UNIVERSITAT POLITÈCNICA DE CATALUNYA(UPC)
UNIVERSITAT DE BARCELONA(UB)
UNIVERSITAT ROVIRA I VIRGILI(URV)

MASTER OF ARTIFICIAL INTELLIGENCE THESIS

Spatial Networks, Thresholded Random Geometric Graphs, and Applications in Electric Vehicle Infrastructure Networks

---

Author:
Cole MacLean

Supervisors:
Sergio Gomez
Joan T Matamalas
Dept. Enginyeria Informàtica
i Matemàtiques
Universitat Rovira i Virgili

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Artificial Intelligence
in*

Complex Networks
Barcelona School of Informatics

June 16, 2017

UNIVERSITAT POLITÈCNICA DE CATALUNYA(UPC)
UNIVERSITAT DE BARCELONA(UB)
UNIVERSITAT ROVIRA i VIRGILI(URV)

Abstract

Artificial Intelligence

Barcelona School of Informatics

Master of Artificial Intelligence

Spatial Networks, Thresholded Random Geometric Graphs, and Applications in Electric Vehicle Infrastructure Networks

by Cole MacLean

The importance of spatial aspects in real-world complex networks has become widely studied in recent years, with a number of new models codifying spatial embeddings into their definitions and entire conferences dedicated to individual Spatial Network models *RGG Conference 2016*. In this paper, we explore the recent advances in Spatial Network modeling and provide 3 main contributions: 1. A new classification of existing Spatial Network models based on their shared parameterizations 2. The introduction of a new Spatial Network model - Thresholded Random Geometric Graphs to better capture the constraints of some real-world system 3. The real-world application and comparison of advanced Spatial Network models to the Electric Vehicle infrastructure network of Tesla's North American Supercharger stations. We show that our new model outperforms existing Spatial Network models in predicting the future structure of the Tesla Supercharger network, and claim that it is better suited for modeling networks of similar structure where a physically limiting maximum distance exists that prevents any connections past that distance, while maximizing some utility of the entire weighted network.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Goals

The motivation for this work is routed in my background as a Chemical Engineer. More concretely, as a *Process* Engineer, granted the power to wield and control complex, dynamic systems and mold them in such a way as to create real and valuable products out of apparent chaos. As any good engineer should, I'm constantly on the lookout for new tools, models and concepts to expand the types of systems I can understand and tame. This constant search led me into the discipline of Artificial Intelligence and is why I enrolled into the M.Sc. of Artificial Intelligence program at the UPC, which further led be into the world of Complex Networks, a domain of a near-infinite power to express almost any system. The malleability of Complex Networks, their successful ability to model real-world systems that I find interesting and compelling to study, and their intuitive formulation and structure make these models critical tools for any engineer's toolbelt.

My focus has always been towards the applied side of scientific study, only skimming the mathematical details and instead jumping straight to solving actual real-world problems, and that's how the core work of this paper began. I've long been a fan of Elon Musk and his many companies, especially Tesla, his electric car company. Tesla has a key differentiator, an electric car Supercharging network for long distance travel across continents. Being an avid member of the community, I began to see many people wonder when the Tesla Supercharger network would expand to their city and realized I could solve this with the skills I was developing in my Master's degree, setting off the exploration, results and analysis for the attempt to solve this problem as the main topic of this paper.

Using the techniques of Complex Networks, I originally probed the structure of the Tesla Supercharger Network in an extensive Exploratory Data Analysis. This work gave me the intuition needed to develop a predictive model using Utility Theory taught in a class I took on Multi-Criteria Decisions Support Systems, and managed to produce promising predictive results. Although seemingly impressive at the time, the resulting model has not been able to continue its performance now that the Supercharger network has grown considerably from the 200 at the time of that previous work to the 418 that exist today. Identifying that more robust tools are required to truly capture the growing structure of the Supercharger network, I've once again turned to Complex Networks in the hopes of finding the right hammer for the job. The models developed, experiments observed and theory discussed in this paper are all derived from this goal of discovering what is possible to predict about the future of the Supercharger network given its historical and current structure. This goal of prediction given a set of data is a shared theme across many of the courses in the M.Sc. of Artificial Intelligence program of the UPC, and I've drawn from my collective experience in this Masters to perform the work of this paper.

The ultimate goal of this paper is to develop a mode capable of capturing the underlying structure of the Tesla Supercharger network and provide the ability to predict the future structure. This main objective comes with 2 necessary sub-tasks that are also goals of this study, which are to identify and explore existing models that may be capable of achieving the main goal and the development of a new model and theory with the potential of outperforming existing models in this task. The following chapters are presented as the realization of these goals.

1.2 Organization

This paper has been organized into 7 separate chapters, including this brief introduction as Chapter 1. Chapter 2 provides the background in Complex Network theory needed and heavily relied upon in future chapters. Chapter 3 introduces the Spatial Networks sub-domain of Complex networks which is the main focus of this paper, and is where our new classification of these Spatial Network models can be found. Chapter 4 provides a discussion on some of the most exiting application of Spatial Network models and provides further motivation for their study. Chapter 5 introduces the new model presented in this paper, Thresholded Random Geometric Graphs and provides analytical and experimental conditions for connectivity of this new model. Chapter 6 is the realization of the final goal of this paper, providing a full analysis of the Tesla Supercharger network and shows the final results of predicting its future structure using the proposed model of this paper compared to existing models across multiple evaluation metrics. Chapter 7 concludes the work with a final discussion of the results and future work.

Chapter 2

Background

The domain of Complex Networks has been inspired by the study of real-world systems that can be represented by a set of nodes or vertices connected by edges with non-trivial structural properties. Systems taking the form of Complex Networks abound, including: the Internet, social networks, organizational networks, metabolic networks, food webs, roadways, infrastructure networks, brain connectivity and neural networks (Newman, 2003).

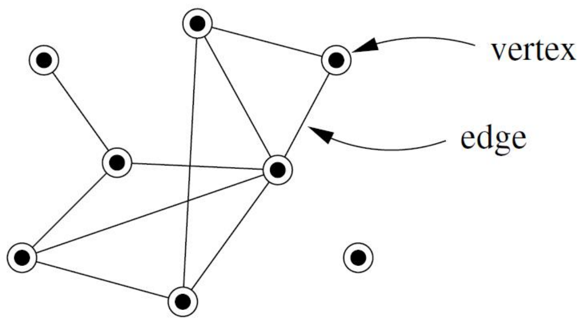

FIGURE 2.1: A small example network with eight vertices and ten edges Newman, 2003

The closely related and overlapping field of Random Graphs was initiated in 1959-1960 in "On the evolution of random graphs" 1960 with the definition of the Erdős-Rényi (ER) Random Graph, where a pair taken from N vertices are connected by an edge with probability, p.

In this section, we present an overview for the key concepts in Complex Networks and Random Graphs used throughout this paper.

2.1 Basic Definitions

Here we define some key definitions and concepts of various terms used in Complex Networks and this paper.

- Vertex or Nodes are the fundamental units of a networks and can be used to represent any entity and can have one or many attributes.
- Edges are the links that connect nodes and can be undirected or directed if there exists directionality in the representation of connected nodes
- Degree is count of the number of edges connected to a node
- A Component of the network is the set of nodes that can be reached from each other node by paths running along edges of the graph
- The Giant Component is the component of the network having the largest count of connected nodes

2.2 Degree Distributions and Scale-Free Networks

The key characteristic of degree for individual nodes in a network can be expanded to produce a distribution of degrees for all nodes in the network which is known as the Degree Distribution of the network. Formally, we can define the degree distribution by defining p_k to be the fraction of vertices in the network that have degree k. Equivalently, p_k is the probability that a vertex chosen uniformly at random has degree k. The histogram produced by the degree of each node is the degree distribution (Newman, 2003).

A networks degree distribution can be used as a property of comparison amongst many other networks and investigate similarities between real-world networks from diverse domains. Many real-world networks have been shown to have power-law degree distributions, which are often referred to as Scale-Free networks, where the degree, k of a node is drawn from a probability distribution, p_k parameterize by a power-law exponent, γ shown in Equation 2.1 (Barabási, 2009).

$$p_k \propto k^{-\gamma} \tag{2.1}$$

One of the most well studied models proposed to explain the growth of a network in producing power-law degree distributions is the Barabási-Albert Preferential Attachment model (Barabási and Albert, 1999). Known as "the rich get richer" mechanism, the model proposes a growth model of networks were the probability for a node, N to connect to a node, m that is already in the network increases with the degree of node m, resulting in power-law degree distributions for networks generated with this model.

2.3 Average Degree and Connectivity

The Average Degree of a network,$\langle k \rangle$, is defined as the summation of all degrees for each node,k_i, in the network divided by the total number of nodes in the network, N, as described in Equation 2.3.

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{2.2}$$

A network is said to be Connected when all N nodes of the network exist in the single component of the network. The percentage of nodes within the giant component of a network is an often used measure to describe how well connected a network is.

2.4 The Erdős-Rényi Random Graph

The Erdős-Rényi Random Graph is a simple model of a network where N number of nodes are taken and connect each pair (or not) with probability, p. Many properties of the Erdős-Rényi Random Graph are exactly solvable in the limit of large graph size, with the model having a Poisson degree distribution shown in Equation **??** (Newman, 2003)

$$p_k = \binom{N}{k} p^k (1-p)^{n-k} \tag{2.3}$$

The expected structure of the random graph varies with the value of the probability of connection parameter,p and the critical behavior of this parameter has been well studied.

2.5 Phase Transitions and Critical Phenomena

The study of Phase Transitions and Critical Phenomena in Complex Networks, where a sharp transition in a system property exists as a function of one of its parameters, has been a cornerstone of studies in Random Graphs and Complex Networks since the introduction of classical random graphs by Erdős and Rényi who described the structural phase transition of the emergence of a giant connected component (Erdős and Rényi, 1961). Erdős and Rényi demonstrated that the random graph possesses a phase transition, from a low-density, low-p state in which there are few edges and all components are small to a high-density, high-p state in which an extensive fraction of all vertices are joined together in a single giant component. This phase transition is depicted in Figure 2.2.

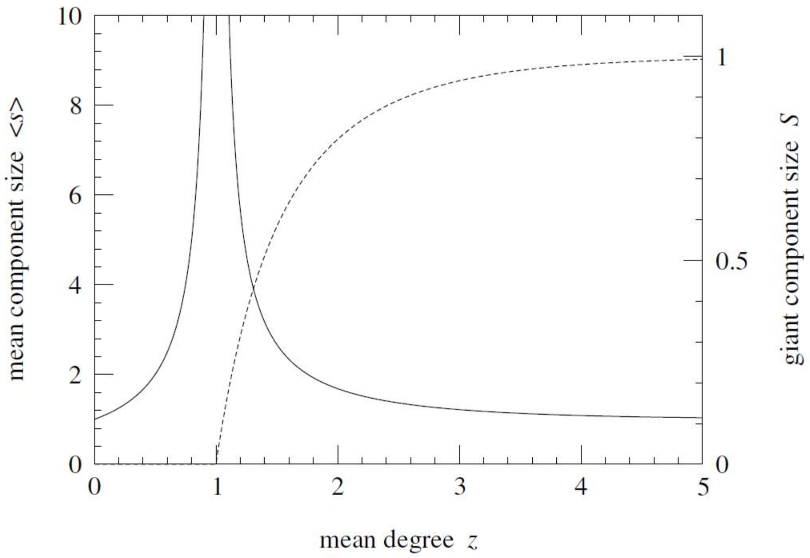

FIGURE 2.2: The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the ER random graph (Newman, 2003)

Critical phenomena in networks include a wide range of issues: structural phase transitions, the emergence of critical scale-free network architectures, various percolation phenomena, and epidemic thresholds (Dorogovtsev, Goltsev, and Mendes, 2008).

2.6 Critical Scaling and Strong vs Weak Zero-One Laws

For convenience we introduce the notation of $P(N, \Pi_N)$ to indicate the probability for any graph, G of size N parameterized by scaling Π to be connected.

A Scaling, Π is defined as any mapping $\Pi : \mathbb{N}_0 \to \mathbb{R}_+$, and can be used in finding conditions for the probability of being connected on such scalings to ensure either $\lim_{N\to\infty} P(N, \Pi) = 1$ or $\lim_{N\to\infty} P(N, \Pi) = 0$. Typically there exist scalings, deemed critical, which act as a boundary in the space of scalings between these two extremes (Makowski and Yagan, 2013). The terminology of Strong and Weak Zero-One laws was first developed by McColm for random graphs on a line segment (McColm, 2004) and adapted by Makowski for Random Threshold Graphs. A strong zero-one law is said to hold (for graph connectivity) with critical scaling Π^* if for any scaling Π satisfying:

$$\lim_{N\to\infty} \frac{\Pi_N}{\Pi_N^*} = c \tag{2.4}$$

for some c > 0, we have

$$\lim_{N\to\infty} P(N, \Pi_N) = \begin{cases} 1 & \text{if } 0 < c < 1 \\ 0 & \text{if } 1 < c \end{cases} \tag{2.5}$$

Any scaling Π^* satisfying Equations 2.4 and 2.5 is called a strong critical scaling.

Alternatively, a weak zero-one law is said to hold (for graph connectivity) with critical scaling Π^* if for any scaling Π we have:

$$\lim_{N\to\infty} P(N, \Pi_N) = \begin{cases} 1 & \text{if } \lim_{N\to\infty} \frac{\Pi_N}{\Pi_N^*} = 0 \\ 0 & \text{if } \lim_{N\to\infty} \frac{\Pi_N}{\Pi_N^*} = \infty \end{cases} \tag{2.6}$$

Any scaling Π^* satisfying Equation 2.6 is called a weak critical scaling.

Figure 2.3 depicts an example strong and weak critical scaling phenomena for the connectivity of Random Threshold Graphs for different weight distributions with the model parameter, θ being substituted for Π as the scaling parameter . The weak nature of the zero-one law in the right pane of 2.3 is evident from the figure since $P(N, c\theta_N^*) = 1$ only for 0 < c < 0.3, becoming close to zero only after c > 8 – Contrast this with the strong zero-one laws observed in the left pane (Makowski and Yagan, 2013).

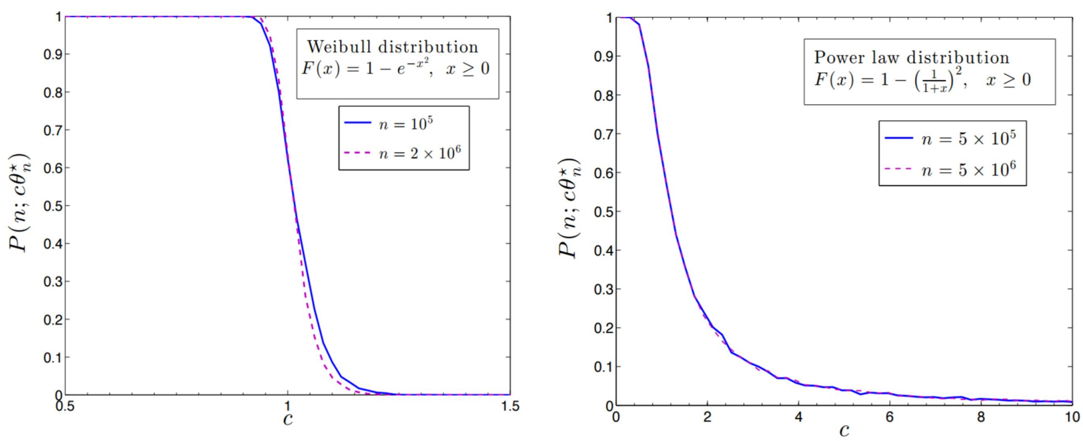

FIGURE 2.3: Example strong (left) and weak (right) zero-one thresholds from (Makowski and Yagan, 2013) for critical scaling of connectivity in Random Threshold Graphs

Chapter 3

Spatial Networks

A sub-domain of Complex Networks known as Spatial Networks attempts to better model real-world networks by allowing the addition of spatiality into the models, a consideration not present in conventional Random Graphs. Many real-world complex systems have spatial components constraining the network structures these types of systems can produce. Infrastructure networks such as transportation, electrical, and telecommunication systems, social networks and even our own synaptic networks are all embedded in physical space. Spatial Networks provide a framework for network models having spacial elements, where nodes are embedded in space and a metric is incorporated that influences the probability of connection between nodes. Typically, the probability of connection is a decreasing function of the metric, with most models assuming Euclidean distance in 2-dimensions or 3-dimensions. The intuition of most Spatial Network models propose that there exists an increasing cost of connection between nodes that are further apart, which is a likely element for most spatially embedded systems, such as infrastructure or biological networks. Figure 3.1 highlights the spatial embeddings of real networks.

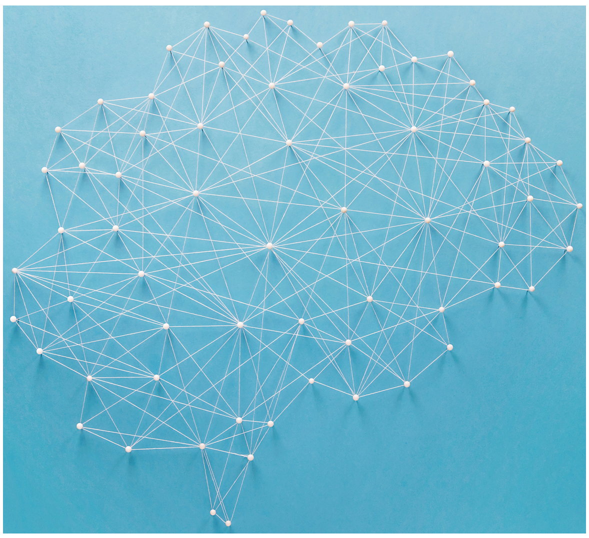

FIGURE 3.1: Artist rendering of the human brain, highlighting the spatial embedding of complex networks

3.1 Characteristics of Spatial Networks

Although there has been a recent surge in interest for Spatial Networks, likely caused by the availability of datasets for large networks and the computational power required to analyze them, Spatial Networks were actually long ago the subject of many studies in quantitative geography and many modern questions in the complex system field are actually at least 40 years old (Barthélemy, 2011). These past few decades have seen the development of many Spatial Network models capable of constructing a surprising variety of network structures with characteristics distinctly different than those of traditional random graphs. Interesting properties arise by the addition of spatial components to network models that are common across some of the most important classes of Spatial Network models.

The addition of spatial constraints limits the tendency for global hubs, where a few nodes are highly connected with far reaching connections across the network, a property utilized in the Barabási-Albert preferential attachment model used as a mechanism to explain the emergence of scale-free powerlaw degree distributions of many real-world networks (Barabási and Albert, 1999). However, many Spatial Network models have proofs for the conditions on the model that do generate scale-free powerlaw degree distributions even with the limiting hub generation tendency of Spatial Networks, such as in the introduction paper for the Geometric Threshold Graph model (Masuda, Miwa, and Konno, 2004). The usual underlying assumption of increasing connection cost with distance also favors the generation of cliques, causing large clustering coefficients in many Spatial Network models, as close nodes generally tend to become fully attached to each other. This tendency can also lead to large average shortest paths between nodes as local nodes are highly connected, but the allowance for long-range shortcuts in many Spatial Networks can recover the small average shortest path of small-world networks (Barthélemy, 2011). These unique properties of Spatial Networks, the ability to reconcile these models with empirical real-world network data and their physical realism make them an exciting and interesting domain of study.

3.2 Spatial Network Models

The potential application of Spatial Networks to such a wide variety of real-world systems has motivated substantial research into these networks, with many unique but closely related models being proposed with theoretical proofs for many of their network properties. The Spatial Networks review article by (Barthélemy, 2011) provides a comprehensive overview of the field and reviews many of the most important theoretical proofs for many Spatial Network models. Here we introduce and define some of the most common and recent Spatial Network models developed so far. We first define some terms and notation used in the Spatial Network models.

$P(X)$ - The spatial distribution dictating how the coordinates of the nodes are sample for placement onto the spatial embedding

R - The maximum connection distance

$P(d_{ij})$ - The probability of edge connection as a function of the distance, d_{ij}, between nodes i,j where i $\neq$ j

θ - The node weight threshold for connection

E_{ij} indicates an edges exists between nodes i and j.

3.2.1 Random Geometric Graphs (RGG) - R

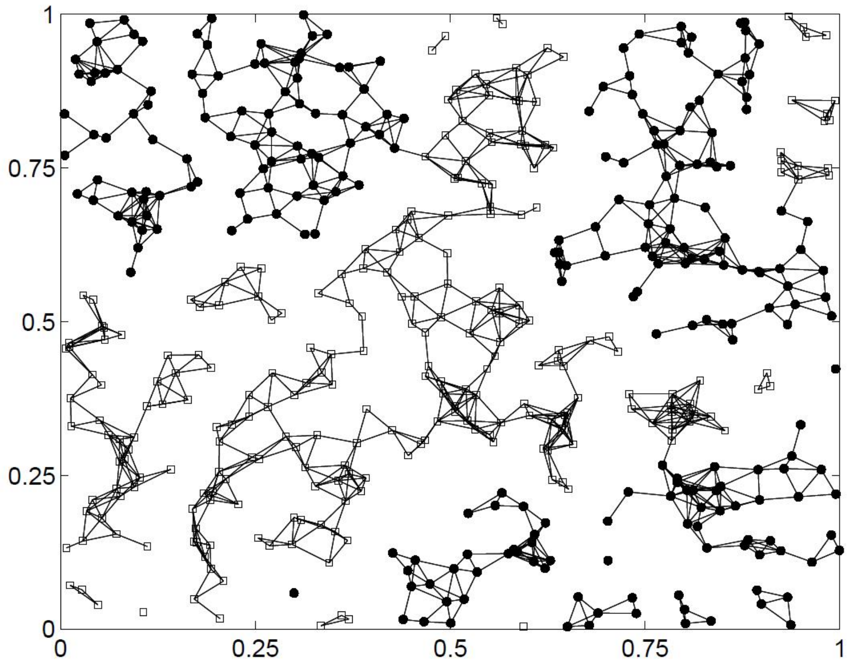

FIGURE 3.2: A 2D Random Geometric Graph with N = 500 and $\langle k \rangle$ = 5 (Dall and Christensen, 2002)

A d-dimensional Random Geometric Graph is a graph where each of the N nodes is assigned random coordinates in the box $[0, 1]^d$, and only nodes 'close' to each other are connected by an edge.(Dall and Christensen, 2002). Any node within or equal to the maximum connection distance, R, is a connected node and the structure of the network is fully defined by R. RGGs, similar to Unit Disk Graphs (Clark, Colbourn, and Johnson, 1990), have been widely used to model ad-hoc wireless networks (Nemeth and Vattay, 2003). An edge, E_{ij} exists according to Equation 3.1

$$E_{ij} = d_{ij} \leq R \tag{3.1}$$

One of the simplest and most well studied of the Spatial Networks, Random Geometric Graphs have found many real-world applications and entire conferences are now detected to this single Spatial Network model *RGG Conference 2016*.

3.2.2 Waxman Graphs - $P(d_{ij})$

Waxman Graphs are the spatial generalization of ER random graphs, where the probability of connection of nodes depends on a function of the distance between them (Waxman, 1988). The original edge probability function proposed by Waxman is exponential in d_{ij}, providing two connection probability tuning parameters, α and β provided in Equation 3.2.

$$E_{ij} = P(d_{ij}) = \beta e^{-d_{ij}/L\alpha} \tag{3.2}$$

Where L is the maximum distance between each pair of nodes. The shape of the edge probability function, P(d_{ij}), plays the key role in determining the structure of a Waxman graph.

3.2.3 Random Threshold Graphs (RTG) - θ

A simple graph G is a threshold graph if we can assign weights, drawn from a weight distribution, $f(w)$, to the vertices such that a pair of distinct vertices are connected exactly when the sum of their assigned weights is or exceeds a specified threshold, θ (Reilly and Scheinerman, 2009). Threshold Graphs are not themselves Spatial Networks, as they do not incorporate a specific geometry or metric, but they introduce the ability to consider node weights as part of the network model which is utilized by other Spatial Network models such as Geometric Threshold Graphs.

$$E_{ij} = (w_i + w_j) \geq \theta \tag{3.3}$$

3.2.4 Geometric Threshold Graphs (GTG) - $P(d_{ij})$, θ

Geometric Threshold Graphs are the geographical generalization of Random Threshold Graphs, where a pair of vertices with weights w_i, w_j drawn from a weight distribution $f(w)$, and a distance d_{ij} are connected if and only if the product between the sum of weights w_i and w_j with the edge connection function, $P(d_{ij})$, is greater than or equal to a threshold value, θ (Masuda, Miwa, and Konno, 2005).

$$E_{ij} = (w_i + w_j)P(d_{ij}) \geq \theta \tag{3.4}$$

3.2.5 Soft Random Geometric Graphs (SRGG) - R, $P(d_{ij})$

A recent extension of Random Geometric Graphs couples the influence of distance between nodes that are within the maximum connection distance, R, to better model real-world systems where node proximity does not necessarily guarantee a connection between 'close' nodes. In Soft Random Geometric Graphs, the probability of connection between nodes i and j is a function of their distance, d_{ij}, if $d_{ij} \leq R$. Otherwise, they are disconnected (Penrose, 2016).

$$E_{ij} = \begin{cases} P(d_{ij}) & \text{if } d_{ij} \leq R \\ 0 & \text{if } d_{ij} > R \end{cases} \tag{3.5}$$

3.3 Shared Model Parameter Classification

Here we investigate the relationships between the introduced Spatial Network models and propose a novel classification of these models by considering their shared parameterization. Specifically, we show that the Spatial Network models are obtained by combinations of only 3 parameters, R, $P(d_{ij})$ and θ plus the spatial distribution for node embedding, $P(X)$. Figure 3.3 shows the relationships between Spatial Network Models connected by their shared parameterizations and highlights with dashed lines the new model proposed in this paper.

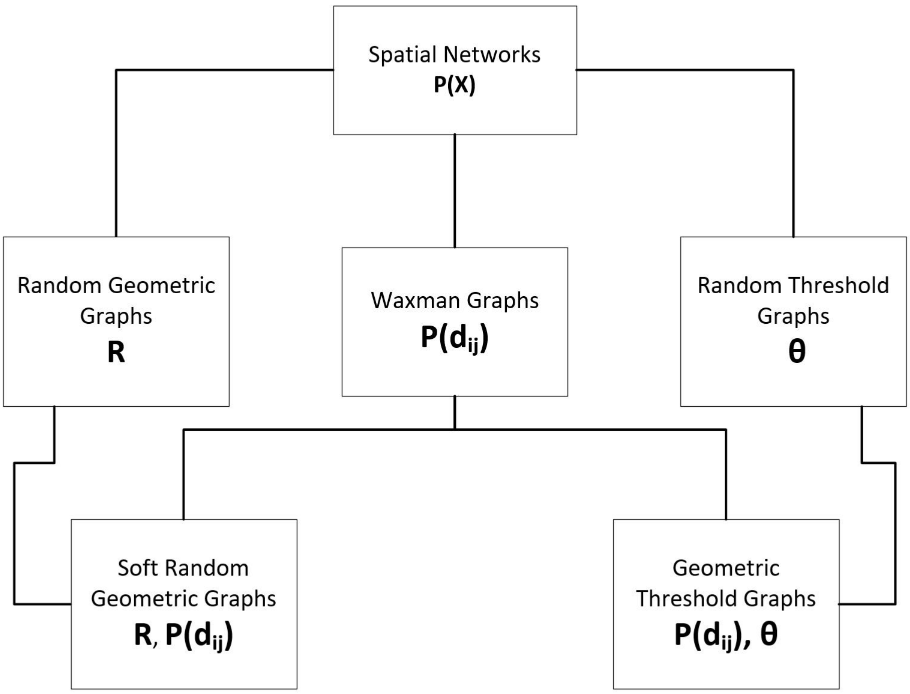

FIGURE 3.3: Spatial Network Model Classification

The Spatial Network classification of Figure 3.3 provides researchers an intuitive introduction into the field of Spatial Networks and highlights patterns in their formulations. It also presents the identification of potential gaps in current model formulations, including the consideration for a model combining the maximum distance parameter, R with the influence of connection thresholds, θ, which is the parameterization for the model introduced in this paper in Chapter 5.

Chapter 4

Applications of Spatial Networks

With the definitions of common Spatial Networks enumerated, we explore the current and most successful applications of these models to real-world systems, highlighting the motivation for studying these useful mathematical constructs and the need for further development of these models.

4.1 Transportation and Infrastructure Networks

We start with one of the most commonly studied class of networks, and the category for the motivating example of this paper discussed in Section 6.1 are Transportation and Infrastructure networks, where junctions are dotted around towns, cities, countries or even globally and are physically connected to provide some means for the transfer of atoms, bits or people from one place to another. The spatial nature of Infrastructure Networks is immediately evident by their definition of connecting distant locations, and the underlying assumption of many Spatial Network models where the cost of connection increases is intuitive in many real-world Transportation Networks, such as the cost of building roads or train lines in highway and rail networks.

The classic way of representing a transportation networks and the construction of their topological structures was discussed by Kurant and Thiran, who proposed a simple representation of considering the stations as nodes and the connections as the edges of the network (Kurant and Thiran, 2006). A well studied example using this representation is that of Airline networks, where the nodes are airports and the links are the different direct air traffic routes provided by particular airlines. Airline networks present interesting spatial aspects, as they have unique local national structure, but have grown to international scales, which have been shown to be scale-free structures but with anomalous centrality caused by more than geographic constraints alone (Guimera et al., 2005).

Another common example for the use of Spatial Networks in transportation systems is their application in the study and development of road and highway networks.

Recent empirical studies have discovered surprising similarities in road networks between different cities. In (Lämmer, Gehlsen, and Helbing, 2006), the others discover powerlaw distribution in the road networks of Germans 20 largest cities. In (Cardillo et al., 2006), the authors use Spatial Networks to analyze the structural properties of 21 world city and were able to classify that cities of the same class, ex. grid-iron or medieval, exhibit roughly similar properties. In (Crucitti, Latora, and Porta, 2006), the authors use Spatial Networks to provide extended visualization and characterization of the city structure and propose the interesting results that self-organized cities exhibit scale-free properties similar to those found in non-spatial networks, while planned cities do not. The application of Spatial Networks in urban and road design has been one of the most notorious use cases which is likely to continue and expand.

Only two of the well studied classes of Transportation and Infrastructure Networks using Spatial Networks have been discussed here, but many others including subway and metropolitan public transport networks, rail, electric and utility grids and the countless other systems that makeup the modern world are using the models and theory discussed in this paper.

4.2 Neuroscience

FIGURE 4.1: Networks of the Brain Textbook Cover (Sporns, 2010)

Governments, research institutions and private companies are spending billions to study and understand the human brain, with the hope of discovering the biological mechanisms that contribute to devastating brain diseases, inspire new computational paradigms and open the door to technologies only dreamed of in science fiction, such as digital-brain Neural Lace interfaces (*Elon Musk enters the world of* *brain-computer interfaces*). The rewards of these research efforts are sure to revolutionize humanity, but the humbling complexity of the brain created over evolutionary timescales, puts this project beyond some of the most challenging faced in our history, and the scope of the problem continues to grow as we discover new complexities at every layer. Only recently, studies have shown that the accepted dogma that every cell in an individual has the same DNA may not be true, with every neuron potentially having a unique mutation, which has the potential to multiply our current estimates for the complexity of the human brain by orders of magnitude (*Scientists Surprised to Find No Two Neurons Are Genetically Alike*), inspiration and tools from a wide variety of disciplines. One of these tools that is gaining popularity in the field of neuroscience is the application of Spatial Networks in understanding the Complex Network structures of the human brain. Figure 4.2 highlights some of this complexity and the Spatial Network dimension that is the focus of this paper.

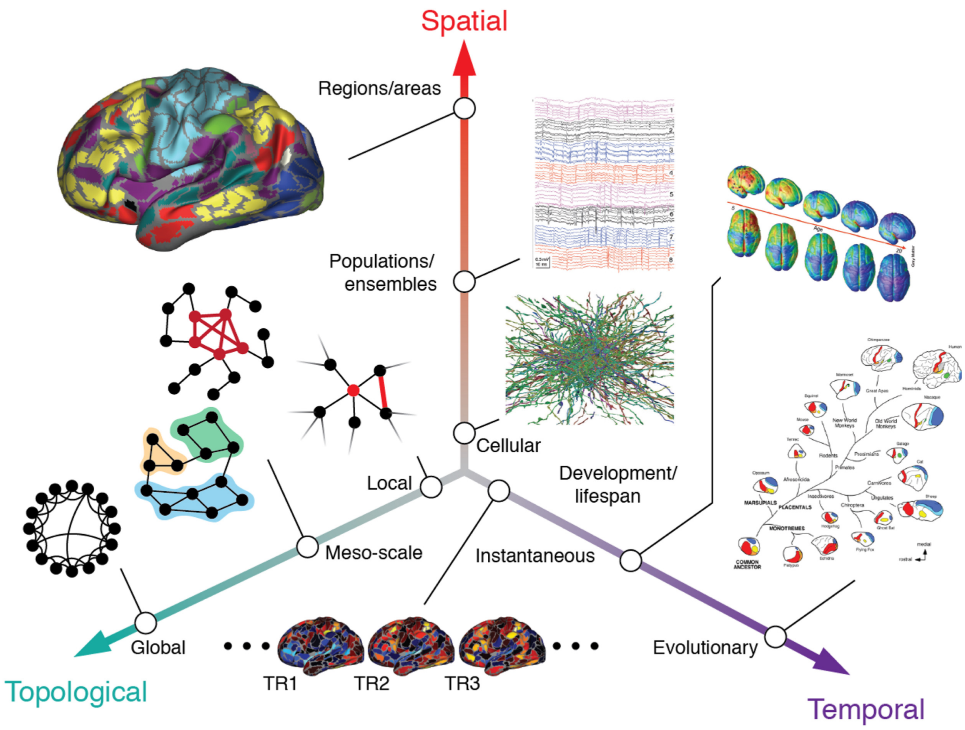

FIGURE 4.2: The multi-scale brain (Betzel and Bassett, 2016)

The application of Complex Network theory to neuroscience has been around since humans first began studying the brain, with the first clear, recognizably scientific representations of the human brain being prototype anatomical maps (Papo et al., 2014). Recent successes of using Complex Networks to understand the brain include identifying rich club neurons of locomotor circuits in Caenorhabditis elegans, indicating the importance of these specific neurons and suggesting that this may be a general and scale-invariant principle of brain network organization (Towlson et al., 2013), topological revelations of hierarchically organized structural systems in the functional connectivity of the brain (Stephan et al., 2000) and defining the community structure of the brain connectome (Reus et al., 2014).

This model of connection probability as a function of distance inherent in many Spatial Network models is equivalent to the "wiring-cost" model of brain connectivity, an extensively studied optimization model of neuronal networks (Bullmore and Sporns, 2012). Recent studies have shown the importance of considering spatial relations between cortical regions of the brain, and have used spatial patterns to discover similarities between monkey, human and mouse cortex (Song, Kennedy, and Wang, 2014).

Being the simplest model, Random Geometric Graphs have found many applications in neuroscience. In (O'Dea, Crofts, and Kaiser, 2013), the use of the RGG model is applied to detailed neuro-imaging data to define a network whose spatial embedding represents accurately the folded structure of the cortical surface of a rat brain and investigate the propagation of activity over this network under simple spreading and connectivity rules and conclude that studies which omit physiological network structure risk simplifying the dynamics in a potentially significant way, highlighting the importance of Spatial Networks as models of brain functions. Another paper comparing the ability of different random network models in explaining the functional brain network structure of an fMRI dataset composed of 908 individuals diagnosed with autism and Asperger, showed the RGG model's ability to competitively discriminate the real network structure compared to traditional random network models (Fujita, Vidal, and Takahashi, 2017). Other studies have shown the high clustering coefficient of structural brain connectivity that can be at least partly attributed to the spatial embedding of the brain and which RGGs can better describe than traditional random network models (Gastner and Ódor, 2016). All of these papers have been published in the past 5 years, and there has been an acceleration in papers describing the importance of spatial embedding in network models of brain connectivity.

The excellent review article by Bullmore and Sporns, (Bullmore and Sporns, 2012) the authors detail the high cost of long neural links in different brain regions, and evaluate hypothetical benefits of creating these links. This type of long distance spatial consideration in link connection is precisely the underlying assumption of many Spatial Network models. Two closely related papers inspired by the work of Bullmore and Sporn, discover the so called "Exponential Distance Rule" (EDR), where neuronal connections exhibit an exponential decay as function of the projection distance between cortical areas, mimicking the exponential form of connection probability originally proposed by Waxman. Both papers highlight the empirical evidence supporting the need for network models to consider spatial embeddings (Ercsey-Ravasz et al., 2013) (Horvát et al., 2016). In (Alexander-Bloch et al., 2012), the authors directly studied the impact of different connection probabilities as functions of euclidean distance in the resting-state functional magnetic resonance imaging brain networks of 20 healthy volunteers and 19 patients with childhood-onset schizophrenia and argue that the data are consistent with the interpretation that spatial and topological disturbances of functional network organization could arise from excessive “pruning” of short-distance functional connections in schizophrenia. The empirical evidence of decreasing connection probability as a function of distance in structural and functional brain networks, the application to brain wiring cost models and the success of Waxman and Geographical Threshold Graphs in modeling similar network structures is proving it to be a powerful tool in modeling many areas of the brain.

Complex Network theory is a young field, and the extension to Spatial networks is even less mature. The adoption of these tools is only now beginning to accelerate in the domain of neuroscience. In the extensive Brain Graphs review by Bullmore and Basset, they comment "there are many methods by which topological and geometric measures could be combined; for example, we can weight topological edges by physical distance between nodes for weighted network analysis. We consider that this topo-physical mapping of brain networks is likely to become of considerable interest in the future, although it has not yet been much developed." (Bullmore and Bassett, 2011), mimicking the intuition of the more advance Spatial Network models introduced in recent years, including the new model introduced in Section **??** of this paper.

The need to consider the spatial embedding of brain networks is becoming increasingly important in discovering physically realistic models of structural and functional brain connectivity. The recent and ongoing advances in Spatial Network theory and their adoption in neuroscience is sure to continue, and these diverse disciplines will continue to grow together. Both disciplines are entering a stage of maturity where their combination will accelerate the advances of both fields, as the application to neuroscience informs promising areas for theoretical Spatial Network research, and emerging theoretical models will propel new advances in neuroscience. The complexity challenge in neuroscience is daunting, but the advancement of powerful tools like Spatial Networks provide footholds to continue peeling away and understanding all layers of the human brain.

4.3 Other Applications

The most obvious and well studied applications of Spatial Networks are probably the ones discussed above, but their is a diverse set of industries and application using Spatial Networks. In (Lambiotte et al., 2008), as a measure of social ties, Lambiotte used mobile phone data for 3.3 million customers in Belgium to measure the impact of geography on social ties. A similar study by (Liben-Nowell et al., 2005) use Spatial Networks to find a powerlaw decay in social ties as a function of distance, highlighting the need for considering spatiality in social networks. Further afield, Random Geometric Graphs have been used in robot planning and navigation

(Solovey and Halperin, 2016). Spatial Networks are even find applications in the newly emerging "sharing economy", with a study on the structure of bicycle sharing networks (Austwick et al., 2013). The use-cases and applications of Spatial Networks is clearly valuable and diverse, and highlights the motivation of this work and informs the development and application of a new Spatial Network model presented in the follow chapters.

Chapter 5

Thresholded Random Geometric Graphs

5.1 Model Introduction and Definition

We present a new Spatial Network model extending Random Geometric Graphs to incorporate the threshold parameter of Random Threshold Graphs, θ, termed Thresholded Random Geometric Graphs (TRGG). The motivations for developing this model are two-fold: maintaining the practical simplicity and physical realism of the Random Geometric Graph model, while providing the ability to add the influence of network weights to the model to allow a more faithful representation of real-world systems. The underlying hypothesis of this model is that there exists a maximum distance at which connections between nodes can be made, and within that maximum distance the connection is determined by the quality or fitness of a node in making a good connection, measured by its respective weight. Pairs of nodes that have sufficient sum total weights above the threshold parameter, θ, are connected. The introduction of the new model parameterized by the combination of the Spatial Network parameters R and θ fills in a gap of the classification presented in Figure 3.3 and is highlighted with dotted-lines in the updated classification of Figure 5.1.

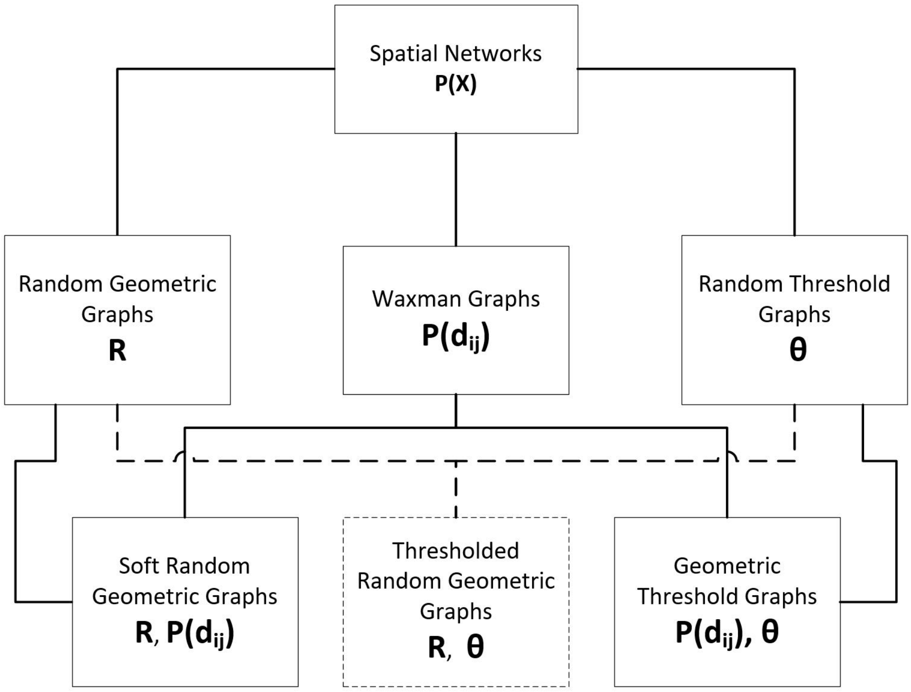

FIGURE 5.1: Spatial Network Model Classification

The types of real-world systems motivating the Thresholded Random Geometric Graphs model are those where the probability of connection does not depend on the distance between the connecting nodes, but a physically limiting maximum distance exists that prevents any connections past that distance, while the system seeks to maximize some utility of the entire weighted network. Examples of systems having these properties include those that have on-board energy sources such as fuel-tanks or batteries, that are depleted by displacement to some maximum range like automobiles or trains and require refueling stations in desirable locations. Other systems are industrial processes limited by physics, such as pipeline or HVAC systems where pressure losses are incurred to some maximum distance before pumping stations are required. Even the motivating example for Random Geometric Graphs, Ad-Hoc Wireless Networks by (Nemeth and Vattay, 2003), can be represented with this new model to allow the consideration of which nodes to make connections for optimal routing given their respective connection utility weights.

Equation 5.1 formalizes the definition of Thresholded Random Geometric Graphs, where the edge, E_{ij}, exists if the distance, d_{ij} for nodes i and j is less than or equal to the maximum distance, R, and their summed weights, w_i, w_j, are greater than or equal to θ.

$$E_{ij} = \begin{cases} (w_i + w_j) \geq \theta & \text{if } d_{ij} \leq R \\ 0 & \text{if } d_{ij} > R \end{cases} \tag{5.1}$$

5.2 Properties of Thresholded Random Geometric Graphs

In this section, we study the connectivity of the newly proposed model and comment on the critical scaling behavior for various parameterizations of the model and provide estimates for the connectivity of a TRGG under the specified conditions of this section. For all results in this section, and exponential and Pareto weight distribution described in Equations 5.2 and 5.3 are assumed and nodes are uniformly embedded onto the unit square exemplified in Figure 5.2. Distances, d_{ij} between nodes i and j are Euclidean. Non-uniform spatial distributions or considered in subsequent sections.

$$f(w) = \lambda e^{-\lambda w} \tag{5.2}$$

$$f(w) = \frac{\lambda}{w^{\lambda+1}} \tag{5.3}$$

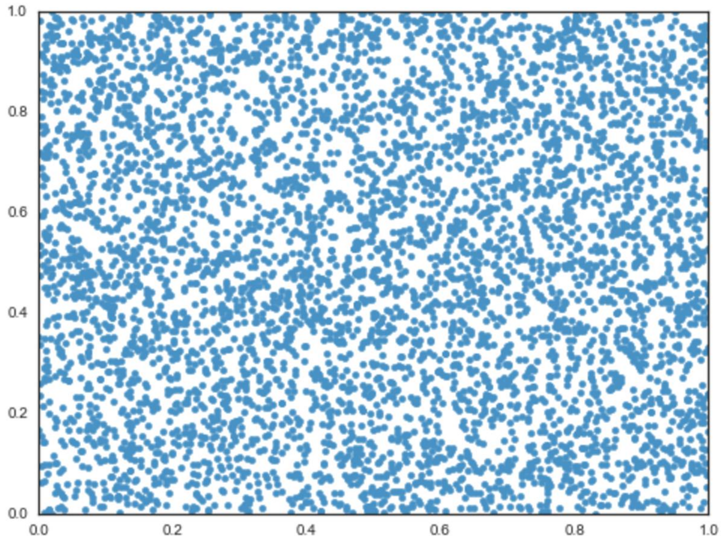

FIGURE 5.2: Example embedding of N = 5000 uniformly distributed nodes on the unit square

5.2.1 Connectivity of Random Geometric Graphs

We first revisit the connectivity theory of Random Geometric Graphs, since the minimum connection distance, R, where Random Geometric Graphs first begin to become fully connected will be the same for Thresholded Random Geometric Graphs.

In (Gupta and Kumar, 1998), the authors prove the critical minimum value of $R = R_c$ that ensures connectivity of the network as a function of N is of the form described in Equation 5.4.

$$R_c = \sqrt{\frac{log(N)}{\pi N}} \tag{5.4}$$

Figure 5.3 depicts the phase transition in connectivity of Random Geometric Graphs for different networks of size N, highlighting the distinct relationship between the critical R value for the emergence of the giant component as a function of N as described in Equation 5.4.

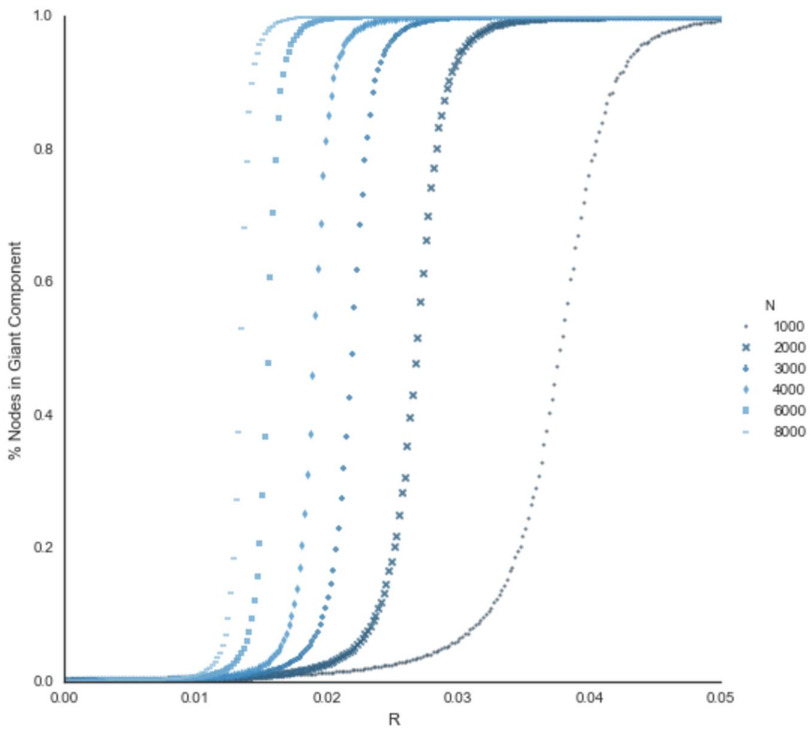

FIGURE 5.3: Connectivity Phase Transition of Random Geometric Graphs at Various Networks of Size N

5.2.2 Connectivity of Threshold Graphs

At the upper limit of R equal to the maximum length scale of the spatial embedding, which for the unit square is $\sqrt{2}$, all vertices of the network are within the maximum distance, R, from each other and the model devolves into the Random Threshold Graph introduced in Section 3.2.3. The condition for connectivity of a Random Threshold Graph is dictated by whether the summation of the minimally and maximally weighted vertex exceed the threshold parameter, θ. If the minimally weighted vertex does not exceed θ with the maximally weighted vertex, it fails to with all other vertices and remains an isolated vertex, rendering the network disconnected (Reilly and Scheinerman, 2009). When N is sufficiently large, the weight w uniquely determines the vertex degree k for the given weight distribution, $f(w)$, by Equation 5.5 and the degree distribution is described by Equation 5.6 (Masuda, Miwa, and Konno, 2004).

$$k = N[1 - F(\theta - w)] \tag{5.5}$$

$$p(k) = f(w)\frac{dw}{dk} = \frac{f[\theta - F^{-1}(1 - \frac{k}{N})]}{Nf[F^{-1}(1 - \frac{k}{N})]} \tag{5.6}$$

In (Makowski and Yagan, 2013), the authors prove the existence of strong and weak zero-one scaling laws for threshold graphs, with exponential weight distributions having strong zero-one laws and powerlaw weight distributions having only a weak zero-one laws a critical scalings, θ^*_n, according to Equations 5.7 and 5.8 respectively.

$$\theta^*_N = c\frac{log(N)}{\lambda} \tag{5.7}$$

$$\theta^*_N = cN^{\frac{1}{\lambda}} \tag{5.8}$$

5.2.3 Connectivity of Thresholded Random Geometric Graphs

The theory developed for Random Geometric Graphs and Random Threshold Graphs introduced in the above sections is extended to characterize the connectivity bounds for Thresholded Random Geometric Graphs. It is trivial to show that the lower bound of $R = R_c$, for connectivity in Random Geometric Graphs described in Equation 5.4 will be equivalent to Thresholded Random Geometric Graphs, as setting the threshold parameters, θ, to 0 exactly recovers the Random Geometric Graph model, and any value of $\theta > 0$ will reduce network connectivity. Networks with R values above the critical value of Random Geometric Graphs have more interesting connectivity characteristics, and are the focus of this section.

Above R_c, the condition for connectivity becomes a combination of influences of the model parameters R and θ. The same condition for connectivity of Random Threshold Graphs, where the summation of the minimally and maximally weighted vertex exceed θ, will apply for Thresholded Random Geometric Graphs, but this condition needs to be applied locally for every circle of Area = πR^2 surrounding each vertex. The degree distribution for any local network within πR^2 of any node can be obtained by substituting the local node count, $N\rho$, where ρ is the local node density, for N in Equation 5.6 as shown in Equation 5.9.

$$p(k_{local}) = \frac{f[\theta - F^{-1}(1 - \frac{k}{N\rho})]}{N\rho f[F^{-1}(1 - \frac{k}{N\rho})]} \tag{5.9}$$

With a uniform distribution of nodes embedded onto the unit square, we can estimate a constant density, $\rho = \frac{\pi R^2}{1^2}$, will exist in each circle and Equation 5.10 is obtained, which for uniformly distributed nodes and constant R is equivalent to the degree distribution of the entire network. The estimate for ρ using the assumption of uniform distribution everywhere will not hold for nodes along the edges of the unit square, as the density of nodes beyond the embedding edges will be 0, reducing the average density of nodes with connection areas reaching beyond these edges. Knowing this limitation, we propose 5.10 as an estimate for the true degree distribution.

$$p(k_{local}) = p(k) \approx \frac{f[\theta - F^{-1}(1 - \frac{k}{N\pi R^2})]}{N\pi R^2 f[F^{-1}(1 - \frac{k}{N\pi R^2})]} \tag{5.10}$$

Here we explore the connectivity of Thresholded Random Geometric Graphs numerically, with empirical simulation results characterizing the connectivity properties of the model.

Similar to the critical value R_c for Random Geometric Graphs, the critical value θ_c for Thresholded Random Geometric Graphs, where a sharp transition in network connectivity exists, is a function of N as depicted in Figure 5.4.

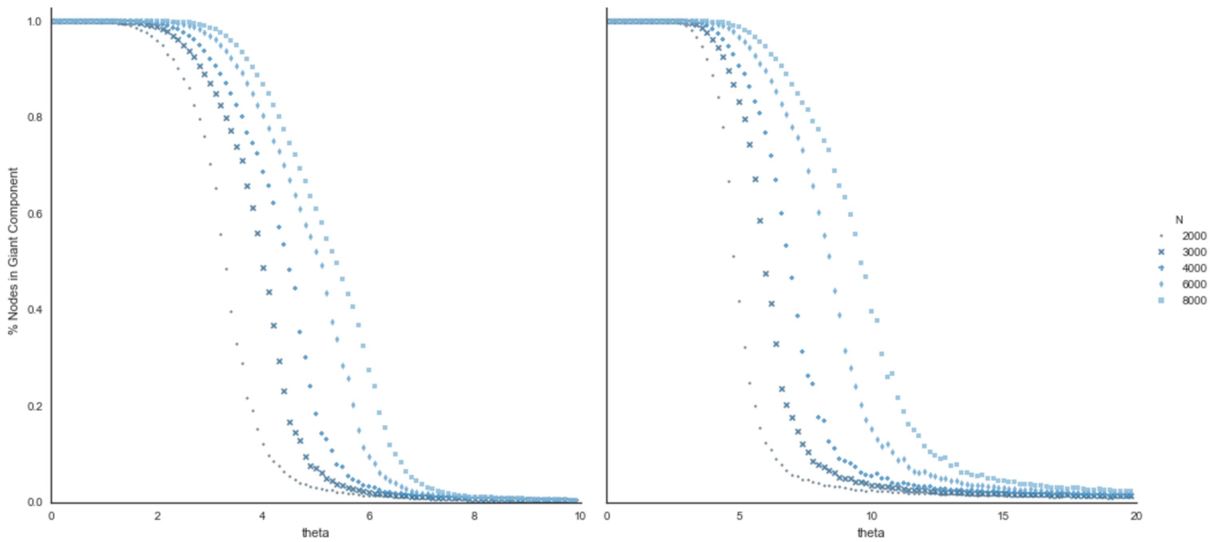

FIGURE 5.4: Connectivity Phase Transition of Thresholded Random Geometric Graphs at Various Networks of Size N and R = 0.05. Left - Exponential weight distribution with λ = 1, Right - Pareto weight distribution with λ = 2

Unlike the critical value R_c for Random Geometric Graphs, the critical value θ_c for Thresholded Random Geometric Graphs is unbounded and depends not only on N, but also the weight distribution $f(w)$ and R. The unbounded nature of θ_c makes generalizing a critical value range difficult, and the theory for scaling laws of Threshold Graphs developed in (Makowski and Yagan, 2013) is utilized. Using the constant density, $\rho = \pi R^2$, of the uniformly distributed nodes on the unit square, we substitute N in Equations 5.7 and 5.8 with $N\pi R^2$ in a similar fashion as the Degree Distribution of Equation 5.10, resulting in Equations 5.11 and 5.12. Again, this will produce a limited estimate of the true scaling using the assumption of uniformly distributed density of nodes everywhere, which will not exist for nodes near the edges of the unit square.

$$\theta_N^* \approx c\frac{log(N\pi R^2)}{\lambda} \tag{5.11}$$

$$\theta_N^* \approx c(N\pi R^2)^{\frac{1}{\lambda}} \tag{5.12}$$

Figure 5.5 depicts the scaled connectivity thresholds of Figure 5.4 using the scaling laws of Equations 5.11 and 5.12. The overlapping lines indicate the substituted scaling laws for Threshold Graphs directly apply for Thresholded Random Geometric Graphs at constant values of R.

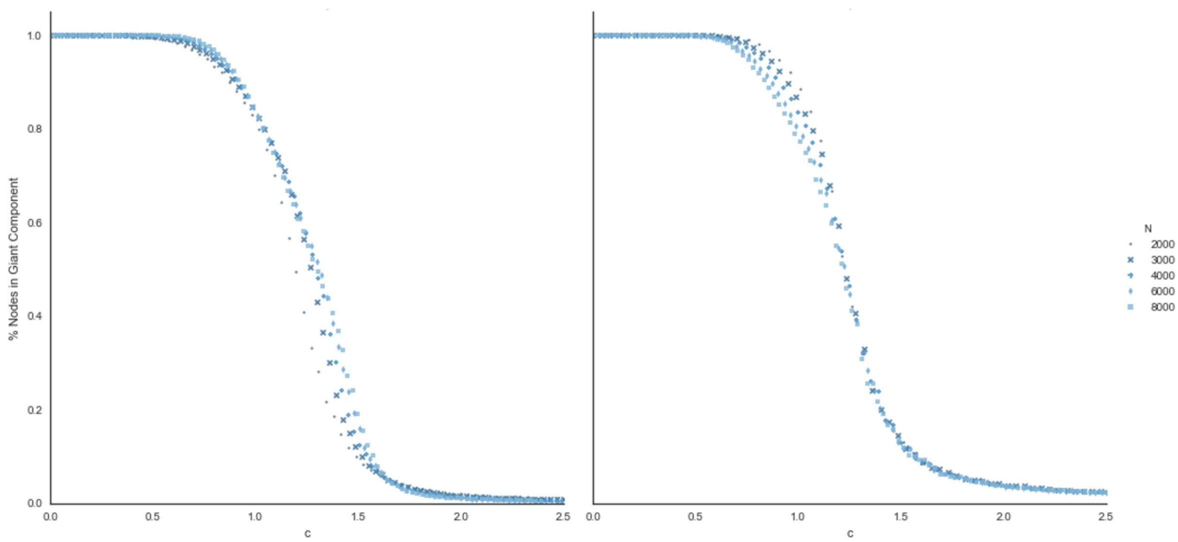

FIGURE 5.5: Scaled Connectivity Phase Transition of Thresholded Random Geometric Graphs at Various Networks of Size N and R = 0.05 using the substituted scaling laws of (Makowski and Yagan, 2013) Left - Exponential weight distribution with λ = 1, Right - Pareto weight distribution with λ = 2

The seemingly strong scaling law for the exponential case, with the transition regions bounded closely to 1, and the weak scaling law of the power-law distribution, with the transition continuing past 2.5 depicted in Figure 5.5 agree with the proven properties of Threshold Graphs in (Makowski and Yagan, 2013).

In order to fully characterize the conditions for connectivity of the model and to confirm the estimates of Equations 5.11 and 5.12, the effect of varying R for a network of size N is also studied. Figure 5.6 depicts the connectivity for both weight distributions for a network of size N = 2000 at R values of 0.05,0.1,0.15,0.2 and 0.25 and shows a clear relationship between network connectivity and R, which for the exponential distribution, is very similar to the relationship between connectivity and N, but strikingly different for the Pareto distribution, an expected result of Equations 5.11 and 5.12 due to the non-linear relationship between the model parameters N and R and the weight distribution parameter λ for the Pareto distribution critical scaling.

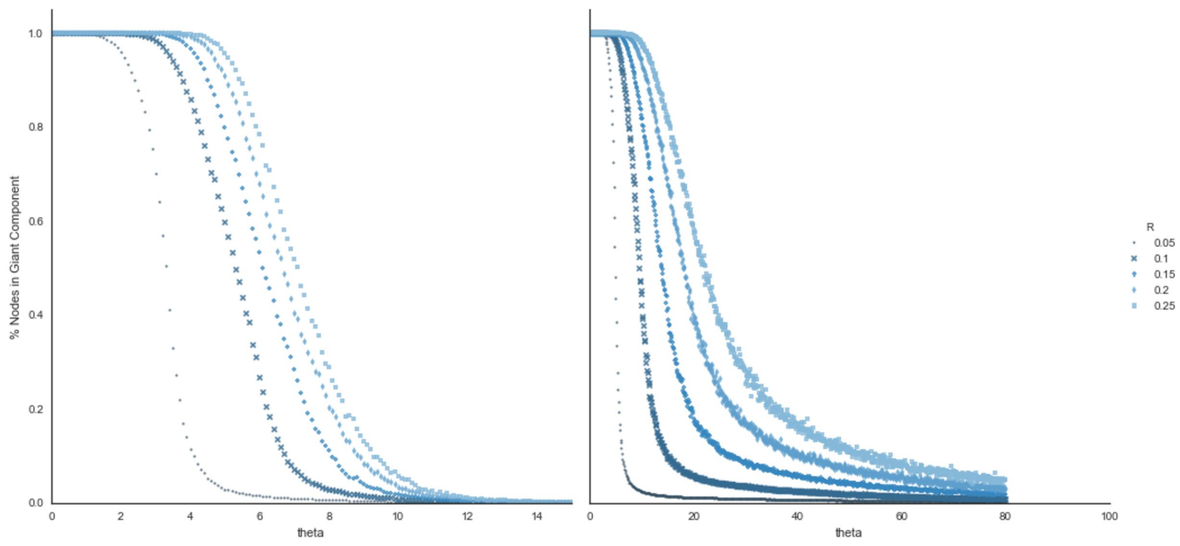

FIGURE 5.6: Scaled Connectivity Phase Transition of Thresholded Random Geometric Graphs with a networks of Size N = 2000 and various R. Left - Exponential weight distribution with λ = 1, Right - Pareto weight distribution with λ = 2

To confirm the substituted critical scaling estimates of Equations 5.11 and 5.12, we apply the scaling to Figure 5.6 in a similar fashion as Figure 5.5. Figure 5.7 shows the critical scaling for the connectivity of Thresholded Random Geometric Graphs for a network of N = 2000 parametrized with various values of R.

The consistent strong zero-one connectivity transition and overlap for the exponential distribution indicates the scaling law of Equation 5.11 fully generalizes the conditions for connectivity in Thresholded Random Geometric graphs with exponential weight distributions, and a constant transition threshold value for c exists for all parameterizations. The characteristics of the Pareto distribution of Figure 5.7 depict consistent scaling, but indicates a transition of a continue weakening of the zero-one scaling law as the maximum distance parameter, R is increased, extending fully into the weak zero-one scaling law of Threshold Graphs. This is a very interesting property of Thresholded Random Geometric Graphs, providing the ability to change the rate of connectivity transition as a function of R, a characteristic that could prove

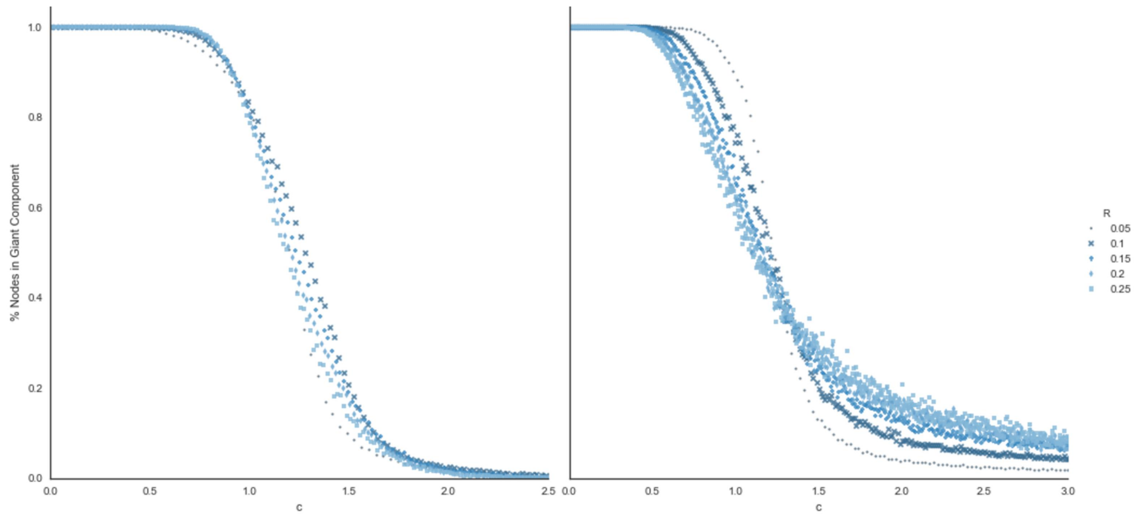

FIGURE 5.7: Scaled Connectivity Phase Transition of Thresholded Random Geometric Graphs with a networks of Size N = 2000 and various R using the substituted scaling laws of (Makowski and Yagan, 2013) Left - Exponential weight distribution with $\lambda = 1$, Right - Pareto weight distribution with $\lambda = 2$

useful in real applications. The formal proofs of these properties and the exact values of c for each distribution are left for future work, but the unique characteristics shown here highlight the potential for future studies and applications of the proposed model.

5.2.4 Non-Uniform Spatial Embedding Distributions

The underlying assumption facilitating the estimation for the degree distribution and critical scaling laws of Thresholded Random Geometric Graphs in the previous section is the uniformity of distribution onto th unit square of nodes, leading to a constant density, ρ, everywhere. This assumption may be overly restrictive, with many real-world systems displaying non-uniform spatial distribution such as the bias of global cities to lie near non-uniformly distributed coastlines or riverbeds. In this section, we investigate numerically the impacts of non-uniform spatial distributions on the connectivity characteristics of Thresholded Random Geometric Graphs and confirm that the estimates of Equations 5.7 and 5.8 fall apart under these conditions.

Here we explore the impacts on network connectivity using a Mixture of Gaussians spatial distribution with means, μ = 0.2 and 0.8 and a common standard deviation, σ = 0.1 for both the x and y coordinate of nodes embedded onto the unit square which is exemplified in Figure 5.8.

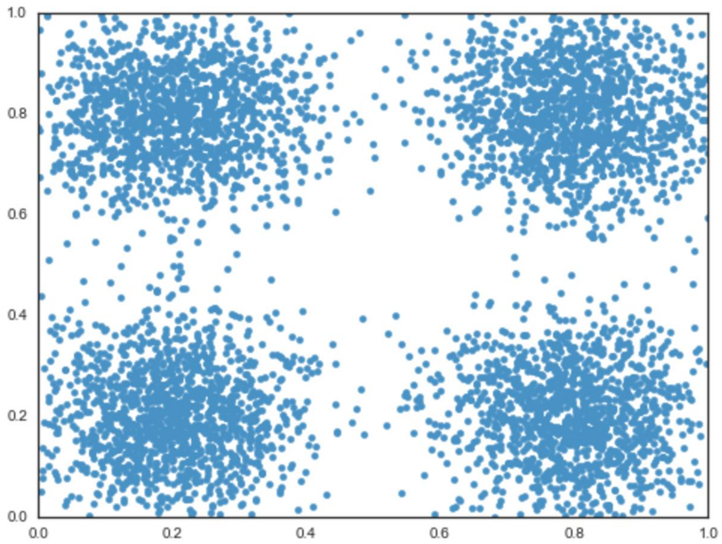

FIGURE 5.8: Example embedding of N = 5000 Mixture of Gaussian with μ = 0.2 and 0.8 and σ = 0.1 for both the x and y coordinates distributed nodes on the unit square

Figure 5.9 depicts the connectivity for both weight distributions for a network of size N = 1000 at R values of 0.05,0.1,0.15,0.2 and 0.25 for the non-uniform double Gaussian distribution shown in 5.8.

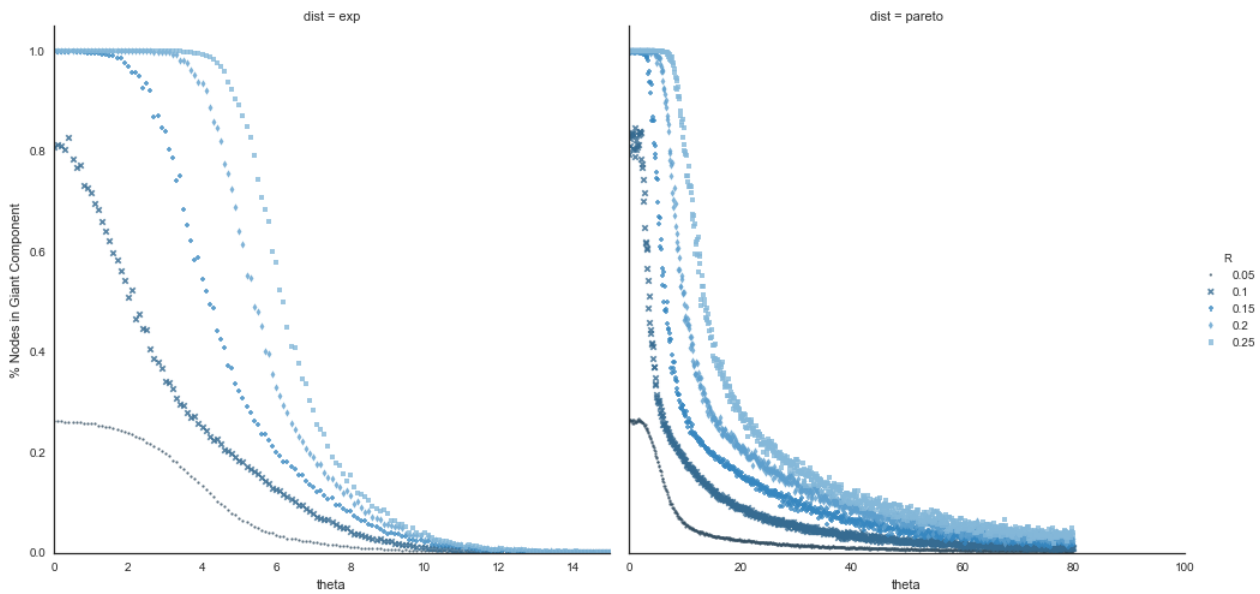

FIGURE 5.9: Connectivity Phase Transition with networks of Size N = 1000 and various R with non-uniform spatial distribution. Left - Exponential weight distribution with $\lambda = 1$, Right - Pareto weight distribution with $\lambda = 2$

Although the clear relationship between connectivity and R persists for our non-uniform spatial distribution similar to the uniform case, there are obvious highly nonlinear dynamics occurring, especially at the lower values of R. Applying the scaling estimations of Equations 5.7 and 5.12, we see in Figure 5.10 that these estimations clearly do not hold for the case of non-uniform spatial distributions, as was expected from the underlying uniformity assumptions of the estimates.

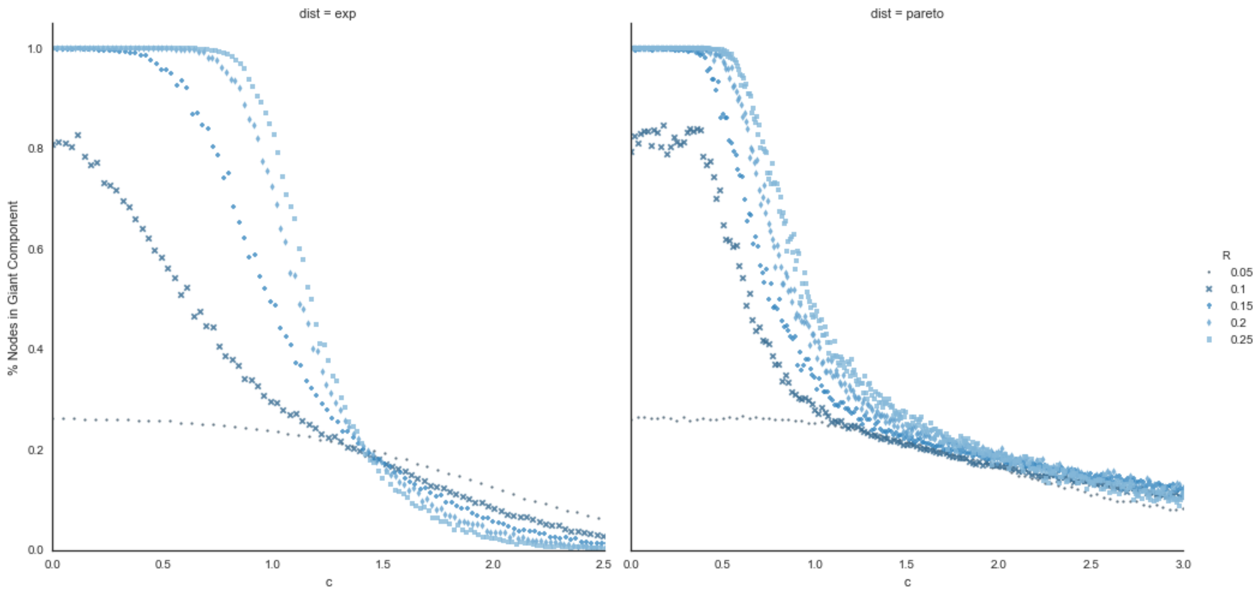

FIGURE 5.10: Scaled Connectivity Phase Transition with networks of Size $N = 1000$ and various R with non-uniform spatial distribution. Left - Exponential weight distribution with $\lambda = 1$, Right - Pareto weight distribution with $\lambda = 2$

The limitations in our analysis for failing to capture the dynamics of non-uniform spatial distributions, which are likely to exist in most real-world applications, suggests the need for future work in the theory and mathematics to fully characterizing the general conditions for connectivity in the newly present Thresholded Random Geometric Graphs. These diverse dynamics under simple parameterizations also suggest the extensive expressiveness capabilities of this new model, and efforts in extending their underlying theory are likely to generate substantial contributions in the field of Spatial Networks.

Chapter 6

A Motivating Example - Telsa Supercharger Network

6.1 Motivating Example Introduction

The motivation for the study of Spatial Networks in this paper is grounded in the real-world application of Spatial Network models to the Tesla Supercharger electric vehicle charging infrastructure network.

Tesla is an Electric Vehicle (EV), storage and panel manufacturer that recently surpassed GM to become the most valuable vehicle manufacturer in the United States (*Gaurdian*). One of Tesla's key differentiators is its Supercharger network of fast electric vehicle charging stations, allowing users to travel large distances with electric vehicles using this infrastructure network.

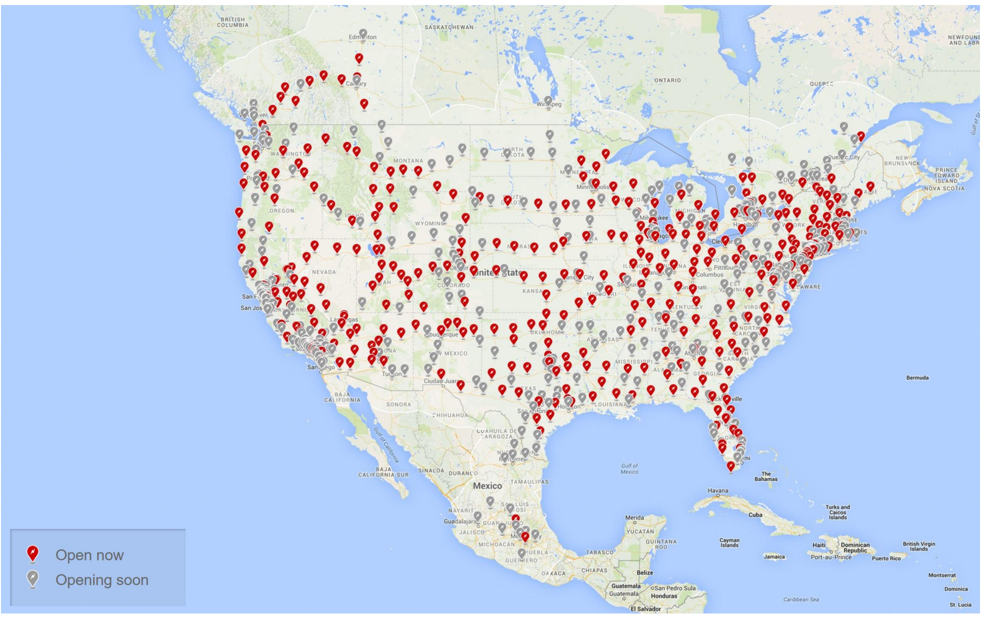

FIGURE 6.1: Tesla's Supercharger Network - 2017 (*Tesla*)

The Supercharger network was selected as our motivating example for 3 key reasons:

- Admissible - The infrastructure system and spatial nature of the Tesla Supercharger network lends itself directly to its study using Spatial Networks.
- Valuable - The large amount of current and future investment into these types of systems makes it a pragmatic and valuable system to study.
- Practical - The availability of data and the existence of multiple Supercharger networks across different continents to compare and discover fundamental patterns from makes it a feasible system to study.

6.1.1 Supercharger Network Dataset

The Supercharger data used in this paper is obtained from (*supercharger.info*), filtered for the Canadian and American Supercharger locations, totaling 386 Opened Superchargers as of April 2017. Distances between Supercharger and city locations are obtained via the Google Directions API. The collected data has been structure into a Networkx Graph (Hagberg, Schult, and Swart, 2008) which is made up of nested dictionaries keyed on the geohash of each Superchargers GPS coordinates. See Appendix A for an example node's data structure. See Appendix B for an overview of the geohashing algorithm used extensively in the source code of this analysis.

6.1.2 Network Model Definition

The Supercharger Network is modeled with the nodes being the cities of the network region, embedded onto the unit square by normalizing the GPS coordinates of each city and weighted by the population of the city, as a percent of the datasets total population. Connections between nodes are made for cities that have Superchargers and are within the maximum base range of a Tesla model 3 (215miles - 346km) (*Tesla*) as determined by the driving distance computed by Google directions API between Supercharger GPS locations. Figure 6.2 depicts the representation of Figure 6.1 according to this network model definition, excluding Mexico and "Opening Soon" Superchargers.

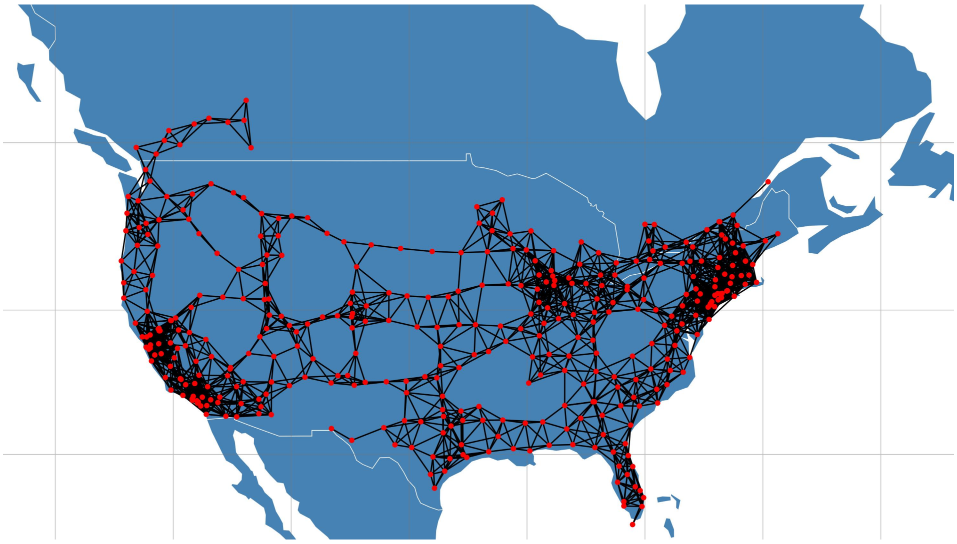

FIGURE 6.2: Tesla Supercharger Network Model

Using the Albers Equal-Area Conic Projection (Snyder, 1987) with parameters enumerated in Table 6.1, the Supercharger Network can be projected onto the unit square as depicted in Figure 6.3. The same projection is applied to the top 5,000 cities by population for the United States and Canada depicted in Figure 6.4, providing the network expansion space available to simulate the growth of the Supercharger network on.

TABLE 6.1: Parameterization for Supercharger Network Albers Projection

Parameter	**Value**	**Definition**
lat_1	40.0	First Standard Parallel
lat_2	60.0	Second Standard Parallel
lon_0	-97.0	Central Point - Longitude
lat_0	47.0	Central Point - Latitude
Width	6000000	Width of the Projection
Height	4500000	Height of the Projection

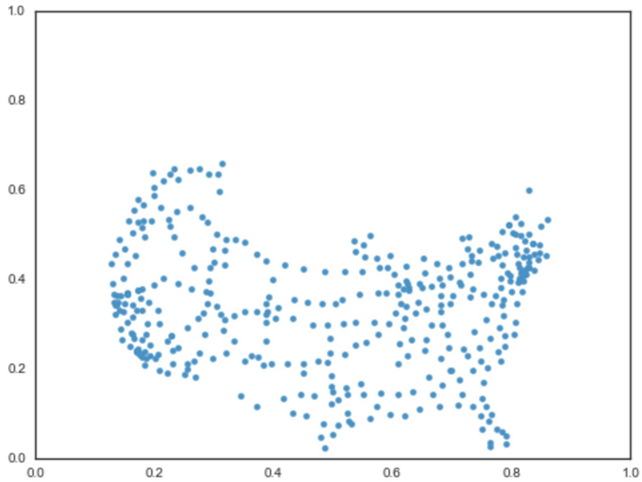

FIGURE 6.3: Telsa Supercharger Network Albers Projection onto the Unit Square

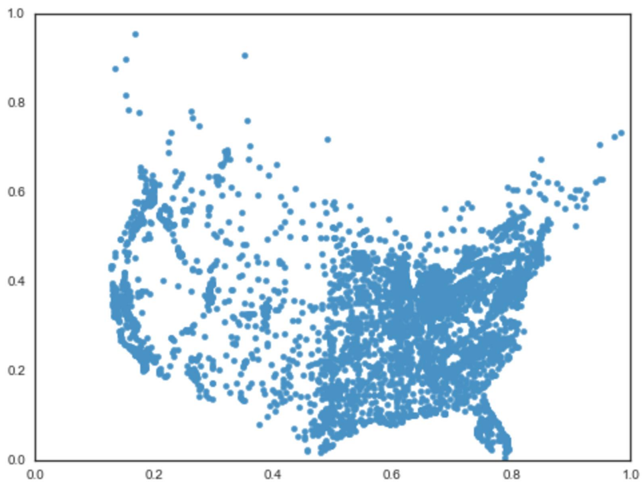

FIGURE 6.4: Top 5,000 Canadian and American Cities by population Albers Projection onto the Unit Square

With this definition of the Supercharger Network model and network expansion space, the network can be analyzed and modeled using the Spatial Network models of Section 3.2.

6.2 Supercharger Network Parameterization

In this section we review some of the properties of the current Tesla Supercharger network and discover the Spatial Network parameters, λ, R, $P(d_{ij})$ and θ, specific to the Tesla Supercharger network.

6.2.1 Supercharger Network - λ

The first parameter we fit to the real data of the Tesla Supercharger network is λ, the powerlaw exponent of Equation 5.3 that characterizes the weight distribution of nodes embedded on the unit square, which are city populations in our case. Using the Powerlaw python package by (Alstott, Bullmore, and Plenz, 2014), we approximate the powerlaw fit with the packages Maximum Likelihood Estimate (MLE) of the top 5,000 Canadian and American city populations data. The obtained MLE for λ is 2.3 with x_{min} = 48,433 which agrees almost exactly with the estimates for similar city population data performed in (Newman, 2005), with λ = 2.3 with x_{min} = 40,000. Figure 6.5 shows the real distribution of the city population data versus the fitted distribution.

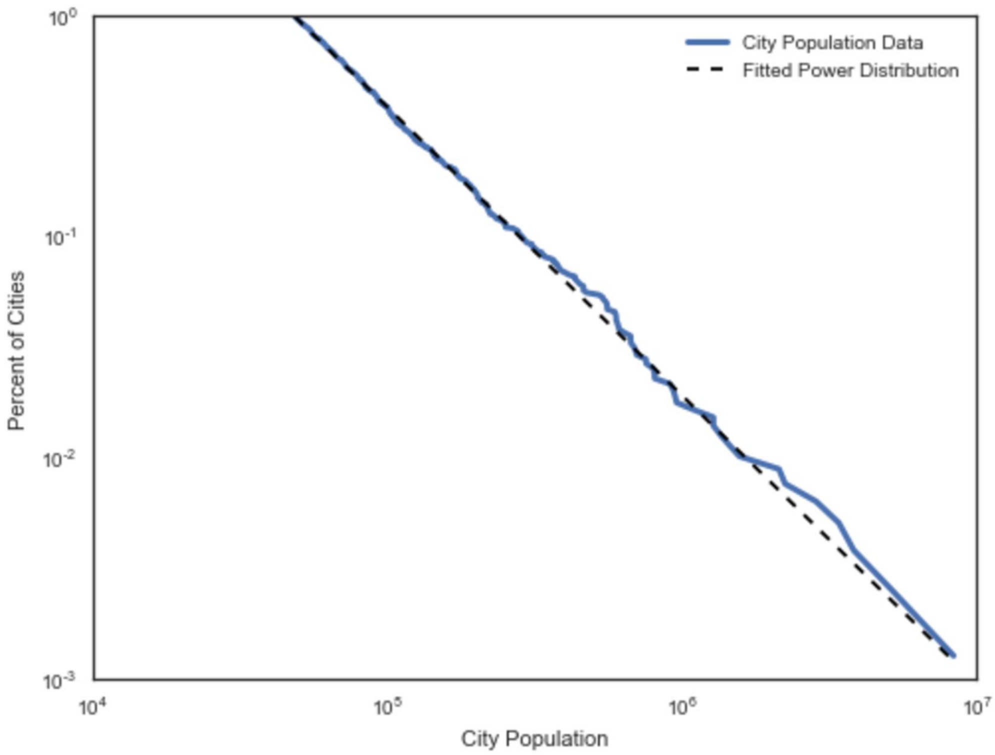

FIGURE 6.5: Top 5,000 Canadian and American City Populations Powerlaw fit λ = 2.3

6.2.2 Supercharger Network - R

The next parameter needed for the Spatial Network models is R, the maximum connection distance. The maximum connection distance for the Supercharger network is provided in the model definition of Section 6.1.2, namely the maximum base range of a Tesla model 3 (215miles - 346km), when scaled to the model's unit square projection is R = 0.0461.

6.2.3 Supercharger Network - $P(d_{ij})$

The distribution for the probability of connection as a function of distance, $P(d_{ij})$ can be directly estimated from the actual distribution of the connection distances in the current Supercharger network. Figure 6.6 plots the Supercharger network's connection counts at each distance with a fitted Gaussian Kernel Density Estimate (KDE) of the distribution.

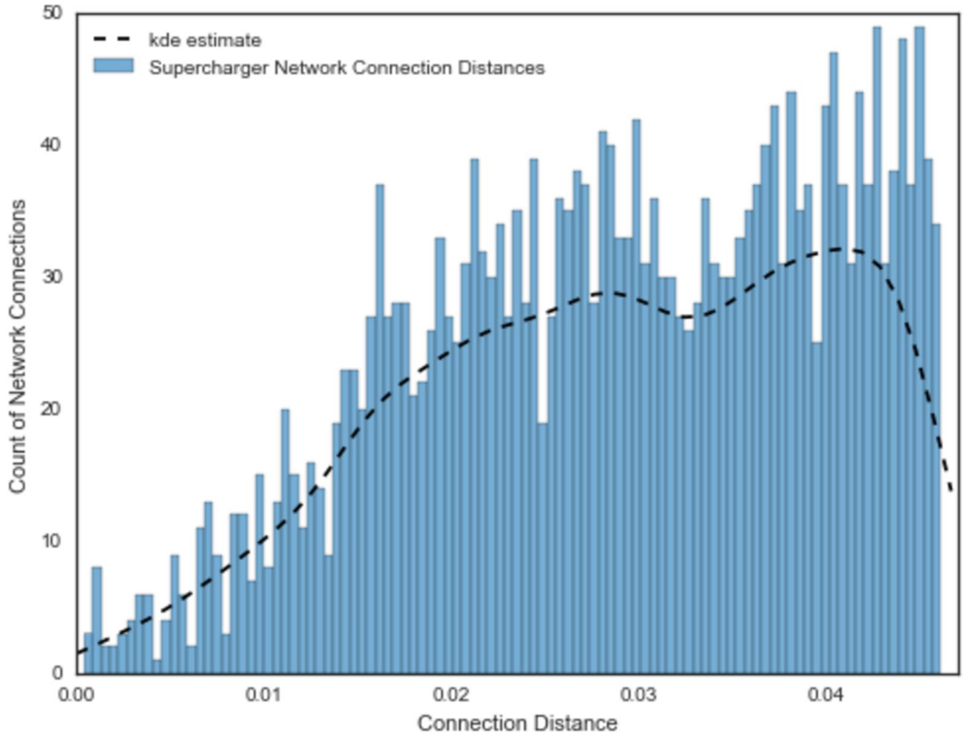

FIGURE 6.6: Supercharger Network Connection Probability Distribution, $P(d_{ij})$ with Gaussian KDE

An interesting observation of Figure 6.6 is the *increasing* probability of connection as a function of distance, which is in stark contradiction to the base assumption of monotonically *decreasing* probabilities as a function of distance assumed in many Spatial Network models (Barthélemy, 2011). In the context of the Supercharger network, an increasing connection probability as a function of distance to some maximum distance makes intuitive sense, as an underly driving force of the network must be at least partially to expand broadly, which requires a bias towards longer distanced connections. Other networks are likely to display a similar property and the observation of Figure 6.6 supports the claim of this paper that Thresholded Random Geometric Graphs may be better suited at modeling networks of this type.

6.2.4 Supercharger Network - θ

The final parameter needed for Spatial Network models of the Supercharger network is the minimum connection threshold, θ. Again, θ can be estimated directly from the data of the current Supercharger network by inspecting the distribution of weighted connections of the network and selecting the minimum. The distribution of connection weights is depicted in Figure 6.7, and the lower range is highlighted in Figure 6.8, with a minimum connection weight taken from the data being $\theta = 0.0044$.

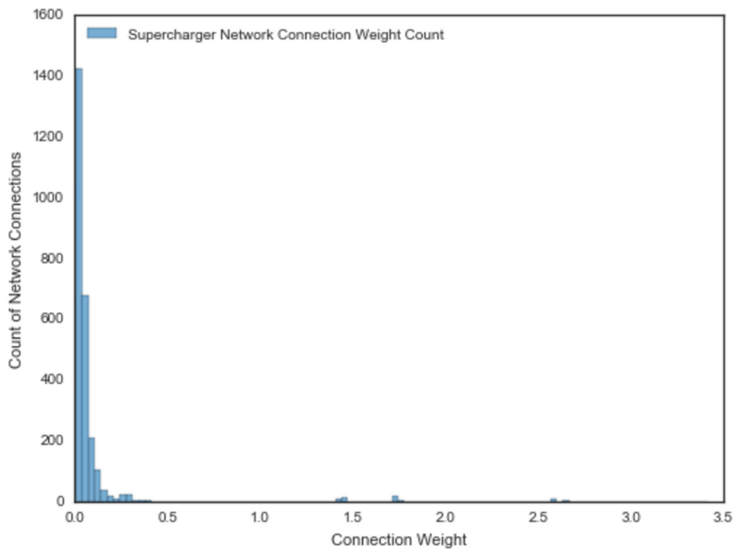

FIGURE 6.7: Supercharger Network Connection Weight Distribution

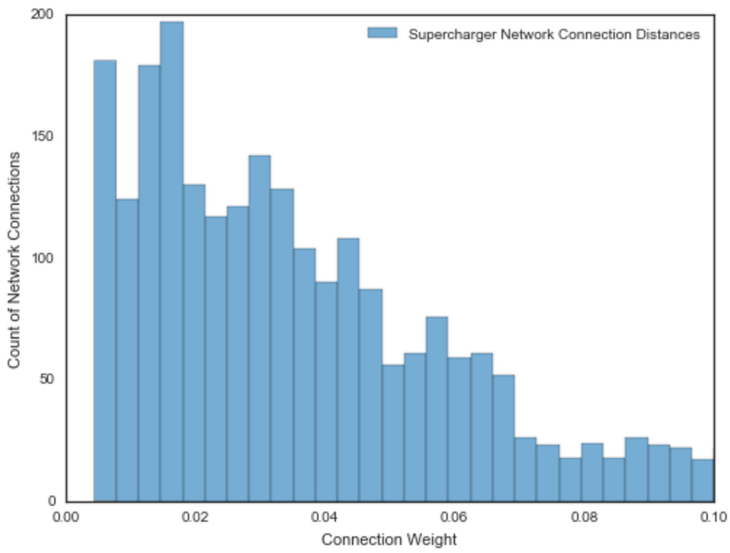

FIGURE 6.8: Supercharger Network Connection Weight Distribution Lower Bound

6.2.5 Supercharger Network - Parameter Summary

Table 6.2 summarizes the Spatial Network parameter estimates for the Tesla Supercharger Network. These parameters can be used to model the Supercharger network with various Spatial Network models and compared which model best represents the underlying structure of our motivating example.

TABLE 6.2: Tesla Supercharger Network Spatial Network Parameter Estimates

Parameter	**Value**
λ	2.3
R	0.0461
$P(d_{ij})$	KDE
θ	0.0044

6.3 Supercharger Network Spatial Network Modeling

With the Spatial Network parameter estimates obtained in Section 6.2, we can parameterize the Spatial Network models of Section 3.2. Here we focus our analysis to the recent advanced Spatial Network models, namely Soft Random Geometric Graphs, Geometric Threshold Graphs and the Thresholded Random Geometric Graphs model presented in this paper, and benchmark their performance against randomly selected nodes and the RGG model. To validate which model best captures the underlying structure of the Supercharger network, 10,000 examples of each model are generated using the model definition and parameterization of the above sections, and each models ability to predict the 32 newest Superchargers not included in the models' parameterization estimates are compared.

6.3.1 Network Evaluation Metrics

In order to compare each Spatial Network model's ability to discover the underlying structure of the Supercharger network and correctly generalize to future Supercharger locations, we define a few network evaluation metrics that capture the distinct properties of the Tesla Supercharger Network. The 3 evaluation metrics are population Coverage, geographic Breadth and network Efficiency,defined here.

- Coverage - The percent of population represented in the main connected giant component of the network

- Proximity - The Earth Movers Distance (EMD) (Pele and Werman, 2009) of the simulated network to the real network
- Efficiency - The geographic diameter of the network divided by the average shortest path length

6.3.2 Spatial Network Model Validation

Each Spatial Network model is initialized with the seed network of 386 Opened Superchargers and new Supercharger locations are added according to the parameterization of each model until the network has 418 Superchargers, simulating the addition of the 32 held-out Superchargers. Each model is run for 10,000 simulations, and the most frequently occurring simulated network is selected, representing the models final predicted Supercharger network. Each evaluation metric of Section 6.3.1 is computed and the similarity to the real network values are summarized in Table 6.3.

TABLE 6.3: Spatial Network Model Evaluation Metrics Similarities to Real Network Metric Values

Metric	Real Network	TRGG	GTG	SRGG	RGG	Random
Coverage	0.697	97.3	86.7	63.7	52.3	73.4
Proximity	1.0	90.7	91.5	90.4	89.7	88.0
Efficiency	0.348	93.9	95.0	88.5	90.6	20.4
Average	N/A	**93.4**	91.1	80.9	77.5	60.6

The results in Table 6.3 indicate that the TRGG model is able to best capture the the structure of the real Supercharger network across the 3 measured metrics, with an average similarity to the real Network of 93.4%. The GTG model's performance is very similar to the TRGG model, and slightly outperforms the TRGG model in both the Proximity and Efficiency metrics. This comparable performance makes intuitive sense, as both models attempt to capture the same underlying structure: the influences of both spatial constraints and weight thresholds on the network. It is interesting to note that the parameterization of the GTG model is much more complex than that of the TRGG model, as a Gaussian KDE of the connection distance distribution was directly computed and used, where the TRGG model only utilized 2 real valued numbers as parameters. This also explains the GTG models ability to closely capture the spatial distribution of the 32 new Supercharger locations and its high Proximity metric, as it directly utilizes the closely matching spatial distance distribution of the initial 386 Supercharger locations. This extra complexity in the GTG model compared to the TRGG model makes in more likely to overfit to the data, and as the Supercharger network continues to expand the generalization performance of the

GTG model may degrade compared to the simpler TRGG model. It would be interesting to revisit this analysis in the future using the same parameterizations learned here and compare performances on a larger set of future Supercharger locations. The other Spatial Network models that do not incorporate network weights have much poorer performance, especially in the Coverage metric which measures the total network weight, and indicate the importance of including networking weighting parameters for accurately modeling the Supercharger network. These results highlight the ability of the newly proposed TRGG model to maintain the simplicity of the RGG model while accurately capturing the underlying structure of Spatial Networks like the Tesla Supercharger network, and we claim it will have similar comparable performance when applied to the classes of networks described in Section 5.1.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Motivated by the concrete problem of predicting where future Superchargers would or should be constructed, the results in this paper have taken us through the long and sometimes winding path of exploring the field of Spatial Networks, the survey of its state-of-the-art models and their applications, the discovery of a gap in current model representation power, the need to develop a new model to fill that gap including results for the conditions of connectivity, and the direct and successful application of this new model to achieve the original motivating goal.

Complex Networks are wonderful tools for representing all sorts of physical systems, but the need to consider spatial constraints in the development of these networks is proving to be crucially important, and the domain of Spatial Networks charged with this consideration is continuing to see increasing study. Many simple models have been proposed with recent proposals increasing relative model complexity and these most common Spatial Network models share common parameters, and a novel classification using this shared parameterization has been presented in this paper to help new researchers quickly intuit the relationships between some of these common Spatial Network models. Another hope for this new classification is to facilitate the identification of gaps in current model parametrization to inform the creation of new and useful models similar to the approach taken in this paper.

The true motivation behind the work in this paper was the discovery and understanding of a powerful tool that has proven to be useful in a wide range of engineering problems, and Spatial Networks continues to prove itself as one of those tools. Finding successes in fields as diverse as Transportation and Infrastructure Networks, Neuroscience, robotics and social network analysis, Spatial Networks are deserving of the amount of research and theory being developed for them, as they continue to help solve real-world challenges.

Having domain knowledge about the structure of the Tesla Supercharger network that proved to be useful in the development of a model capturing its structure, the work of this paper discovered a gap in the existing Spatial Network models in adequately incorporating that domain knowledge. This inspired the proposal of the new Spatial Network model, Threshold Random Geometric Graphs, which is the core work of this paper. Here we saw the diverse dynamics presented by this model, the conditions for connectivity under some conditions and the identification for future work applied to this new model likely to find further results in real-world applications.

The successful application of this new model to a real-world use-case and dataset came with the usual challenges of Data Science, with much of the difficult work hidden behind the scenes in data collection, munging, validation and structuring. With the final performance of the newly presented TRGG model being shown to outperform a selection of other Spatial Network models for our use-case in the Tesla Supercharger network, we hope that walking through the method of parameterization of these models discussed in this paper might be used as a template for the analysis of other real-world systems and to ultimately provide a classification for which Spatial Network models are best suited to particular classes of real-world systems. We also use the results of our new model on the Supercharger network to claim that it may be better than current models in modeling the types of systems discussed in the model introduction in Section 5.1.

As the young field of Complex Networks and the sub-domain of Spatial Networks enter maturity, new quantifications for their proposed models are needed. In this paper, we propose a new classification of Spatial Networks to ease the introduction of the field to new researchers and capture the relationships between common models. A new model, Thresholded Random Geometric Graphs is proposed to better capture the constraints of some real-world systems and applied to a real network with the aim of being a template for future applied studies using these powerful Spatial Network models. The rich representation of Spatial Networks to real-world systems is proving to be an exciting field of study and the continued research into their structure and applications is posed to help accelerate the discovery of breakthroughs in many vastly diverse domains.

7.2 Future Work

The expressiveness of Spatial Network models and their ability to intuitively capture physical properties of real systems has accelerated the number of theoretical and applied papers studying their use. Here we mention some of the potentially fruitful future work using Thresholded Random Geometric Graphs, and Spatial Networks more generally, in both theoretical and applied domains.

7.2.1 Future Work - Theoretical

As is typical when a new model is proposed in any scientific domain, the total combination of previous analytical studies to past models becomes immediately applicable to the new model. In the field of Complex Networks, these typical studies include investigations and proofs for the conditions of connectivity, degree distributions, clustering, diameter and many other well studied Complex Network properties, all of which can be investigated for Threshold Random Geometric Graphs. Beyond the typical Complex Network properties, the unique property of changing phase transition rates as a function of a parameter of the model deserves special attention, and quantifying the exact mathematical constructs needed to describe this phenomena and its use-cases in tuning real-world networks to these types of structures could lead to interesting results. The theory supporting even the oldest and simplest Spatial Network models continues to expand, even while new and more complex models are being proposed and the theoretical future work for these models showing no sign of slowing down.

7.2.2 Future Work - Applied

The model proposed in this paper was inspired by a specific application, modeling the Tesla Supercharger Network where the fundamental constraints of the network were not adequately met by existing models. In Section **??**, we propose other real-world systems that we claim are well modeled by Thresholded Random Geometric Graphs, but have only investigated a single instance in this paper. Many similar studies should be conducted to confirm the new model's ability to adequately model the suggested systems. For Spatial Networks more generally and the plethora of existing models, a quantification for which model is best suited for which types of systems is needed. In this paper, we have relied on intuition for describing which types of real-world systems are best captured by our proposed model, but an expansive study comparing common characteristics that lend themselves towards being modeled by a particular Spatial Network model would be hugely beneficial to the field. Finally, of particular interest is the application of Spatial Network models to the study of the human brain and the analogous structures of deep learning neural networks, which is only now starting to be adopted, and further exploration of these models in those domains are likely to discover interesting and fruitful results.

Appendix A

Networkx Supercharger Data Structure


```
{'City': 'Dayton',
 'Country': 'USA',
 'Elev': '258',
 'GPS': '39.85675, -84.276458',
 'GPS_lon_lat': [-84.276458, 39.85675],
 'Open Date': '2014-11-11 00:00:00',
 'SC_data': {'Chargers': ['8'],
  'Charging': ['8 Supercharger stalls, available 24/7'],
  'Lodging': ['Comfort Inn and Suites'],
  'Restaurants': ['Chipotle', "McDonald's", "Steak 'n Shake (open 24 hrs)"],
  'Restrooms': ['Meijer',
   'Chipotle',
   "McDonald's",
   "Steak 'n Shake (open 24 hrs)"],
  'WiFi': ["McDonald's", 'Comfort Inn and Suites']},
 'SC_index': '133',
 'Stalls': 8.0,
 'State': 'OH',
 'Status': 'Open',
 'Street Address': '9200 N Main St',
 'Supercharger': 'Dayton, OH',
 'Tesla': 'http://www.teslamotors.com/findus/location/supercharger/daytonsupercharger',
 'Thread': 'http://supercharge.info/service/supercharge/discuss?siteId=241',
 'Zip': '45445',
 'city_state': 'Dayton_OH',
 'geohash': 'dph4dpxy2hwn',
 'lat': 39.85675,
 'lon': -84.276458,
 'population': 7719014.0}
```


FIGURE A.1: Example Networkx Data Supercharger Structure

Appendix B

Geohashing

The concept of geohashing was utilized extensively in this work to facilitate the efficient geographical search over Supercharger locations. Geohashing is a technique of encoding geographical locations (ie. GPS coordiantes) as strings that represent rectangle bounding boxes of customizable precision. The longer the encoded string is, the more precise the geographic bounding box (Lee et al., 2014). This specialized encoding and the attribute of customizable precision is very useful in identifying if 2 locations are "close" to each other. We use this concept to efficiently search geographically for items that fall within a certain distance from one another and reduce the need to use the Google Directions API for every pair of Supercharger locations in our networks.

Illustrated in Figure B.1 are example expanded bounding boxes for the Denver Supercharger at geohash precisions 2,3 and 4

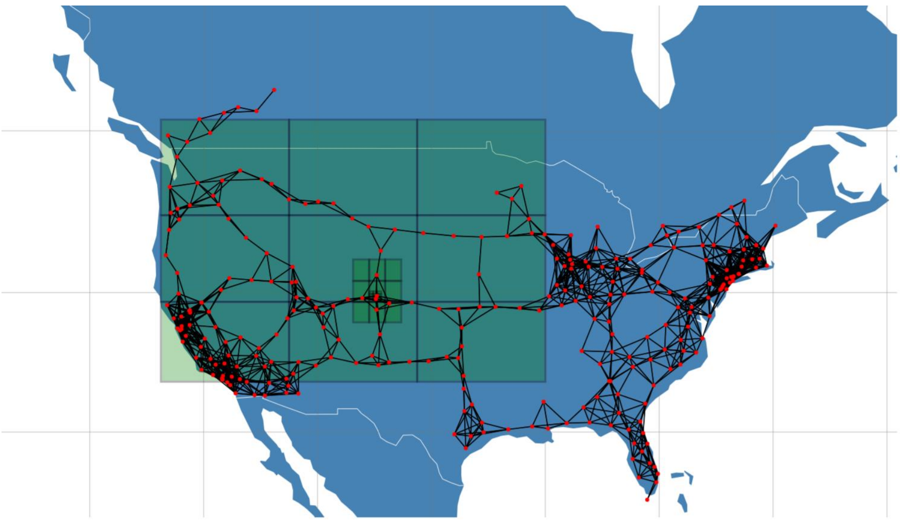

FIGURE B.1: Example Geohash Bounding Boxes for Denver Supercharger at precisions 2,3 and 4

One draw back to this technique is the large differences in precision between precisions 2, 3 and 4. However, using iterative expansions of any desired precision, we can create more refined bounding boxes of any size. This concept is illustrated in Figure B.2.

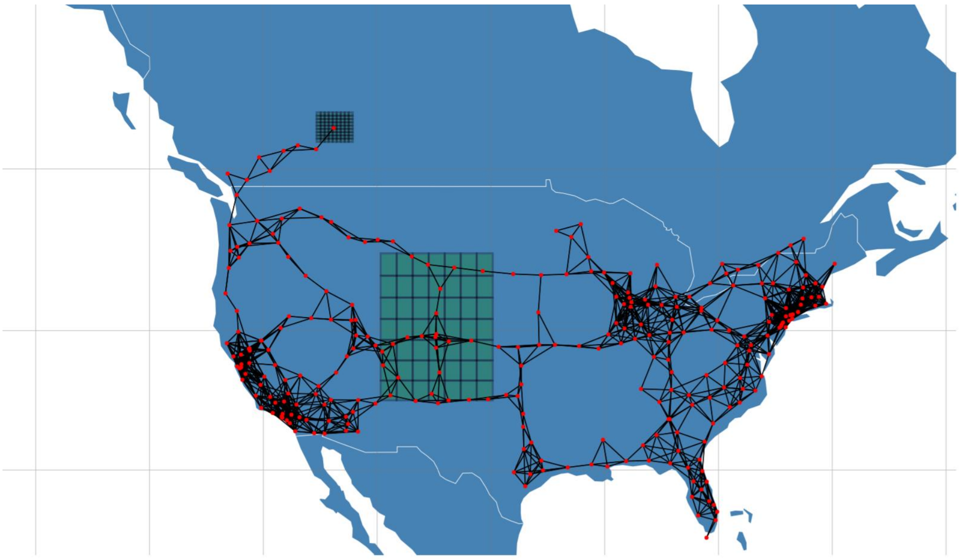

FIGURE B.2: Example Geohash Bounding Boxes with arbitrary precision

Using the above techniques, we can find all Supercharger that are "close" (for any arbitrary definition of close), to any other Supercharger efficiently as depicted in Figure B.3, drastically reducing the computational (and financial) costs of querying all pairs of Supercharger locations through the Google Directions API.

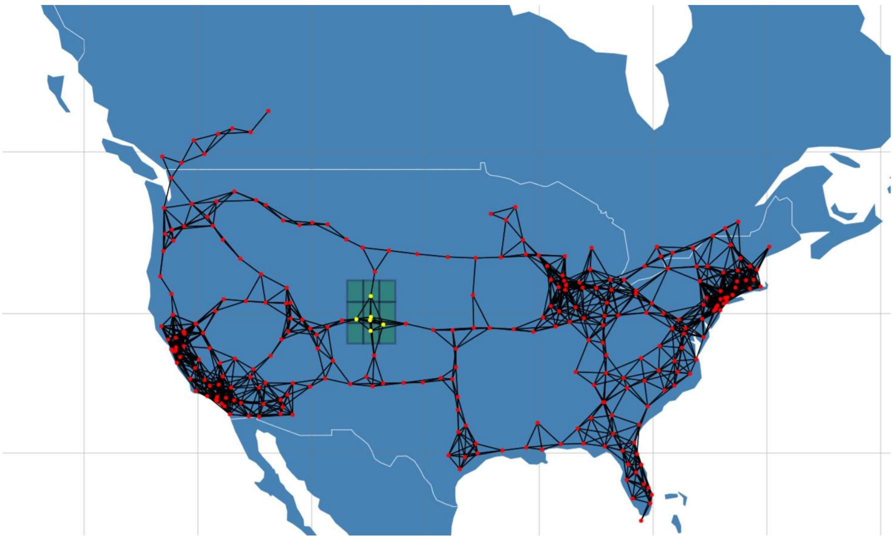

FIGURE B.3: Example Geohash Nearest Neighbors Search

The bounding boxes for all nodes in the current network can then be combined to establish all reachable locations of the current network, providing the expansion search space of the Supercharger network which is show in Figure B.4.

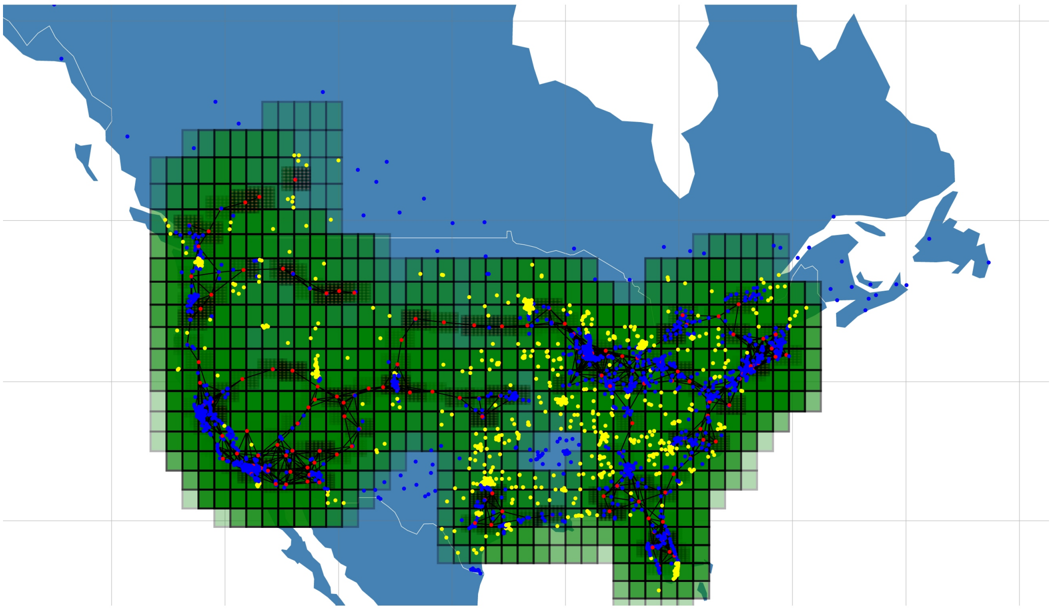

FIGURE B.4: Supercharger Network Geohashed Expansion Search Space

Bibliography

Alexander-Bloch, Aaron F et al. (2012). "The anatomical distance of functional connections predicts brain network topology in health and schizophrenia". In: *Cerebral cortex*, bhr388.

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz (2014). "powerlaw: a Python package for analysis of heavy-tailed distributions". In: *PloS one* 9.1, e85777.

Austwick, Martin Zaltz et al. (2013). "The structure of spatial networks and communities in bicycle sharing systems". In: *PloS one* 8.9, e74685.

Barabási, Albert-László (2009). "Scale-free networks: a decade and beyond". In: *science* 325.5939, pp. 412–413.

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *science* 286.5439, pp. 509–512.

Barthélemy, Marc (2011). "Spatial networks". In: *Physics Reports* 499.1, pp. 1–101.

Betzel, Richard F and Danielle S Bassett (2016). "Multi-scale brain networks". In: *NeuroImage*.

Bullmore, Ed and Olaf Sporns (2012). "The economy of brain network organization". In: *Nature Reviews Neuroscience* 13.5, pp. 336–349.

Bullmore, Edward T and Danielle S Bassett (2011). "Brain graphs: graphical models of the human brain connectome". In: *Annual review of clinical psychology* 7, pp. 113–140.

Cardillo, Alessio et al. (2006). "Structural properties of planar graphs of urban street patterns". In: *Physical Review E* 73.6, p. 066107.

Clark, Brent N, Charles J Colbourn, and David S Johnson (1990). "Unit disk graphs". In: *Discrete mathematics* 86.1-3, pp. 165–177.

Crucitti, Paolo, Vito Latora, and Sergio Porta (2006). "Centrality measures in spatial networks of urban streets". In: *Physical Review E* 73.3, p. 036125.

Dall, Jesper and Michael Christensen (2002). "Random geometric graphs". In: *Physical review E* 66.1, p. 016121.

Dorogovtsev, Sergey N, Alexander V Goltsev, and José FF Mendes (2008). "Critical phenomena in complex networks". In: *Reviews of Modern Physics* 80.4, p. 1275.

Ercsey-Ravasz, Mária et al. (2013). "A predictive network model of cerebral cortical connectivity based on a distance rule". In: *Neuron* 80.1, pp. 184–197.

Erdős, Paul and Alfréd Rényi (1961). "On the strength of connectedness of a random graph". In: *Acta Mathematica Hungarica* 12.1-2, pp. 261–267.

Fujita, André, Maciel C. Vidal, and Daniel Y. Takahashi (2017). "A Statistical Method to Distinguish Functional Brain Networks". In: *Frontiers in Neuroscience* 11, p. 66. ISSN: 1662-453X. DOI: 10.3389/fnins.2017.00066. URL: http://journal.frontiersin.org/article/10.3389/fnins.2017.00066.

Gastner, Michael T and Géza Ódor (2016). "The topology of large Open Connectome networks for the human brain". In: *Scientific reports* 6.

Gaurdian. https://www.theguardian.com/technology/2017/apr/10/tesla-most-valuable-car-company-gm-stock-price. Accessed: 2017-04-10.

Guimera, Roger et al. (2005). "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles". In: *Proceedings of the National Academy of Sciences* 102.22, pp. 7794–7799.

Gupta, Piyush and Panganamala R Kumar (1998). "Critical power for asymptotic connectivity". In: *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on*. Vol. 1. IEEE, pp. 1106–1110.

Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart (2008). "NetworkX". In: *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA USA, pp. 11–15.

Horvát, Szabolcs et al. (2016). "Spatial embedding and wiring cost constrain the functional layout of the cortical network of rodents and primates". In: *PLoS Biol* 14.7, e1002512.

Kurant, Maciej and Patrick Thiran (2006). "Extraction and analysis of traffic and topologies of transportation networks". In: *Physical Review E* 74.3, p. 036114.

Lambiotte, Renaud et al. (2008). "Geographical dispersal of mobile communication networks". In: *Physica A: Statistical Mechanics and its Applications* 387.21, pp. 5317–5325.

Lämmer, Stefan, Björn Gehlsen, and Dirk Helbing (2006). "Scaling laws in the spatial structure of urban road networks". In: *Physica A: Statistical Mechanics and its Applications* 363.1, pp. 89–95.

Lee, Kisung et al. (2014). "Efficient spatial query processing for big data". In: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, pp. 469–472.

Liben-Nowell, David et al. (2005). "Geographic routing in social networks". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.33, pp. 11623–11628.

Makowski, Armand M and Osman Yagan (2013). "Scaling laws for connectivity in random threshold graph models with non-negative fitness variables". In: *IEEE Journal on Selected Areas in Communications* 31.9, pp. 573–583.

Masuda, Naoki, Hiroyoshi Miwa, and Norio Konno (2004). "Emergence of scale-free networks from thresholding of vertex weights". In: *Phys. Rev. E* 70.cond-mat/0403524, p. 036124.

Masuda, Naoki, Hiroyoshi Miwa, and Norio Konno (2005). “Geographical threshold graphs with small-world and scale-free properties”. In: *Physical Review E* 71.3, p. 036108.

McColm, Gregory L (2004). “Threshold functions for random graphs on a line segment”. In: *Combinatorics, Probability and Computing* 13.03, pp. 373–387.

Nemeth, G and G Vattay (2003). “Giant clusters in random ad hoc networks”. In: *Physical Review E* 67.3, p. 036110.

Newman, Mark EJ (2003). “The structure and function of complex networks”. In: *SIAM review* 45.2, pp. 167–256.

– (2005). “Power laws, Pareto distributions and Zipf’s law”. In: *Contemporary physics* 46.5, pp. 323–351.

O’Dea, Reuben, Jonathan J Crofts, and Marcus Kaiser (2013). “Spreading dynamics on spatially constrained complex brain networks”. In: *Journal of The Royal Society Interface* 10.81, p. 20130016.

“On the evolution of random graphs” (1960). In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1, pp. 17–60.

Papo, David et al. (2014). *Complex network theory and the brain*.

Pele, Ofir and Michael Werman (2009). “Fast and robust earth mover’s distances”. In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 460–467.

Penrose, Mathew D et al. (2016). “Connectivity of soft random geometric graphs”. In: *The Annals of Applied Probability* 26.2, pp. 986–1028.

Reilly, Elizabeth Perez and Edward R Scheinerman (2009). “Random threshold graphs”. In: *the electronic journal of combinatorics* 16.1, R130.

Reus, Marcel A de et al. (2014). “An edge-centric perspective on the human connectome: link communities in the brain”. In: *Phil. Trans. R. Soc. B* 369.1653, p. 20130527.

RGG Conference 2016. http://www.birs.ca/events/2016/5-day-workshops/16w5095. Accessed: 2017-06-03.

Simon Makin. *Scientists Surprised to Find No Two Neurons Are Genetically Alike*. "" https://www.scientificamerican.com/article/scientists-surprised-to-find-no-two-neurons-are-genetically-alike/s". Accessed: 2017-06-03.

Snyder, John Parr (1987). *Map projections–A working manual*. Vol. 1395. US Government Printing Office.

Solovey, Kiril and Dan Halperin (2016). “Sampling-based bottleneck pathfinding with applications to Fréchet matching”. In: *arXiv preprint arXiv:1607.02770*.

Song, H Francis, Henry Kennedy, and Xiao-Jing Wang (2014). “Spatial embedding of structural similarity in the cerebral cortex”. In: *Proceedings of the National Academy of Sciences* 111.46, pp. 16580–16585.

Sporns, Olaf (2010). *Networks of the Brain*. MIT press.

Stephan, Klaas E et al. (2000). “Computational analysis of functional connectivity between areas of primate cerebral cortex”. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 355.1393, pp. 111–126.

supercharger.info. https://supercharge.info/. Accessed: 2017-04-15.

Tesla. https://www.tesla.com/supercharger. Accessed: 2017-05-31.

Tesla. https://www.tesla.com/model3. Accessed: 2017-06-03.

The Economist. *Elon Musk enters the world of brain-computer interfaces*. "http://www.economist.com/news/science-and-technology/21719774-do-human-beings-need-embrace-brain-implants-stay-relevant-elon-musk-enters". Accessed: 2017-06-03.

Towlson, Emma K et al. (2013). "The rich club of the C. elegans neuronal connectome". In: *Journal of Neuroscience* 33.15, pp. 6380–6387.

Waxman, Bernard M (1988). "Routing of multipoint connections". In: *IEEE journal on selected areas in communications* 6.9, pp. 1617–1622.