove that W converges in total variational distance (and

hence in distribution) to $Po(\mathbb{E}[W])$, one simply has to show that these b_i terms tend to zero in some limit. One of the main benefits of this method, is that it depends only on the first two moments of W (unlike e.g. the method of moments approach).

When applying this method to RGGs, it is obvious to see that for a hard RGG, if the size of the neighbourhood is set to be sufficiently larger than the connection radius, then the b_3 term is zero and this term can be ignored. However, when analysing the soft RGG, this cut off does not exist and hence some care must be taken in the calculation of this term. We will use this approach in Chapter 4 to show that the number of isolated nodes in our 1-D soft RGG model (under certain conditions) can be well approximated by a Poisson distribution.

2.3 Mobility Models: Random Waypoint

The two main applications of the work done in this thesis are to vehicular networks and swarm robotics, two types of communication network in which mobility is a key component. Although the mathematical analysis done here on VANETs looks at a snapshot in time, the work done on swarms incorporates mobility into the research. To allow for a baseline to test the efficacy of our algorithm against, we used completely random motion by modelling the movements of the robots via the Random Waypoint (RWP) mobility model.

2.3.1 Random Waypoint Model

In the RWP mobility model, each node chooses a destination uniformly at random in some deployment region Q with each node's choice being completely independent of each other. The node then travels to this destination with constant velocity $v \in [v_{\min}, v_{\max}]$, where v is chosen Uniformly at random, and $v_{\min} > 0$ and $v_{\max} < \infty$. Once the node arrives at this location it remains static for a random pause time t_p chosen from $[0, t_{p,\max}]$ with $t_{p,\max} < \infty$. Once an original distribution of the node locations is defined, this model completely describes the movement of all of the users in the network. Figure 2.3 shows an example of the first six movements of a single node in a 2-dimensional state space.

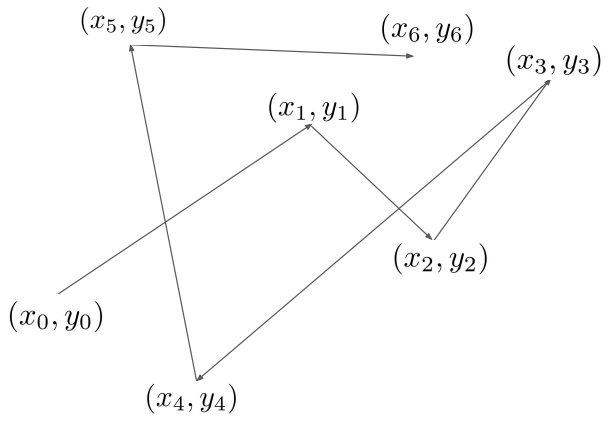


Figure 2.3: Example of node's first six movements in a 2-dimensional RWP mobility model.

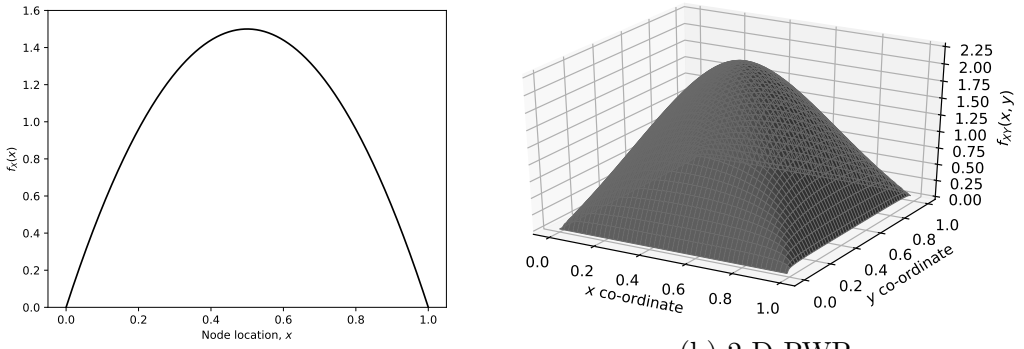
One object of interest is the distribution of the node locations in space. This will aid in the understanding of how the nodes will interact with each other and also which areas will have higher densities of nodes. It was shown in [BRS03] that under the assumption of no pause time (i.e. $t_p = 0$) and no nodes that remain static forever, whatever the starting deployment of the nodes (whether it be uniform or some other distribution) the node locations become non-uniform over time. This behaviour can be understood through the following:

- Nodes tend to cross the centre of Q with relatively high frequency.
- For a long running time, the stochastic distribution of the nodes converges to an asymptotically stable distribution with the maximum node density in the middle of Q.

In [BRS03], under the assumptions of no pause time and no permanently static nodes, the asymptotically stable pdf of the location of a mobile node moving on a line segment $[0, a]$ and on the unit square $[0, 1]^2$ were calculated. The 1-dimensional case gives this pdf to be

$$f_X(x) = \begin{cases} -\frac{6}{a^3}x^2 + \frac{6}{a^2}x, & \text{for } 0 < x < a, \\ 0, & \text{otherwise.} \end{cases}$$

The pdf's for the 1- and 2-D cases are illustrated in Figure 2.4.



(a) 1-D RWP

(b) 2-D RWP

Figure 2.4: The asymptotically stable pdf of the location of a mobile node moving on (a) the unit line segment $[0, 1]$, and (b) on the unit square $[0, 1]^2$, under the assumptions of no pause time and no permanently static nodes.

2.3.2 Markovian Random Waypoint Model

The RWP mobility model as a baseline for comparing against, however in the context of swarm robotics, it is not a very realistic one due to the completely random destination choice. One extension of this model is to allow for the destination choice of the user to be weighted such that certain areas have a higher probability of being chosen. Such examples are the Markovian Random Waypoint model [HLV06] in which the waypoints constitute a Markov process and given the current waypoint, the next waypoint is chosen according to a conditional pdf. The application stated in this work is to the creation of "typical routes" of the users in the network. Another (almost identical) model, which the authors called the Weighted Random Waypoint model, was studied in [HMS+05]. Here, the space was broken down into a discrete set of potential locations and probabilities assigned to the movement between any two pairs of locations which depended on both their geographical position as well as the time at which the user arrived into a segment. Here the authors were comparing this model with the RWP mobility model to see how this effected network performance at a University campus. Both of these extensions allow for a more accurate model of how users in a swarm network may move around an area: rather than choosing their destinations at random, a decision would be made based upon where they currently are and what their task is.

Chapter 3

Soft Random Geometric Graphs: Isolation

When studying connectivity in graphs, the natural first hurdle is to understand *isolated nodes* i.e. nodes which have no edges to any other nodes in the network. If a node in a communication network is isolated, then it is unable to send messages to other users as well as being unable to receive them, thus leading to the network not being fully connected. In higher dimensional versions of this model it has been shown that, under certain constraints on the connection function, in the ultra-dense limit isolated nodes are the only obstruction to having a fully connected network [MA12; Pen16].

In this chapter, we begin by defining the model upon which our work is based, namely the one-dimensional soft RGG. Simulations are then studied to give an understanding of the connectivity properties of this graph as well as introduce a conjecture on what is causing the lack of full connectivity. This then leads to us deriving bounds on the probability of isolated nodes occurring in the network and finding the scaling regime in which isolated nodes occur with high probability.

The work in this chapter is based on the paper [WDG20] for which I was a co-author. This paper was written in collaboration with Professor Carl Dettmann and Dr. Ayalvadi Ganesh. All of the simulation work was conducted by myself using the NetworkX [HSS08] package within Python along with some of my own code[1]. The theoretical analysis was primarily conducted by myself with CD and AG contributing ideas to this analysis along with checking calculations and suggesting proof methods.

[1]This code along with the data analysed can be found in my GitHub repository here https://github.com/mw12747/1D_Soft_RGG.

3.1 The Model

Node locations in RGGs are typically modelled by point processes [Hae12], and most commonly by a Poisson point process, as defined in Chapter 2. In our model, the nodes of the RGG are the points of a homogenous PPP, $\mathcal{P}$, of intensity $\lambda = 1$ on an interval $[0, L]$, where $L \in (0, \infty)$. The edge between points x and y in $\mathcal{P}$ is present with probability $H(|x - y|)$, independent of the point configuration and of all other edges, thus defining a soft RGG. Here, $|x-y|$ denotes the Euclidean distance between x and y, and $H : \mathbb{R}_+ \to [0, 1]$ is the connection function.

The model is parametrised by the scalar L and the function $H(\cdot)$. We impose mild restrictions on H in Section 3.3 as well as in Chapters 4 and 5, but do not assume a specific functional form. There is no loss of generality in assuming that $\lambda = 1$ as this simply defines the unit of length. Consequently, λL is a dimensionless quantity rather than having units of length. Another common scaling in the literature is to set $L = 1$, while making λ a free parameter. These are equivalent up to a suitable rescaling of $H(\cdot)$ [MA11b].

The choice of a Poisson process is made primarily for analytical tractability, but is realistic for traffic in its free flow state [RPM04], and hence for VANETs. The most widely used connection functions in the wireless communications literature are of the form

$$H(r) = \exp\left(-(r/r_c)^\eta\right), \tag{3.1}$$

as was derived in Eqn. (2.2). As before, $r_c > 0$ specifies the link range, while $\eta > 0$ is related to the path loss exponent. The most common examples are the Waxman and Rayleigh connection functions, which correspond to $\eta = 1$, and $\eta = 2$ respectively:

$$H_{\text{Wax}}(r) = e^{-(r/r_c)}, \quad H_{\text{Ray}}(r) = e^{-(r/r_c)^2}. \tag{3.2}$$

The hard RGG is recovered in the limit of η tending to infinity. Figure 3.1 depicts edge probabilities as a function of distance, for different connection functions from this class. More general connection functions, many of which have the same general shape, may be found in [DG16].

Our goal is to derive expressions for the probability that a soft RGG generated by our model is fully connected. This appears intractable, so we restrict attention to two specific modes of disconnection, namely *isolated nodes* and *uncrossed gaps*. These are defined in the next section, and shown to account for most disconnections in extensive simulations. But even for these events, exact expressions for their probabilities

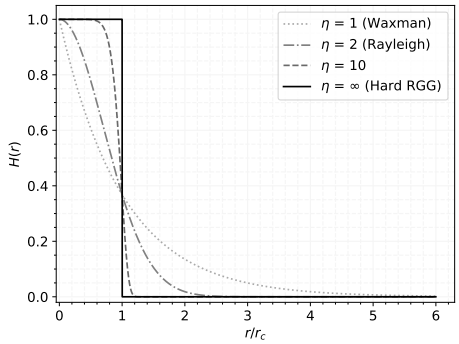


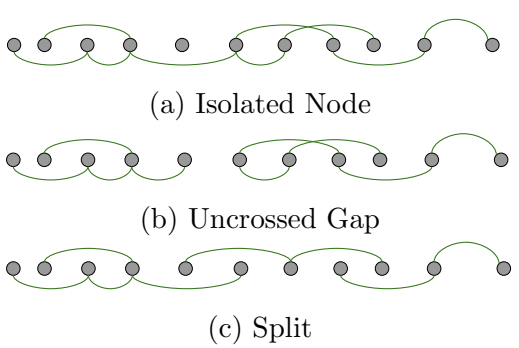
Figure 3.1: The probability that two nodes, a distance r apart, are connected for different values of η as defined in Eqn. (3.1) with $r_c = 1$.

cannot be obtained in closed form. Hence, we study an asymptotic regime where L tends to infinity while the connection function is rescaled as $H^L(\cdot) = H(\cdot/R_L)$; we seek to identify a scaling regime R_L, tending to infinity at a specified rate with L, which is critical for the emergence of isolated nodes or uncrossed gaps.

It is known in two or more dimensions that, for natural analogues of our model, the existence of isolated nodes exhibits a sharp change at a scaling of $R_L = \ln L$ (as discussed in Section 2.1.3). That is, for R_L growing significantly faster than $\ln L$ there are longer range connections and no isolated nodes with high probability, whilst for R_L growing significantly slower than $\ln L$ there are infinitely many isolated nodes in the limit. We show in Section 3.3 that a similar change at a $\ln L$ scaling is also observed in 1-D if the connection function is integrable. In Section 5.1, we show, under the additional condition that the connection function is monotonically decreasing and has unbounded support, that uncrossed gaps have a vanishingly small probability of occurring in this scaling regime. This is true even for rapidly decaying connection functions, and is in stark contrast to the hard RGG model, where uncrossed gaps are at least as likely as isolated nodes.

3.2 Simulations

A soft RGG in one dimension may fail to be connected in many ways. We conjecture that the two main obstructions to full connectivity are the presence of isolated nodes or uncrossed gaps, defined below. In the following, V denotes a realisation of $\mathcal{P}$, i.e., a point configuration comprising the node set of the graph.



(a) Isolated Node

(b) Uncrossed Gap

(c) Split

Figure 3.2: Three different disconnection modes.

Definition 3.2.1 (Isolated Node)**.**
A node $x \in V$ is said to be isolated if there is no edge between the point x and any point $y \in V \setminus \{x\}$.

Definition 3.2.2 (Uncrossed Gap)**.**
An uncrossed gap is said to occur at $x \in V$ if $V \cap (x, L]$ is non-empty and there are no edges between $V \cap [0, x]$ and $V \cap (x, L]$.

Notice that in a hard RGG, if $v \in V$ is isolated and u is the rightmost point of V in $[0, v)$ (if there is one), then there is an uncrossed gap at u, and another one at v. Thus, a hard RGG may have uncrossed gaps without isolated nodes but not conversely; it is disconnected exactly when there is an uncrossed gap. The situation for soft RGGs is more complicated. It may be disconnected even if there are no isolated nodes or uncrossed gaps, as illustrated in Figure 3.2c. Nevertheless, we conjecture that isolated nodes and uncrossed gaps together account for most of the probability of the graph being disconnected, converging to all of it in a suitable limiting regime. While we have been unable to prove this, we present some evidence below from simulations.

In order to gain an understanding of the causes of disconnection, simulations were run for different system sizes and link ranges, for both Waxman and Rayleigh connection functions. We denote by P_{iso} the proportion of simulation runs (for a fixed set of parameter values) in which the graph contained an isolated node, by P_{ucg} the proportion with an uncrossed gap and by $P_{\text{iso}\cup\text{ucg}}$ the proportion with either. Comparing these with P_{dis}, the proportion of simulation runs in which the network fails to be fully connected, yields insights into the primary causes of disconnection.

The simulations were run for two different system sizes, $L = 1\,000$ and $L = 10\,000$; since the Poisson process of node locations has unit intensity, L is the expected

number of nodes. The link range r_c is varied, and the proportions of simulations exhibiting isolated nodes, uncrossed gaps or disconnection are plotted against the mean node degree (i.e. the average number of edges per node) corresponding to that value of r_c. (We chose not to plot the results against r_c as the meaning of this parameter is somewhat opaque, whereas the mean node degree is intuitive.) The findings are displayed in Figure 3.3 for Waxman and Rayleigh connection functions as described in Eqn. (3.2). The plots are based on 5 000 simulations for each set of parameter values. In both figures, we observe a fairly sharp transition between mean degrees for which disconnection (and its individual causes) have probability close to 1 and those for which it has probability close to 0. For isolated nodes, this transition occurs close to a mean degree of $\ln L$, exactly as for Erdős-Rényi random graphs.

Figure 3.3a shows that in a system with 1 000 nodes on average and a Waxman connection function, isolated nodes and uncrossed gaps are approximately equally prevalent. But as the system size increases to 10 000 nodes on average, Figure 3.3b shows that isolated nodes become dominant. We also see that $P_{\text{iso}\cup\text{ucg}}$ is almost exactly the same as P_{dis}, supporting our intuition that node isolation and uncrossed gaps are the dominant modes of disconnection.

Figures 3.3c and 3.3d show that, for a Rayleigh connection function, the main obstruction to full connectivity appears to be uncrossed gaps. As the hard RGG is obtained in the limit as η tends to infinity, and uncrossed gaps are the cause of disconnection in hard RGGs, it seems intuitive that they should dominate for large enough values of η. This intuition is supported by the simulations. Nevertheless, it is at odds with the analysis presented in Section 5.1, which shows that isolated nodes dominate asymptotically whenever the connection function, $H(r)$, has unbounded support, i.e. it has infinite range. It appears then that the asymptotics don't kick in even at the large system sizes simulated. This includes simulations not shown in this thesis looking at networks with up to 100 000 nodes. This discrepancy is discussed in further detail in Chapter 5 with some more detailed calculations given in Section 5.3 with a more detailed discussion about finite system sizes, especially with respect to communication networks, being given in Section 7.1.2. Very large system sizes appear to be necessary to see this behaviour and hence it remains an open question to resolve this apparent discrepancy. Once again, comparing $P_{\text{iso}\cup\text{ucg}}$ with P_{dis} shows that the main contribution to disconnection comes from isolated nodes and uncrossed gaps. This thesis will therefore focus on a theoretical analysis of these two events.

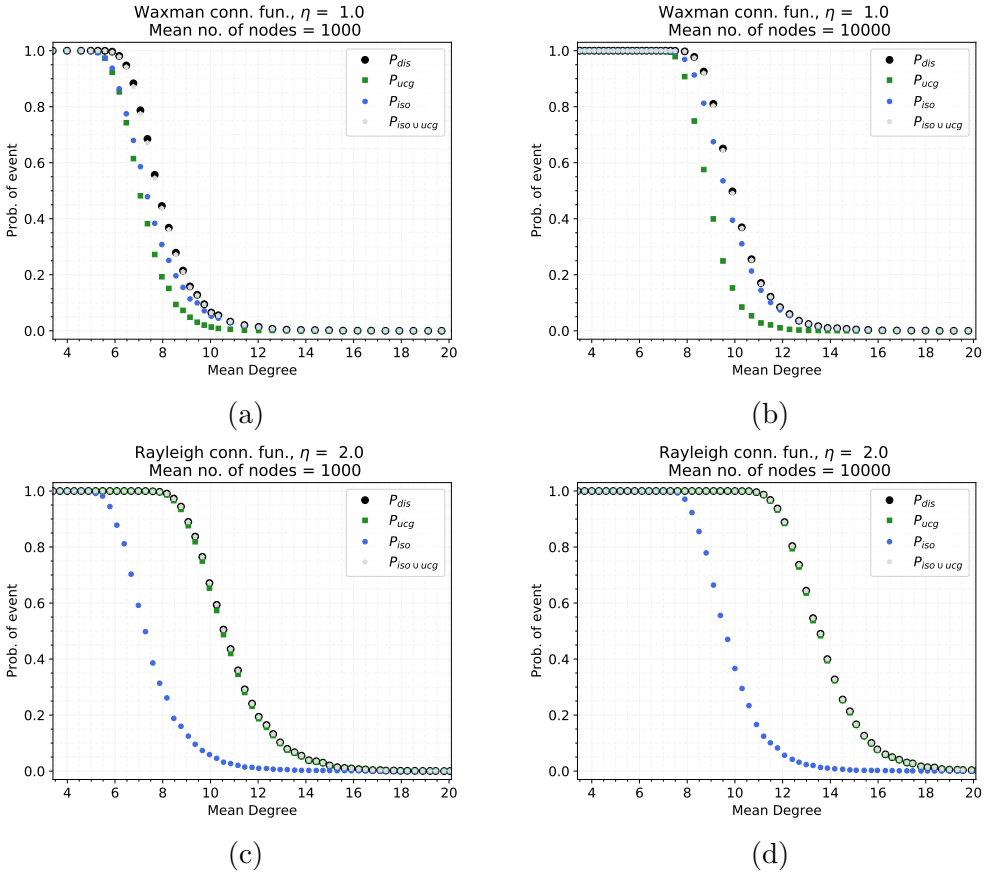


Figure 3.3: The proportion of simulations in which the network is disconnected (P_{dis}), compared with the proportion of simulations suffering from isolated nodes (P_{iso}), uncrossed gaps (P_{ucg}), or either of the two ($P_{\text{iso}\cup\text{ucg}}$). Each plot is based on 5 000 simulations of a soft RGG whose nodes are generated according to a unit rate PPP on a line segment of length $L = 1\,000$ or $L = 10\,000$, and whose edges are created according to a Waxman or Rayleigh connection function (Eqn. (3.2)).

3.3 Scaling Regimes

In this section, we derive bounds on the probability that there are isolated nodes, namely, nodes which have no edge to any other node. We then explore an asymptotic regime in which L, the expected number of nodes in the system, tends to infinity, while the range of the connection function grows as a function of L. We seek to identify a scaling regime for the connection function at which a sharp threshold for the existence of isolated nodes can be discerned.

In order to eliminate boundary effects and simplify calculations, we modify our model slightly by identifying the endpoints of the line segment $[0, L]$, thereby enforcing periodic boundary conditions (PBCs). This is equivalent to modifying the line segment to be a one-dimensional torus (or a circle). The effect of this modification on the number of isolated nodes is negligible, as it only affects nodes close to the boundary. We do not formalise this assertion here as this would add significant complexity but without much insight.

Denote by $\rho(x, y)$ the circular (or toroidal) distance between x and y, and by $\rho_\infty(x, y) = |x - y|$ the Euclidean distance between them. Define

$$h = H \circ \rho,$$

where $\circ$ denotes composition of functions; $h(x, y)$ is the probability that two points at locations x and y on the line segment with PBCs are connected. Note that $h(x, y)$ is symmetric, and invariant to translations on this line segment, i.e. $h(x, y) = h(y, x)$ and $h(x, y) = h(x - t, y - t)$ for some translation t. Denote by P_x the Palm measure conditional on the Poisson process having a point at x, i.e., $\mathrm{P}_x(E)$ is the probability of an event E conditional on there being a node at x. The Palm measures conditional on any finite set of points are denoted analogously. We write $\mathbb{E}_x, \mathbb{E}_{x,y}$ etc. for the corresponding expectations. We make the following assumption about the connection function throughout this section.

Assumption A: The function $H : \mathbb{R}_+ \mapsto [0, 1]$ is integrable, i.e.,

$$\|H\|_1 = \int_0^\infty H(x)dx < \infty.$$

We now derive expressions for the expectation and variance of the number of isolated nodes. We will use these, along with Markov's and Chebyshev's inequalities, to obtain probability bounds on the existence of isolated nodes. (See Sec 2.2.1 for

more on this topic).

We start with the expected number of isolated nodes. Let $\chi(x)$ denote the indicator that there is a point of the Poisson process at x and that it is isolated. Now, conditional on there being a point at x, the remaining points constitute a unit rate Poisson process on $[0, L]$ by Slivnyak's theorem (Theorem 2.1.3). Consequently, the set of neighbours of x (the points to which an edge is present) constitute an inhomogenous Poisson process, with intensity measure $H(\rho(x, \cdot)) = h(x, \cdot)$. The event that the point at x is isolated is the event that this Poisson process is empty. Recall (Defn. 2.1.7) that for a Poisson process X_ν with intensity $\nu(\cdot)$ on $[0, L]$, the probability that this process is empty is given by $\mathbb{P}(X_\nu = \emptyset) = e^{-\int_0^L \nu(x)dx}$. The probability that the node at x is isolated is the expectation of the indicator of this event:

$$\begin{aligned} \mathbb{E}_x[\chi(x)] &= \mathbb{P}_x(\chi(x) = 1) \\ &= e^{-\int_0^L h(x,y)dy} = e^{-\int_0^L h(0,y)dy}, \end{aligned} \tag{3.3}$$

which does not depend on x, as expected by the symmetry in the model.

Let N_{iso} denote the number of isolated nodes. By integrating Eqn. (3.3) over $x \in [0, L]$ with respect to the intensity of the Poisson process, which was assumed to be unity, we get

$$\mathbb{E}[N_{\text{iso}}] = \int_0^L \mathbb{E}_x(\chi(x))dx = Le^{-2\int_0^{L/2} H(y)dy}, \tag{3.4}$$

where the last equality holds because we are working with toroidal distance. As L tends to infinity, the exponent tends to $-2\|H\|_1$, which is a finite constant. Hence, $\mathbb{E}[N_{\text{iso}}]$ scales linearly in L. This leads us to ask if the random variable N_{iso} does so as well, i.e., if it behaves as an extensive quantity, in the language of thermodynamics. The answer depends on whether correlations decay quickly enough, and can be addressed by calculating the variance of N_{iso}.

As a first step towards computing the variance, we condition on there being points at x and y; by Slivnyak's theorem again, the Palm measure $\mathrm{P}_{x,y}$ corresponds to a unit rate Poisson process on $[0, L] \setminus \{x, y\}$. For any finite $\mathcal{A} \subset \mathbb{R}$ and any $y \in \mathbb{R}$ with $y \notin \mathcal{A}$ we denote

$$\begin{aligned} \phi_h(y, \mathcal{A}) &= 1 - \prod_{x \in \mathcal{A}} (1 - h(x, y)) \\ &= \mathbb{P}(y \text{ is non-isolated in } \mathcal{A}) \end{aligned} \tag{3.5}$$

The set of points that are neighbours of either x or y constitute an inhomogenous

Poisson process of intensity $\phi_h(\cdot, \{x, y\})$. The event that the points at both x and y are isolated is the event that this point process is empty, and that there is no edge between x and y. Hence,

$$\begin{aligned} \mathbb{E}_{x,y}[\chi(x)\chi(y)] &= (1 - h(x,y))e^{-\int_0^L \phi_h(z,\{x,y\})dz} \\ &\leq e^{-\int_0^L \phi_h(z,\{x,y\})dz}, \end{aligned} \tag{3.6}$$

since $0 \leq h(x, y) \leq 1$. Using that the expected number of ordered pairs of distinct isolated nodes is found by integrating Eqn. (3.6) with respect to x and y weighted by the joint density of the node pairs (equal to unity here), along with translation invariance,

$$\begin{aligned} \mathbb{E}[N_{\text{iso}}(N_{\text{iso}} - 1)] &= \int_0^L \int_0^L \mathbb{E}_{x,y}[\chi(x), \chi(y)]dxdy \\ &= L \int_0^L \mathbb{E}_{0,x}[\chi(0)\chi(x)]dx \\ &\leq L \int_0^L e^{-\int_0^L \phi_h(z,\{0,x\})dz} dx. \end{aligned} \tag{3.7}$$

Recall that the squared coefficient of variation of a random variable, denoted c_V^2, is defined as the ratio of its variance to its squared mean. Hence, we obtain from Eqns. (3.3), (3.4), (3.6), (3.7) that

$$\begin{aligned} c_V^2(N_{\text{iso}}) - \frac{1}{\mathbb{E}[N_{\text{iso}}]} &= \frac{1}{L} \int_0^L \frac{\mathbb{E}_{0,x}[\chi(0)\chi(x)]}{\mathbb{E}_0[\chi(0)]^2} dx - 1 \\ &\leq \frac{1}{L} \int_0^L \left(\frac{e^{-\int_0^L \phi_h(z,\{x,y\})dz}}{e^{-2\int_0^L h(0,z)dz}} - 1 \right) dx. \end{aligned} \tag{3.8}$$

Since this holds for any y, we set y to be 0 for arithmetic ease. Expanding out $\phi_h(z, \{x, 0\})$, Eqn. (3.8) becomes

$$c_V^2(N_{\text{iso}}) - \frac{1}{\mathbb{E}[N_{\text{iso}}]} = \frac{1}{L} \int_0^L \left(\frac{e^{-\int_0^L [h(0,z)) + h(x,z) - h(0,z)h(x,z)]dz}}{e^{-2\int_0^L h(0,z)dz}} - 1 \right) dx.$$

Finally, due to the periodic boundary conditions, $\int_0^L h(x, z)dz = \int_0^L h(0, z)dz$ and

we finally arrive at

$$\begin{aligned} c_V^2(N_{\text{iso}}) - \frac{1}{\mathbb{E}[N_{\text{iso}}]} &= \frac{1}{L}\int_0^L \left(\frac{e^{-2\int_0^L h(0,z) + \int_0^L h(0,z)h(x,z)dz}}{e^{-2\int_0^L h(0,z)dz}} - 1 \right) dx \\ &= \frac{1}{L}\int_0^L \left(e^{\int_0^L h(0,z)h(x,z)dz} - 1 \right) dx. \end{aligned} \tag{3.9}$$

Since h is bounded above by 1, we have

$$\begin{aligned} \int_0^L h(0,z)h(x,z)dz &\leq \int_0^L h(0,z)dz \\ &\leq 2\int_0^\infty H(z)dz < \infty. \end{aligned}$$

In other words, the above integral is bounded, uniformly in x. Hence, there is a finite constant C, which does not depend on x or L, such that

$$e^{\int_0^L h(0,z)h(x,z)dz} \leq 1 + C\int_0^L h(0,z)h(x,z)dz.$$

It follows that

$$\begin{aligned} \frac{1}{L}\int_0^L \left(e^{\int_0^L h(0,z)h(x,z)dz} - 1 \right) dx &\leq \frac{1}{L}\int_0^L \left(C\int_0^L h(0,z)h(x,z)dz \right) dx \\ &= \frac{C}{L}\int_0^L h(0,z)\left(\int_0^L h(x,z)dx \right) dz \\ &= \frac{4C}{L}\int_0^{L/2} H(z)dz \int_0^{L/2} H(y)dy \\ &\leq \frac{4C}{L}\|H\|_1^2. \end{aligned}$$

The interchange of the order of integration in the third line is justified by Tonelli's theorem. Substituting the above in Eqn. (3.9), we obtain that

$$c_V^2(N_{\text{iso}}) \leq \frac{1}{\mathbb{E}[N_{\text{iso}}]} + \frac{4C}{L}\|H\|_1^2. \tag{3.10}$$

We already noted that $\mathbb{E}[N_{\text{iso}}]$ scales linearly in L. Therefore, the right hand side above tends to zero as L tends to infinity. Now, by Chebyshev's inequality (Theorem

2.2.2), we have for any $\epsilon > 0$ that

$$\mathbb{P}\left(|N_{\text{iso}} - \mathbb{E}[N_{\text{iso}}]| > \epsilon \mathbb{E}[N_{\text{iso}}]\right) \leq \frac{c_V^2(N_{\text{iso}})}{\epsilon^2},$$

which tends to zero as L tends to infinity. Thus, N_{iso} concentrates around its expected value, and scales linearly with L, making it an extensive quantity. Therefore, as we let L tend to infinity there will asymptotically almost surely be an isolated node in the graph.

We now turn to the question of identifying a scaling regime where the probability of seeing isolated nodes in the graph exhibits a sharp transition.

3.3.1 Scaled Connection Function

Consider a family of 1-D soft RGGs as above, indexed by L, and with scaled connection functions $H^L(z) = H(z/R_L)$, where R_L is an increasing function of L. For now, the only assumption we make is that R_L tends to infinity and R_L/L tends to zero as L tends to infinity. Let $h^L = H^L \circ \rho$. We shall study these graphs in the asymptotic regime $L \to \infty$.

We begin by rewriting Eqn. (3.4) as

$$\begin{aligned}\mathbb{E}[N_{\text{iso}}] &= L \exp\left(-2\int_0^{L/2} H^L(y)dy\right) \\ &= L \exp\left(-2\int_0^{L/2} H(y/R_L)dy\right)\end{aligned}$$

Making the change of variables $u = y/R_L$, we get

$$\mathbb{E}[N_{\text{iso}}] = L \exp\left(-2R_L\int_0^{L/2R_L} H(u)du\right). \tag{3.11}$$

Since R_L/L was assumed to tend to zero, $\int_0^{L/2R_L} H(u)du$ tends to $\|H\|_1$ as L tends to infinity, and we get

$$\lim_{L\to\infty} \mathbb{E}[N_{\text{iso}}] = \lim_{L\to\infty} Le^{-2R_L\|H\|_1},$$

provided the limit on the right exists.

We see from the above expression that the scaling required is R_L growing loga-

rithmically in L. Indeed, taking $R_L = \gamma \ln L$, we get

$$\mathbb{E}[N_{\text{iso}}] \xrightarrow{L\to\infty} \begin{cases} 0, & \text{if } 2\gamma\|H\|_1 > 1, \\ 1, & \text{if } 2\gamma\|H\|_1 = 1, \\ \infty, & \text{if } 2\gamma\|H\|_1 < 1. \end{cases} \tag{3.12}$$

Thus, the mean number of isolated nodes exhibits a sharp transition at $\gamma = 1/2\|H\|_1$. We wish to show that the random variable denoting the number of isolated nodes does so as well.

It is immediate from Markov's inequality (Theorem 2.2.1), which states that $\mathbb{P}(X \geq 1) \leq \mathbb{E}[X]$ for any non-negative random variable X, that

$$\mathbb{P}(N_{\text{iso}} \geq 1) \xrightarrow{L\to\infty} 0 \quad \text{if} \quad 2\gamma\|H\|_1 > 1. \tag{3.13}$$

In words, such a choice of γ ensures that large networks have a vanishingly small chance of containing isolated nodes.

To investigate the behaviour for γ smaller than this threshold, we must return to the coefficient of variation. We can rewrite Eqn. (3.9) as

$$c_V^2(N_{\text{iso}}) - \frac{1}{\mathbb{E}[N_{\text{iso}}]} \leq \frac{1}{L}\int_0^L \left(e^{\int_0^L h^L(0,z)h^L(x,z)dz} - 1\right)dx. \tag{3.14}$$

Defining

$$g^L(x) = \int_0^L h^L(0,z)h^L(x,z)dz,$$

we have

$$\begin{aligned} \int_0^L g^L(x)dx &= \int_0^L \int_0^L h^L(0,z)h^L(x,z)dzdx \\ &= \int_0^L h^L(0,z)\left(\int_0^L h^L(x,z)dx\right)dz \\ &= \left(\int_0^L h^L(0,y)dy\right)^2 \\ &= \left(2\int_0^{L/2} H\left(\frac{y}{R_L}\right)dy\right)^2 \\ &= \left(2R_L\int_0^{L/2R_L} H(y)dy\right)^2 \\ &\leq 4R_L^2\|H\|_1^2. \end{aligned} \tag{3.15}$$

The interchange of the order of integration in the second line is justified by Tonelli's theorem. For $\alpha > 0$, define the set $\Psi_\alpha = \{x \in [0, L) : g^L(x) > \alpha\}$. Since g^L is non-negative, it follows from Eqn. (3.15) that the Lebesgue measure of the set Ψ_α, denoted $m(\Psi_\alpha)$, is bounded as follows:

$$m(\Psi_\alpha) \leq \frac{4R_L^2 \|H\|_1^2}{\alpha}. \tag{3.16}$$

Now, if x is not in Ψ_α, then $g^L(x) \leq \alpha$, and so there is a constant $C_\alpha > 0$ such that $e^{g^L(x)} \leq 1 + C_\alpha g^L(x)$. Hence,

$$\begin{aligned} \int_{\Psi_\alpha^c} \Big(e^{g^L(x)} - 1\Big) dx &\leq C_\alpha \int_{\Psi_\alpha^c} g^L(x) dx \\ &\leq 4C_\alpha R_L^2 \|H\|_1^2, \end{aligned} \tag{3.17}$$

where Ψ_α^c denotes the complement of Ψ_α in $[0, L]$. We have used Eqn. (3.15) and the non-negativity of g^L to obtain the last inequality.

Integrating $e^{g^L(x)} - 1$ over Ψ_α, we have

$$\begin{aligned} \int_{\Psi_\alpha} \Big(e^{g^L(x)} - 1\Big) dx &= \int_{\Psi_\alpha} \Big(\exp\Big(\int_0^L h^L(0,z) h^L(x,z)\Big) - 1\Big) dx \\ &\leq \int_{\Psi_\alpha} \Big(\exp\Big(\int_0^L h^L(0,z)\Big) - 1\Big) dx \\ &= \int_{\Psi_\alpha} \Big(\exp\Big(2R_L \int_0^{L/2R_L} H(z) dz\Big) - 1\Big) dx \\ &\leq m(\Psi_\alpha)\big(e^{2R_L \|H_1\|} - 1\big) \\ &\leq \frac{4R_L^2 \|H\|_1^2}{\alpha}\big(e^{2R_L \|H_1\|} - 1\big). \end{aligned} \tag{3.18}$$

We have used the fact that $h^L(\cdot, \cdot) \leq 1$ to obtain the inequality on the third line, and Eqn. (3.16) to obtain the last inequality. Now, substituting Eqns. (3.17) and (3.18) into eqn. (3.14), and noting that $\Psi_\alpha \cup \Psi_\alpha^c = [0, L]$, we get

$$c_V^2(N_{\text{iso}}) \leq \frac{1}{\mathbb{E}[N_{\text{iso}}]} + \frac{4C_\alpha R_L^2 \|H\|_1^2}{L} + \frac{4R_L^2 \|H\|_1^2}{\alpha L}\big(e^{2R_L \|H_1\|} - 1\big).$$

Thus, for the scaling $R_L = \gamma \ln L$, we have

$$c_V^2(N_{\text{iso}}) \leq \frac{1}{\mathbb{E}[N_{\text{iso}}]} + \frac{4C_\alpha \gamma^2 \|H\|_1^2 \ln^2 L}{L} + \frac{4\gamma^2 \|H\|_1^2 \ln^2 L}{\alpha L}\big(e^{2\gamma \|H_1\| \ln L} - 1\big).$$

Suppose $2\gamma\|H\|_1 < 1$. Then, as L tends to infinity, $\mathbb{E}[N_{\text{iso}}]$ tends to infinity by Eqn. (3.12), while the second and third terms in the sum on the RHS tend to zero. Hence,

$$c_V^2(N_{\text{iso}}) \xrightarrow{L\to\infty} 0, \text{ if } 2\gamma\|H\|_1 < 1. \tag{3.19}$$

But, by Chebyshev's inequality,

$$\mathbb{P}(N_{\text{iso}} = 0) \leq \mathbb{P}(|N_{\text{iso}} - \mathbb{E}N_{\text{iso}}| \geq \mathbb{E}N_{\text{iso}}) \leq c_V^2(N_{\text{iso}}),$$

and so, it tends to zero as L tends to infinity, for $\gamma < 1/2\|H\|_1$. Combining this with the result established in Eqn. (3.13), we conclude that

$$\mathbb{P}(N_{\text{iso}} = 0) \xrightarrow{L\to\infty} \begin{cases} 0, & \text{if } 2\gamma\|H\|_1 < 1, \\ 1, & \text{if } 2\gamma\|H\|_1 > 1. \end{cases} \tag{3.20}$$

Remarks.

1. The mean degree of the nodes, given by $\int_0^L h^L(0,z)dz$, is asymptotic to $2\gamma\|H\|_1 \ln L$, while the mean number of nodes in the interval $[0, L]$ is equal to L. Thus, the theorem states that there is a sharp threshold for the existence of isolated nodes when the mean degree is equal to the natural logarithm of the mean number of nodes. This is shown in Figures 4.1(a) and (b) by the dashed vertical line. This threshold is exactly the same as for the hard RGG (see, e.g, [AR02; HM06]).

2. As was stated above, a superficially different scaling is seen in the densification regime. Here, the line segment remains fixed as $[0, 1]$, whilst the intensity λ of the PPP increases to infinity, and the scaling parameter $R(\lambda)$ decreases to zero. The results given here can be translated to this regime using a suitable rescaling of $H(.)$. See [MA11b] for a more detailed discussion.

3. The threshold for the emergence of isolated nodes is insensitive to the connection function H and depends only on its integral, which is the expected number of connections. This is demonstrated in Figures 4.1(a) and (b) by the fact that the simulations for both the Waxman and Rayleigh connection functions overlap with each other. This lack of sensitivity is also seen in the threshold for soft RGGs in two dimensions which depends only on the integral of H as was shown in [MA11b].

3.4 Concluding Remarks

In this chapter, we presented empirical evidence from simulations that the main modes of disconnection of our graphs are the presence of isolated nodes, and the presence of uncrossed gaps which partition the network into two or more disjoint clusters. Motivated by this evidence, in the next two chapters we provided a rigorous mathematical analysis of the probabilities of isolated nodes or uncrossed gaps being present, in a suitable asymptotic scaling regime. Our random graph model is restricted to point sets generated according to a homogeneous Poisson process, but allows for very general connection functions with the only assumption being that the connection function is integrable.

The analysis in this chapter showed that the probability of the occurrence of isolated nodes shows an abrupt transition from nearly 1 to nearly 0 as the mean node degree increases. The transition is insensitive to the connection function, and occurs when the mean node degree is close to the natural logarithm of the mean number of nodes.

The threshold phenomenon arises in the limit of large system sizes (with the connection function scaling suitably). It does not provide explicit probabilities for the occurrence of isolated nodes in finite-sized networks. Moreover, the bounds provided by Markov's and Chebyshev's inequalities are rather weak. Finally, the limit result does not say what happens at $\gamma = 1/2\|H\|_1$.

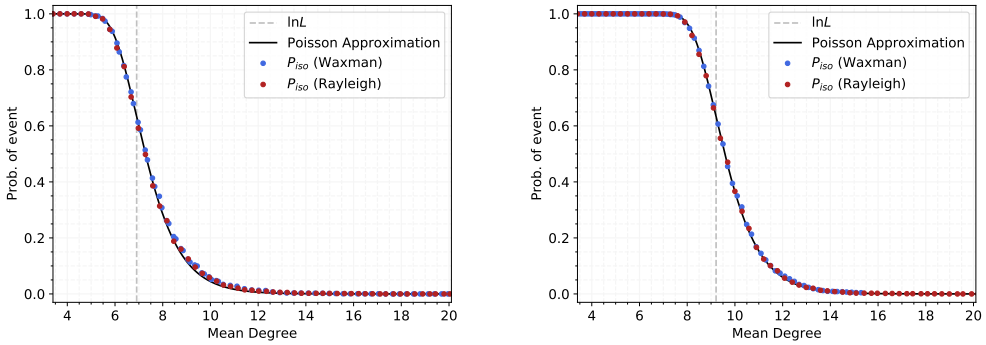
The behaviour at this critical value of γ will be the focus of the following chapter.

Chapter 4

Soft Random Geometric Graphs: Poisson Approximation

A naturally occurring phenomenon in random graphs is that the number of isolated nodes can be well approximated by a Poisson distribution [ER60; Pen97; Pen16]. Since this appears in almost all random graph models, and especially since it has been shown to be true in the higher dimensional analogues of our model, there is no reason to believe that this will not also be the case in the 1-D model.

In [Pen16], it was shown for soft RGGs in two or more dimensions with an underlying vertex set generated from a PPP, and for a large class of connection functions, that the number of isolated nodes can be well approximated by a Poisson random variable. In particular, it was proven that the probability that there are no isolated nodes is well approximated by $e^{-\lambda}$, where λ is the mean of this Poisson distribution. Evidence in support of this phenomenon also being true in our model is presented in Figure 4.1, where we have compared the proportion of simulations with no isolated nodes with this Poisson approximation for systems with $L = 1\,000$ and $L = 10\,000$. The proportion of $5\,000$ simulations in which at least one isolated node was present is plotted against the mean node degree, for both the Waxman (blue circles) and Rayleigh (red circles) connection functions. The plots demonstrate that the probability of isolated nodes appearing in the network is insensitive to whether the connection function is Waxman or Rayleigh, and depends only on the mean degree. (This can be seen by the fact that the simulations for both connection functions overlap with each other in the figures.) The plots also show that the probability of occurrence of isolated nodes rises sharply as the mean degree falls below $\ln L$, and that the transition becomes sharper as the system size grows. Finally, the solid black



(a) (b)

Figure 4.1: Probability of isolated nodes being present: comparison of simulations (circles) with Poisson approximation (black solid line). Each circle is a proportion from 5 000 simulation runs.

line in the plots depicts the function $1 - \exp\left(-Le^{-\bar{k}}\right)$, where $\bar{k}$ is the mean degree. This function is the Poisson approximation for the probability of at least one isolated node being present. The plots show that the Poisson approximation is very close to the observed prevalence of isolated nodes in the simulations. This simulation based evidence therefore gives us reason to believe that the number of isolated nodes in our model can be well approximated by a Poisson distribution.

This chapter presents a more formal statement of the theorem of a Poisson approximation which is followed by a proof based on the Chen-Stein method for Poisson approximation. The work in this chapter is as yet unpublished. Most of the theoretical analysis was conducted by myself with CD and AG aiding in the proof methods and finalising the technical details of the calculations.

4.1 The Model

As before, we will concentrate on the unit rate ($\lambda = 1$) homogeneous Poisson point process on an interval $[0, L]$ for $L \in (0, \infty)$. We denote by $\mathcal{P}_L$ the points of this PPP which lie within the line segment $[0, L]$ which defines our underlying vertex set. An edge between two nodes x and $y \in \mathcal{P}_L$ is present with probability

$$H^L(|x - y|) = H\left(\frac{|x - y|}{R_L}\right),$$

where $R_L = \ln(\tau L)/2\|H\|_1$ for some positive constant τ, and $\|H\|_1 = \int_0^\infty H(z)dz$. We denote by $\mathcal{G}_{H^L}(\mathcal{P}_L)$ the graph with vertex set $\mathcal{P}_L$ and edges created according to the connection function H^L. To avoid triviality of the problem, we assume that $H(\cdot)$ is not identically 0 everywhere and hence $\|H\|_1 > 0$. We define the set of connection functions $\mathcal{H}$ such that for any $H \in \mathcal{H}$:

Condition A: The function $H : \mathbb{R}_+ \to [0, 1]$ is strictly monotonically decreasing and continuous, i.e.,

$$H(x) > H(y) \quad \text{for all } x < y,$$

and in particular, for any constant $\delta > 0$ there exists some $\epsilon > 0$ such that

$$H(\delta) \leq 1 - \epsilon.$$

Condition B: The function $H : \mathbb{R}_+ \to [0, 1]$ has unbounded support, i.e.,

$$\int_r^\infty H(x)dx > 0 \quad \text{for all } r \geq 0.$$

Condition C: The function $H : \mathbb{R}_+ \to [0, 1]$ is integrable, i.e.,

$$\|H\|_1 = \int_0^\infty H(z)dz < \infty.$$

We now define the toroidal line segment; this alleviates the difficulties in analysing the boundary effects which add significant complexity without much insight. We denote by $\rho(x, y)$ the circular (or toroidal) distance between x and y, and by $\rho_\infty(x, y) = |x - y|$ the Euclidean distance between them. A point x in the torus can connect to y either directly or via the identification of the end points. This means that $\rho(x, y) = \min(\rho_\infty(x, y), L - \rho_\infty(x, y))$. Define

$$h = H \circ \rho,$$

and

$$h^L = H^L \circ \rho,$$

where $\circ$ denotes composition of functions; $h(x, y)$ is the probability that two points at locations x and y on the torus are connected. Note that $h(x, y)$ is symmetric, and invariant to translations on this line segment, i.e. $h(x, y) = h(y, x)$ and $h(x, y) = h(x - t, y - t)$ for some translation t. We are now able to state the main theorem of

this chapter.

Theorem 4.1.1. *Fix $\tau \in \mathbb{R}_+$ and consider the soft RGG $\mathcal{G}_{h^L}(\mathcal{P}_L)$, where $R_L = \ln(\tau L)/2\|H\|_1$. Let N_{iso} denote the number of isolated nodes in $\mathcal{G}_{h^L}(\mathcal{P}_L)$, whose dependence on L is implicit but omitted for notational simplicity. Then, in the limit as L tends to infinity, the quantity N_{iso} converges in total variational distance to a Poisson random variable with mean $1/\tau$. In particular, $\mathbb{P}(N_{iso} = 0)$ tends to $e^{-1/\tau}$.*

4.2 Proof of Theorem 4.1.1

We begin this section by introducing a truncated version of the connection function. We claim that in the limit as L tends to infinity, the total variational distance between the number of isolated nodes in the original graph, and the number of isolated nodes in the truncated version tends to 0. We write the total variational distance between two random variables X and Y (or more formally between their distributions) as $d_{TV}(X, Y)$.

Lemma 4.2.1. *Truncate h^L by setting $\tilde{h}^L = h^L(x, y)\mathbb{1}\{\rho(x, y) \leq R_L^{1+1/\alpha}\}$. Denote by N_{iso} the number of isolated nodes in $\mathcal{G}_{h^L}(\mathcal{P}_L)$, and by $\widetilde{N}_{iso}$ the number of isolated nodes in $\mathcal{G}_{\tilde{h}^L}(\mathcal{P}_L)$. Then,*

$$\lim_{L\to\infty} d_{TV}(N_{iso}, \widetilde{N}_{iso}) = 0.$$

Proof. We are able to naturally couple $\mathcal{G}_{h^L}(\mathcal{P}_L)$ and $\mathcal{G}_{\tilde{h}^L}(\mathcal{P}_L)$ by firstly generating $\mathcal{G}_{h^L}(\mathcal{P}_L)$, and then removing any edges of Euclidean length at least $R_L^{1+1/\alpha}$. Denoting by $\widetilde{N}_{\text{iso}}$ the number of isolated nodes in $\mathcal{G}_{\tilde{h}^L}(\mathcal{P}_L)$, we observe that $\widetilde{N}_{\text{iso}} \geq N_{\text{iso}}$ since we are only removing edges. Therefore, it is immediate from Markov's inequality, which states that for any non-negative random variable X, $\mathbb{P}(X \geq 1) \leq \mathbb{E}[X]$, that

$$\mathbb{P}(|N_{\text{iso}} - \widetilde{N}_{\text{iso}}| \geq 1) \leq \mathbb{E}[|N_{\text{iso}} - \widetilde{N}_{\text{iso}}|] = \mathbb{E}[\widetilde{N}_{\text{iso}}] - \mathbb{E}[N_{\text{iso}}]. \tag{4.1}$$

From Eqn. (3.4), we can calculate the asymptotic expectation of N_{iso} as

$$\begin{aligned}\mathbb{E}[N_{\text{iso}}] &= L\exp\left(-2R_L\int_0^{L/2R_L} H(r)dr\right)\\ &= L\exp\left(-\frac{\ln(\tau L)}{\|H\|_1}\int_0^{L/2R_L} H(r)dr\right) \xrightarrow{L\to\infty} \frac{1}{\tau}.\end{aligned} \tag{4.2}$$

And that of $\widetilde{N}_{\text{iso}}$ as

$$\begin{aligned}\mathbb{E}[\widetilde{N}_{\text{iso}}] &= L\exp\left(-2R_L\int_0^{L/2R_L}\widetilde{H}(r)dr\right)\\ &= L\exp\left(-\frac{\ln(\tau L)}{\|H\|_1}\int_0^{R_L^{1+1/\alpha}}H(r)dr\right)\xrightarrow{L\to\infty}\frac{1}{\tau}.\end{aligned} \tag{4.3}$$

The total variational distance between two random variables is defined as the maximal possible difference that the two probability distributions can assign to the same event. Therefore, to complete the proof, we combine, Eqns. (4.1) - (4.3) to see

$$\mathbb{P}(|N_{\text{iso}}-\widetilde{N}_{\text{iso}}|\geq 1)\xrightarrow{L\to\infty}0.$$

□

We will henceforth only be working with the truncated connection function, $\tilde{h}$. To prove Theorem 4.1.1 we will use the Chen-Stein method of Poisson approximation [AGG+89; BHJ92; MA12; Pen16]. We discretise the torus into mL segments of width $1/m$ and denote the ith segment by A_i. We denote by $\Gamma=\{1,2,...,mL\}$ the complete set of index values that i can take. Let $\mathcal{P}_L$ denote the points of a unit rate Poisson point process which fall within the line segment $[0,L]$, and let $\mathcal{I}$ denote the set of isolated nodes in the (truncated) graph $\mathcal{G}_{\tilde{h}^L}(\mathcal{P}_L)$. We define

$$\begin{aligned}J_i &:= \mathbb{1}\left\{\mathcal{P}_L(A_i)=1\right\},\\ I_i &:= \mathbb{1}\left\{\mathcal{P}_L(A_i)=1 \text{ and this point is in } \mathcal{I}\right\},\end{aligned}$$

for $i\in\Gamma$. We denote the centre of the segment A_i by x_i. (Although A_i, I_i, J_i, and x_i all implicitly depend on m, this dependence is omitted for notational simplicity.) We will make use of the following asymptotic notation:

$$f(x)\sim_x g(x)\quad \text{ if } \lim_{x\to\infty}\frac{f(x)}{g(x)}=1.$$

For any finite $\mathcal{A}\subset\mathbb{R}$ and any $y\in\mathbb{R}$ with $y\notin\mathcal{A}$ we denote

$$\begin{aligned}\phi_{\tilde{h}}(y,\mathcal{A}) &= 1-\prod_{x\in\mathcal{A}}\Big(1-\tilde{h}(x,y)\Big)\\ &= \mathbb{P}(y \text{ is non-isolated in } \mathcal{A})\end{aligned} \tag{4.4}$$

Lemma 4.2.2. *For fixed $x, y \in [0, L]$, we denote by A_i and A_j, respectively, the segments in which they live. The probability that exactly one node exists within A_i satisfies*

$$\mathbb{P}(J_i = 1) \sim_m \frac{1}{m}. \tag{4.5}$$

The probability that exactly one isolated node exists within A_i satisfies

$$\mathbb{P}(I_i = 1) \sim_m \frac{1}{m} \exp\left(-2 \int_0^{L/2} \widetilde{H}^L(z) dz\right). \tag{4.6}$$

The probability that exactly one isolated node exists within both A_i and A_j satisfies

$$\mathbb{P}(I_i I_j = 1) \sim_m \frac{1}{m^2}(1 - \tilde{h}^L(x, y)) \exp\left(-\int_0^L \phi_{\tilde{h}^L}(z, \{x, y\}) dz\right). \tag{4.7}$$

The proof of Lemma 4.2.2 is given in Section 4.3. For a set of indicator random variables I_i, $i \in \Gamma$, we define

$$p_i := \mathbb{E}[I_i], \quad p_{ij} := \mathbb{E}[I_i I_j],$$

$$W_{m,L} := \sum_{i \in \Gamma} I_i.$$

For each $i \in \Gamma$, we choose $B_i \subset \Gamma$ such that I_i is (nearly) independent of all I_j for $j \notin B_i$. We think of B_i as being a "ball of dependence" for i [AGG+89]. Let x_i denote the centre of the segment A_i, then for $\alpha > 0$ we define

$$B_i = \left\{ j \in \Gamma : \rho(x_i, x_j) \leq 3R_L^{1+1/\alpha} \right\},$$

and,

$$b_1 := \sum_{i \in \Gamma} \sum_{j \in B_i} p_i p_j, \quad b_2 := \sum_{i \in \Gamma} \sum_{j \in B_i \setminus \{i\}} p_{ij}, \quad b_3 := \sum_{i \in \Gamma} \mathbb{E}\left[\left| \mathbb{E}\left[I_i | (I_j : j \in \Gamma \setminus B_i) \right] - p_i \right| \right].$$

As described in [AGG+89], b_1 measures the neighbourhood size, b_2 measures the expected number of neighbours of a given occurrence, and b_3 measures the dependence between an event and the number of occurrences outside of its neighbourhood.

Lemma 4.2.3. *For h continuous and denoting by $S\left(x, R_L^{1+1/\alpha}\right)$ the line segment of*

width $2R_L^{1+1/\alpha}$ *and centred at* x*, define the integral* K *by*

$$K := 2\int_0^L \int_{S\left(x,R_L^{1+1/\alpha}\right)} \exp\left(-\int_0^L \phi_{\tilde{h}^L}(z,\{x,y\})dz\right) dydx.$$

Then,

$$d_{TV}\left(\widetilde{N}_{iso}, Po\left(\mathbb{E}[\widetilde{N}_{iso}]\right)\right) \leq \min\left(1, 1/\mathbb{E}[\widetilde{N}_{iso}]\right) K.$$

Proof. From [BHJ92, Th. 1.A], and [MA12, Th. 12]) we have the following bound on the total variational distance between a sum of indicator functions and a Poisson random variable,

$$d_{TV}\left(W_{m,L}, Po\left(\mathbb{E}[W_{m,L}]\right)\right) \leq \min\left(1, \frac{1}{\mathbb{E}[W_{m,L}]}\right)(b_1+b_2+b_3), \tag{4.8}$$

where $Po(\mathbb{E}[W_{m,L}])$ denotes a Poisson random variable with mean $\mathbb{E}[W_{m,L}]$. The b_3 term is describing the dependence between the event that a node at location x_i is isolated and the event that a node outside of the ball of dependence around x_i is isolated. By the definition of $W_{m,L}$ there are no edges of length greater than $R_L^{1+1/\alpha}$. Therefore, in the limit as $m \to \infty$ there is no dependence between the event that a node in segment A_i is isolated and the event that a node outside of B_i is isolated and hence $\lim_{m\to\infty} b_3 = 0$.

Following from Eqns. 4.6 and 4.7,

$$\begin{aligned} b_2 &:= \sum_{i\in\Gamma}\sum_{i\neq j\in B_i} p_{ij} = \sum_{i\in\Gamma}\sum_{i\neq j\in B_i} \mathbb{P}(J_iJ_j=1)\mathbb{P}(I_iI_j=1|J_iJ_j=1) \\ &\sim_m \frac{1}{m^2}\sum_{i\in\Gamma}\sum_{i\neq j\in B_i} \mathbb{P}(I_iI_j=1|J_iJ_j=1) \\ &\sim_m \int_0^L \int_{S\left(x,R_L^{1+1/\alpha}\right)} \left(1-\tilde{h}^L(x,y)\right) e^{-\int_0^L\left(\phi_{\tilde{h}^L}(z,\{x,y\}\right)dz} dydx. \end{aligned}$$

Therefore,

$$\begin{aligned} K_2 := \lim_{m\to\infty} b_2 &= \int_0^L \int_{S\left(x,R_L^{1+1/\alpha}\right)} \left(1-\tilde{h}^L(x,y)\right) e^{-\int_0^L\left(\phi_{\tilde{h}^L}(z,\{x,y\}\right)dz} dydx \\ &\leq \int_0^L \int_{S\left(x,R_L^{1+1/\alpha}\right)} e^{-\int_0^L\left(\phi_{\tilde{h}^L}(z,\{x,y\}\right)dz} dydx =: K/2. \end{aligned} \tag{4.9}$$

In a similar vein, we can derive

$$K_1 := \lim_{m\to\infty} b_1 = \int_0^L \int_{S\left(x,3R_L^{1+1/\alpha}\right)} \exp\left(-\int_0^L \tilde{h}^L(x,z)dz - \int_0^L \tilde{h}^L(y,z)dz\right) dydx. \tag{4.10}$$

By the union bound,

$$\phi_{\tilde{h}^L}(z,\{x,y\}) \leq \tilde{h}^L(x,z) + \tilde{h}^L(y,z), \tag{4.11}$$

and hence $K_1 \leq K_2 \leq K$. Recalling Eqn. 4.6, and noting that since we are working on the torus, the indicator I_i does not depend on the subscript i,

$$\begin{aligned}\lim_{m\to\infty} \mathbb{E}[W_{m,L}] &= \lim_{m\to\infty} \sum_{i\in\Gamma} \frac{1}{m}\mathbb{E}[I_i] = L \lim_{m\to\infty} \mathbb{E}[I_i] \\ &= L\exp\left(-2\int_0^{L/2} \widetilde{H}^L(z)dz\right) = \mathbb{E}[\widetilde{N}_{\text{iso}}],\end{aligned} \tag{4.12}$$

where the final equality comes from Eqn. (4.3). Thus combining Eqns. (4.8)-(4.12) and the fact that $\lim_{m\to\infty} b_3 = 0$,

$$\begin{aligned}\lim_{m\to\infty} \min\left(1, \frac{1}{\mathbb{E}[W_{m,L}]}\right)(b_1+b_2+b_3) &= \min\left(1, \frac{1}{\mathbb{E}[\widetilde{N}_{\text{iso}}]}\right)(K_1+K_2) \\ &\leq \min\left(1, \frac{1}{\mathbb{E}[\widetilde{N}_{\text{iso}}]}\right) K.\end{aligned} \tag{4.13}$$

$W_{m,L}$ counts the number of segments for which there is exactly one node of $\mathcal{P}_L$ and this node is isolated. Therefore, there is no way for $W_{m,L}$ to over count the number of isolated nodes in $\mathcal{G}_{\tilde{h}^L}(\mathcal{P}_L)$, it can only under count, hence $\widetilde{N}_{\text{iso}} \geq W_{m,L}$. Combining Eqns. (4.3) and (4.12) with Markov's inequality once again,

$$\lim_{m\to\infty} \mathbb{P}(|\widetilde{N}_{\text{iso}} - W_{m,L}| \geq 1) \leq \lim_{m\to\infty} \mathbb{E}[\widetilde{N}_{\text{iso}}] - \mathbb{E}[W_{m,L}] = 0.$$

Thus,

$$\lim_{m\to\infty} d_{TV}(N_{\text{iso}}, W_{m,L}) = 0. \tag{4.14}$$

Noting that total variational distance adheres to the triangle inequality, we see that

for all m,

$$d_{TV}\left(\widetilde{N}_{\text{iso}}, Po\left(\mathbb{E}[\widetilde{N}_{\text{iso}}]\right)\right) \leq d_{TV}\left(\widetilde{N}_{\text{iso}}, W_{m,L}\right) + d_{TV}\left(W_{m,L}, Po\left(\mathbb{E}[\widetilde{N}_{\text{iso}}]\right)\right).$$

Since this holds for all m, we can look at the right hand side of this inequality in the limit as m tends to infinity, which, combining Eqns. (4.8), (4.13), and (4.14), is bounded above by $\min\left(1, 1/\mathbb{E}[\widetilde{N}_{\text{iso}}]\right) K$. Therefore,

$$d_{TV}\left(\widetilde{N}_{\text{iso}}, Po\left(\mathbb{E}[\widetilde{N}_{\text{iso}}]\right)\right) \leq \min\left(1, \frac{1}{\mathbb{E}[\widetilde{N}_{\text{iso}}]}\right) K.$$

□

Lemma 4.2.4. *Under the assumption that $\tilde{h}$ is monotonically decreasing, there exists an $\omega > 0$ such that for all $x \in \left[0, 3R_L^{1+1/\alpha}\right]$ and for L sufficiently large,*

$$\int_0^L \tilde{h}^L(x,z)(1-\tilde{h}^L(0,z))dz \geq \omega R_L. \tag{4.15}$$

Proof. Since the connection function, $\tilde{h}$, is monotonically decreasing and continuous up to the cut off point of $R_L^{1+1/\alpha}$, for any constant $\delta > 0$ there exists an $\epsilon > 0$ such that for L sufficiently large, $\tilde{h}(0,\delta) \leq 1-\epsilon$, and hence

$$\tilde{h}^L(0, \delta R_L) \leq 1 - \epsilon. \tag{4.16}$$

For L sufficiently large, there also exists a constant $c \in \left(\delta, R_L^{1+1/\alpha}\right)$ such that

$$\widetilde{H}(c) > 0.$$

Since $\tilde{h}^L(x,z) - \tilde{h}^L(0,z)\tilde{h}^L(x,z) \geq 0$,

$$\begin{aligned}
\int_0^L \left(\tilde{h}^L(x,z) - \tilde{h}^L(0,z)\tilde{h}^L(x,z)\right) dz &\geq \int_{\delta R_L}^{L/2} \tilde{h}^L(x,z)\left(1-\tilde{h}^L(0,z)\right) dz \\
&\geq \epsilon \int_{\delta R_L}^{L/2} \tilde{h}^L(x,z) dz \\
&= \epsilon \int_{\delta R_L}^{L/2} \widetilde{H}\left(\frac{|z-x|}{R_L}\right) dz,
\end{aligned} \tag{4.17}$$

where the second inequality follows from Eqn. (4.16). For $x \in [0, \delta R_L]$, we note that

$z \geq x$ and make the substitution $s = \frac{z-x}{R_L}$ in Eqn. (4.17),

$$\epsilon \int_{\delta R_L}^{L/2} \widetilde{H}\left(\frac{z-x}{R_L}\right) dz = \epsilon R_L \int_{\delta - \frac{x}{R_L}}^{\frac{L}{2R_L} - \frac{x}{R_L}} \widetilde{H}\left(s\right) ds. \tag{4.18}$$

For L sufficiently large, $L/2R_L - \delta \geq R_L^{1+1/\alpha} > c$, hence

$$\epsilon R_L \int_{\delta - \frac{x}{R_L}}^{\frac{L}{2R_L} - \frac{x}{R_L}} \widetilde{H}\left(s\right) ds \geq \epsilon R_L \int_{\delta}^{c} \widetilde{H}(s) ds \geq \omega_1 R_L, \tag{4.19}$$

where $\omega_1 = \epsilon(c-\delta)\widetilde{H}(c) > 0$.

For $x \in [\delta R_L, 3R_L^{1+1/\alpha}]$,

$$\begin{aligned} \epsilon \int_{\delta R_L}^{L/2} \widetilde{H}\left(\frac{|z-x|}{R_L}\right) dz &= \epsilon \int_{\delta R_L}^{x} \widetilde{H}\left(\frac{x-z}{R_L}\right) dz + \epsilon \int_{x}^{L/2} \widetilde{H}\left(\frac{z-x}{R_L}\right) dz \\ &\geq \epsilon \int_{x}^{L/2} \widetilde{H}\left(\frac{z-x}{R_L}\right) dz. \end{aligned} \tag{4.20}$$

As before,

$$\begin{aligned} \epsilon \int_{x}^{L/2} \widetilde{H}\left(\frac{z-x}{R_L}\right) dz &= \epsilon R_L \int_{0}^{\frac{L}{2R_L} - \frac{x}{R_L}} \widetilde{H}\left(s\right) ds \\ &\geq \epsilon R_L \int_{0}^{c} \widetilde{H}\left(s\right) ds \geq \omega_2 R_L, \end{aligned} \tag{4.21}$$

where $\omega_2 = \epsilon c \widetilde{H}(c) > 0$. Setting $\omega = \min\{\omega_1, \omega_2\} = \omega_1$, we complete the proof. □

We are now able to prove Theorem 4.1.1.

Proof of Theorem 4.1.1. Using the triangle inequality, we see that for all L,

$$d_{TV}(N_{\text{iso}}, Po(1/\mathbb{E}[N_{\text{iso}}])) \leq d_{TV}(N_{\text{iso}}, \widetilde{N}_{\text{iso}}) + d_{TV}\left(\widetilde{N}_{\text{iso}}, Po\left(1/\mathbb{E}[N_{\text{iso}}]\right)\right).$$

Therefore, combining Eqns. (4.2) and (4.3) with Lemmas 4.2.1 and 4.2.3,

$$\begin{aligned} \lim_{L\to\infty} d_{TV}(N_{\text{iso}}, Po(1/\tau)) &\leq \lim_{L\to\infty} d_{TV}(N_{\text{iso}}, \widetilde{N}_{\text{iso}}) + d_{TV}\left(\widetilde{N}_{\text{iso}}, Po\left(1/\tau\right)\right) \\ &\leq \min\left(1, \tau\right) \lim_{L\to\infty} K, \end{aligned}$$

where $K := 2\int_0^L \int_{S\left(x, R_L^{1+1/\alpha}\right)} \exp\left(-\int_0^L \phi_{\tilde{h}^L}(z, \{x, y\}) dz\right) dy dx$. Without loss of generality, since we are working on the torus, we can assume that x is at the origin.

Thus by the definition of K,

$$\begin{aligned}
K &= 2L \int_{S\left(0,R_L^{1+1/\alpha}\right)} \exp\left(-\int_0^L \phi_{\tilde{h}^L}(z,\{0,y\})dz\right) dy \\
&= 4Le^{-2\int_0^{L/2}\tilde{H}^L(z)dz} \int_0^{3R_L^{1+1/\alpha}} \exp\left(-\int_0^L \left(\tilde{h}^L(y,z)(1-\tilde{h}^L(0,z))\right) dz\right) dy \\
&\leq 12Le^{-2\int_0^{L/2}\tilde{H}^L(z)dz} R_L^{1+1/\alpha} \exp\left(-\omega R_L\right) dy,
\end{aligned}$$

where this final line comes from Lemma 4.2.4. Thus, by Eqn. (4.3),

$$\lim_{L\to\infty} K \leq \lim_{L\to\infty} \frac{12}{\tau} R_L^{1+1/\alpha} \exp(-\omega R_L) = 0. \tag{4.22}$$

□

4.3 Additional Calculations

In this section, we prove Lemma 4.2.2 by obtaining the asymptotic expressions seen in Eqns. (4.5) - (4.7).

Proof of Lemma 4.2.2. We begin with Eqn. (4.5). $\mathcal{P}_L$ is a unit rate Poisson point process defined on $[0, L]$ and hence the probability that there is exactly one point of this process within $A_i \in [0, L]$ is given by

$$\mathbb{P}(J_i = 1) = \frac{1}{m} e^{-1/m} \sim_m \frac{1}{m}. \tag{4.23}$$

Now, notice that, conditional on $J_iJ_j = 1$, and on the locations x and y of the single nodes in the segments A_i and A_j respectively, the set of all nodes with an edge to either the node at x or y comprise a Poisson point process (PPP) on $[0, L] \setminus A_i$, with intensity $\phi_{\tilde{h}^L}(\cdot, \{x, y\})$. The event that the nodes at x and y are isolated is the event that this PPP is empty and the nodes at x and y are not connected to each other. Thus, denoting by $A_{i,j}(x, y)$ the event that the node in A_i is located at x and the node in A_j is located at y,

$$\mathbb{P}(I_iI_j = 1 | J_iJ_j = 1, A_{i,j}(x, y)) = (1 - \tilde{h}^L(x, y))e^{-\lambda_{x,y}}, \tag{4.24}$$

where

$$\begin{aligned}\lambda_{x,y} &= \int_{[0,L]\setminus(A_i\cup A_j)} \phi_{\tilde{h}^L}(z,\{x,y\})dz \\ &= \int_0^L \phi_{h^L}(z,\{x,y\})dz - \int_{A_i} \phi_{h^L}(z,\{x,y\})dz - \int_{A_j} \phi_{h^L}(z,\{x,y\})dz.\end{aligned} \tag{4.25}$$

By the assumed monotonicity of H,

$$\int_{A_i} \tilde{h}^L(x,z)dz \leq \frac{H(0)}{m},$$

which tends to zero as m tends to infinity. Hence,

$$\begin{aligned}\lim_{m\to\infty} \int_{A_i} \phi_{\tilde{h}^L}(z,\{x,y\})dz &= \lim_{m\to\infty} \int_{A_i} \left(\tilde{h}^L(x,z) + \tilde{h}^L(y,z) - \tilde{h}^L(x,z)\tilde{h}^L(y,z)\right) dz \\ &= 0.\end{aligned} \tag{4.26}$$

An equivalent statement can be made for $\int_{A_j} \phi_{\tilde{h}^L}(z,\{x,y\})dz$. It is well-known for a PPP that, conditional on $J_i = 1$, the single node in A_i is uniformly distributed within this interval. Hence, it follows from (4.24) that

$$\mathbb{P}(I_iI_j = 1 | J_iJ_j = 1) = m^2(1 - \tilde{h}^L(x,y)) \int_{A_i}\int_{A_j} e^{-\lambda_{x,y}} dxdy. \tag{4.27}$$

It is straightforward from (4.25), (4.26) and (4.27) that

$$\mathbb{P}(I_iI_j = 1 | J_iJ_j = 1) \sim_m (1 - \tilde{h}^L(x,y)) \exp\left(-\int_0^L \phi_{\tilde{h}^L}(z,\{x,y\})dz\right). \tag{4.28}$$

Combining this with the fact that

$$\mathbb{P}(J_iJ_j = 1) = \frac{1}{m^2} e^{-2/m} \sim_m \frac{1}{m^2},$$

we obtain the expression for $\mathbb{P}(I_iI_j = 1)$ given in Eqn. (4.7). The expression for Eqn. (4.6) is found in a very similar fashion and hence is omitted for brevity. □

4.4 Concluding Remarks

In this chapter, we used the Chen-Stein method for Poisson approximation to show that the distribution of the number of isolated nodes in our (toroidal) 1-D soft RGG converges in total variational distance to a Poisson random variable (Theorem 4.1.1). Our work is constricted to the underlying point process being a homogeneous Poisson point process but allows for very general connection functions with our only assumptions being that the connection function is continuous, strictly monotonically decreasing with unbounded support, and integrable.

The main result of this chapter is not unexpected due to the fact that it has been shown to be true in higher dimensional versions of the model as well as in the 1-D original (hard) RGG. However, this is the first time it has been proven for the 1-D soft RGG. One reason for this is due to the previous lack of proof that isolated nodes are a dominant factor in the connectivity of these graphs in the lower dimensional model. This is a consequence of the fact that the graph can split into two large disconnected clusters. However, in the following chapter, we will see that the isolated nodes (or a lack thereof) are more important in determining the connectivity of the network than an *uncrossed gap* in the limit as the system size grows to infinity. It is therefore conjectured that isolated nodes are in fact the dominant mode of disconnection in this regime even in the lower dimensional model; hence the analysis given in this chapter is crucial to understanding the connectivity of these graphs.

Through this and the previous chapter, we have developed a good understanding of the behaviour of the isolated nodes within a one-dimensional soft RGG. However, as mentioned, we know that they are not the only barrier to full connectivity. Therefore the subsequent chapter investigates *uncrossed gaps* in order to develop a deeper understanding of this object and its effect on the connectivity of these graphs.

Chapter 5

Soft Random Geometric Graphs: Uncrossed Gaps

In Chapter 3, we showed that the probability that isolated nodes exist exhibits a sharp transition, from being close to 1 to being close to 0, as the mean degree is increased. Moreover, the point at which this transition occurs depends only on the mean degree, and is insensitive to the connection function.

Our primary interest is in the probability of disconnection, not just in whether isolated nodes exist. We noted in Section 3.2 that there are other modes of disconnection (Figure 3.2), and conjectured that uncrossed gaps are the other main mode. Hence, we now turn to estimating the probability that there are uncrossed gaps. We will focus on the scaling regime identified in the Chapter 3, namely $R_L = \gamma \ln L$, and $H^L(\cdot) = H(\cdot/R_L)$.

The work in this chapter is based on the paper [WDG20] for which I was a co-author. This paper was written in collaboration with Professor Carl Dettmann and Dr. Ayalvadi Ganesh. The theoretical analysis was primarily conducted by myself with CD and AG contributing ideas to this analysis along with checking calculations and suggesting proof methods.

5.1 Uncrossed gaps at the critical scaling for isolated nodes

In this section, we need to assume the following about the connection function.

Assumption A: The function $H : \mathbb{R}_+ \mapsto [0,1]$ is integrable, i.e.,

$$\|H\|_1 = \int_0^\infty H(x)dx < \infty.$$

Assumption B The connection function $H : \mathbb{R}_+ \to [0,1]$ is monotone decreasing and has unbounded support, i.e.,

$$H(x) \geq H(y) \quad \text{for all } x \leq y,$$

and

$$\int_r^\infty H(x)dx > 0 \quad \text{for all } r \geq 0.$$

Our main result in this section is that the probability of seeing uncrossed gaps becomes negligible as L tends to infinity, for $R_L = \gamma \ln L$ and any $\gamma > 0$. In particular, it holds for $\gamma = 2\|H\|_1$, when isolated nodes begin to emerge. Consequently, the main cause of disconnection of soft RGGs in 1-D is isolated nodes, just as in higher dimensions; 1-D hard RGGs are anomalous in this regard.

We remark that, while we believe the assumption about unboundedness of the support of H to be essential (indeed, our result does not hold for hard RGGs), the assumption of monotonicity appears to be an artefact of our proof technique.

In order to avoid boundary effects near 0 and L, we extend the Poisson process from $[0, L]$ to the infinite real line. (In this case, it is less convenient to work with the circle, since the existence of a single uncrossed gap on the circle does not guarantee disconnection.) We want to calculate the probability that there are no edges crossing the origin, under the Palm measure corresponding to having a point of the Poisson process at the origin. We denote probabilities and expectations under this measure by P_0 and $\mathbb{E}_0$ respectively, as before.

Let $\mathcal{P}$ denote a Poisson point process on $\mathbb{R}$ of unit intensity. For $A \subset \mathbb{R}$, we write $\mathcal{P}_A$ to denote the restrictions of $\mathcal{P}$ to A. Denote by X_0 the indicator of the event that there is a point at the origin and that it marks an uncrossed gap, i.e., there are no edges between $\mathcal{P}_{(-\infty,0)} \cup \{0\}$ and $\mathcal{P}_{(0,\infty)}$. We wish to compute $\mathbb{E}_0[X_0]$. As it is not amenable to exact calculation, we obtain a bound on it below.

Fix $\alpha, \delta > 0$ and set $n_L = \lfloor \delta R_L \rfloor$. Denote by X_0^α the indicator of the event that there are no edges between $\mathcal{P}_{(-\alpha R_L,0)} \cup \{0\}$ and $\mathcal{P}_{(0,\infty)}$. Clearly, $X_0 \leq X_0^\alpha$. Moreover, conditional on $\mathcal{P}_{(-\alpha R_L,0)}$, the set of points on $(0,\infty)$ which have an edge to some

point in $(-\alpha R_L, 0]$ constitute a Poisson process with intensity

$$\begin{aligned}\lambda_{\mathcal{P}_{(-\alpha R_L,0)}}(y) &= 1 - \bar{h}^L(0,y) \prod_{z\in\mathcal{P}_{(-\alpha R_L,0)}} \bar{h}^L(z,y) \\ &= 1 - \bar{H}^L(y) \prod_{z\in\mathcal{P}_{(-\alpha R_L,0)}} \bar{H}^L(y-z),\end{aligned} \tag{5.1}$$

where $\bar{h}^L(0,y) = 1 - h^L(0,y)$ and $\bar{H}^L(y) = 1 - H^L(y)$. Hence, the number of such points is a Poisson random variable with mean $\int_0^\infty \lambda_{\mathcal{P}_{(-\alpha R_L,0)}}(y)dy$. Since X_0^α is the indicator that this random variable takes the value zero, we obtain that

$$\mathbb{E}_0\big[X_0^\alpha \;\big|\; \mathcal{P}_{(-\alpha R_L,0)}\big] = \exp\left(-\int_0^\infty \lambda_{\mathcal{P}_{(-\alpha R_L,0)}}(y)dy\right). \tag{5.2}$$

We also have by Eqn. (5.1) and Assumption B that

$$\lambda_{\mathcal{P}_{(-\alpha R_L,0)}}(y) \geq 1 - \bar{H}^L(y+\alpha R_L)^{1+|\mathcal{P}_{(-\alpha R_L,0)}|}, \tag{5.3}$$

where $|\mathcal{P}_A|$ denotes the number of points of $\mathcal{P}$ in the set A. Hence, if we condition further on there being at least n_L points of $\mathcal{P}$ in $(-\alpha R_L, 0)$, then we obtain by substituting Eqn. (5.3) in Eqn. (5.2) that

$$\begin{aligned}\mathbb{E}_0\big[X_0^\alpha \;\big|\; |\mathcal{P}_{(-\alpha R_L,0)}| \geq n_L\big] &\leq \exp\Big(-\int_0^\infty \big(1 - \bar{H}^L(y+\alpha R_L)^{1+n_L}\big)dy\Big) \\ &= \exp\Big(-R_L\int_0^\infty \big(1 - \bar{H}(z+\alpha)^{1+n_L}\big)dz\Big).\end{aligned} \tag{5.4}$$

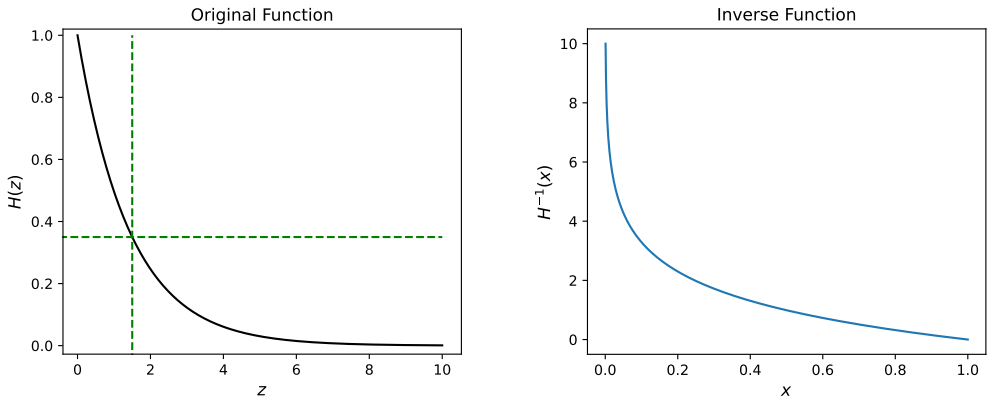
We have used the change of variables $z = y/R_L$ to obtain the last equality.

Define $H^{-1}(x) = \sup\{z : H(z) \geq x\}$ to be the generalised inverse of H; H^{-1} maps $(0, H(0)]$ to $[0,\infty)$. To see this more clearly, an example of this function is illustrated below in Figure 5.1.

Since $R_L = \gamma \ln L$ and $n_L = \lfloor \delta R_L \rfloor$ for fixed constants $\gamma, \delta > 0$, n_L tends to infinity as L tends to infinity. Consequently,

$$H^{-1}\left(\frac{1}{1+n_L}\right) \to \infty \text{ as } L \to \infty, \tag{5.5}$$

because the support of H is unbounded by Assumption B. Since H is monotone decreasing, we have for all $y \leq H^{-1}\left(\frac{1}{1+n_L}\right)$ that $H(y) \geq \frac{1}{1+n_L}$, and hence, $\bar{H}(y)^{1+n_L} \leq$



(a) (b)

Figure 5.1: An example of an original connection function, $H(z)$ in (a) and its inverse in $H^{-1}(z)$ in (b). The horizontal green line in (a) represents an example x value, and the vertical line represents the corresponding $H^{-1}(x)$ value. It is important to note that the inverse function is decreasing in x and tends to infinity as x tends to 0.

e^{-1}. Therefore,

$$\begin{aligned} R_L \int_0^{\infty} \big(1 - \bar{H}(z+\alpha)^{1+n_L}\big) dz &\geq R_L \int_0^{H^{-1}(\frac{1}{1+n_L})} \big(1 - \bar{H}(z+\alpha)^{1+n_L}\big) dz \\ &\geq (1-e^{-1})\gamma H^{-1}\left(\frac{1}{1+n_L}\right) \ln L \\ &= \omega(\ln L), \end{aligned}$$

where the last equality follows from Eqn. (5.5). We use the notation $f(x) = \omega(g(x))$ to denote, for functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$, that $g(x)/f(x)$ tends to infinity as x tends to infinity. Substituting the above expression into Eqn. (5.4), we obtain that

$$\mathbb{E}_0\big[X_0^{\alpha} \;\big|\; |\mathcal{P}_{(-\alpha R_L, 0)}| \geq n_L\big] = o(L^{-\kappa}), \tag{5.6}$$

for arbitrary $\kappa > 0$.

Next, we bound the probability of the event that $|\mathcal{P}_{(-\alpha R_L,0)}|$ is smaller than n_L. Since $|\mathcal{P}_{(-\alpha R_L,0)}|$ is a Poisson random variable with mean αR_L, a standard use of the Bernstein inequality (also known as Chernoff's inequality, Corollary 2.2.3) yields

that

$$\begin{aligned}\mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| < n_L) &\leq \mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| \leq \delta R_L) \\ &\leq \inf_{\theta \leq 0} e^{-\theta\delta R_L}\mathbb{E}\big[e^{\theta|\mathcal{P}_{(-\alpha R_L,0)}|}\big] \\ &= \exp\big(-\sup_{\theta\leq 0}\big[\theta\delta R_L - \alpha R_L(e^\theta - 1)\big]\big) \\ &= \exp(-\gamma I_\alpha(\delta)\ln L),\end{aligned} \tag{5.7}$$

where $I_\alpha(\delta) = -\delta\ln\frac{\alpha}{\delta} + \alpha - \delta$. It is easy to see that, for fixed $\alpha > 0$, $I_\alpha(\delta)$ tends to α as δ decreases to zero. Hence, given arbitrary $\kappa, \gamma > 0$, we can choose α and δ such that $\gamma I_\alpha(\delta) > \kappa$. Then,

$$\begin{aligned}\mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| < n_L) &\leq e^{-\kappa\ln L} \\ &= L^{-\kappa}.\end{aligned} \tag{5.8}$$

Combining Eqns. (5.6) and (5.8), and noting that X_0^α is a $\{0,1\}$-valued random variable, we obtain that

$$\begin{aligned}\mathbb{E}_0[X_0^\alpha] &= \mathbb{E}_0\big[X_0^\alpha \;\big|\; |\mathcal{P}_{(-\alpha R_L,0)}| \geq n_L\big]\mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| \geq n_L) \\ &\quad + \mathbb{E}_0\big[X_0^\alpha \;\big|\; |\mathcal{P}_{(-\alpha R_L,0)}| < n_L\big]\mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| < n_L) \\ &\leq \mathbb{E}_0\big[X_0^\alpha \;\big|\; |\mathcal{P}_{(-\alpha R_L,0)}| \geq n_L\big] + \mathbb{P}(|\mathcal{P}_{(-\alpha R_L,0)}| < n_L) \\ &\leq L^{-\kappa}(1 + o(1)).\end{aligned}$$

Finally, the expected number of uncrossed gaps on $[0, L]$, which we denote $\mathbb{E}[N_{\text{ucg}}]$, is obtained by integrating $\mathbb{E}_0[X_0]$ over $[0, L]$ with respect to the unit intensity measure of the Poisson process of node locations. Since $X^0 \leq X_0^\alpha$, we conclude that

$$\mathbb{E}[N_{\text{ucg}}] = L\mathbb{E}_0[X_0] \leq L^{1-\kappa}(1 + o(1)).$$

As $\kappa > 0$ is arbitrary, we see by choosing $\kappa > 1$ that $\mathbb{E}[N_{\text{ucg}}]$ tends to zero as L tends to infinity. Thus, the probability of an uncrossed gap is vanishing in the limit $L \to \infty$ for any choice of $\gamma > 0$ and $R_L = \gamma \ln L$. This implies in particular that, at the critical scale of the connection function at which isolated nodes begin to appear, the probability of seeing an uncrossed gap is negligible. In other words, the primary mechanism responsible for causing disconnection in 1-D soft RGGs is the isolation of individual nodes.

As this is in sharp contrast to the situation for 1-D hard RGGs, in which disconnection is always due to uncrossed gaps, we comment briefly on the intuition behind

the result. When the connection function is scaled as above, the *isolation* of a node is primarily determined by what happens in an interval of order R_L around it. In the case of a hard RGG, this interval has to be empty. In the case of a soft RGG, it is not empty, but contains fewer nodes than expected, only ϵR_L, for some ϵ much smaller than 1. Isolation is then achieved by all ϵR_L of these potential edges being absent. Both these events (having fewer nodes in the interval, and having no edges to these nodes) have probability which is exponentially small in R_L; at the scaling $R_L = \gamma \ln L$, this works out to a probability of order $1/L$ (of isolation per node). However, for there to be an *uncrossed gap* at the index node, $(\epsilon R_L/2)^2$ potential edges between the $\epsilon R_L/2$ nodes on each side of the index node have to be absent. This event has probability decaying exponentially in R_L^2, and is hence much less likely than the event of node isolation. The intuition sketched out above is fleshed out in the following sections.

5.2 Conditional probability of uncrossed gaps

We now look at the intuition behind why isolated nodes dominate uncrossed gaps as the cause of disconnection in soft RGGs, whereas the opposite is true of hard RGGs. We do this by conditioning on there being an isolated node at the origin, and computing the conditional probability that it also marks an uncrossed gap. In order to simplify calculations, we work with the Poisson process on the infinite real line.

Denote by P_0 the Palm probability corresponding to the presence of a point at the origin, and by $\hat{\mathrm{P}}_0$ the measure obtained by conditioning further on this point being isolated. The corresponding expectations are denoted $\mathbb{E}_0$ and $\hat{\mathbb{E}}_0$. Then, under $\hat{\mathrm{P}}_0$, the remaining points constitute an inhomogenous Poisson process, with intensity function

$$\mu(x) = \bar{h}(0, x) = \bar{H}(|x|),\ x \in \mathbb{R},$$

where $\bar{h}(x, y) = 1 - h(x, y)$, and $\bar{H}(x) = 1 - H(x)$ respectively. We now follow the same approach as in the previous section. Conditional on $\mathcal{P}_{(-\infty,0)}$, the point process restricted to the negative real line, the set of points on $(0, \infty)$ which are connected

to some point in $\mathcal{P}_{(-\infty,0)}$ constitute an inhomogenous Poisson process with intensity

$$\begin{aligned}\Lambda(x) &= \bar{H}(x)\Big(1 - \prod_{z\in\mathcal{P}_{(-\infty,0)}} \bar{h}(z,x)\Big),\\ &= \bar{H}(x)\Big(1 - \exp\Big(\sum_{z\in\mathcal{P}_{(-\infty,0)}} g(z,x)\Big)\Big),\end{aligned} \tag{5.9}$$

where $g(x,y) = \ln(\bar{h}(x,y))$. The total number of points is thus a Poisson random variable, with mean equal to $\int_0^\infty \Lambda(x)dx$. The event that the origin marks an uncrossed gap is the event that this Poisson random variable takes the value zero, which has probability

$$\begin{aligned}\mathbb{P}(0 \text{ is uncrossed}) &= \hat{\mathbb{E}}_0\Big[\exp\Big(-\int_0^\infty \Lambda(x)dx\Big)\Big]\\ &\geq \exp\Big(-\int_0^\infty \hat{\mathbb{E}}[\Lambda(x)]dx\Big).\end{aligned} \tag{5.10}$$

We have used Jensen's inequality to obtain the inequality above and the interchange of integral and expectation is justified by Tonelli's theorem. We have also replaced $\hat{\mathbb{E}}_0$ by $\hat{\mathbb{E}}$, as the conditioning on having a point of the PPP at the origin has been taken into account in the expression for Λ.

Now, by invoking Campbell's formula for the Laplace functional of the inhomogenous Poisson process $\mathcal{P}_{(-\infty,0)}$ (Theorem 2.1.1 and Defn. 2.1.3), we obtain from Eqn. (5.9) that, for $x > 0$,

$$\begin{aligned}\hat{\mathbb{E}}[\Lambda(x)] &= (1-H(x))\Big(1 - \exp\Big(\int_{-\infty}^0 \big(e^{g(z,x)} - 1\big)\big)(1-h(0,z))dz\Big)\Big)\\ &= (1-H(x))\Big(1 - \exp\int_{-\infty}^0 -h(z,x)(1-h(0,z))dz\Big)\\ &= (1-H(x))\Big(1 - \exp\Big(-\int_0^\infty (1-H(z))H(x+z)dz\Big)\Big),\end{aligned} \tag{5.11}$$

where we have now returned to the original notation of $(1-H(\cdot))$ rather than $\bar{H}$ to make the derivation clearer. In the scaling regime in which the connection function $H(\cdot)$ is replaced by $H^L(\cdot) = H(\cdot/R_L)$, with the substitution $s = z/R_L$ we can rewrite the above as

$$\hat{\mathbb{E}}[\Lambda(x)] = (1-H(x/R_L))\Big(1 - \exp\Big(-R_L\int_0^\infty (1-H(s))H((x/R_L)+s)ds\Big)\Big),$$

and so with the substitution $y = x/R_L$,

$$\int_0^\infty \hat{\mathbb{E}}[\Lambda(x)]dx = R_L \int_0^\infty (1 - H(y))\big(1 - e^{-R_L \int_0^\infty (1-H(s))H(y+s)ds}\big)dy. \tag{5.12}$$

Define $H^{-1}(x) = \inf\{y \geq 0 : H(y) \leq x\}$, and note that H is integrable. By the assumption that H is monotone decreasing, $1 - H(s) \geq 1 - x$ for all $s > H^{-1}(x)$. Hence,

$$\begin{aligned}
\int_0^\infty (1 - H(s))H(y+s)ds &= \int_0^{H^{-1}(x)} (1 - H(s))H(y+s)ds \\
&\qquad + \int_{H^{-1}(x)}^\infty (1 - H(s))H(y+s)ds \\
&\geq \int_0^{H^{-1}(x)} (1 - H(s))H(y+s)ds + \\
&\qquad + (1-x)\int_{H^{-1}(x)}^\infty H(y+s)ds \\
&\geq (1-x)\int_{H^{-1}(x)}^\infty H(y+s)ds.
\end{aligned}$$

By setting $x = 1/2$,

$$\int_0^\infty (1 - H(s))H(y+s)ds \geq \frac{1}{2}\int_{H^{-1}(1/2)}^\infty H(y+s)ds.$$

Substituting this in Eqn. (5.12), invoking the inequality $1 - H(y) \geq 1/2$ for all $y \geq H^{-1}(1/2)$ once more, and using the monotonicity of $H(\cdot)$, we get

$$\begin{aligned}
\int_0^\infty \hat{\mathbb{E}}[\Lambda(x)]dx &\geq R_L \int_{H^{-1}(1/2)}^\infty \frac{1}{2}\big(1 - e^{-(R_L/2)\int_{H^{-1}(1/2)}^\infty H(y+s)ds}\big)dy \\
&\geq \frac{R_L}{2}\int_{H^{-1}(1/2)}^{2H^{-1}(1/2)} \Big(1 - \exp\Big(-\frac{R_L}{2}\int_{3H^{-1}(1/2)}^\infty H(s)ds\Big)\Big)dy.
\end{aligned} \tag{5.13}$$

Now, the integral in the exponent is a strictly positive constant, by the assumption that H has unbounded support. Hence,

$$\exp\Big(-\frac{R_L}{2}\int_{3H^{-1}(1/2)}^\infty H(s)ds\Big) \to 0 \text{ as } R_L \to \infty,$$

and it follows that

$$\begin{aligned}\liminf_{R_L\to\infty} \frac{1}{R_L}\int_0^\infty \hat{\mathbb{E}}[\Lambda(x)]dx &\geq \frac{1}{2}\int_{H^{-1}(1/2)}^{2H^{-1}(1/2)} 1dy \\ &= \frac{1}{2}H^{-1}(1/2) = \nu.\end{aligned} \tag{5.14}$$

Since H has unbounded support, ν is a strictly positive constant, and we conclude that $\int_0^\infty \hat{\mathbb{E}}[\Lambda(x)]dx$, the expected number of nodes in $[0,\infty)$ which have a neighbour in $(-\infty,0)$, tends to infinity as R_L tends to infinity. This does not prove that there is at least one such node with high probability, but it is at least strongly suggestive of it. Thus, Eqn. (5.14) gives us strong reason to believe that the point at the origin, which was conditioned to be isolated, has very small probability of marking an uncrossed gap. This provides some partial intuition for why uncrossed gaps are rare in soft RGGs at the scaling threshold for the emergence of isolated nodes, whereas they are more prevalent that isolated nodes in hard RGGs.

5.3 Scaling for emergence of uncrossed gaps

In Section 5.2, we provided intuition for why uncrossed gaps are rare in the scaling regime at which isolated nodes emerge in 1-D soft RGGs. Here, we seek to identify the scaling regime at which uncrossed gaps emerge. We begin by calculating a bound on the expected number of uncrossed gaps within an interval $[0,L]$, in a soft RGG whose nodes are placed according to a unit rate PPP on the infinite real line; edges are then created independently, with probability $H(r)$ for nodes that are distance r apart. We suppose that Assumption B (monotone decreasing and unbounded support) from Section 5.1 continues to hold.

Define X_0 to be the indicator that there is a point at the origin and that it marks an uncrossed gap, i.e., there are no edges between $\mathcal{P}_{(-\infty,0)} \cup \{0\}$ and $\mathcal{P}_{(0,\infty)}$. We wish to calculate $\mathbb{E}_0[X_0]$, where $\mathbb{E}_0$ denotes expectation under the Palm measure conditional on the PPP having a point at the origin; thus, $\mathbb{E}_0[X_0]$ is the probability that the point at the origin constitutes an uncrossed gap. The expected number of uncrossed gaps in $[0,L]$ is then given by $\mathbb{E}[N_{\text{ucg}}] = \int_0^L \mathbb{E}_0[X_0]dx = L\mathbb{E}_0[X_0]$.

The set of points on $(0,\infty)$ which have an edge to some point in $(-\infty,0]$ constitute a Cox process, which we denote by $\mathcal{C}(\cdot)$; conditional on $\mathcal{P}_{(-\infty,0)}$, they constitute a

Poisson process with intensity

$$\Lambda(y) = 1 - (1 - h(0, y)) \prod_{z \in \mathcal{P}_{(-\infty,0)}} (1 - h(z, y)) = 1 - \exp\big(g(0, y) + \textstyle\sum_{z \in \mathcal{P}_{(-\infty,0)}} g(z, y)\big), \tag{5.15}$$

where $g(x, y) = \log(1 - h(x, y))$. We use a capital letter to denote the intensity to make it explicit that the intensity is random, as it is a function of the random measure $\mathcal{P}$.

Now, the event that there are no edges from $(-\infty, 0]$ to $(0, \infty)$ is precisely the event that the Cox process of points in $(0, \infty)$ reached by such edges is empty, i.e., that $|\mathcal{C}((0, \infty))| = 0$. This is exactly the event whose indicator we defined as X_0. Hence, we obtain using the tower rule (law of iterated expectation) that

$$\mathbb{E}_0[X_0] = \mathbb{E}_0[\mathbb{E}_0[X_0|\mathcal{P}_{(-\infty,0)}]] = \mathbb{E}_0\Big[\exp\Big(-\int_0^\infty \Lambda(y)dy\Big)\Big], \tag{5.16}$$

where $\Lambda(\cdot)$ is given by Eqn. (5.15).

A lower bound on this expectation is easy to compute using Jensen's inequality, which states that, if X is a random variable and f is a convex function, then

$$f(\mathbb{E}[X]) \leq \mathbb{E}(f(X)).$$

Applying this to Eqn. (5.16) with $f(x) = e^{-x}$, we get

$$\mathbb{E}_0[X_0] \geq \exp\Big(-\mathbb{E}_0\Big[\int_0^\infty \Lambda(y)dy\Big]\Big) = \exp\left(-\int_0^\infty \mathbb{E}[\Lambda(y)]dy\right), \tag{5.17}$$

where, as in Eqn. (5.10), the interchange of integral and expectation is justified by Tonelli's theorem and we have replaced $\mathbb{E}_0$ by $\mathbb{E}$, as the conditioning on having a point of the PPP at the origin has been taken into account in the expression for Λ.

The expectation of $\Lambda(y)$ can now be calculated using Campbell's formula for the characteristic functional of the Poisson point process [Hae12, Definition 4.7]. Letting

$g(x,y) = \ln(1-h(x,y))$, we obtain from Eqn. (5.15) that

$$\begin{aligned}
\mathbb{E}[\Lambda(y)] &= 1 - (1-h(0,y))\mathbb{E}\big[\exp\big(\textstyle\sum_{z\in\mathcal{P}_{(-\infty,0)}} g(z,y)\big)\big] \\
&= 1 - (1-h(0,y))\exp\left(\int_{-\infty}^{0}\left(e^{g(z,y)}-1\right)dz\right) \\
&= 1 - (1-h(0,y))\exp\left(-\int_{-\infty}^{0} h(z,y)dz\right) \\
&= 1 - (1-H(y))\exp\left(-\int_{y}^{\infty} H(z)dz\right).
\end{aligned} \tag{5.18}$$

Integrating this over $y \in (0,\infty)$, we obtain

$$\begin{aligned}
\int_0^\infty \mathbb{E}[\Lambda(y)]dy &= \int_0^\infty 1 - (1-H(y))\exp\left(-\int_y^\infty H(z)dz\right)dy \\
&= \int_0^\infty \left(1 - e^{-\int_y^\infty H(z)dz}\right)dy + \int_0^\infty H(y)e^{-\int_y^\infty H(z)dz}dy \\
&= \int_0^\infty \left(1 - e^{-\int_y^\infty H(z)dz}\right)dy + 1 - e^{-\int_0^\infty H(z)dz}.
\end{aligned} \tag{5.19}$$

Now, considering the scaled version of the connection function, $H^L(z) = H(z/R_L)$, we can rewrite the above as

$$\begin{aligned}
\int_0^\infty \mathbb{E}[\Lambda(y)]dy &= \int_0^\infty \left(1 - e^{-\int_y^\infty H^L(z)dz}\right)dy + 1 - e^{-\int_0^\infty H^L(z)dz} \\
&= \int_0^\infty \left(1 - e^{-\int_y^\infty H(z/R_L)dz}\right)dy + 1 - e^{-\int_0^\infty H(z/R_L)dz} \\
&\le \int_0^\infty \left(1 - e^{-R_L\int_{y/R_L}^\infty H(z)dz}\right)dy + 1.
\end{aligned} \tag{5.20}$$

Note that $1 - e^{-x} \le x$ for all $x \in \mathbb{R}$, and so,

$$1 - e^{-R_L\int_{y/R_L}^\infty H(z)dz} \le \min\Big\{1, R_L\int_{y/R_L}^\infty H(z)dz\Big\}.$$

Hence, for arbitrary $x > 0$, we have

$$\begin{aligned}\int_0^\infty 1 - e^{-R_L \int_{y/R_L}^\infty H(z)dz} dy &\leq R_L x + \int_{R_L x}^\infty R_L \int_{y/R_L}^\infty H(z)dzdy \\ &= R_L x + R_L \int_x^\infty \int_{R_L x}^{R_L z} H(z)dydz \\ &\leq R_L x + R_L^2 \int_x^\infty z H(z)dz.\end{aligned} \tag{5.21}$$

We shall henceforth restrict attention to the generalised Rayleigh connection function, $H(r) = \beta \exp(-(r/r_c)^\eta)$, for a fixed $\eta > 0$. For this connection function, we rewrite the last integral on the right hand side of Eqn. (5.21) as

$$\begin{aligned}\int_x^\infty z H(z)dz &= \int_x^\infty \beta z e^{-(z/r_c)^\eta} dz \\ &= \frac{\beta r_c^2}{\eta} \int_{\left(\frac{x}{r_c}\right)^\eta}^\infty u^{\frac{2}{\eta}-1} e^{-u} du \\ &= \frac{\beta r_c^2}{\eta} \Gamma\left(\frac{2}{\eta}, \left(\frac{x}{r_c}\right)^\eta\right),\end{aligned} \tag{5.22}$$

where the second line comes from the substitution $u = \left(\frac{z}{r_c}\right)^\eta$ and $\Gamma(\alpha, y)$ denotes the incomplete Gamma function,

$$\Gamma(\alpha, y) = \int_y^\infty z^{\alpha-1} e^{-z} dz.$$

Observe that

$$\begin{aligned}\frac{\Gamma(\alpha, y)}{y^{\alpha-1}e^{-y}} &= \int_y^\infty \left(\frac{x}{y}\right)^{\alpha-1} e^{-(x-y)} dx \\ &= \int_0^\infty \left(1 + \frac{z}{y}\right)^{\alpha-1} e^{-z} dz,\end{aligned}$$

which tends to 1 as y tends to infinity, by the Dominated Convergence Theorem. Hence, taking $x = r_c(\ln R_L)^{1/\eta}$ in Eqn. (5.22), we see that

$$\int_{r_c(\ln R_L)^{1/\eta}}^\infty z H(z)dz = \frac{\beta r_c^2}{\eta} \Gamma\left(\frac{2}{\eta}, \ln R_L\right) \sim_{R_L} \frac{\beta r_c^2}{\eta} \frac{(\ln R_L)^{\frac{2}{\eta}-1}}{R_L}.$$

Here, for functions f and g on $\mathbb{R}_+$, we write $f(x) \sim_x g(x)$ as $x \to \infty$ to denote that $f(x)/g(x)$ tends to 1 as x tends to infinity. Substituting the above in Eqn. (5.21),

we get

$$\int_0^\infty 1 - e^{-R_L \int_{y/R_L}^\infty H(z)dz} dy \leq r_c R_L (\ln R_L)^{1/\eta} + (1 + o(1)) \frac{\beta r_c^2}{\eta} R_L (\ln R_L)^{\frac{2}{\eta}-1}.$$

Combining the above with Eqns. (5.17) and (5.20), we obtain the following lower bound on the probability that there is an uncrossed gap at the origin:

$$\mathbb{E}_0[X_0] \geq \exp\Big(-C R_L (\ln R_L)^{\max\{\frac{1}{\eta}, \frac{2}{\eta}-1\}}\Big),$$

where $C > 0$ is a fixed constant that does not grow with R_L.

The expected number of uncrossed gaps in $[0, L]$ is given by $\mathbb{E}[N_{\text{ucg}}] = L\mathbb{E}_0[X_0]$. Hence, it follows from the equation above that

$$\mathbb{E}[N_{\text{ucg}}] \geq L \exp\big(-C R_L (\ln R_L)^\theta\big) \text{ where } \theta = \max\left\{\frac{1}{\eta}, \frac{2}{\eta} - 1\right\}. \tag{5.23}$$

The above expression motivates us to consider the scaling regime

$$R_L = \gamma \frac{\ln L}{(\ln \ln L)^\theta}. \tag{5.24}$$

A straightforward calculation shows that

$$\begin{aligned} C R_L (\ln R_L)^\theta &= \gamma C \ln L \Big(1 + \frac{\theta \ln \ln \ln L + \ln \gamma}{\ln \ln L}\Big)^\theta \\ &= \gamma C \ln L (1 + o(1)). \end{aligned}$$

If we conjecture that the inequality in Eqn. (5.23) is an approximate equality, then it follows that the expected number of uncrossed gaps exhibits a sharp threshold at $\gamma = 1/C$, in the sense that

$$\mathbb{E}[N_{\text{ucg}}] \to \begin{cases} 0, & \gamma > 1/C, \\ +\infty, & \gamma < 1/C, \end{cases}$$

as L tends to infinity.

Thus, our calculations lead us to conjecture that uncrossed gaps appear when the connection range scales as $R_L = \ln L/(C(\ln \ln L)^\theta)$; here θ is the related to the power law in the exponent, η of the generalised Rayleigh connection function, while C is a constant that depends in a complicated way on the parameters of the connection

function. We now make a few remarks about this scaling regime. Firstly, C and θ depend on details of the connection function, and are not universal, in contrast to the threshold for isolated nodes. Secondly, R_L is significantly smaller than the $\ln L$ threshold for isolated nodes, confirming that uncrossed gaps appear only when connection functions are of much shorter range than required for the appearance of isolated nodes. Finally, the hard RGG model is a limiting case of the generalised Rayleigh connection function as η tends to infinity; correspondingly, $\theta = 1/\eta$ tends to zero, and the threshold value of R_L tends to a constant multiple of $\ln L$. Thus, we recover the $\ln L$ scaling for the emergence of uncrossed gaps in the hard RGG in the limit.

We also comment here on the statements made at the end of Section 3.2 regarding the fact that we don't expect to see these results in simulations unless very large system sizes are analysed. For clarity, we will denote by R_L^{ucg} the scaling regime in which there is a sharp threshold for the appearance of uncrossed gaps, and we denote by R_L^{iso} the relevant scaling regime for isolated nodes. These have been shown in this thesis to be

$$R_L^{ucg} = \frac{\ln L}{C(\ln \ln L)^\theta},$$

and

$$R_L^{iso} = \frac{\ln L}{2\|H\|_1}.$$

Therefore, to find the value of L (i.e the system size) at which there is a crossover between uncrossed gaps being the more important factor when discussing connectivity and when isolated nodes are more important, we need to find the value of L for which $R_L^{ucg} = R_L^{iso}$, which we denote by L^*. This occurs at

$$\frac{\ln L^*}{2\|H\|_1} = \frac{\ln L^*}{C(\ln \ln L^*)^\theta}.$$

Rearranging this we see that

$$L^* = \exp\left(\exp\left(\left(\frac{2\|H\|_1}{C}\right)^{1/\theta}\right)\right).$$

For $\eta \geq 1$, $\theta = \frac{1}{\eta}$ and hence this becomes

$$L^* = \exp\left(\exp\left(\left(\frac{2\|H\|_1}{C}\right)^{\eta}\right)\right). \tag{5.25}$$

The calculation of the value of C is an interesting problem in its own right and is hence left as future work. However, Eqn. (5.25) is able to tell us that L^* is delicately related to this value of C, and could be extremely large.

The analysis in this chapter showed that uncrossed gaps have negligible probability in the scaling regime in which isolated nodes exhibit their transition. In other words, if any uncrossed gaps are present, then isolated nodes are present in abundance, but not conversely. Consequently, between the two, it is the presence of isolated nodes that is pivotal for determining connectivity. This was shown, not just for specific connection functions, but any function satisfying the mild assumptions of monotonocity, integrability, and unbounded support. The assumption of unbounded support is crucial to our method of proof as it is necessary in both Eqns.(5.3) and (5.5), however, we believe that this is due to our method of proof, rather than being a necessity when analysing this problem.

We also analysed the conditional probability of uncrossed gaps to understand the intuition behind why isolated nodes dominate the uncrossed gaps as the cause of disconnection. Although this analysis is not rigorous it gives the reader hints as to why this is the case. Following from this, the scaling for the emergence of uncrossed gaps is also investigated for a specific form of the connection function, namely the *generalised Rayleigh* connection function. Via this analysis an appropriate scaling is conjectured. It is noted that for smaller system sizes, it appears that uncrossed gaps are more important for determining connectivity and so a "crossover" system size for which isolated nodes become more important is also conjectured. A more detailed discussion about finite system sizes, especially with respect to communication networks, is given in Section 7.1.2.

For the final chapter of the thesis, we now move to a different application of networks of autonomous entities, namely *Swarm Robotics*.

Chapter 6

Swarm Robotics

Another application of ad-hoc communication networks is to *Swarm Robotics*. These swarms can solve tasks that are impossible or too hazardous for single robots. For example, following a nuclear radiation leak, a user may wish to establish a distributed communication chain that partly extends into the most dangerous areas to gather new information. The benefit of using a swarm for this task being that the robots are low cost and the network is designed in such a way that it is robust to the loss of users. The challenge is to create long chains while maintaining chain connectivity ('connected reach'), where those at the distant end of the chain are more likely to be disconnected. The ideas of isolation and connectivity in these networks will raise different questions depending on the type of application that they are being used for. It may be that having a continuously fully connected network is not as important since the aim is to explore a large area, or it may be that the goal of the network is to remain fully connected at all costs if the area to be covered is small but continually changing and hence an up-to-date knowledge of the environment is necessary.

In this chapter we look at a specific example of utilising these swarms of robots to model a biological system: social spiders. Here, we look at dynamic 'boldness' levels of a group of spiders to explore such a risky environment for which this behaviour adapts dependent on the size of the group. This is behaviour that has been seen to occur within clusters of *Stegodyphus* social spiders. Boldness is implemented as a variable associated with the risk appetite of individuals to explore regions more distant from a central base. We present a decentralised mechanism for robots, based on the frequency of their social interactions, to adaptively take on 'bold' and 'shy' behaviours, with lower interaction rates leading individuals to remain close to a central base, and higher interaction rates leading to riskier exploratory behaviour that

may result in disconnection from the swarm. Using this new bio-inspired algorithm, which we call SPIDER, swarms are shown to adapt rapidly to the loss of bold individuals by regenerating a suitable shy–bold distribution, with fewer bolder individuals in smaller groups. This allows them to dynamically trade-off the benefits and costs of long chains (information retrieval versus loss of robots) and demonstrates the particular advantage of this approach in hazardous or adversarial environments.

This work is taken from the paper [HJW+20] for which I was a co-author. My contribution to the work was to aid in both the understanding of mathematically defined mobility models such as the Random Waypoint movement (RWP) model and in helping to create the scoring algorithm used to test the algorithm performances. I also helped in editing the final paper that was submitted and have made more edits to the work in the process of including it in this thesis. The main edits made here were to enhance the clarity of the work and expand on concepts that had to be shortened for the original submission of the paper. The work was done in collaboration with George Jenkinson, Dr. Edmund Hunt, and Dr. Sabine Hauert with the simulation work all being done by GJ and EH.

6.1 SPIDER: A bio-inspired swarm algorithm for adaptive risk taking

In biology there is increasing awareness of the ecological and evolutionary significance of individual differences in behaviour within groups; such differences include that owing to 'personality' variation [DHM04]. The social spider *Stegodyphus dumicola* lives in shared nests and engages in collection predation; group living helps their survival [BCB+07]. Recent research found that social interactions between group members shape individual 'personality' types (shy or bold) such that colonies can regenerate a suitable distribution of behaviours within the group following a perturbation [HMF+18]. Essentially, the density of the group provides a proxy for how much risk the group can afford to take – how many bold individuals it ought to sustain. This is an example of phenotypic plasticity which could be instructive for minimal robot swarms [Hun20]. Here, we take bio-inspiration from *S. dumicola* to develop swarm robot controllers that can achieve a decentralised assignment of individual-level risk appetite that is appropriate to swarm-level capacity to sustain potential losses. We present a new algorithm which we call SPIDER (**S***warm* **P***ersonal***I***ty for* **DE***nsity* **R***esponse*), for allocating risk appetite – propensity to explore – to a swarm of agents

in a hazardous environment. This responds to a risk–return trade-off, whereby robots on the periphery of a swarm are both more likely to gain new environmental information, while also being more likely to become disconnected from the swarm or encounter hazards.

Engineering a robot swarm with the ability to dynamically set a risk appetite (boldness) role distribution that is adaptive to the task at hand has real-world applications, especially for difficult and dangerous scenarios [SUS+20]. These include the establishment of ad-hoc networks in, or exploration of, unknown, unpredictable, dangerous areas. E.g. following a natural disaster like an earthquake or an anthropogenic disaster such as a nuclear radiation leak.

In our scenario a swarm begins with a deployment around a central 'base' area. The swarm task is to establish connected chains of robots (with only very limited communication range) into more radially distant areas. With very simple robots that have severely limited navigation abilities, as robots become more radially distant they are more likely to become isolated from the swarm, or even lost altogether. Chain formation is a common mechanism in swarm robotics for searching the environment and facilitating navigation [ND06; STN11; GB18].

A homogeneous swarm may be able to perform its tasks more effectively if its members have a division of labour into different work roles. Pre-allocating roles to a proportion of the swarm as suitable to the given task can be an appropriate approach [KB00; YR14]. However, this compromises the ability of the swarm to generalise to unexpected environments, or to adapt to robot failure or malfunction. Thus, methods by which decentralised agents can coordinate desired emergent distributions of roles adaptive to the task at hand are desirable. Here, our algorithm adapts boldness dynamically to achieve an allocation of near or distant 'place roles' in an emergent communication chain.

Exchange of information within a swarm relies on individuals maintaining regular, if not constant, contact with the bulk of the swarm; minimising the time that agents spend disconnected from the bulk of the swarm ensures a high level of swarm cohesion [HWZ+08; HZF09]. Maintaining connectivity is usually achieved by attaching a cost to becoming disconnected: varying this cost over swarm agents stratifies the swarm with individuals who occupy different functional niches or roles. The cost can be adapted online to reflect the depletion of energy or battery levels [BR14; LSB19]. All of this poses a crucial trade-off for exploratory swarms: new areas are best discovered by robots that actively seek areas they have not covered yet. This may however lead to them losing communication with the swarm. Exploration–connectivity trade-offs

are a common challenge for swarm robotics [HMK+14].

We develop a decentralised swarm algorithm that approaches the exploration–connectivity trade-off in a novel way, using local swarm density as a proxy for the capacity of the swarm to engage in risky behaviour. Our approach may be particularly suited to hazardous environments where agents are removed either temporarily or permanently.

6.2 Methods

6.2.1 Robots and experimental arena

The Kilobot [RAN12] has become a popular swarm robotics research platform that allows conceptual demonstration of swarm algorithms in very large numbers of robots. It is relatively small, with a diameter of 33mm and height of 34mm, and has two vibrating motors instead of wheels. It has a communication range of up to 10cm, and very limited sensing and localisation capabilities. Therefore, for the purposes of facilitating radial navigation the arena is marked with a series of concentric rings the width of a Kilobot that the robots can detect as being labelled 1, 2, or 3. This provides a gradient to count up or down to a desired travel distance, which ranges in our experiments from 0 to 26 rings. An example layout is illustrated in Fig. 6.1. In real-world environments, an environmental gradient could be provided by received signal strength intensity (RSSI) from a base station radio beacon or other robots (e.g. [MWT+19]).

Robots are initialised in a ring around a central 'base' robot and move radially inwards or outwards depending on their boldness.

Kilobox [JSH+18] is a 2-Dimensional simulator adapted from Box2D written in C++ demonstrated to give accurate and fast simulations of Kilobots [RAN12]. All experimental data presented in this study has been obtained through Kilobox simulations.

Experimental trials were all simulated for the equivalent of 5000s. This provides ample time for the swarm to reach a steady state, and is a realistic amount of time for an experiment on real Kilobots with respect to battery life.

6.2.2 Task and performance metric

Main Task

The task for the swarm is to arrange connected chains of robots from a central 'base' area into more radially distant positions in the arena. It does this by arriving at a stable distribution of boldness-related roles that balances the need to occupy the 'safe' central region with the opportunity to cover more distant regions at and beyond the periphery of the swarm. It is important to manage the risk that peripheral robots lose contact with the swarm for extended periods of time, which would impair swarm cohesion. To manage this risk, the dynamic, self-organised swarm boldness distribution is associated with a spatial swarm distribution over a circular arena. The key principle is that a larger swarm can afford to be more risky in how it allocates boldness roles, because the disconnection of one robot (temporarily or permanently) is on balance less likely to compromise the swarm's connected reach.

Performance metric: maximum 'connected reach'

We assign a reward to connected chains of robots reaching into distant regions, which is inherently penalised (reduced) if robots become disconnected from the swarm. To assess the performance of the task we measure the maximum distance reading received at the swarm centre from peripheral robots. Each member of the swarm broadcasts its current distance from the central circle as a message towards the centre of the swarm, which percolates inwards via connected robots. Every 0.5s, the robots reset their broadcast message to convey their current distance, and overwrite this if they receive a message from another robot that conveys a larger distance, such that they act as a signal repeater for any robot that is further out. In a given time frame, the base robot at the centre of the swarm records the distance of the message which has arrived from the furthest point to the centre (Figure 6.1). This simple performance metric can be recorded on-board the 'base' Kilobot for experimental analysis, though it is not required for the controller.

Shy, medium and bold behaviours

For the purposes of our analysis, it is useful to stratify swarm members into 3 categories: shy, medium, and bold, as in [HMF+18]. The three behaviours can be differentiated as follows: shy behaviour manifests as remaining within a region that has a very low chance of becoming disconnected, and the geometry of the arena and

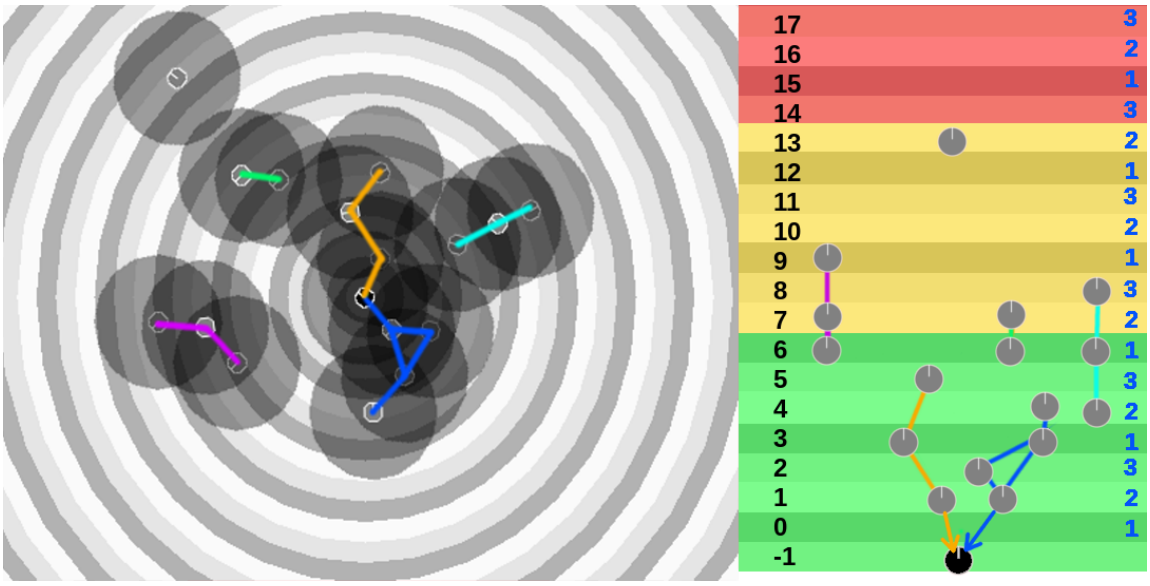


Figure 6.1: Swarm attempting to optimise 'connected reach' in 2D arena (left) and network graph representation (right) of this example swarm layout. The green, yellow and red regions are indicative of the risk of disconnection from the swarm for a given swarm size. Coloured lines show different connected clusters of Kilobots. The blue numbers on the right hand side show the gradient available to the Kilobots' sensors as they move between the grey-scale rings, which they use to track their position. The 'base' Kilobot (black) increments the performance metric (furthest connected reach) every 0.5s for the duration of the experiment, which in this instance would add 5 to the score total.

the dynamics of the boldness mean that in general this is in the central region. Bold robots venture beyond the periphery of the group, into regions that are unlikely to be connected, but that result in a higher incremental performance score (successful transmission of a distant message back to base) if they are connected at any time. Medium boldness robots straddle these two behaviours, often contributing a moderately high score increment themselves and also enabling bolder robots to score occasionally by providing a semi-stable link to the central cluster.

In larger swarms robots are more likely to be connected to the central region even at larger radial distances. This is because there tend to be higher densities in the central region, and hence the swarm would benefit from shifting its boldness distribution to include more bolder robots. In smaller swarms, even at a moderate distance a robot has a risk of being disconnected from the swarm. Therefore, a larger swarm is expected to be able to accommodate more bold individuals because this poses little risk to its connected reach, whereas a small swarm will require a higher proportion of shy individuals in order to remain connected with the base station at all.

We now introduce the two main (and novel) algorithms used in this work to incorporate social interactions into the swarm behaviour, namely, *SPIDER-density* and *SPIDER-boldness*.

6.2.3 A basic boldness mechanism: SPIDER-density

The boldness-based movement behaviour of the SPIDER algorithm is described in pseudo-code in Algorithm 1. We describe two variants of the algorithm, a simple, density-based version (SPIDER-density) and a refined version of the algorithm that includes sensitivity to neighbour boldness (SPIDER-bold).

The desired radial distance (measured in number of rings) from the centre r_d relates to an individual robot's boldness, which ranges from 0 to 255. This radial distance is then given by

$$r_d = \left\lfloor \frac{bold}{10} \right\rfloor .$$

For example, a bolder robot with boldness 150 will have a travel distance of 15 rings from the centre, whereas a shy robot of boldness 53 will seek a distance of 5. The current ring is denoted by r_c. Note that r_d cannot be updated unless and until the destination is reached. To calibrate boldness levels to the arena geometry, distances were specified such that a robot with maximum boldness set its distance goal to

Algorithm 1 SPIDER-density algorithm pseudo-code

1: **procedure** SPIDER-DENSITY(a, b) ▷ input parameters *Setup*: (for individual robots separately)
2: $r_c \leftarrow$ robots placed in ring formation
3: $bold \leftarrow$ set random boldness $\in [0, 255]$
4: $r_d \leftarrow \text{floor}(bold/10)$
5: *Loop*: (every $0.5s$)
6: $density \leftarrow$ number of neighbours within range
7: $bold = bold + a \times density - b$
8: $bold = \text{limitrange}(bold, [0, 255])$
9: $r_c \leftarrow$ travel according to gradient ascent/descent
10: **if** $r_c == r_d$
11: $r_d \leftarrow \text{floor}(bold/10)$
12: **endif**

the outermost region, and a robot with minimum boldness set its travel distance to the distance with index 0. The intermediate boldness levels are then linearly spaced between them.

A simple way to introduce a density-induced boldness dynamic to the swarm is to have a boldness that increases when a robot experiences a 'social interaction' (receives a message at close proximity), and otherwise decreases at a fixed rate. Because the Kilobots send messages with the same constant frequency, the message reception frequency is proportional to how many neighbours are within communication range, and therefore the local density.

Following Algorithm 1, robots move outwards from densely populated central areas towards the outer regions, which are likely to be more sparsely populated, resulting in a drop in their boldness level. Due to the steadily applied boldness decrease, robots are always drawn back toward the centre of the swarm until they encounter at least one other robot. These countervailing behaviours ensure the maintenance of a good level of connectivity throughout the swarm, and avoid over-packing of the central region. The boldness score changes over time in the following way: it increases proportionally to the constant a at each time-step according to how many neighbours are within its communication range, and decreases at each time step by the constant b (as seen in line 7 of Algorithm 1). These can be optimised for swarm performance according to different user requirements, as discussed in Section 6.2.7.

6.2.4 A refined boldness mechanism: SPIDER-boldness

In this section, additional refinements are made to the controller, which we refer to as SPIDER-boldness. Its mechanism is very similar to the pseudo-code of Algorithm 1 but with the changes as described in the following sections. This controller includes each of following four features, which variously make the swarm controller more or less responsive to certain environmental conditions.

Feature 1: Boldness increase with bold neighbours

Following behaviour observed in *S. dumicola* [HMF+18], the controller was adapted so that robots only increase their boldness when in the presence of bolder individuals. This was set to take effect when a Kilobot's boldness level was below the mean of all its neighbours within communication range. As in all other trials, the robots are initialised with random boldness and boldness is dynamic.

Feature 2: 'Addictive' boldness increases

Also inspired by behaviour observed in *S. dumicola* [HMF+18], the behaviour was updated to include an 'addictive' term, where robots that had recently increased their boldness were more likely to increase it again by a larger amount. Feature 1 means that a robot's boldness can only increase if it is more bold than its immediate neighbours. Therefore, within a cluster it is harder for the boldness levels of the robots to increase too quickly before they inevitably move away from their current cluster. However, the addictive increases in boldness provides a way for the swarm to avoid a falling ceiling on its overall boldness.

Feature 3: Relative boldness increase/decrease

To stabilise the distribution of shy and bold individuals over the swarm, the rate at which an individual altered its boldness level was set to be a function of the individual's current boldness level.

In relation to Algorithm 1, the boldness update was amended in SPIDER-boldness

to be:

$$\begin{aligned}
&\textbf{if } mean(neighbour_boldness) > boldness \textbf{ then:}\\
&\quad bold = bold + f(bold) \times addict - g(bold)\\
&\quad addict = addict \times c\\
&\textbf{else:}\\
&\quad bold = bold - g(bold); \quad addict = 1,
\end{aligned}$$

where, for positive constants a, b, and c, f and g are given by

$$f(bold) = \frac{a}{bold},$$

and

$$g(bold) = \frac{b}{bold}.$$

a and b relate to the rate of boldness increase and decay, and c gives the robot's 'addiction rate'. The choices of f and g presented here meant that individuals had a tendency to remain bold for longer once their boldness level was significant. Choosing other forms for these functions would yield different behaviour. The values for a and b can be interpreted as having the effect of changing the rate at which the boldness of a robot changes. $a >> 1$ and $b << 1$ result in stability when the robot is bold, $a << 1$ and $b >> 1$ when it is shy, and intermediate combinations resulting in different stable boldness levels. Higher values for a and b tend to manifest in faster changing and more variable boldness scores.

Feature 4: Updating desired distance en-route

As a robot moves throughout the arena, it will continue to receive information on its changing local neighbourhood density and could act on this to prevent it from straying too far from the swarm and into a very sparse region, or persisting in attempting to reach the centre of the swarm when it is prohibitively crowded. In order to update its behaviour dynamically, we allow the robot to change its desired distance from the centre in transit to accommodate new information. This may increase the cohesion of the swarm and stops overcrowding in the centre as the robots are not bound to completing the task of reaching their waypoint destination.

We set a threshold difference between the travel distance and the current boldness, which when exceeded causes it to update the desired distance based on the current

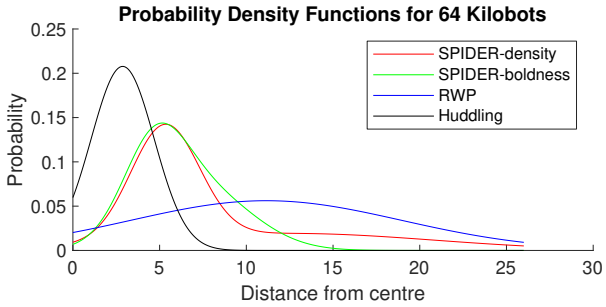


Figure 6.2: The probability density for robot locations, across the 4 controllers, for a swarm size of 64. Curves are fitted to data collected from 10 simulations using Matlab's Gaussian fit function with 1 peak or 2 peaks for the SPIDER controllers (central bulk and congestion effects). Distance from the centre is measured in navigational ring widths (33mm). It is important to note that in the RWP model, the robots are randomly assigned a number between 0 and 26 (minimum and maximum ring values) and aim to travel to this ring directly (e.g. to the nearest part of the 10th ring to them). Therefore, rather than the peak of the curve being at the centre of the arena (ring 0), it is now at the central ring between 0 and 26 (i.e. ring 13).

boldness. In these experiments the threshold was set to update the distance when the boldness had diverged from the that associated with the goal by more than 2.

6.2.5 Swarms without a boldness mechanism

Two movement behaviours were used to provide a baseline against which to test the effectiveness of the SPIDER algorithm. These had no boldness dynamic or communication between agents, and are random waypoint choice (RWP), where each robot sets a random travel distance which is reset upon arrival [APR+18]; and a 'huddling' behaviour where all agents had their distance goal fixed at the centre of the arena.

6.2.6 Testing adaptability and robustness

Balancing connectivity and reach

Different tasks, or robots with different capabilities, may prioritise the stable connectivity of chains over the extent of their reach into distant areas, or vice versa. For example, maintaining a communication network for emergency services in a difficult to reach, unknown terrain demands a constant, stable, connection, whereas a team of robots searching a less dangerous environment for a resource can operate using

only infrequent communication. For this reason it is important that the swarm can easily be adapted to suit the balance of connectivity and reach required.

We anticipate that a swarm prioritising connectivity over exploration is likely to have a boldness that decays very fast when in the presence of few or no neighbours, such that bold robots are infrequent and quickly become shy again; with a slow boldness decay for a swarm that prioritises exploration over constant connectivity. This hypothesis is readily tested by comparing a short scoring period for the central, 'base' robot, to a long one. We examined a scoring period of 20s rather than 0.5s. This meant that the swarm would increment its score by the most distant connection made between the centre and the periphery at any point over the 20s period. This rewards less stable, but more distant connections. This is discussed in more detail in Section 6.3.5.

Robustness testing – population catastrophes

A desired advantage of a decentralised role distribution mechanism is that it should be robust and consistent in a changing or even hazardous environment. Therefore, for a given environment, the distribution of roles throughout the swarm should stabilise to a similar state that optimises the reach–connectivity trade-off from a variety of starting conditions. This ensures that the swarm is resilient to noise and can be decentralised in its operation. The swarm should also act to redistribute its roles as appropriate to a changing population, for example following the destruction of agents.

To examine whether the swarms were resilient to extreme change, they underwent a simulated catastrophic population event. This could correspond to the malfunction, removal, or destruction of a significant section of the robots. This motivates the need for a control mechanism suitable for simple, expendable (replaceable) agents. Such an event is simulated by programming half of the swarm to cease communication with the rest of the robots and remove themselves from the arena after a given amount of time. This catastrophe was repeated over the swarm twice over the course of an experiment, changing the population from 128 to 64 to 32, at intervals of 1500s. In a robust system, the remaining swarm should redistribute itself over the boldness space to optimise for connectivity and reach after each catastrophe.

Because the boldness distribution changes depending on the swarm size, 20 repeats of 5000 second simulations were run in order to find the stable sub-population distribution that each swarm size converged towards given enough time. The same

values of a, b, and c were used for each population size and were those that had proven to be effective for a population of 64 Kilobots (how these were obtained is explained in Section 6.2.7). The results of these simulations were then used as a measure of how well the swarm recovered from catastrophic failure by comparing how the sub-population densities (namely those of the shy, medium, and bold populations) varied over time after each catastrophe. A measure of the distance of the swarms behaviour from equilibrium is calculated as the difference between the current sub-population density and the stable sub-population density at 50s intervals. A measure of the volatility of the swarm behaviour is calculated by firstly calculating the standard deviation of each of the sub-population densities over a 50 second period and then summing these together. The time taken for the swarm to settle after a population catastrophe is then given by the time taken for the distance measure to fall below 0.3 and remain there, and for the volatility measure to fall below 0.1 and 0.2 and remain there for the first and second catastrophes respectively. This change in the volatility threshold is implemented in order to reflect the reduced population size.

Random waypoint choice

In order to have a base level algorithm to make comparisons with, a random waypoint mobility model is also introduced. Here, the target distances were chosen randomly, rather than according to a boldness level. Each individual chooses its destination as a random distance from the centre, roughly corresponding to a 1-Dimensional Random Waypoint (RWP) Mobility Model in which the agents in the network choose a random location along a line as their destination; the distribution of agent locations over a line segment is well described in [BRS03]. Although swarm agents are in a 2-D environment, the controllers presented here act in the radial direction with no consideration for azimuthal distribution, and hence the 1-D model is relevant. In practise, an individual Kilobot chooses a ring to travel to uniformly at random. It then begins its journey towards this ring as described in Section 6.2.1. The fact that it simply travels to its destination ring and not to a specific location in space is why the 1-dimensional version of the RWP mobility model is more relevant to this scenario. By keeping track of individual agent's locations along the radial line, we are able to produce a probability density function (pdf) of where an individual would be at a random time. This is done by integrating the agent's locations over time. This distribution is bell shaped and symmetric about the distance halfway between the centre and the circumference, as is to be expected when analysing the RWP mobility

model. This can be intuitively understood by considering that the random choice of waypoint is expected to give a uniform distribution of distances along the radius, but as the individual has to travel from its current position to its target, it must cross all of the distances between its current location and its destination, which is likely to include regions near the half-radius distance, and unlikely to include regions near the centre or edge of the circle.

As more individuals are introduced, collisions become more likely near the central regions compared to the regions near the edge, because the circular geometry means that regions closer to the centre cannot accommodate as many robots as those regions that are further out. The collisions significantly slow the robots, and so for higher populations of swarms choosing their destination randomly, the spatial distribution of the swarm skews towards the centre of the arena. This skewed distribution of the swarm at high populations suggests that the emergent behaviour of the group will maintain good connectivity at the centre of the arena, however the exploration of the arena beyond the central cluster is likely to be carried out by disconnected individuals in a fashion similar to the particles described in [LTH+92].

6.2.7 Automated parameter optimisation

Given the complex relationship between input model parameters for individual-level robot behaviour, and overall swarm performance for different swarm sizes and user connection stability requirements, we optimise parameters a, b and c using an evolutionary algorithm (EA) [FM08]. A simple EA was developed using a population size of 40 where an individual in the population is a chosen combination of values of a, b and c. Rank-based selection on swarm-level performance was used as measured in Section 6.2.2.

The EA worked over 20 generations on the 3 'genes' (i.e. a, b, and c) with elitism, and a combination of crossover without contamination, and crossover with contamination (the 'genes' are broken down and mixed up), all performed with small consistent mutations (value jitter of $\pm 10\%$). Elitism in this scenario was implemented by only using the top ten performing 'genes' to generate the the pool of 40 for the subsequent generation, meaning that the worst performing 'genes' are discarded. The scores here are calculated by running simulations as described in Section 6.2 with all 40 combinations of the input parameters. Contamination is used here as a way to pseudo-randomly explore the state space in more detail such that the algorithm explores solutions far away from those that have performed well along with solutions

close to those that have performed well. This is explained in the following example. If the first generation includes the combinations of a, b, and c given by

$$\{a_1 = 5, b_1 = 2, c_1 = 17\}$$
$$\{a_2 = 12, b_2 = 9, c_2 = 24\}.$$

If crossover is implemented without contamination, then the second generation may include something like

$$\{a_1 = 5, b_1 = 9, c_1 = 17\}$$
$$\{a_2 = 12, b_2 = 2, c_2 = 24\},$$

in which crossover has meant that the values associated with b_1, and b_2 from the first generation have been switched for the second generation. If crossover is implemented with contamination, then the second generation may instead have included

$$\{a_1 = 5, b_1 = 18, c_1 = 17\}$$
$$\{a_2 = 12, b_2 = 1, c_2 = 24\},$$

where this time as well as the swapping of the inputs, they have also been multiplied by a random factor.

The second generation of input parameters is thus made up of the top ten performing 'genes' replicated with jitter, then ten 'genes' constructed from these but with crossover without contamination implemented, and the final 20 being constructed with crossover with contamination being implemented. This is shown pictorially in Figure 6.3.

Since we are working under the assumption that our algorithm does not get stuck in a sub-optimal local maxima, we are unable to claim that the values produced using this implementation provide globally optimal solutions. However, the EA provides a reliable method by which to find input parameters that produce swarm-level behaviour that scores well on the performance metric. The performance metric itself can be adjusted depending on the nature of the swarm task. The standard performance metric scored the distance of the furthest connected node every 0.5 seconds, but a slower scoring frequency allows for less reliable connectivity, as examined in Section 6.3.5.

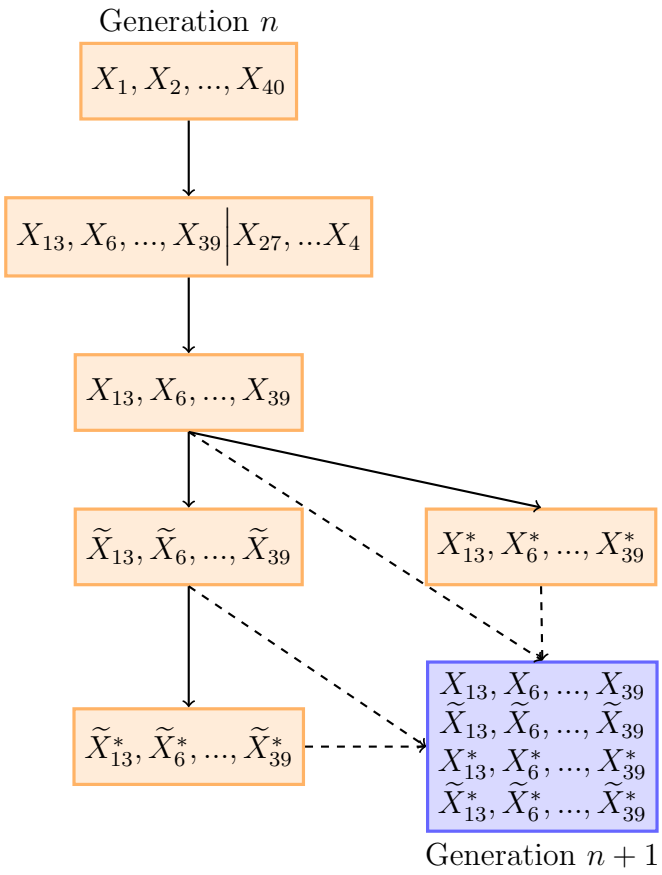


Figure 6.3: A pictorial example of what the algorithm is doing at each generation. The 40 combinations of input parameters are firstly scored and ranked, with only the top ten being kept. The top ten are kept for the following generation (with a value jitter of $\pm 10\%$) as well as being used to create 20 more combinations, ten from crossover without contamination (marked by a tilde) and ten from crossover with contamination (marked with a star). A final ten combinations are also created by performing crossover with contamination on the combinations previously being constructed from crossover without contamination (marked with a tilde and a star).

6.3 Results[1]

For the purpose of analysis swarm density is calculated by dividing the combined area reached by the robots' (here, Kilobots') communication range by the area of the arena with overlapping of communication ranges being ignored. The swarms were scored for populations of 8, 16, 32, 64, and 128, corresponding to densities of 0.0329, 0.0657, 0.1315, 0.2630, and 0.5260 respectively.

6.3.1 Random waypoint choice

Up to a certain population density, robots randomly choosing their desired distance produced a symmetric bell shaped probability density function (pdf) when the individuals' positions were concatenated and integrated over time (Figure 6.2). As the population approached a swarm density of 0.5 (population size approaching 128), collisions and crowding at the centre significantly skewed the distribution towards the centre of the arena. Due to this skew and higher robot density, the performance score of larger populations of robots following the RWP behaviour continued to increase as the number of robots in the swarm increased, thus leading to the performance not differing significantly from swarms with a boldness dynamic (Figure 6.4). Thus, at high swarm densities, more complex control strategies are no more effective than a simple approach as was also noted in [HJH19].

6.3.2 Huddling behaviour

With robot boldness set to a minimum, there was consistent connectivity in the central arena region at the expense of not reaching into further radial distances. Nevertheless, this was sufficient for performance scores comparable with boldness-dependent behaviours at very low populations. As swarm size grew, the performance of risk-taking exploratory behaviours quickly overtook this 'huddling' strategy. However, it is a reasonable approach to reaching a small amount of the area if connectivity is paramount to the task and only a small population is available. It outperforms the random waypoint behaviour up to a swarm density of approximately 0.25 (Figure 6.4). We suspect that this is owing to the stable connectivity facilitated by the increased density about the centre of the arena.

[1]All data analysed here is available on GitHub: https://github.com/geojenks/adaptive-risk-appetite

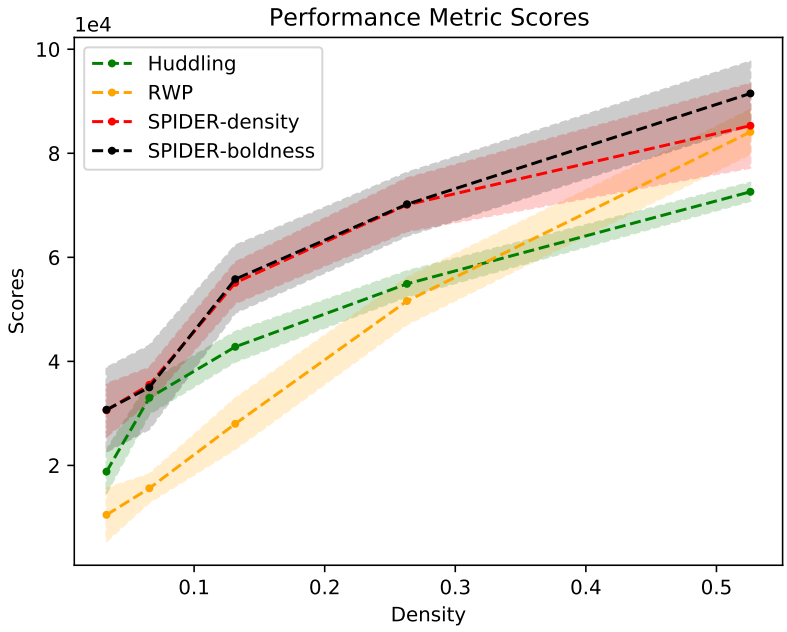


Figure 6.4: Performance metric score comparison for the 4 controllers with the shaded regions showing the 95% confidence intervals (assuming a normal distribution of the simulation results). See Fig. 6.1 for the scoring method: it rewards successful transmission of distant messages back to base. At interim densities the boldness-based controllers outperform huddling and RWP strategies, but this advantage diminishes as the density increases.

Performance Metric Scores of SPIDER Algorithms								
Population	Huddling		RWP		SPIDER-density		SPIDER-boldness	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
8	18800	2100	10500	2590	30600	2580	**30700**	4110
16	33000	1580	15600	1380	**35500**	1670	35000	4100
32	42800	1470	28000	2480	55100	2010	**55800**	3300
64	54900	1250	51600	2290	70100	2600	**70200**	3100
128	72600	904	84100	2170	85300	4130	**91500**	3140

Table 6.1: Scores of performance metrics for the four different control algorithms and varying population sizes with the highest score for each population size given in bold. Recall that every 0.5s the scores for a particular swarm are calculated and this happens for the 5000s for which the simulations are running, giving a total of 10,000 time periods. The scores from each of these time periods are then summed together to give the overall score for a simulation run.

6.3.3 SPIDER-density controller performance

The controller parameters selected by the EA for small swarms tended to favour low risk strategies whereby robots remain close to the central region where they are likely to be connected to the main cluster. As the population grew, the selected parameters produced swarms containing robots much more likely to be found in riskier regions further from the arena centre (N=64 shown in Figure 6.2).

Performance metric score

The SPIDER-density controller outperforms the random waypoint (RWP) behaviour beyond a confidence level of 95% up to and including a population size of 64 robots. Beyond this, the RWP and SPIDER-density behaviours score more closely. At a population density of 0.5, the mean score of the SPIDER-density swarm does not fall outside of a single standard deviation of that of the RWP. However, it did outperform the huddling behaviour across the higher range of population densities but not significantly at the lowest densities (Figure 6.4). This indicates that for small populations, the strategy that the EA evolves is rather risk-averse.

Population catastrophe

The SPIDER-density controller sometimes resulted in moderately fluctuating subpopulations of each boldness class, as can be seen in the initial stages of Figure 6.5A. This may have had the effect of increasing the speed at which the swarm can react to a change such as a population catastrophe; being the fastest controller to stabilise, the swarm took an average of 542.9s to stabilise at a new distribution (Figures 6.5A and B).

6.3.4 SPIDER-boldness controller performance

The SPIDER-boldness controller was particularly effective for larger swarms. This is likely because as the population increases, it becomes more difficult to travel to or from the central region due to blocking, so being able to abandon the destination and select a new desired distance allowed robots to leave the overcrowded central area. The distribution of the robots over the environment is less positively skewed into distant regions than for the more simple SPIDER-density controller. This is a reflection of the decaying local swarm density (Figure 6.2), as a robot will not continue to move outwards far beyond the periphery of the swarm if it does not

encounter another agent. The peak of the pdf of the agent locations moves through 3.88, 4.21, 5.52, 8.41, as the population redoubles from 16 to 128, thus indicating that as the population size increases the most visited regions are increasingly far away from the centre of the arena. The regions toward the edge of the arena are relatively unexplored, despite the swarm having a considerable bold proportion. This is likely because as robots travel toward far regions, their boldness falls and so they adjust their distance to be closer to the centre before they reach their previous distance goal. This can be seen in Figure 6.2 when comparing the pdf of SPIDER-density with SPIDER-boldness and noticing the much more rapid decay seen in SPIDER-boldness as the distance from the centre of the arena becomes more extreme.

Performance metric score

The SPIDER-boldness controller allows the swarm to significantly outperform random waypoint (RWP) behaviour up to a population density of 0.5. For low populations, the swarm-wide behaviour was similar to a swarm of minimal boldness, as most robots were required to be close to the centre to maintain any connection at all. The swarms controlled by this behaviour outperformed the RWP behaviour beyond a confidence level of 95% up to a density of 0.5, with its mean score outperforming RWP by a factor of 2.9, 2.2, 2.0, 1.4 at the respective population densities. Once the population size reached 128 robots, its scores were only slightly better than those of completely random choice, its mean score outperforming that of the RWP behaviour by a factor of 1.1 and a confidence level of 90%. The SPIDER-boldness behaviour significantly outperformed the huddling behaviour with a confidence level above 95% for population sizes above 16, its mean score being better by a factor of approximately 1.3 at each of the swarm densities tested (Figure 6.4).

Population catastrophe

The sub-populations gradually approached their stable proportion after a catastrophe and did not overshoot or oscillate heavily (Figure 6.5C). The swarm stabilised after an average of 615s following the population being halved (Figure 6.5C and D). We believe that this reliable recovery is due to the combination of the stabilising effects of increasing robots' boldness levels only when they were surrounded by bolder neighbours at a rate relative to their current level, and the ability adapt their goal en-route.

Distance Travelled by Kilobot								
Population	Huddling		RWP		SPIDER-density		SPIDER-boldness	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
16	2.31	0.26	9.57	0.96	5.22	0.43	2.98	1.07
32	2.30	0.14	9.67	0.46	4.47	0.18	4.11	0.27
64	2.31	0.07	9.04	0.22	8.23	0.32	6.40	0.44
128	2.33	0.05	7.37	0.15	4.95	0.15	6.12	0.05

Table 6.2: Distance travelled by the Kilobots for each behaviour over the range of populations. Each data point is from 10 repeat experiments. Identifying the most energy-efficient behaviour is not possible from this data without introducing more parameters.

Population Catastrophe Recovery Time (s)						
N=Pre:Post	SPIDER-density			SPIDER-boldness		
	mean	s.d.	success	mean	s.d.	success
128:64	**62.5**	76.8	100%	361.5	89.3	100%
64:32	**40.0**	21.0	100%	110.5	49.9	100%
128:64:32	**51.3**	48.9	100%	236.0	69.6	100%

Table 6.3: Swarms' abilities to stabilise their shy-medium-bold distribution after two population catastrophes. Means are from 10 repeat experiments. Both SPIDER controllers stabilised the role distribution after a catastrophe. The fastest recovery times are emboldened.

6.3.5 Suitability for less connection-critical tasks

The EA was used to optimise the SPIDER controllers at a longer scoring time interval of 20s. By recording the furthest successfully relayed message over this longer time period, there was scope to tolerate longer interim time periods where less distant or no messages were received at the central base. This resulted in a distribution of roles which was bolder than when scoring over a shorter time interval of 0.5s (Table 6.5). This demonstrates that the boldness dynamic can be altered to be suitable for different tasks, or for robots with different capabilities, depending on how the user prioritises connectivity of chains or further reach – exploration and information gathering – from peripheral regions. This is an important finding when considering the applications of this work since it shows that our algorithm is able to adapt to different tasks with differing end goals.

Average Boldness of SPIDER Algorithms				
Population	SPIDER-density		SPIDER-boldness	
	mean	s.d.	mean	s.d.
16	0.19	0.088	0.19	0.045
32	0.20	0.008	0.21	0.036
64	0.25	0.057	0.25	0.070
128	0.25	0.040	0.31	0.046

Table 6.4: Average boldness scores of the Kilobots sustained over the last 1500s of 5000s simulations, from the best scoring swarms for SPIDER-density and SPIDER-boldness. Both controllers increase the mean boldness over the swarm for higher populations.

Boldness in Swarms of Size N=64 For Two Scoring Periods				
Score Period	SPIDER-density		SPIDER-boldness	
	mean	95% range	mean	95% range
0.5s (fast)	0.243	0.231:0.255	0.262	0.239:0.284
20s (slow)	0.30	0.273:0.327	0.342	0.319:0.365
slow/fast	1.23		1.31	

Table 6.5: When scoring the swarm performance metric over longer time periods, the boldness dynamics are optimised for a more risk-taking swarm, because it permits short-term connectivity in the pursuit of a further reach.

6.3.6 Overall comparison of SPIDER variations

Both the SPIDER-density and SPIDER-boldness controllers increased the average boldness scores over the swarm as the population grew (Table 6.4). The similar stable boldness distributions for the two mechanisms show that the boldness dynamic is a reliable way to reach a role distribution. It also allows the swarm to effectively react to its current state in order to stabilise or alter its role distribution.

The SPIDER-density controller was capable of adapting to population changes far more quickly than the SPIDER-boldness controller (Table 6.3). The robots controlled by SPIDER-boldness changed their boldness levels relative to their current level, meaning that bold robots remained bold – and in isolation – for much longer than the simple density-based behaviour. Therefore they took longer to react to population reduction and the need for more shy robots. This is also why SPIDER-boldness was able to regain its stable population when recovering from an all-shy population very quickly, but not as fast from an all-bold population (Table 6.3). The utility of swarm role stability versus adaptability is task-specific, and having an unstable, but highly adaptive role distribution, as demonstrated by the basic behaviour, is indicated as

more energetically demanding (Table 6.2).

The two controller variants achieve similar scores up to a swarm density of 0.263, yet the SPIDER-boldness controller executes the task while consistently travelling significantly less distance, on average covering 85% of the linear distance, and so expending less energy (Table 6.2). The en-route updating means that an agent's behaviour was closely linked to its boldness. By comparison agents using the SPIDER-density controller often had long periods where their behaviour had been set at a time when their boldness was significantly different to its present level. Therefore, it is likely that using a SPIDER-boldness controller with input parameters a, b, c that have been evolved, or otherwise developed, to minimise boldness while achieving a high score, provides a path to minimising the swarm's energy consumption. The fact that SPIDER-boldness was able to out score SPIDER-density whilst having a much smaller proportion of 'bold' Kilobots (as seen in Figure 6.5) and noting that the Kilobots travelled significantly less distance is a good indication of this.

6.4 Concluding Remarks

Our bio-inspired boldness dynamic provides the foundational theory for a computationally cheap, decentralised mechanism by which a swarm of robots can arrive at a stable distribution of roles in relation to risk appetite (boldness or willingness to explore). This boldness dynamic, which we call the SPIDER algorithm, significantly improves the swarm's ability to reach from a base into a potentially hazardous distant area with suitable levels of connectivity.

Demonstrating these boldness mechanisms in real robots – Kilobots – is a potential next step in validating their effectiveness. Equally, although we have examined very simple robots, the SPIDER algorithm could be implemented in robots with more sophisticated navigational and localisation abilities. There is also an opportunity to deploy the SPIDER algorithm in real-world field trials in hazardous environments, to demonstrate the effectiveness of a density-based risk appetite controller for optimising the connectivity–reach trade-off. This decentralised swarm approach is especially relevant when significant numbers of agents are at risk of loss or destruction, such as in adversarial or disaster response scenarios.

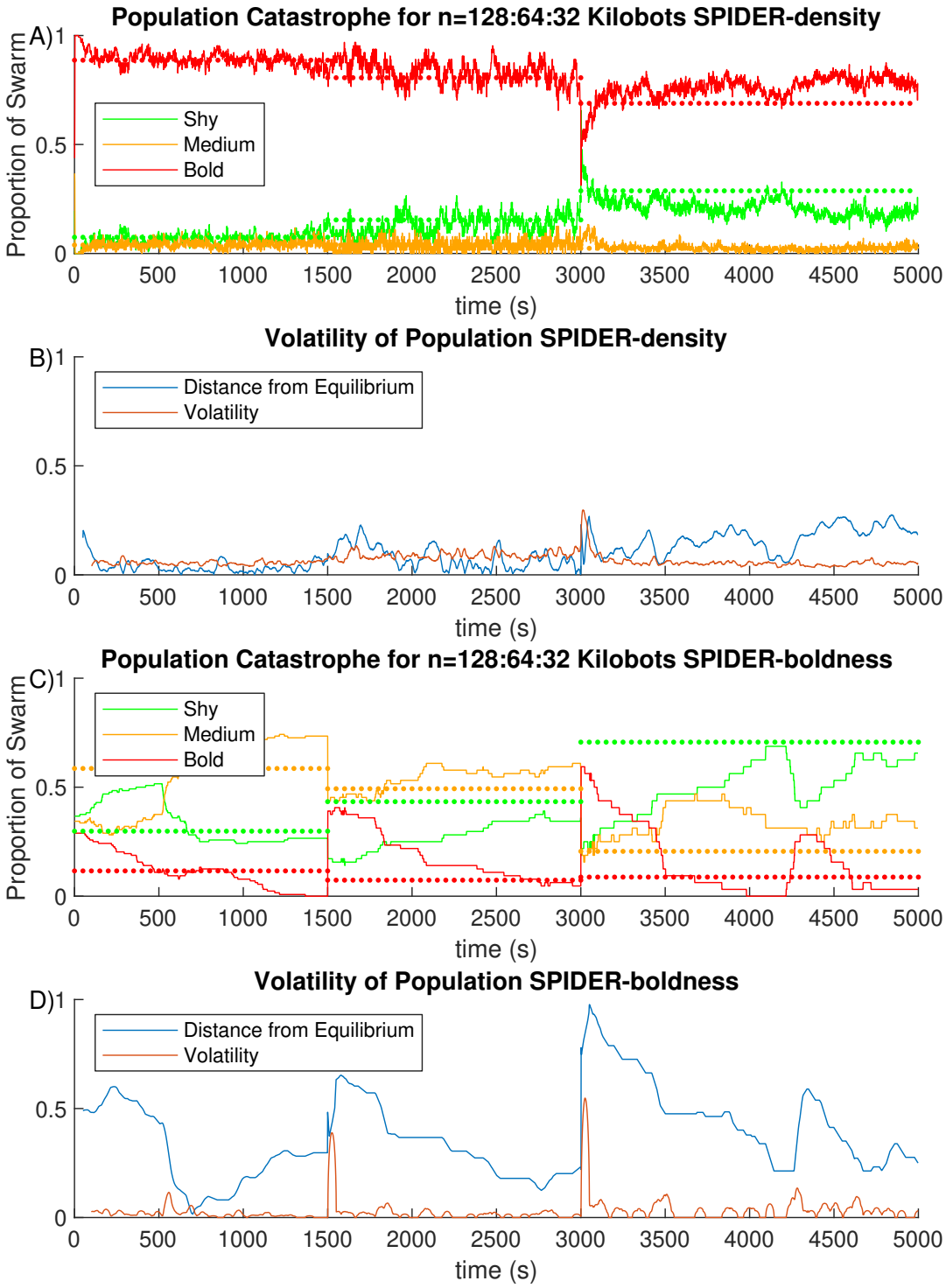


Figure 6.5: Example trials of the SPIDER-density (panes A and B) and SPIDER-boldness (C and D) controllers with a 'population catastrophe'. The population is halved twice from 128 to 64 to 32 robots at 1500s and 3000s. Dotted lines show the sub-population proportions towards which longer simulations converge.

Chapter 7

Conclusions and future work

Due to their low-cost, or even non-existent infrastructure, and their ability to adapt to a wide range of different applications ad-hoc networks are fast becoming a major part of modern day communication networks. Vehicular networks and swarm robotics are just two examples in which they have been used but are also highly important areas for future technologies. With autonomous vehicles fast becoming a reality with road tests and new advancements being announced all the time, the design and manufacture of the networks and infrastructure needed to support their development is becoming increasingly important. For large swathes of the public to begin using this technology, there also has to be wide-spread belief in the safety and efficiency of these systems. The work done in this thesis gives a mathematical framework for modelling and beginning to understand the behaviours of these networks using probabilistic tools. The work gives insight into how and why these networks may become disconnected, leading to important information not reaching its desired target.

Furthermore, as humans we are constantly trying to deepen our understanding of the world around us. Natural systems have evolved to be extremely efficient and bio-inspired computing aims to replicate these efficient behaviours to solve laborious tasks. Swarm robotics is one of the major tools in our arsenal to model these natural systems through its ability to accurately imitate large groups of living creatures. In this thesis we have outlined how the use of a decentralised algorithm can be used to imitate one such natural system in which social interactions between agents inflates their confidence and allows them to explore an environment more freely. The ability to mimic this behaviour allows for us to use these evolutionary advantages in human activities such as exploring unknown or dangerous environments.

In this chapter, we discuss the work done in this thesis with respect to its mathematical discoveries but also its applications to communication networks. We then discuss the future directions in which this work can be taken to further these discoveries.

7.1 1-D Soft Random Geometric Graphs

Although soft RGGs have been widely studied, and especially in the context of communication networks, the 1-Dimensional model has been widely overlooked. This is in part due to the technical difficulties involved in understanding the *uncrossed gaps*, but also in part due to the fact that many networks inhabit higher dimensional space. However, both as a mathematical object as well as in their application to vehicular networks, there are still myriad unsolved problems and new directions in which to take the work.

7.1.1 Future work: Mathematical directions

In Chapter 3 of this thesis, we began by conjecturing that the two key barriers to full connectivity are *isolated nodes* and *uncrossed gaps* and this conjecture was backed up by simulation data. We therefore began by studying isolated nodes in 1-D soft RGGs and showed that their distribution can be well approximated by that of a Poisson distribution. Along with the work done in Chapter 5 we showed that in the limit as the system size grows to infinity, at the point at which uncrossed gaps begin to appear, isolated nodes already exist within the network with high probability. This lead us to conjecture that isolated nodes are the main barrier to connectivity in this limit. However, there are still two key problems for which a deeper understanding is needed.

Firstly, we want to understand how important other modes of disconnection, the event that we have termed a *split*, are when discussing the full connectivity of the network. In Section 3.2 we put forward the conjecture that the two main obstructions to full connectivity are isolated nodes and uncrossed gaps. This is backed up by simulation data however no formal proof of this is given. One way to approach this would be to follow the same idea as in the higher dimensional version and show (as in [Pen16]) that in the limit as the system size grows to infinity, either a vertex is part of the giant component, or it is an isolated vertex. Unfortunately, a direct analysis of all of the ways in which a network could *split* seems to be out of reach.

Secondly, we want to understand the behaviour of the graph for finite system sizes, a very important factor in the analysis of this problem in the context of real-world communication networks. As was remarked at the end of Section 3.2 and then discussed in more detail at the end of Section 5.3, the question of whether isolated nodes or uncrossed gaps are more important with respect to full connectivity is not so clear for finite sized systems, and in particular it seems that for smaller systems with faster decaying connection functions that uncrossed gaps become more important. Therefore, to get a full understanding of the behaviour of these systems, it is of paramount importance to understand these uncrossed gaps. One way to do this is to follow the same idea as that of our analysis of the isolated nodes and try to calculate the probability of one existing in the network. If, as was the case for the isolated nodes, the number of uncrossed gaps can be well approximated by a Poisson distribution, then a good approximation of the probability of the absence of an uncrossed gap only requires calculating the expected number of uncrossed gaps in the network. In the 1-D hard RGG, it has been shown (e.g. [HM08]) that this Poisson approximation holds, and so there is no reason to believe that this would not be the case for the soft RGG. In Figure 7.1 the number of simulations in which an uncrossed gap occurs (circle marked P_{ucg} in the legend) is compared with $1-\exp(-n_{\text{ucg}})$ where n_{ucg} is the total number of uncrossed gaps summed over all simulations for one set of simulation parameters (i.e. an approximation of the expected number of uncrossed gaps). The expression $1-\exp(-n_{\text{ucg}})$ is used since this is the probability that a Poisson process of rate n_{ucg} is empty. Each plot is based on 5 000 simulations where the underlying point process is a unit rate (homogeneous) PPP. Figure 7.1a is for a line segment of length $L = 1\,000$ and Figure 7.1b for a line segment of length $L = 10\,000$. As can be seen in the figures, for both system sizes and for both connection functions considered (Waxman and Rayleigh), the Poisson distribution seems to be a good approximation for the distribution of the number of uncrossed gaps and hence we conjecture that

$$\mathbb{P}(N_{\text{ucg}} = 0) \approx 1 - \exp(-\mathbb{E}[N_{\text{ucg}}]),$$

where N_{ucg} is the number of uncrossed gaps in our network. Along with the conjecture that the two main obstructions to full connectivity are isolated nodes and uncrossed gaps, we are able to completely describe the probability of connectivity via the expectations of these two quantities. If full connectivity is defined only by the lack of isolated nodes and uncrossed gaps and these quantities can be well approximated

by Poisson distributions, then,

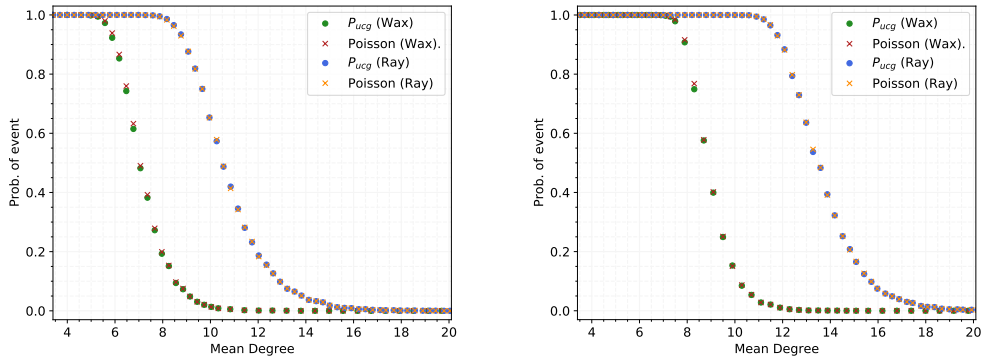
$$\begin{aligned} P_{\text{fc}} &\approx 1 - \mathbb{P}(N_{\text{iso}} = 0)\mathbb{P}(N_{\text{ucg}} = 0) \\ &\approx 1 - \exp\left(\mathbb{E}[N_{\text{iso}}] + \mathbb{E}[N_{\text{ucg}}]\right). \end{aligned}$$

Thus, the future mathematical work in this area should be focused on two main ideas: understanding the behaviour of the uncrossed gaps in the network, and looking at different point distributions and mobility. Uncrossed gaps are very technically difficult to analyse directly as was shown in Chapter 5 in which we were only able to derive a bound on the expectation of the number of uncrossed gaps. In this chapter, the generalised Rayleigh connection function was analysed, and the critical scaling for the emergence of uncrossed gaps was investigated. It was hypothesised that for a 1-D soft RGG on the line segment $[0, L]$, uncrossed gaps appear when the connection function scales as

$$R_L^{ucg} = \frac{\ln L}{C(\ln \ln L)^{\theta}},$$

where C is a constant that depends in a complicated way on the parameters of the connection function and θ is related to the path loss exponent, η. It was also noted that this constant C gives information about the system sizes for which isolated nodes or uncrossed gaps are more important for studying the connectivity of the network. Therefore, studying the quantity C will give insights into network connectivity.

The second direction is to look at different point distributions. It has been shown that when traffic is in its free-flow state that the distributions of the vehicles in the network can be well approximated by a Poisson distribution [RPM04], however, this is not always the most accurate way to model these networks. Other point processes such as the Matérn hard-core point process (as introduced in Chapter 2) have been used to model vehicular networks [TLH+16] along with Cox processes [CD18]. An interesting question to ask would be how much of the insight gained from investigating the PPP still holds in these more complex point processes. The analysis given in this thesis assumes that the network is being looked at at a snapshot in time and hence mobility is ignored. Applying the ideas of this work to different point processes, as well as incorporating mobility into the models, will allow the use of this research in real-world studies of vehicular networks.



(a) $L = 1000$ (b) $L = 10000$

Figure 7.1: The proportion of simulations for which an uncrossed gap was present, P_{ucg}, compared with $1 - \exp(-n_{\text{ucg}})$ where n_{ucg} is the total number of uncrossed gaps summed over all simulations for one set of simulation parameters. The comparisons are done for a Waxman connection function (green circles and red crosses) and a Rayleigh connection function (blue circles and orange crosses) and the underlying point processes are unit rate PPPs on a line segment of length (a) $L = 1000$ and (b) $L = 10000$.

7.1.2 Results: Communication Networks

Within the context of real-world vehicular communication networks, the size of the network is likely to be on the scale of 10s, or 100s (or possibly 1000s depending on how the infrastructure is designed). This is much smaller than the scales simulated in this thesis, where the minimum simulated network size was around 1000 nodes. Having not done much work on these smaller scales, it is difficult to comment on the behaviour of the systems, however we can make educated guesses based on the results we have seen. It seems that as we look at smaller network sizes, the importance of uncrossed gaps becomes more prevalent (with this even being evident in large network simulations where $\eta \geq 2$). Therefore, as mentioned, an understanding of uncrossed gaps in soft RGGs will play a key role in analysing connectivity of these networks. It also becomes unclear how things like Poisson approximations will be affected by the smaller system size as in general these only hold in the limit of the system size tending to infinity. But at these smaller scales it is also easier to do brute force calculations due to the smaller number of nodes. Alongside this, the effects of the boundaries will also become more important due to the larger proportion of nodes being "close" to the boundary. Whilst a highly interesting topic, unfortunately this will have to be left as future work as discussed in the following section.

7.1.3 Future work: Communication Networks

The main aim of this work was to study full connectivity in communication networks to understand where issues may arise in their design. The locations of road-side units (RSUs) will be paramount to maintaining a robust and well-connected network and the findings of this work will aid in this via an understanding of what causes disconnection in these networks. However full connectivity of the network may not always be possible, and may also not always be necessary. One area of research currently being considered in this field is assessing the $k-$hop connectivity between a vehicle and a RSU. The $k-$hop connectivity between two users is the minimum number of steps to get from one to the other via intermediate connected users. For example, if two users are directly connected, they are said to have 1-hop connectivity. This idea has been investigated in [KKL+20] in which the expected number of vehicles with 2-hop connectivity to a RSU under Rayleigh fading conditions is calculated.

Another future direction for this work could be to combine the investigations into full connectivity to studies on how information propagates through VANETs. Understanding how information is spread through these networks poses some very interesting research questions: What effect does it have on the network if some users have latent information? How long can a vehicle be disconnected from the network and still act efficiently? How does the connectivity of the network affect the routing of information? These types of questions are very important to understand for future applications in which hybrid systems of both autonomous (self-driving cars) and non-autonomous vehicles will combine.

7.2 Swarm Robotics

Swarms of robots are becoming a very widely used tool in the study of communication networks, especially in environments in which there is a lack of infrastructure and even more so if this environment is too dangerous for humans to explore. They offer a low cost way to explore unknown areas and carry out tasks previously impeded by the high costs of the robots used. One such example of a low cost robot is the Kilobot upon which the simulations examined in this thesis were based. The work presented here was solely evaluated through simulations performed in the Kilobox simulator. It was shown in Chapter 6 that the SPIDER-boldness algorithm described significantly improves the swarm's ability to reach from a base into a potentially hazardous distant area with suitable levels of connectivity. The obvious next step would

be to implement this code on a real Kilobot as a way of validating the effectiveness of the boldness mechanisms. Equally, Kilobots are very simple robots and the SPIDER algorithm could also be tested on much more complex robots with more sophisticated navigational and localisation abilities. It would also be of interest to deploy the SPIDER algorithm in real-world field trials in hazardous environments, testing the effectiveness of a density-based risk appetite controller for optimising the connectivity–reach trade-off. When in adversarial or disaster response scenarios where it may be too dangerous for humans, or there is a high risk of loss or destruction of robots, this decentralised swarm approach is of particular interest. Therefore, testing their performance in real-world trials will be highly important.

7.3 Closing Remarks

Full connectivity in ad-hoc networks is a highly desirable property, especially for safety critical cases such as VANETs, and is often seen as a pre-requisite of these systems. Random Geometric Graphs, and especially soft RGGs, allow for the in-depth study of these networks without the costs associated with physically building and testing them.

This thesis gives an in depth study of the properties of the 1-D soft RGG and discusses its use as a tool for modelling vehicular networks to aid with the design and implementation of this exciting technology. As well as being of interest to the engineering world, the 1-D soft RGG inhabits a rare space in the mathematical world since it appears to be a more complicated object than its higher dimensional counterparts. It therefore also offers a challenge to mathematicians to understand its behaviours and this thesis leaves many open questions for the interested reader. Along with the promising work carried out to test a bio-inspired algorithm, we have ventured into the world of autonomous systems: a crucial instrument for the future of smart cities.

The future world of smart cities and interconnected devices offers incredible flexibility and the potential for far reaching applications, but also needs as many tools as possible at its disposal to plan and design these networks. Stochastic Geometry, and RGGs in particular, offer a great mathematical toolbox to be used for this purpose and also to understand the limitations and, more importantly, the great potential of ad-hoc networks.

Bibliography

[ANB11] VK Muhammed Ajeer, PC Neelakantan, and AV Babu. "Network connectivity of one-dimensional vehicular ad hoc network". *2011 International Conference on Communications and Signal Processing*. IEEE. 2011, pp. 241–245.

[AOE13] Nabeel Akhtar, Oznur Ozkasap, and Sinem Coleri Ergen. "VANET topology characteristics under realistic mobility and channel models". *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2013, pp. 1774–1779.

[ADC+15] Mani Amoozadeh, Hui Deng, Chen-Nee Chuah, H Michael Zhang, and Dipak Ghosal. "Platoon management with cooperative adaptive cruise control enabled by VANET". *Vehicular communications* vol. 2, no. 2, (2015), pp. 110–123.

[AR02] Martin JB Appel and Ralph P Russo. "The connectivity of a graph on uniform points on $[0,1]^d$". *Statistics & Probability Letters* vol. 60, no. 4, (2002), pp. 351–357.

[AGG+89] Richard Arratia, Larry Goldstein, Louis Gordon, et al. "Two moments suffice for Poisson approximations: the Chen-Stein method". *The Annals of Probability* vol. 17, no. 1, (1989), pp. 9–25.

[APR+18] Fidel Aznar, Mar Pujol, Ramon Rizo, Francisco A Pujol, and Carlos Rizo. "Energy-efficient swarm behavior for indoor UAV ad-hoc network deployment". *Symmetry* vol. 10, no. 11, (2018), p. 632.

[BF96] Gutti Jogesh Babu and Eric D Feigelson. "Spatial point processes in astronomy". *Journal of statistical planning and inference* vol. 50, no. 3, (1996), pp. 311–326.

[BB10] François Baccelli and Bartłomiej Błaszczyszyn. *Stochastic geometry and wireless networks*. Vol. 1. Now Publishers Inc, 2010.

[BBS07] Adrian Baddeley, Imre Bárány, and Rolf Schneider. "Spatial point processes and their applications". *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004* (2007), pp. 1–75.

[BGN17] M Bakhshipour, M Jabbari Ghadi, and F Namdari. "Swarm robotics search & rescue: A novel artificial intelligence-inspired optimization approach". *Applied Soft Computing* vol. 57, (2017), pp. 708–726.

[BBS+07] Paul Balister, Bela Bollobas, Amites Sarkar, and Santosh Kumar. "Reliable density estimates for coverage and connectivity in thin strips of finite length". *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. 2007, pp. 75–86.

[BSB08] Paul Balister, Amites Sarkar, and Béla Bollobás. "Percolation, connectivity, coverage and colouring of random geometric graphs". *Handbook of large-scale random networks*. Springer, 2008, pp. 117–142.

[BHJ92] Andrew D Barbour, Lars Holst, and Svante Janson. *Poisson approximation*. Vol. 2. The Clarendon Press Oxford University Press, 1992.

[Bar11] Marc Barthélemy. "Spatial networks". *Physics Reports* vol. 499, no. 1-3, (2011), pp. 1–101.

[BZG18] Danielle S Bassett, Perry Zurn, and Joshua I Gold. "On the nature and use of models in network neuroscience". *Nature Reviews Neuroscience* vol. 19, no. 9, (2018), pp. 566–578.

[Ben88] Gerardo Beni. "The concept of cellular robotic system". *Proceedings 1988 IEEE International Symposium on Intelligent Control*. IEEE Computer Society. 1988, pp. 57–58.

[Ben04] Gerardo Beni. "From swarm intelligence to swarm robotics". *International Workshop on Swarm Robotics*. Springer. 2004, pp. 1–9.

[BR14] O. Bénichou and S. Redner. "Depletion-controlled starvation of a diffusing forager". *Physical Review Letters* vol. 113, no. 23, (2014).

[BRS03] Christian Bettstetter, Giovanni Resta, and Paolo Santi. "The node distribution of the random waypoint mobility model for wireless ad hoc networks". *IEEE Transactions on Mobile Computing* vol. 2, no. 3, (2003), pp. 257–269.

[BCB+07] T. Bilde, K. S. Coates, K. Birkhofer, T. Bird, A. A. Maklakov, Y. Lubin, and L. Avilés. "Survival benefits select for group living in a social spider despite reproductive costs". *Journal of Evolutionary Biology* vol. 20, no. 6, (2007), pp. 2412–2426. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2007.01407.x`.

[Bol01] Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[BFB+13] Manuele Brambilla, Eliseo Ferrante, Mauro Birattari, and Marco Dorigo. "Swarm robotics: a review from the swarm engineering perspective". *Swarm Intelligence* vol. 7, no. 1, (2013), pp. 1–41.

[BKB+13] James von Brecht, Theodore Kolokolnikov, Andrea L Bertozzi, and Hui Sun. "Swarming on random graphs". *Journal of Statistical Physics* vol. 151, no. 1-2, (2013), pp. 150–173.

[BSB15] James von Brecht, Benny Sudakov, and Andrea L Bertozzi. "Swarming on random graphs II". *Journal of Statistical Physics* vol. 158, no. 3, (2015), pp. 699–734.

[CA12] Neelakantan Chandrasekharamenon and Babu Ancharev. "Connectivity analysis of one-dimensional vehicular ad hoc networks in fading channels". *EURASIP Journal on Wireless Communications and Networking* vol. 2012, no. 1, (2012), p. 1.

[Che75] Louis HY Chen. "Poisson approximation for dependent trials". *The Annals of Probability* (1975), pp. 534–545.

[Che20] Vishnu Vardhan Chetlur Ravi. "Stochastic geometry for vehicular networks". PhD thesis. Virginia Tech, 2020.

[CD18] Vishnu Vardhan Chetlur and Harpreet S Dhillon. "Coverage analysis of a vehicular network modeled as Cox process driven by Poisson line process". *IEEE Transactions on Wireless Communications* vol. 17, no. 7, (2018), pp. 4401–4416.

[CSK+13] Sung Nok Chiu, Dietrich Stoyan, Wilfrid S Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[CDG12] Justin Coon, Carl P Dettmann, and Orestis Georgiou. "Full connectivity: Corners, edges and faces". *Journal of Statistical Physics* vol. 147, no. 4, (2012), pp. 758–778.

[DHM04] Sasha R. X. Dall, Alasdair I. Houston, and John M. McNamara. "The behavioural ecology of personality: consistent individual differences from an adaptive perspective". *Ecology Letters* vol. 7, no. 8, (2004), pp. 734–739.

[Dar53] Donald A Darling. "On a class of problems related to the random division of an interval". *The Annals of Mathematical Statistics* (1953), pp. 239–253.

[DG16] Carl P. Dettmann and Orestis Georgiou. "Random geometric graphs with general connection functions". *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* vol. 93, no. 3, (2016), pp. 1–16.

[DGP18] Carl P. Dettmann, Orestis Georgiou, and Pete Pratt. "Spatial networks with wireless applications". *Comptes Rendus Physique* vol. 19, no. 4, (2018), pp. 187–204.

[Dev81] Luc Devroye. "Laws of the iterated logarithm for order statistics of uniform spacings". *The Annals of Probability* vol. 9, no. 5, (1981), pp. 860–867.

[DF14] Luc Devroye and Nicolas Fraiman. "The random connection model on the torus". *Comb. Probab. Comput.* vol. 23, no. 5, (2014), pp. 796–804.

[DBS12] Paolo Di Lorenzo, Sergio Barbarossa, and Ali H Sayed. "Decentralized resource assignment in cognitive networks based on swarming mechanisms over random graphs". *IEEE Transactions on Signal Processing* vol. 60, no. 7, (2012), pp. 3755–3769.

[Dig00] Peter J Diggle. "Overview of statistical methods for disease mapping and its relationship to cluster detection". *Spatial Epidemiology: Methods and Applications* vol. 87, (2000), p. 103.

[Dom89] Cyril Domb. "Covering by random intervals and one-dimensional continuum percolation". *Journal of statistical physics* vol. 55, no. 1-2, (1989), pp. 441–460.

[DZM+09] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. "Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods". *European Physical Journal: Special Topics* vol. 174, no. 1, (2009), pp. 157–179.

[Dro97] Alon Drory. "Exact solution of a one-dimensional continuum percolation model". *Physical Review E* vol. 55, no. 4, (1997), p. 3878.

[EHH13] Hesham ElSawy, Ekram Hossain, and Martin Haenggi. "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey". *IEEE Communications Surveys & Tutorials* vol. 15, no. 3, (2013), pp. 996–1019.

[EBP+14] Richard Gilles Engoulou, Martine Bellaıche, Samuel Pierre, and Alejandro Quintero. "VANET security surveys". *Computer Communications* vol. 44, (2014), pp. 1–13.

[ER60] P Erdős and A Rényi. "On The Evolution Of Random Graphs". *Publ. Math. Inst. Hung. Acad. Sci* vol. 5, no. 1, (1960), pp. 17–60.

[EGK+04] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. "Modelling disease outbreaks in realistic urban social networks". *Nature* vol. 429, no. 6988, (2004), pp. 180–184.

[FM08] Dario Floreano and Claudio Mattiussi. *Bio-inspired artificial intelligence: theories, methods, and technologies*. MIT press, 2008.

[FL04] Chuan Heng Foh and Bu Sung Lee. "A closed form network connectivity formula one-dimensional MANETs". *2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577)*. Vol. 6. IEEE. 2004, pp. 3739–3742.

[FCH+20] Bradley Fraser, Andrew Coyle, Robert Hunjet, and Claudia Szabo. "An analytic latency model for a next-hop data-ferrying swarm on random geometric graphs". *IEEE Access* vol. 8, (2020), pp. 48929–48942.

[FN88] Toshio Fukuda and Seiya Nakagawa. "Approach to the dynamically reconfigurable robotic system". *Journal of Intelligent and Robotic Systems* vol. 1, no. 1, (1988), pp. 55–72.

[GB18] Lorenzo Garattoni and Mauro Birattari. "Autonomous task sequencing in a robot swarm". *Science Robotics* vol. 3, no. 20, (2018), pp. 1–12.

[GDC14] Orestis Georgiou, Carl P. Dettmann, and Justin P. Coon. "Network connectivity: Stochastic vs. deterministic wireless channels". *2014 IEEE International Conference on Communications, ICC 2014*. IEEE, 2014, pp. 77–82.

[Gil61] Edward N Gilbert. "Random plane networks". *Journal of the Society for Industrial and Applied Mathematics* vol. 9, no. 4, (1961), pp. 533–543.

[GJ96] Erhard Godehardt and Jerzy Jaworski. "On the connectivity of a random interval graph". *Random Structures & Algorithms* vol. 9, no. 1-2, (1996), pp. 137–161.

[GK99] Piyush Gupta and Panganamala R Kumar. "Critical power for asymptotic connectivity in wireless networks". *Stochastic analysis, control, optimization and applications*. Springer, 1999, pp. 547–566.

[Hae12] Martin Haenggi. *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[HSS08] Aric Hagberg, Pieter Swart, and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[Han07] Guang Han. "Connectivity analysis of wireless ad-hoc networks". PhD thesis. University of Maryland, 2007.

[HM06] Guang Han and Armand M Makowski. "Very sharp transitions in one-dimensional MANETs". *2006 IEEE International Conference on Communications*. Vol. 1. IEEE. 2006, pp. 217–222.

[HM07] Guang Han and Armand M Makowski. "A very strong zero-one law for connectivity in one-dimensional geometric random graphs". *IEEE Communications Letters* vol. 11, no. 1, (2007), pp. 55–57.

[HM08] Guang Han and Armand M Makowski. "Connectivity in one-dimensional geometric random graphs: Poisson approximations, zero-one laws and phase transitions". *Unpublished manuscript* (2008).

[HSB+17] Hamssa Hasrouny, Abed Ellatif Samhat, Carole Bassil, and Anis Laouiti. "VANet security challenges and solutions: A survey". *Vehicular Communications* vol. 7, (2017), pp. 7–20.

[HMK+14] Sabine Hauert, Sara Mitri, Laurent Keller, and Dario Floreano. "Evolving cooperation: from biology to engineering". *The Horizons of Evolutionary Robotics*. Ed. by Patricia A Vargas, Ezequiel A Di Paolo, Inman Harvey, and Phil Husbands. MIT press, 2014.

[HWZ+08] Sabine Hauert, Laurent Winkler, Jean Christophe Zufferey, and Dario Floreano. "Ant-based swarming with positionless micro air vehicles for communication relay". *Swarm Intelligence* vol. 2, no. 2-4, (2008), pp. 167–188.

[HZF09] Sabine Hauert, Jean-Christophe Zufferey, and Dario Floreano. "Evolved swarming without positioning information: An application in aerial communication relay". *Autonomous Robots* vol. 26, no. 1, (2009), pp. 21–32.

[HA14] D Helen and D Arivazhagan. "Applications, advantages and challenges of ad hoc networks". *Journal of Academia and Industrial Research (JAIR)* vol. 2, no. 8, (2014), pp. 453–457.

[HBS+21] Yassine Hmamouche, Mustapha Benjillali, Samir Saoudi, Halim Yanikomeroglu, and Marco Di Renzo. "New trends in stochastic geometry for wireless networks: A tutorial and survey". *Proceedings of the IEEE* (2021).

[HMD+04] Jeroen Hoebeke, Ingrid Moerman, Bart Dhoedt, and Piet Demeester. "An overview of mobile ad hoc networks: applications and challenges". *Journal-Communications Network* vol. 3, no. 3, (2004), pp. 60–66.

[HMS+05] Wei-jen Hsu, Kashyap Merchant, Haw-wei Shu, Chih-hsin Hsu, and Ahmed Helmy. "Weighted waypoint mobility model and its impact on ad hoc networks". *ACM SIGMOBILE Mobile Computing and Communications Review* vol. 9, no. 1, (2005), pp. 59–63.

[HJW+20] Edmund R Hunt, George Jenkinson, Michael Wilsher, Carl P Dettmann, and Sabine Hauert. "SPIDER: a bioinspired swarm algorithm for adaptive risk-taking". *Artificial Life Conference Proceedings*. MIT Press. 2020, pp. 44–51.

[HJH19] Edmund R Hunt, Simon Jones, and Sabine Hauert. "Testing the limits of pheromone stigmergy in high-density robot swarms". *Royal Society Open Science* vol. 6, (2019), p. 190225.

[HMF+18] Edmund R Hunt, Brian Mi, Camila Fernandez, Brandyn M Wong, Jonathan N Pruitt, and Noa Pinter-Wollman. "Social interactions shape individual and collective personality in social spiders". *Proceedings of the Royal Society B* vol. 285, no. 20181366, (2018), pp. 12–14.

[Hun20] Edmund Robert Hunt. "Phenotypic plasticity provides a bioinspiration framework for minimal field swarm robotics". *Frontiers in Robotics and AI* vol. 7, no. 23, (2020).

[HLV06] E. Hyytia, P. Lassila, and J. Virtamo. "A markovian waypoint mobility model with application to hotspot modeling". *2006 IEEE International Conference on Communications*. Vol. 3. 2006, pp. 979–986.

[Joh96] Don H Johnson. "Point process models of single-neuron discharges". *Journal of computational neuroscience* vol. 3, no. 4, (1996), pp. 275–299.

[JSH+18] Simon Jones, Matthew Studley, Sabine Hauert, and Alan Winfield. "Evolving behaviour trees for swarm robotics". *Proceedings of the 13th International Symposium on Distributed Autonomous Robotic Systems*. Springer International Publishing, 2018, pp. 487–501.

[KKL+20] Alexander P Kartun-Giles, Konstantinos Koufos, Xiao Lu, Nicolas Privault, and Dusit Niyato. "Two-hop connectivity to the roadside in a VANET under the random connection model". *arXiv preprint arXiv:2005.14407* (2020).

[KM02] James Kennedy and Rui Mendes. "Population structure and particle swarm performance". *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*. Vol. 2. IEEE. 2002, pp. 1671–1676.

[KKG+16] Georgie Knight, Alexander P Kartun-Giles, Orestis Georgiou, and Carl P Dettmann. "Counting geodesic paths in 1-D VANETs". *IEEE Wireless Communications Letters* vol. 6, no. 1, (2016), pp. 110–113.

[KB00] Michael J.B. Krieger and Jean Bernard Billeter. "Call of duty: Self-organized task allocation in a population of up to twelve mobile robots". *Robotics and Autonomous Systems* vol. 30, no. 1, (2000), pp. 65–84.

[LTH+92] Hernan Larralde, Paul Trunfio, Shlomo Havlin, H. Eugene Stanley, and George H. Weiss. "Territory covered by N diffusing particles". *Nature* vol. 355, no. 6359, (1992), pp. 423–426.

[LP17] Günter Last and Mathew Penrose. *Lectures on the Poisson process*. Vol. 7. Cambridge University Press, 2017.

[LIB+09] Richard Law, Janine Illian, David FRP Burslem, Georg Gratzer, CVS Gunatilleke, and IAUN Gunatilleke. "Ecological information from spatial patterns of plants: insights from point process theory". *Journal of Ecology* vol. 97, no. 4, (2009), pp. 616–628.

[LC08] Geunho Lee and Nak Young Chong. "A geometric approach to deploying robot swarms". *Annals of Mathematics and Artificial Intelligence* vol. 52, no. 2-4, (2008), pp. 257–280.

[Lév39] Paul Lévy. "On dividing a segment by points chosen at random". *CR Acad. Sci. Paris* vol. 208, (1939), pp. 147–149.

[LSB19] Guannan Li, Ivan Svogor, and Giovanni Beltrame. "Long-term pattern formation and maintenance for battery-powered robots". *Swarm Intelligence* vol. 13, no. 1, (2019), pp. 21–57.

[LLZ+08] Ming Li, Kejie Lu, Hua Zhu, Min Chen, Shiwen Mao, and B Prabhakaran. "Robot swarm communication networks: architectures, protocols, and applications". *2008 Third International Conference on Communications and Networking in China*. IEEE. 2008, pp. 162–166.

[LD15] Wei Lu and Marco Di Renzo. "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation". *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. MSWiM '15. Cancun, Mexico: Association for Computing Machinery, 2015, pp. 179–188.

[LR88] John A Ludwig and James F Reynolds. *Statistical ecology: a primer in methods and computing*. Vol. 1. John Wiley & Sons, 1988.

[MA11a] Guoqiang Mao and Brian DO Anderson. "On the asymptotic connectivity of random networks under the random connection model". *2011 Proceedings IEEE INFOCOM*. IEEE. 2011, pp. 631–639.

[MA11b] Guoqiang Mao and Brian DO Anderson. "Towards a better understanding of large-scale network models". *IEEE/ACM Trans. on Networking* vol. 20, no. 2, (2011), pp. 408–421.

[MA12] Guoqiang Mao and Brian DO Anderson. "Connectivity of large wireless networks under a general connection model". *IEEE transactions on information theory* vol. 59, no. 3, (2012), pp. 1761–1772.

[MA13] Guoqiang Mao and Brian DO Anderson. "Connectivity of large wireless networks under a general connection model". *IEEE Trans. on Info. Theory* vol. 59, no. 3, (2013), pp. 1761–1772.

[MZA10] Guoqiang Mao, Zijie Zhang, and Brian DO Anderson. "Probability of k-hop connection under random connection model". *IEEE communications letters* vol. 14, no. 11, (2010), pp. 1023–1025.

[Mat13] Bertil Matérn. *Spatial variation*. Vol. 36. Springer Science & Business Media, 2013.

[MWT+19] K. N. McGuire, C. De Wagter, K. Tuyls, H. J. Kappen, and G. C. H. E. de Croon. "Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment". *Science Robotics* vol. 4, no. 35, (2019).

[MR96] Ronald Meester and Rahul Roy. *Continuum Percolation*. Vol. 119. Cambridge University Press, 1996.

[MBH14] Mohamed Nidhal Mejri, Jalel Ben-Othman, and Mohamed Hamdi. "Survey on VANET security challenges and possible cryptographic solutions". *Vehicular Communications* vol. 1, no. 2, (2014), pp. 53–66.

[MH+14] Sebastian Meyer, Leonhard Held, et al. "Power-law models for infectious disease spread". *The Annals of Applied Statistics* vol. 8, no. 3, (2014), pp. 1612–1639.

[MSB+11] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. "Self-exciting point process modeling of crime". *Journal of the American Statistical Association* vol. 106, no. 493, (2011), pp. 100–108.

[NBK07] Huu Quynh Nguyen, François Baccelli, and Daniel Kofman. "A stochastic geometry analysis of dense IEEE 802.11 networks". *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*. IEEE. 2007, pp. 1199–1207.

[NVS+13] Vincenzo Nicosia, P. E. Vertes, William R Schafer, Vito Latora, and Edward T Bullmore. "Phase transition in the economically modeled growth of a cellular nervous system". *Proceedings of the National Academy of Sciences* vol. 110, no. 19, (2013), pp. 7880–7885.

[ND06] Shervin Nouyan and Marco Dorigo. "Chain based path formation in swarms of robots". *Ant Colony Optimization and Swarm Intelligence*. Springer Berlin Heidelberg, 2006, pp. 120–131.

[Pen03] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[Pen97] Mathew D Penrose. "The longest edge of the random minimal spanning tree". *The annals of applied probability* (1997), pp. 340–361.

[Pen16] Mathew D Penrose. "Connectivity of soft random geometric graphs". *The Annals of Applied Probability* vol. 26, no. 2, (2016), pp. 986–1028.

[Pra19] Pete Pratt. "Connectivity and mobility in wireless Networks". PhD thesis. University of Bristol, 2019.

[PSS+12] Rachel L Pullan, Hugh JW Sturrock, Ricardo J Soares Magalhaes, Archie CA Clements, and Simon J Brooker. "Spatial parasite ecology and epidemiology: a review of methods and applications". *Parasitology* vol. 139, no. 14, (2012), pp. 1870–1887.

[RBJ+14] Craig Robson, Stuart Barr, Philip James, and Alistair Ford. "Resilience of hierarchical critical infrastructure networks". *International Symposium for Next Generation Infrastructure Conference Proceedings*. UCL STEaPP, 2014, pp. 15–20.

[RPM04] Roger P Roess, Elena S Prassas, and William R McShane. *Traffic Engineering*. Prentice Hall, 2004.

[RAN12] Michael Rubenstein, Christian Ahler, and Radhika Nagpal. "Kilobot: A low cost scalable robot system for collective behaviors". *2012 IEEE International Conference on Robotics and Automation*. Computer Society Press of the IEEE, 2012.

[SD19] Chiranjib Saha and Harpreet S Dhillon. "Machine learning meets stochastic geometry: Determinantal subset selection for wireless networks". *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2019, pp. 1–6.

[Şah04] Erol Şahin. "Swarm robotics: From sources of inspiration to domains of application". *International workshop on swarm robotics*. Springer. 2004, pp. 10–20.

[SUS+20] Melanie Schranz, Martina Umlauft, Micha Sende, and Wilfried Elmenreich. "Swarm robotic behaviors and current applications". *Frontiers in Robotics and AI* vol. 7, (2020), p. 36.

[Sha81] D Shalitin. "Continuous percolation in one dimension". *Journal of Physics A: Mathematical and General* vol. 14, no. 8, (1981), p. 1983.

[Sha48] Claude E Shannon. "A mathematical theory of communication". *The Bell system technical journal* vol. 27, no. 3, (1948), pp. 379–423.

[Sos00] Alexander Soshnikov. "Determinantal random point fields". *Russian Mathematical Surveys* vol. 55, no. 5, (2000), p. 923.

[STN11] Valerio Sperati, Vito Trianni, and Stefano Nolfi. "Self-organised path formation in a swarm of robots". *Swarm Intelligence* vol. 5, no. 2, (2011), pp. 97–119.

[Suk37] Pandurang V Sukhatme. "Tests of significance for samples of the chi-square-population with two degrees of freedom". *Annals of Eugenics* vol. 8, no. 1, (1937), pp. 52–56.

[TLH+16] Zhen Tong, Hongsheng Lu, Martin Haenggi, and Christian Poellabauer. "A stochastic geometry approach to the modeling of DSRC for vehicular safety communication". *IEEE Transactions on Intelligent Transportation Systems* vol. 17, no. 5, (2016), pp. 1448–1458.

[TML+08] Yasser Toor, Paul Muhlethaler, Anis Laouiti, and Arnaud De La Fortelle. "Vehicle ad hoc networks: applications and related technical issues". *IEEE communications surveys & tutorials* vol. 10, no. 3, (2008), pp. 74–88.

[TSZ08] Salvatore Torquato, A Scardicchio, and Chase E Zachary. "Point processes in arbitrary dimension from fermionic gases, random matrix theory, and number theory". *Journal of Statistical Mechanics: Theory and Experiment* vol. 2008, no. 11, (2008), P11019.

[Wax88] Bernard M Waxman. "Routing of multipoint connections". *IEEE Journal on Selected Areas in Communications* vol. 6, no. 9, (1988), pp. 1617–1622.

[WDG20] Michael Wilsher, Carl P Dettmann, and Ayalvadi Ganesh. "Connectivity in one-dimensional soft random geometric graphs". *Physical Review E* vol. 102, no. 6, (2020), p. 062312.

[Wol02] Stephen Wolfram. *A New Kind of Science*. English. Wolfram Media, 2002.

[YOW08] Gongjun Yan, Stephan Olariu, and Michele C Weigle. "Providing VANET security through active position detection". *Computer communications* vol. 31, no. 12, (2008), pp. 2883–2897.

[YR14] Seung Kook Yun and Daniela Rus. "Adaptive coordinating construction of truss structures using distributed equal-mass partitioning". *IEEE Transactions on Robotics* vol. 30, no. 1, (2014), pp. 188–202.