POLYTECHNIC UNIVERSITY OF CATALONIA

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

Department of Informatics
Engineering and Mathematics
Rovira i Virgili University

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

in

Complex Networks
Barcelona School of Informatics

May 31, 2017

Polytechnic University of Catalonia

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 Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too…

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Random Graphs and Complex Networks

1.2 Spatial Networks

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 an obvious element for most spatially embedded systems, such as infrastructure or biological networks.

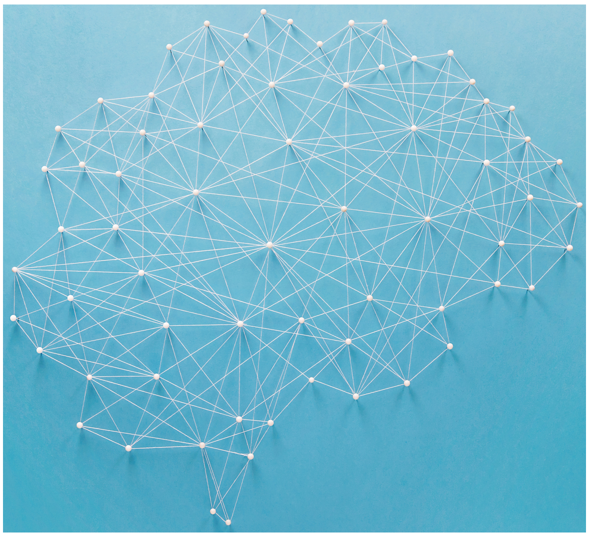

FIGURE 1.1: Artist rendering of the human brain, highlighting the spatial elements of complex networks

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 the most common Spatial Network models. In this paper, we investigate the relationships between the most common Spatial Network models and propose a novel classification of these models by considering their shared parameterization. Specifically, we show that the most common Spatial Network models are obtained by combinations of only 3 parameters. We also propose a new Spatial Network model, Thresholded Random Geometric Graphs, extending Random Geometric Graphs in adopting the threshold parameter, θ.

Shared Model Parameters

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

For all Spatial Network models considered in this work, nodes are uniformly distributed onto the unit square with weights, w, sampled from a power law distribution, $f(w)$, controlled by the additional power law exponent parameter, λ, found in Equation 1.1.

$$f(w) = w^{-\lambda} \tag{1.1}$$

Distance, d_{ij}, is assumed to be the Euclidean distance unless otherwise stated, but in general can be any measure of distance between entities, including non-spatial measures. Figure 1.2 shows the relationships between Spatial Network Models connected by their shared parameterizations and highlights with dashed lines the new model proposed in this paper.

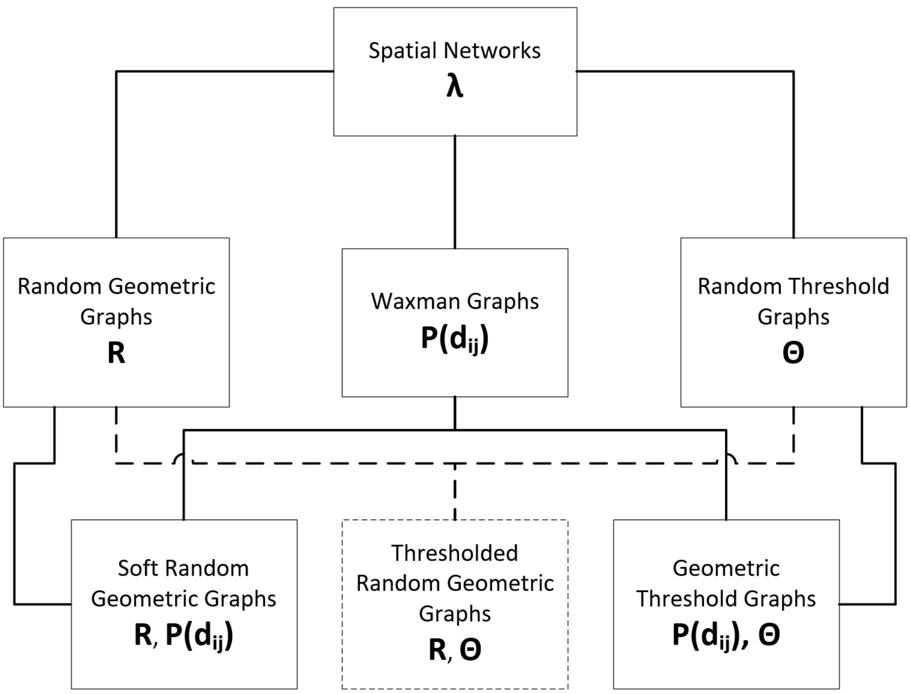

FIGURE 1.2: Spatial Network Model Classification

1.2.1 Spatial Network Models

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. In this section, we introduce the definitions for each of the models depicted in Figure 1. Note that for each definition, the notation E_{ij} indicates an edges exists between nodes i and j.

Random Geometric Graphs - R

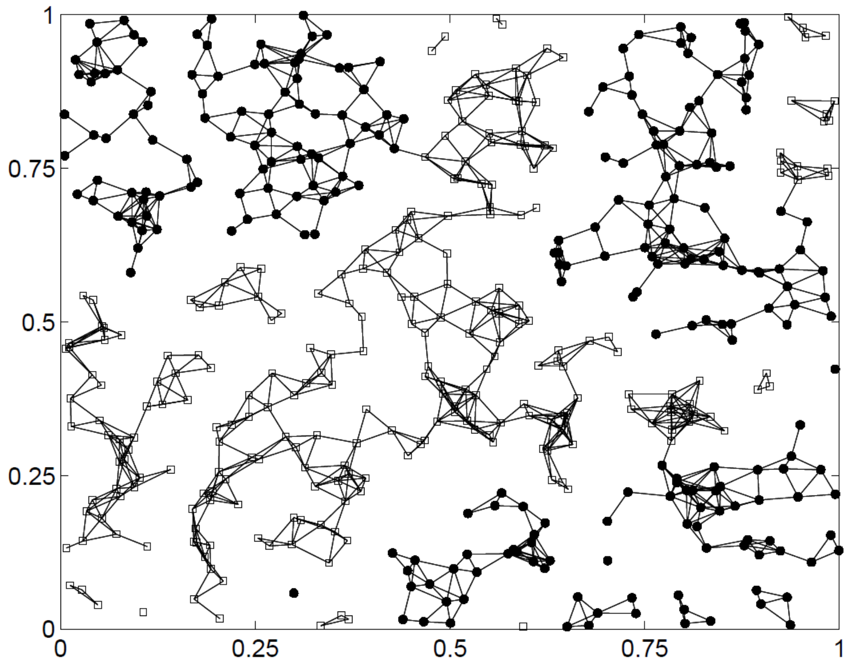

FIGURE 1.3: A 2D Random Geometric Graph with N = 500 and α = 5 Dall and Christensen, 2002

A d-dimensional Random Geometric Graph (RGG) 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 1.2

$$E_{ij} = d_{ij} \leq R \tag{1.2}$$

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 1.3.

$$P(d_{ij}) = \beta e^{-d_{ij}/L\alpha} \tag{1.3}$$

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.

Random Threshold Graphs - θ

A simple graph G is a threshold graph if we can assign weights to the vertices such that a pair of distinct vertices is adjacent 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.

$$E_{ij} = (w_i + w_j) \geq \theta \tag{1.4}$$

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

Geometric Threshold Graphs are the geographical generalization of Random Threshold Graphs, where a pair of vertices with weights w_i, w_j, and 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.

$$E_{ij} = (w_i + w_j)P(d_{ij}) \geq \theta \tag{1.5}$$

Soft Random Geometric Graphs - 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.

$$E_{ij} = \begin{cases} P(d_{ij}) & \text{if } d_{ij} \leq R \\ 0 & \text{if } d_{ij} > R \end{cases} \tag{1.6}$$

Chapter 2

Thresholded Random Geometric Graphs

2.1 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. 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 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 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 Nemeth and Vattay, 2003, can be represented with this new model to allow the consideration for which nodes to make connections for optimal routing given their respective connection utility weights.

Equation 2.1 formalizes the definition of Thresholded Random Geometric Graphs, where the edge, E_{ij}, exists if node i and j are within or equal to the maximum distance, R, and their summed weights, w_i, w_j, are greater than or equal to θ.

$$E_{ij} = \begin{cases} (w_i + w_j) \geq \theta & \text{if } d_{ij} \leq R \\ 0 & \text{if } d_{ij} > R \end{cases} \tag{2.1}$$

2.2 Properties of Thresholded Random Geometric Graphs

In this section, we study the connectivity of the newly proposed model and comment on the unique properties of the model compared the existing Spatial Network models.

2.2.1 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 2.2.

$$R_c = \sqrt{\frac{log(N)}{\pi N}} \tag{2.2}$$

Figure 2.1 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.

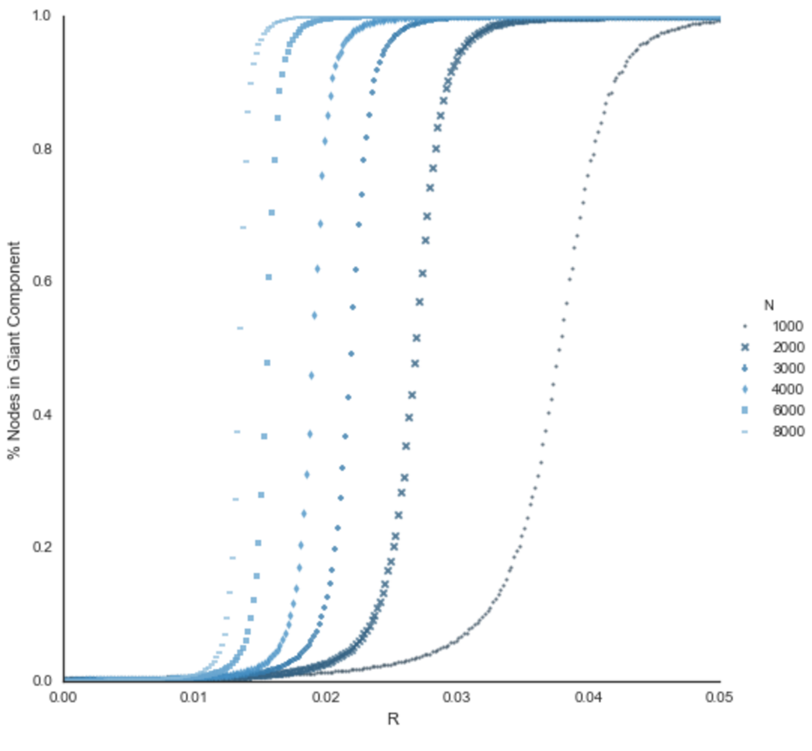

FIGURE 2.1: Connectivity Phase Transition of Random Geometric Graphs at Various Networks of Size N

We confirm the theoretical lower bound of Equation 2.2 numerically by iteratively simulating networks at each value of N until the minimum value of R, R_{min}, resulting in a fully connected network stabilizes over 100,000 runs of networks generated with R values between $[0.999R_{min}, R_{min}]$, which we use as an estimate for

R_c. See Appendix for more details on the numerical lower bounding of R_c. Figure displays the results of the lower bounding simulations, with the slope of the linear trend being $1/\sqrt{\pi}$ theoretically, as per Equation 2.2, which our simulation value of agrees with within percent.

2.2.2 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 networks are within the maximum distance, R, from each other and the model devolves into the Random Threshold Graph introduced in Section 1.2.1. 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 2.3 and the degree distribution is described by Equation 2.4 Masuda, Miwa, and Konno, 2004.

$$k = N[1 - F(\theta - w)] \tag{2.3}$$

$$p(k) = f(w)\frac{dw}{dk} = \frac{f[\theta - F^{-1}(1 - \frac{k}{N})]}{Nf[F^{-1}(1 - \frac{k}{N})]} \tag{2.4}$$

2.2.3 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 2.2 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 for 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 density, ρ for N in Equation 2.4 as shown in Equation 2.5.

$$p(k_{local}) = \frac{f[\theta - F^{-1}(1 - \frac{k}{\rho})]}{\rho f[F^{-1}(1 - \frac{k}{\rho})]} \tag{2.5}$$

With a uniform distribution of nodes embedded onto the unit square, a constant density, $\rho = \frac{\pi R^2}{1^2}$, nodes will exist in each circle and Equation 2.6 is obtained, which for uniformly distributed nodes is equivalent to the degree distribution of the entire network.

$$p(k_{local}) = p(k) = \frac{f[\theta - F^{-1}(1 - \frac{k}{\pi R^2})]}{\pi R^2 f[F^{-1}(1 - \frac{k}{\pi R^2})]} \tag{2.6}$$

Using the degree distribution of Equation 2.6, we can inspect the probability of a node having only a single edge, $k = 1$, as a function of θ and use that as an estimate for the upper bound of θ for connectivity of the network as a function of R, which fully characterizes the conditions of connectivity for Thresholded Random Geometric Graphs.

Chapter 3

A Motivating Example - Telsa Supercharger Network

3.1 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.

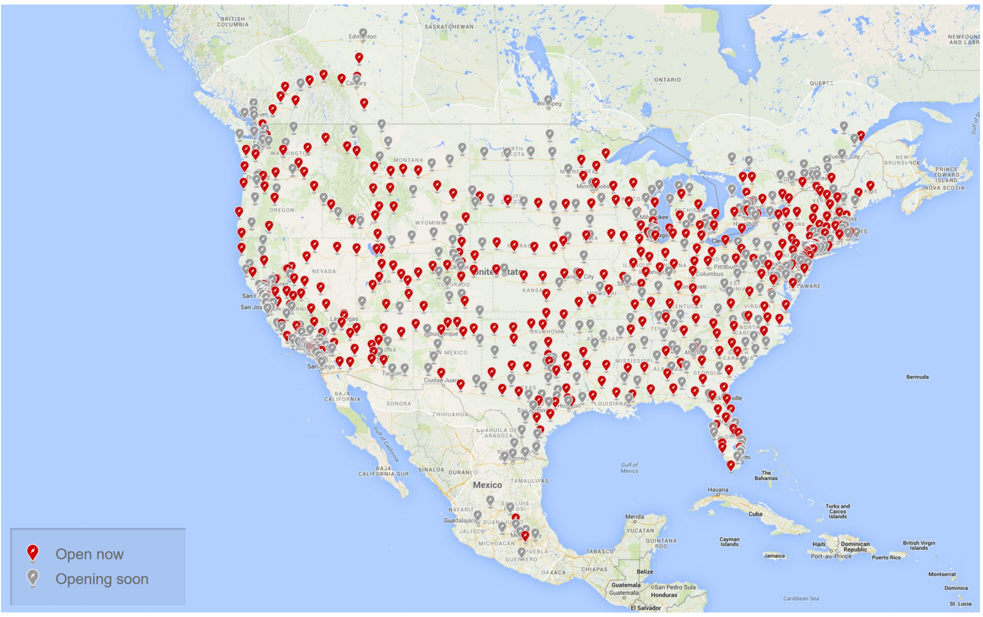

FIGURE 3.1: Tesla's Supercharger Network - 2017 *Tesla*

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.

3.1.1 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 regions 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) as determined by the driving distance computed by Google directions API between Supercharger GPS locations. Figure 3.2 depicts the representation of Figure 3.1 according to this network model definition, excluding Mexico and "Opening Soon" Superchargers.

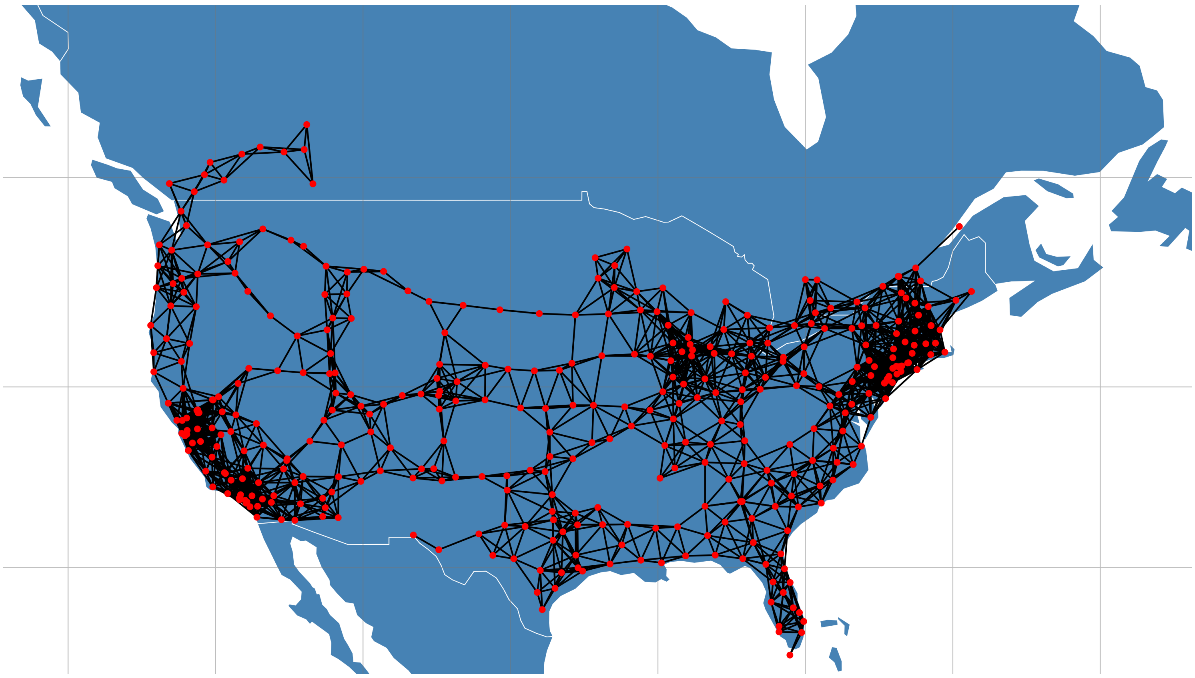

FIGURE 3.2: Tesla Supercharger Network Model

Using the Albers Equal-Area Conic Projection Snyder, 1987 with parameters enumerated in Table 3.1, the Supercharger Network can be projected onto the unit square as depicted in Figure 3.3. The same projection is applied to the top 5,000 cities by population for the United States and Canada depicted in Figure 3.4, providing the network expansion space available to simulate the growth of the Supercharger network on.

TABLE 3.1: 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

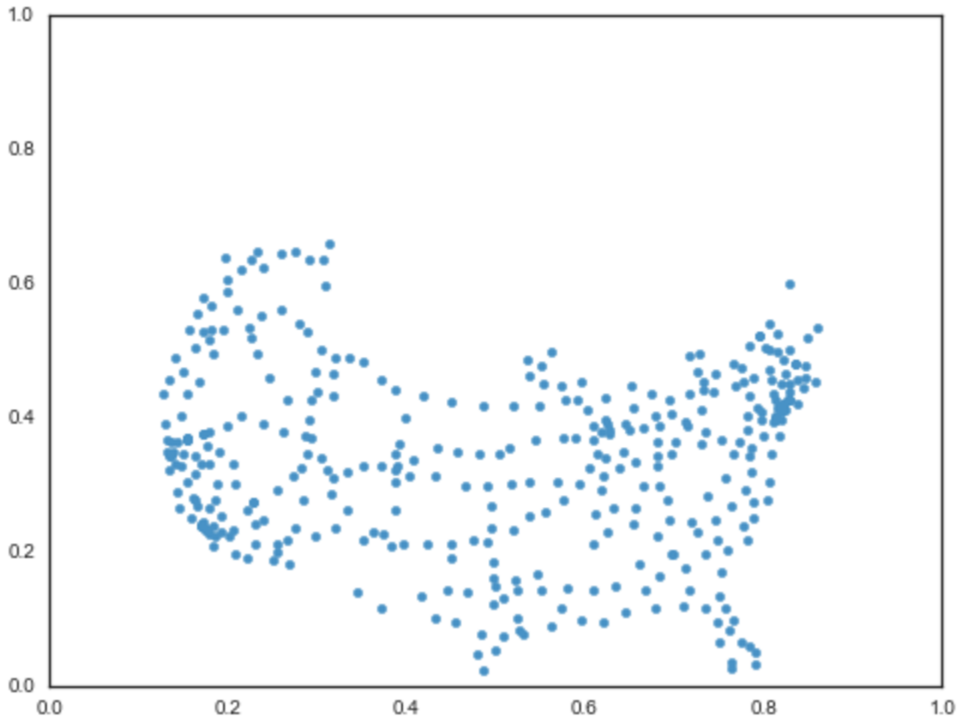

FIGURE 3.3: Telsa Supercharger Network Albers Projection onto the Unit Square

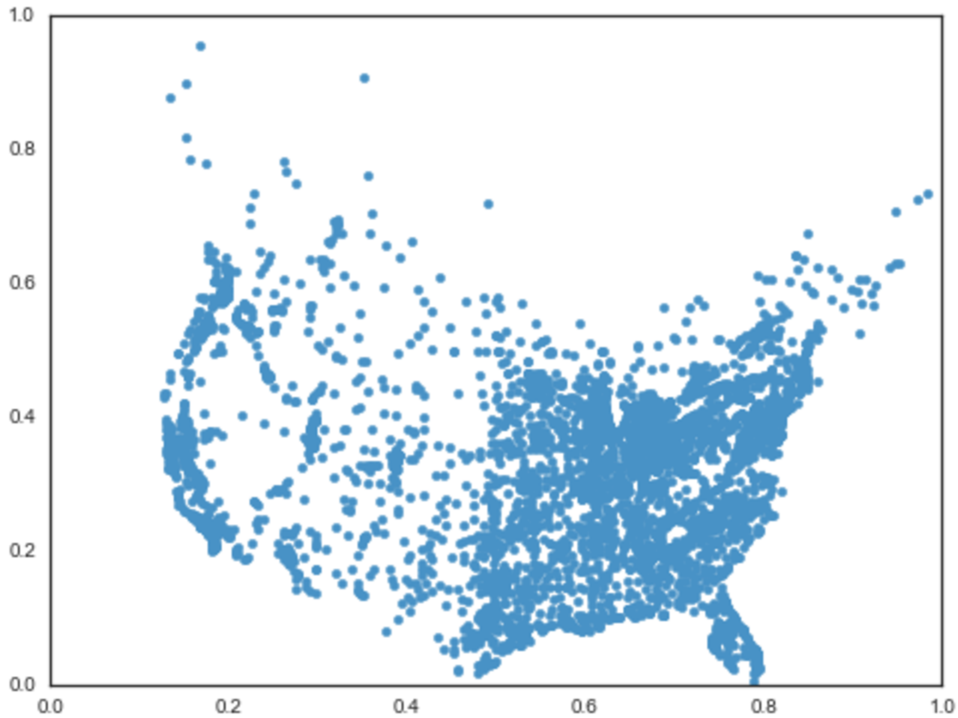

FIGURE 3.4: Top 5,000 Canadian and American Cities by population Albers Projection onto the Unit Square

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 1.2.1.

3.2 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.

3.2.1 Supercharger Network - λ

The first parameter we fit to the real data of the Tesla Supercharger network is λ, the powerlaw exponent of Equation 1.1 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 3.5 shows the real distribution of the city population data versus the fitted distribution.

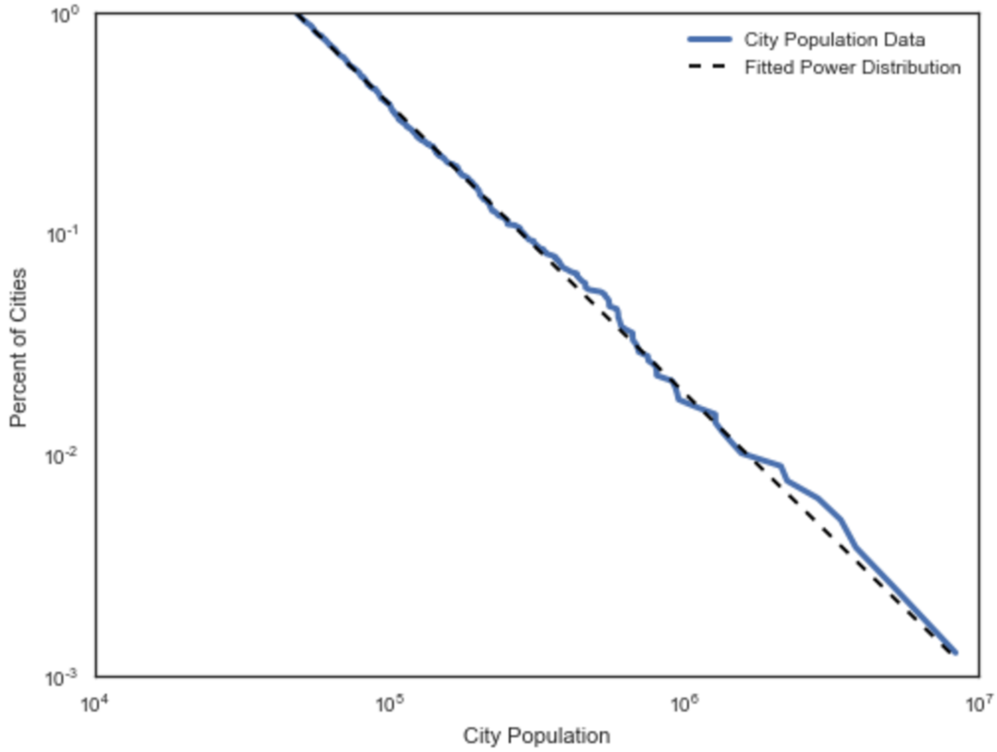

FIGURE 3.5: Top 5,000 Canadian and American City Populations Powerlaw fit $\lambda = 2.3$

3.2.2 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 3.1.1, namely the maximum base range of a Tesla model 3 (215miles - 346km), when scaled to the model's unit square project is $R = 0.0461$.

3.2.3 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 3.6 plots the Supercharger network's connection counts at each distance with a fitted Gaussian Kernel Density Estimate (KDE) of the distribution.

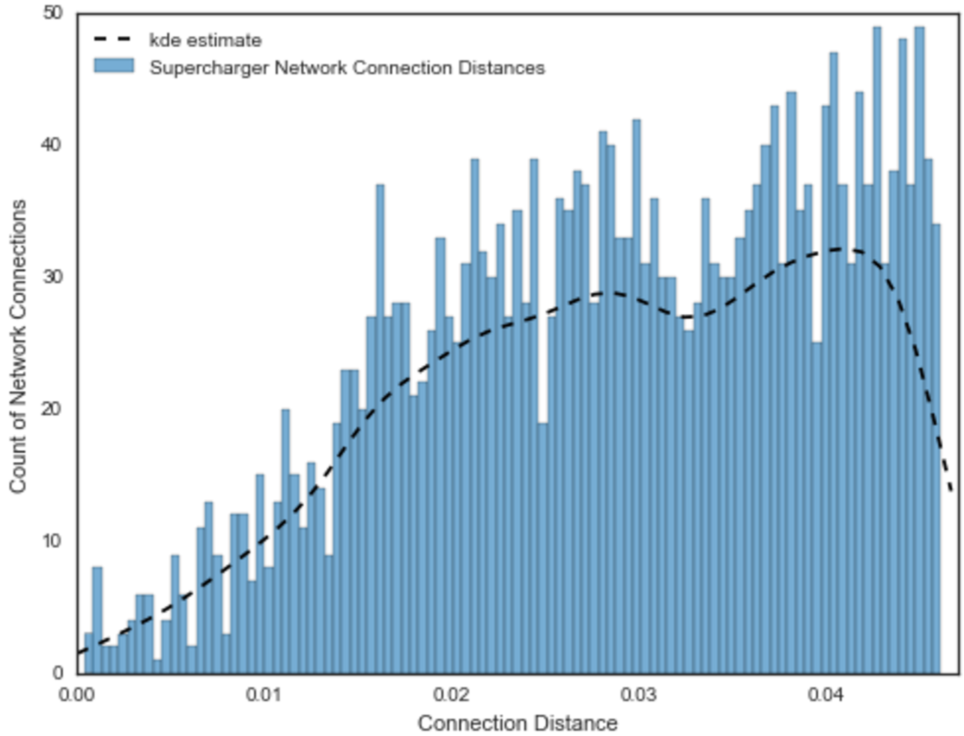

FIGURE 3.6: Supercharger Network Connection Probability Distribution, $P(d_{ij})$ with Gaussian KDE

An interesting observation of Figure 3.6 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 3.6 supports the claim of this paper that Thresholded Random Geometric Graphs may be better suited at modeling networks of this type.

3.2.4 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 selected the minimum. The distribution of connection weights is depicted in Figure 3.7, with a minimum connection weight which we estimate as $\theta = 0.034$.

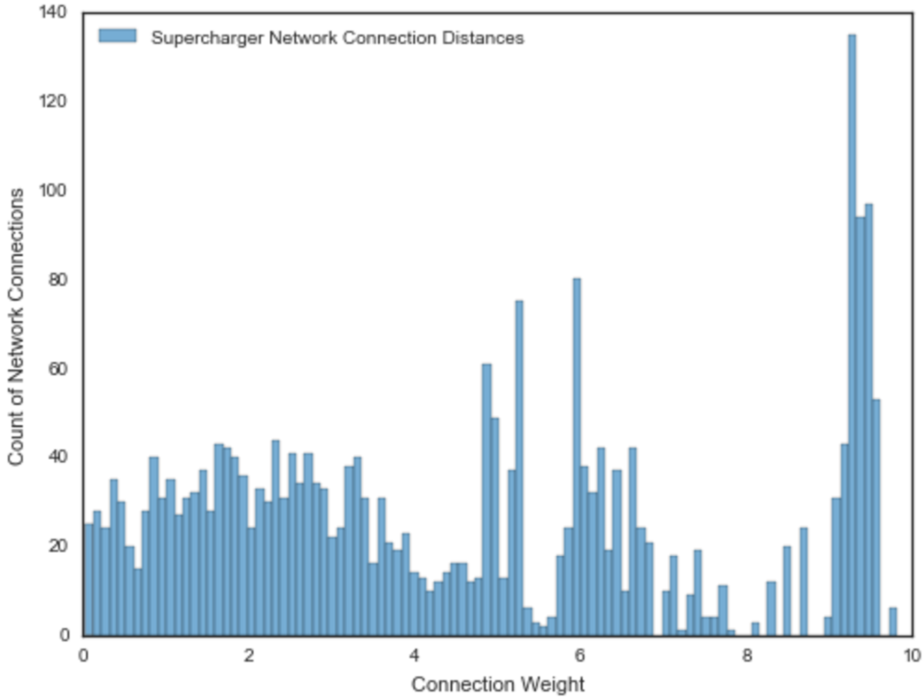

FIGURE 3.7: Supercharger Network Connection Weight Distribution

3.2.5 Supercharger Network - Parameter Summary

Table 3.2 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 3.2: Tesla Supercharger Network Spatial Network Parameter Estimates

Parameter	**Value**
λ	2.3
R	0.0461
$P(d_{ij})$	KDE Estimate
θ	0.034

3.3 Supercharger Network Spatial Network Modeling

With the Spatial Network parameter estimates obtained in Section 3.2, we can parameterize the Spatial Network models of Section 1.2.1. 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. For each of these models, we generate 100 examples using the model definition and parameterization of the above sections, and compare each models ability to generate the real Supercharger network. We then test the models generalization capabilities by comparing how accurate they are in predicting the future proposed Supercharger locations not included in the models' parameterization estimates.

Chapter 4

Future Work and Conclusions

4.1 Main Section 1

Bibliography

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

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

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

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

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

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.

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.

— (2005). "Geographical threshold graphs with small-world and scale-free properties". In: *Physical Review E* 71.3, p. 036108.

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

Newman, Mark EJ (2005). "Power laws, Pareto distributions and Zipf's law". In: *Contemporary physics* 46.5, pp. 323–351.

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.

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

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

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