QUANTIFYING INCONVENIENCE IN INCOMPLETE URBAN STREET NETWORKS: A NEW METRIC

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Ankur Mathur

August 2017

ABSTRACT

Incomplete networks are those in which a road/intersection are unavailable to route through. Information centrality has been used to quantify the efficiency loss due to incompleteness. We propose a new topological method to quantify this by summing the excess distances one must travel. The new metric (SED) is found to be significantly correlated with IC across three representative networks. It is distributed Weibull and we provide a theoretical basis as to why. IC is distributed as a power law with varying exponents. The research then proposes several metrics to rank networks based on different policy questions. From the IC one can rank by the network's inherent inequity. From the SED, one can rank per median/modal SED, percentage of most susceptible nodes, and excess CO_2 emitted. Finally, we propose how SED can be helpful in location setting and theorize the existence of a trade-off between SED and the network's operating cost.

BIOGRAPHICAL SKETCH

I was born in New Delhi, India in 1991 and moved to Mumbai, an overwhelming metropolis that is at once satisfying and horrifying in equal measure. All my life I have lived amidst tall structures, lengthy roads, dense populations and of course, traffic jams. The sound of a car horn and the school bell blur as they ring together in the nostalgia corner of my brain. I had decided then that I shall do something about them – the traffic jams, that is. So, I moved to Nagpur – in the center of a roasting oven that the hinterland of India is in the summers – to study Civil Engineering at a state school; and graduated with top honors, and a broken heart. Never mind that, it had turned me into a part-time poet, part-time editor and a full-time student. I began writing and soon found that the space I needed couldn't be found in the sardine-like atmospheres of the Indian research-scape. So, I packed again and landed in the USA. And this is where I am writing this biographical sketch from. Thinking of it like a prose-poem, but realizing fully well, that *you* have just turned the page.

Dedicated to my parents, Mr. Nishu Simlote and Mrs. Neera Simlote

ACKNOWLEDGMENTS

The footprints of this path have long washed,
Now I remain with memories, awashed
From who I received love thus far,
I leave gratitude's door, ajar.

I hope the quatrain above sums up my immense gratitude to everyone who has been part of this journey. I thank my first professor who chastised me into becoming a student for life, Dr. Ajay Tembhrukar of VNIT. Also, to my Bachelor's research guide, Dr. Landge who taught me the power of ethical research and truth.

I thank Dr. Oliver Gao for letting me pursue a line of thought which was new for me and for being there to force me to ask tough questions of myself.

To Dr. Samaranayake who with his new thinking and frontline research inspired me to think something new every day.

To the Staff in Hollister Hall for their quick conversations that helped me through the day. To the staff at Uris and Olin Libraries who have put in their hours to help me find the resources that resulted in this research.

To Ms. Andrea Cohen (M. Eng. '17) for helping with plotting the risk profile better. To my friends both in Ithaca and in India for existing. To Hassan (M. Eng. '16) for putting up with me and my nocturnal rants.

To my parents who have trusted me in doing something good.

And finally, to you, the reader for vicariously sharing in the pleasure of discovery.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AM	Adjacency Matrix
BM	Base Matrix
DM	Difference Matrix
EM	Euclidean Distance Matrix
IBR	Inverse Base Removal Matrix
IC	Information Centrality
IRM	Inverse Removal Matrix
ISR	Inverse Sum Removed Matrix
MLE	Maximum Likelihood Estimator
OD	Origin - Destination Matrix
PMSN	Percentage of Most Susceptible Nodes
RM	Removal Matrix
SED	Sum of Excess Distances

LIST OF SYMBOLS

C_i^I	Information Centrality of node *i*
d	Number of persons who die per incident
d_{Euclid}	The shortest distance between I and j as the crow flies
d_{ij}	Shortest distance between *i* and *j*
E	Efficiency of graph G
E_i	the expected number of count of values in the i-th bin
F	Frequency of event per annum that kills *d* persons
F_i	The *i*-th plotting point for a distribution
G	A network graph representing urban street topology
k	Shape Factor of a Weibull distribution
L	Likelihood Equation
O_i	The Observed number of Count of Values in the *i*-th bin
P	Cumulative number of nodes in G holding W percentage of the information
W	Cumulative Percentage of information held in P nodes
α	The exponent of a Power Law
λ	The Rate of an Exponential distribution
λ	Scale factor of a Weibull Distribution

PREFACE

Before you, is a thesis titled "Quantifying Inconvenience in Incomplete Urban Street Networks: A New Metric" which has been written to fulfill requirements for the degree of Masters of Science in Transportation Systems Engineering at Cornell University. The basis of this research is a study conducted on three representative street networks using an old and this new metric. I was engaged in writing this thesis from October 2016 to June 2017.

The research question was formulated through several back and forth conversations between me and my very respected advisor, Dr. Oliver Gao who graciously allowed me to pursue my line of thinking and work on a new topic. The research began with a few hiccups – namely finding usable datasets, formulating the idea into a mathematically representative concept and automating the calculations using a computer code. In the end, though, it has presented to me wonderful results that I hope will help you in understanding why we need better designed cities than the ones we live in.

I thank my advisors Messers O. Gao and S. Samaranayake whose continues support and trust has led me, in the words of Tennyson, "to strive, to seek, to find, and not to yield". I am also thankful to everyone at Cornell, at my previous alma mater, VNIT and all my friends whose wonderful and awe inspiring work I have always looked up to.

I think it is also apt to thank Ithaca, a city I have called home through two years and if I have ever been down or dull or uninspired, its gorges and trails and walks and scenes have filled me with hope and faith. I hope you find it an interesting read.

Ankur Mathur.

Ithaca, New York.

June 7, 2017.

CHAPTER 1

INTRODUCTION

What is this life, if full of care, We have no time to stand and stare.
-W. H. Davies

1.1 Urbanization over the Years

Cities have existed from ancient times. Remains from the Indus Valley Civilization show us that the cities of Mohenjo Daro and Harappa have existed between 2600 BCE to 1600 BCE and were known for their urban planning – citadels, towns and dedicated burial areas (Vahia & Yadav, 2010). Other civilizations too, over the years, have had their contributions to city making and planning. The Romans, for example, are famously attributed to have formalized transportation in building the Appian Way.

Cities today have come a long way in keeping populations orderly and with ease. However, every now and then, we do hear of incidences where people must face diversions, road blockages, and the like to do several issues. Some of these are:

a. Violence leading to protection of forensic evidence
b. Road Maintenance
c. Low lying streets are flooded first in a network
d. Special needs, beyond the normal need for law enforcement (City of Indianapolis vs Edmond, 2000)
e. Passage of motorcade of Presidents, or Guests of the State

While these cause inconveniences, there is no doubting that they are part of the driving experience in cities across the world. The result is a longer drive which avoids the affected areas. This, coupled with the fact that more and more, the population of the world is living in cities, makes availability of the road segments an important aspect to investigate in the field of transportation. In this chapter, we discuss the trends, over the years of urbanization and how this trend will develop over the future. We also delve deeper into the phenomenon of violence in the public realm which leads to loss of life and prove that such incidents, albeit of smaller magnitude, happen with a high frequency across the world proving that even if the last four of the points mentioned above do not happen often, there are enough incidents of violence that lead to incomplete networks making this research pertinent.

The United Nations released the World Urbanization Prospects report in 2014 (United Nations, 2014) which is a benchmark publication in the development of cities across the world. It reported that for the first time, more people across the world lived in cities, at 54% and forecasted 66% for 2050. The following figure tells the story of urbanization. We see that the wealthier parts of the world have always been better off than the rest. The most remarkable progress story has been China whose curve covers the most sweep. Although India, the other most rapidly urbanizing country in the world started at a relatively better off position, it's climb has not been as good.

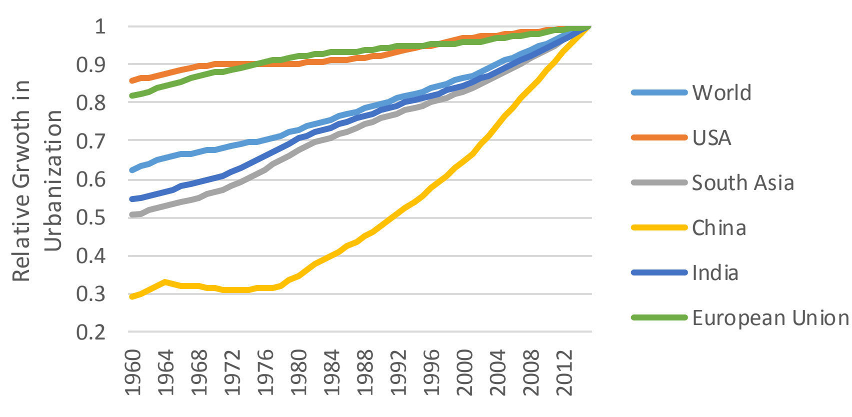


Figure 1: Relative Growth of Major Economies/Blocs[1]

[1] Source: World Bank

When we discuss the sheer percentage of people living in urban areas, the plot is:

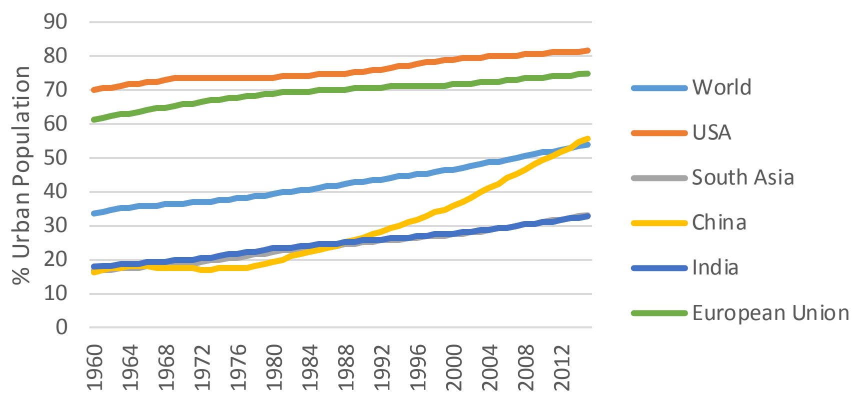
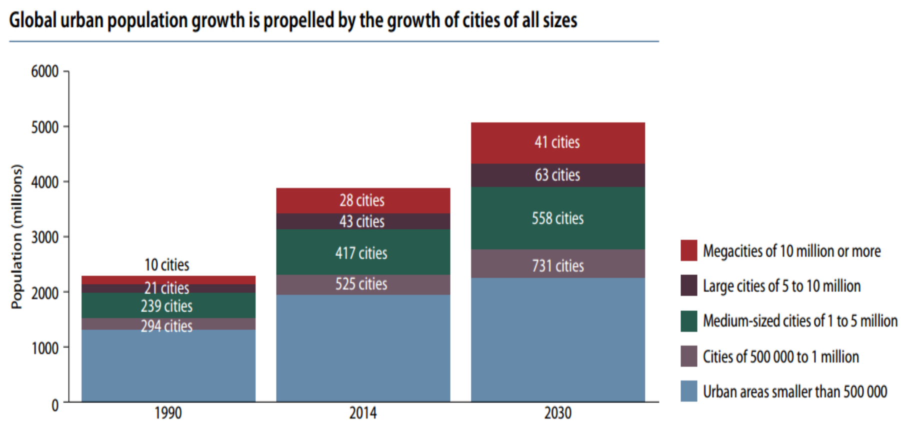


Figure 2: Growth in Urban Population Growth in Major Economies/Blocs[2]

We see again, that the world has had a steady rise in urbanization but China's growth has been exponential, ever since the late 1970s. It surpassed India's in 1990 and became better off than the world average in this decade. The report states further that India, China and Nigeria are expected to add 908 million new urban dwellers in the next 32 years.

This is going to mean that a number of cities that shall be needed to house these people will also have to be built. We see that trend also gets exhibited in the United Nations' forecasts. While there are 28 cities across the world of more than 10 million or more people, there are expected to be 41 such cities by 2030. Other urban centers will have to rise in number too. It could be claimed that cities that will house the population of the future have not been built yet. However, the largest chunk of urban population will still reside in relatively smaller cities of 500 000 or less people. This means that small, compact urban centers shall be the representative city of the future, more than it is now. The following figure from the report (United Nations, 2014) tells the story.

[2] Source: World Bank



[3]

Figure 3: Cities as Growth Centers

1.2 Advantages of Urbanization

This immense push for pushing population out of rural areas and into cities brings with it great economic success. By clustering people into well-defined urban boundaries social services like policing, education, health can be better administered. It also provides the governments with a clearer unit of population where targeted intervention can be introduced and results monitored.

First, we can show that higher percentage of urban population is correlated with a lower poverty percentage across countries. The worldwide data on percentage of people living below the poverty line of $1.9 per day and the percentage of urban population are highly negatively correlated (cf. Figure 4). The correlation is -0.98 and statistically significant. The causality is evident – we would expect that as people move into the cities, their incomes rise and this leads to a reduction in poverty figures.

Further, we can also show that this trend works at the scale of individual countries as well.

[3] Source: United Nations World Urbanization Prospects, 2014

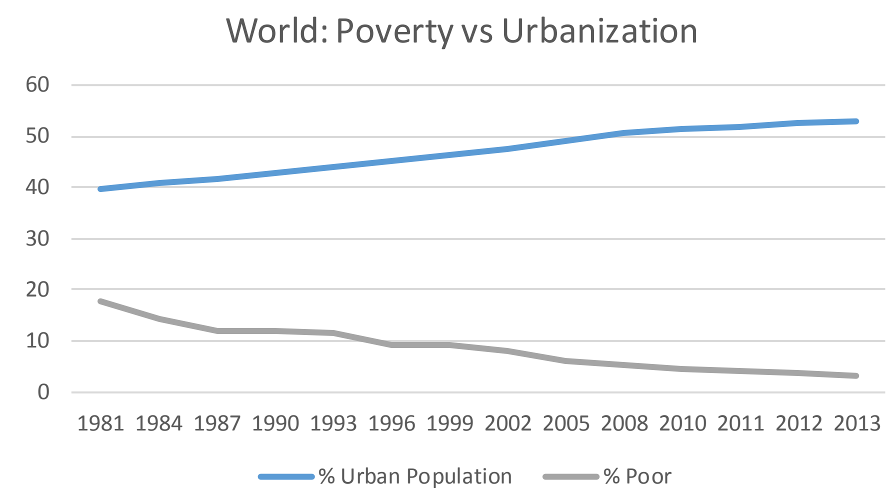


Figure 4: Urbanization and Poverty[4]

We claimed before that the living in cities is also linked to better civic facilities. We shall now explore this aspect of urban life. We see from the plots that follow that education and health parameters also show improvement as the urban population increases.

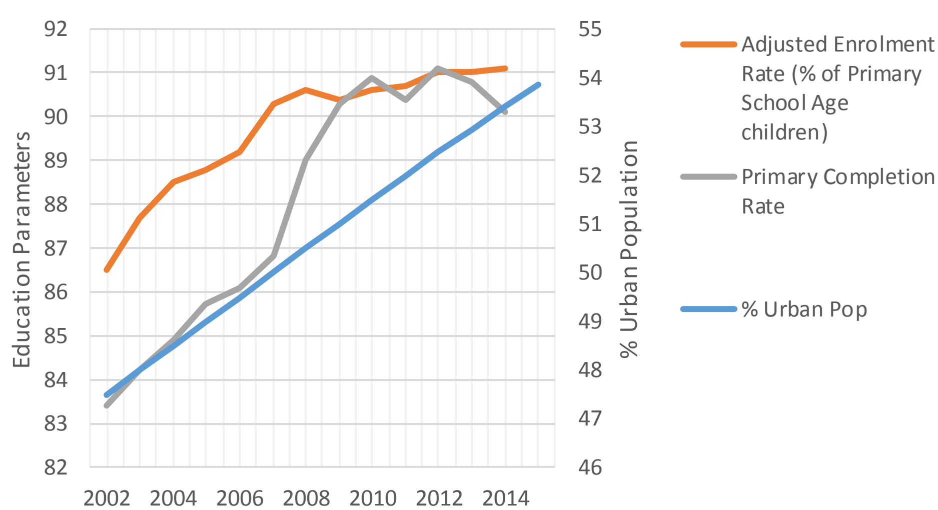


Figure 5: Urbanization and Education Parameters

[4] Source: World Bank and author's calculations based on the data therefrom.

Here's a look at the effect of urbanization on global education. There is no denying that the enrolment and the completion rate in primary schools increases with increasing urban population numbers. Insofar as completion rates are concerned, there is no doubt other factors as work proven by the concavity around the year 2007 which accelerated the rate of primary school completion. Such effects are also seen in parameters of health. For example, we see that Adolescent Fertility Rates (as proxy for access to medical services like medical contraception which are prevalent in urban areas more than in rural areas) is inversely proportional to urbanization.

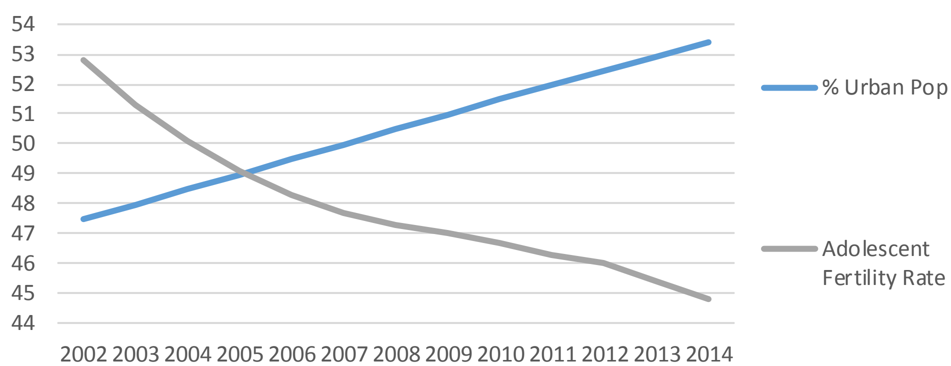


Figure 6: Urbanization and Health Parameter

We conclude from the discussion above that several desired parameters – both social and economic – are closely correlated with urbanization. Therefore, improving cities and livability if cities is of paramount importance.

The Economist's Intelligence Unit, which measures the livability of cities across the world reported that livability in the cities has been declining in a fifth of all those surveyed on five accounts of which infrastructure and stability contributed 45% of the weight (The Economic Intelligence Unit, 2016). Data from the same report also brings to light a stark reality – when compared with the data from 2011, in 2016, the rise in cities improving their rank was much less than the magnitude of degradation for which whose ranking decreased resulting thus, in a net worsening of the situation. (The Data Team, The Economist Intelligence Unit, 2016). This can also be seen from the figure below where the red dominates the blue both in number and spread.

Figure 7: Change in Livability Scores[5]

We now show how urbanization is linked with this. The following plot tells the story that countries with the best cities have higher urbanization than those with lower.

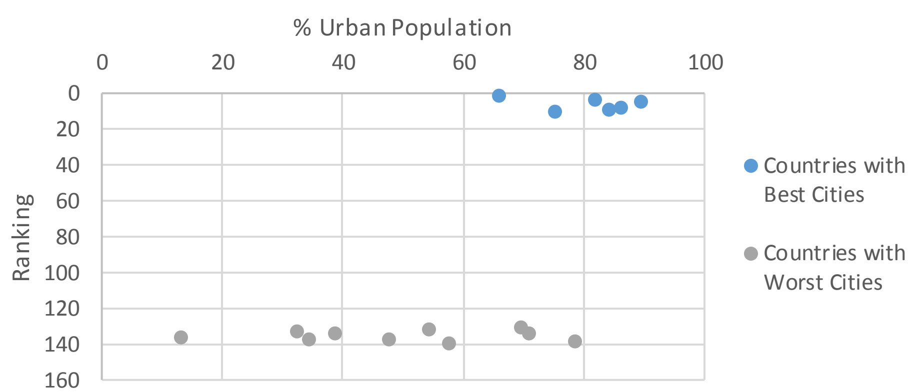


Figure 8: Livability Scores and Urbanization[6]

[5] Source: The Economic Intelligence Unit 2016

[6] Source: Author's calculations based on data therefrom

Their averages (80.5% versus 49.72%) differ statistically significantly ($p = 0.001$).

In all this discussion, there is also the role of security. As indicated initially, a large part of keeping people moving throughout the city is to have roads available to run vehicles on them. We know that populations across the world are rising and settling into cities. And so, the pressure to maintain road availability is only going to increase in the future. We had also indicated several possible reasons. We shall now tackle some of those issues to highlight the extent to which road blockages affect our daily commute.

1.3 Frequency of Small Violent Incidences

How impactful are these incidents? This a question that can be answered by calculating a risk profile of the incidents. For this purpose, the data (maintained at the Database of Worldwide Terrorism Incidents by the Rand National Defense Research Institute Project)[7] on all the incidents was classified into classes denoting the number of deaths caused due to them. Then the frequency of their occurrence was calculated by dividing the total number of incidents in every class by the number of years of data, in this case, 42. Then a probability of the number of people dying, d, in these incidents was calculated and multiplied by the frequency of occurrence. Such products were added corresponding to each value of *d* and plotted against *d* on a log-log plot. The resulting plot is below. We can clearly see that the frequency of small incidences, say, the ones that lead to the death of 1 or more persons is 204 per annum. Similarly, 5 or more deaths in any accident occur with a frequency of 100 incidents per year. I aver that such incidences would lead to law enforcement led road diversions. So, it could be safely

[7] Data can be accessed at: http://smapp.rand.org/rwtid/search_form.php

said that diversions occur at least 100 times per year – or, twice weekly. We must design our cities against such frequent disruptions. This research is in light of such evidence.

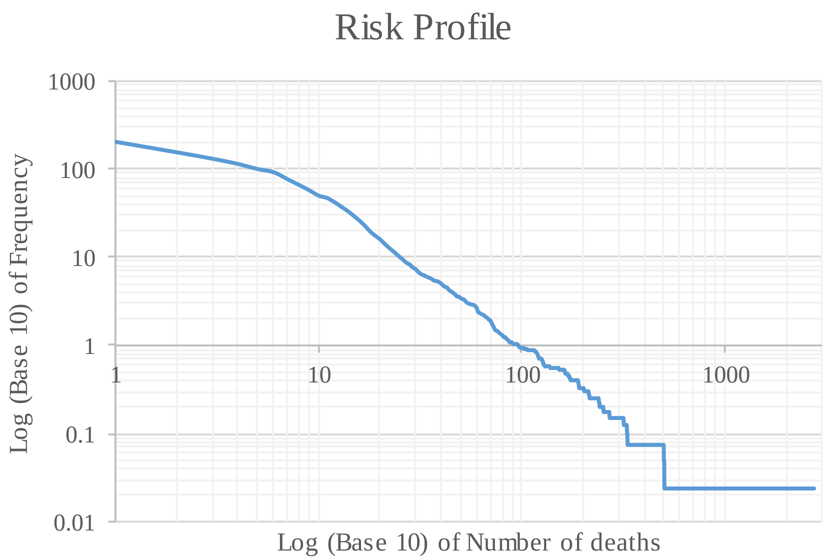


Figure 9: Frequency of small violent attacks[8]

We see, as expected, that as the number of people dying in an incident increases, the frequency of such an incident decrease. Therefore, there are, as expected, a higher number of low intensity incidents fewer higher intensity ones. Flat lines in the graph are due to non-availability of data on incidents for a particular value of d. These flat lines are aberrations in the curve and are only seen at higher values of d where the number of incidents are low. For the lower values, such aberrations are not seen.

Further, for the purposes of modelling this data, we hypothesize that for smaller values of d, the curve of frequency versus d follows a power law of the form: $F = \beta * d^{\gamma}$

We ran a linear regression model taking a further log of both sides to identify the parameters. However, we shall only run this with $d \leq 100$. The results are follows: $\beta =$

[8] Author's calculations based on data from the RAND Corporation

10^3 $and\ \gamma = -1.5$. The p-value of both the parameters is 0 and an adjusted R^2 value is 0.9625.

$$F\ (in\ per\ annum) = 1000 * (d)^{-1.5}$$

The risk profile is a relevant feature in our rationale since these are incidents that affect parts of a city and therefore its road transportation system. Larger incidents affect at a much bigger scale and could require the entire system to be shut down. This research covers only low fatality incidents since they'd cause local incompleteness.

1.4 Data on Road Closures

Some cities in the world provide data on the number and extent of road closures. Toronto in Canada is one of them. The image below shows the locations of all road closures having a "major impact" on traffic. It shows us that such closures are spread over the entire city and are too many for a convenient drive. Some of them are even on the major arterial roads of the city [9].

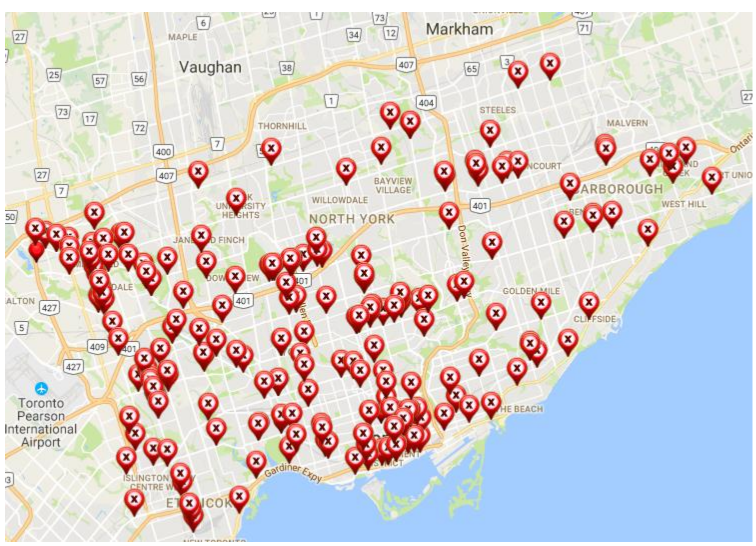


Figure 10: Road Closures in Toronto, Map of

[9] http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=83f6e69ae554e410VgnVCM10000071d60f89RCRD

Further still, data on road closures can be gauged from the list of "Complete Streets" projects listed on the Department of Transportation's website. From there we see that 65 current projects classify as "Complete Streets" and require major changes to the physical infrastructure "…to consider the safe, convenient access and mobility of all roadway users of all ages and abilities" (Department of Transportation, 2011).

Closer home, in Ithaca, we have experienced at least four major road closures in the last year for extended periods of time. College Avenue between Dryden and Oak Streets was closed all last summer leading to all traffic being diverted[10]. Dryden Rd itself has not been available for traffic between College Ave and Bryant Ave due to construction. East Avenue was closed between University Ave and Tower Road in August 2016 for repositioning of the Goldwin Smith Hall Bus Station and adding bike lanes. Tioga St. in Fall Creek was rendered unusable between Court and Farm Streets due to construction of bike lanes (Doolittle, 2016).

From this short anecdotal discussion, it is safe to conclude that road closures affect our lives more often than we think and while routing services like Google Maps can reroute around a closure, it is worthy to note that both public convenience and gas emissions can be reduced by planning cities where such closures affect least the extra amount of distance one must travel.

The second chapter of this dissertation shall propose to objectively define urban topology, network resilience and related vocabulary by citing previous research or texts that have done credible research in this field. It shall also define the metrics we propose to use and state explicitly the mathematical models used to calculate them. A brief

[10] Source: https://transportation.fs.cornell.edu/file/Web-Summer_Impacts_Transportation-2016.pdf

literature survey about the various methods by which network resilience has been quantified are also mentioned and their light, the proposed metrics are justified. In the third chapter, we specify the model used to calculate this *inconvenience* and apply it to some maps of the following cities. In the fourth chapter report the results of the analysis and correlate them with the topological parameters to find any relationship between the two. Here we are testing our hypothesis. The fifth chapter concludes along with a discussion of the shortcomings of this research and avenues for further research.

CHAPTER 2

DEFINITIONS AND LITERATURE SURVEY

Beware of false knowledge; it is more dangerous than ignorance.
-George Bernard Shaw

2.1 Definitions

As is imperative from this quotation from playwright George Bernard Shaw, we must now deal with the problem of definition and stating clearly the terms we shall use in this dissertation.

Sr. No.	***Term***	***Definition***
1.	Topology	the study of geometric properties and spatial relations unaffected by the continuous change of shape or size of figure
2.	Network Topology	It is the arrangement of various elements – links and nodes – in a network. More often used to describe a computer network.
3.	Urban Topology	It is the application of the network topology approach to cities. Here the links are the roads (or, in other cases, alignments of other systems) and the nodes are intersections. This depicts the *physical* part of the network.

4.	Planned Area	An area circumscribed by roads which has been, as it stands now, developed in accordance to a plan which had been decided, ex-ante
5.	Road Network	The set of roads that define our area
6.	Unplanned Area	That area in a city which had been populated before the city plans were drafted to count them as *planned*
7.	Resilience	The resilience of infrastructure systems is their ability to predict, absorb, adapt, and/or quickly recover from a disruptive event such as natural disasters. (NIAC, 2009)
8.	Road Network Resilience	It can be defined as the ability of a road network to come back to its original efficiency after a disruption
9.	Efficiency	It is the time it takes for traffic to pass through a road network.
10.	Information Centrality	Defined for a point i as the relative drop in network efficiency caused by the removal from **G** of the edges incident on i.

Table 1: Definitions of Terms

2.2 Literature Survey

2.2.1 Urban Topology and Metrics

The middle of the last century saw an accelerated growth in research in the field of prediction of transportation flows using topological and geometric parameters of the traffic flow channels (Larson, 1981). The seminal work in the application of networks to neighborhoods, cities and other urban structures was done by Hillier and Hanson (Hillier & Hanson, 1984) and has been consistently done thereafter under the notion of Space Syntax. The need for such a node-link theoretical approach was felt since the complex networks – like those found in sociological studies – were explored and found that *structural centrality* was an overarching theme in them. This would mean that in networks, some places (or points, or persons of reference) would be more important than others. In extending this point to this research, Wilson writes that in urban planning and design, like in economic geography, centrality, called by names such as accessibility, transport cost has entered stressing that some places are more important than others (Wilson, 2000). This becomes our basis for entering into this realm.

There are two ways in which places, in the form of maps, can be represented as networks.

i. *The Primal Approach* – nodes are intersections and edges are streets. This is the most recognizable form of representation that conserves the metric (since the values of the edges are often the real distances between intersections) and the topological (since the relative positions of the nodes are not changed) information.

ii. *The Dual Approach* – nodes are streets and edges are intersections. This form, used in the Space Syntax methodology is useful to conserve completely, the topological information but not the metric. It helps us understand the emergence of power laws in the distribution of space in cities. (Crucitti, Latora, & Porta, 2006)

The primary problem with the SS approach is that it converts the streets into a dimensionless node thereby destroying the metric properties of the network; relations between nodes are turned into merely step distances. Another problem is that SS approach relies on a single centrality index, the overall integration-closeness centrality index, and suffers from end effect which tends to concentrate high centrality values around the geometric center of the image of a map rendering the whole process meaningless (Crucitti, Latora, & Porta, 2006).

They further mention the MCA – Multiple Centrality Assessment which relies on the following:

i. Primal, and not dual representation;

ii. Metric distance, not topologic steps;

iii. Many centrality indices.

Latora and Marchiori, define four important metrics of centrality based on network efficiency – closeness, C^C, betweenness, C^B, straightness, C^S, and information, C^I. Of our interest, here is the information centrality defined above. To formalize the definition, let G be a network represented as a valued graph of N points and K edges. Then

$$C_i^I = \frac{\Delta E}{E} = \frac{E(\boldsymbol{G}) - E(\boldsymbol{G}')}{E(\boldsymbol{G})}$$

Where E(**G**) shall be defined as:

$$E[\boldsymbol{G}] = \frac{1}{N(N-1)} \sum_{i \neq j \in \boldsymbol{G}} \frac{d_{Euclid}}{d_{ij}}$$

Where d_{ij} is the shortest path between i and j and d_{Euclid} is the distance as the crow flies (Latora & Marchiori, 2007).

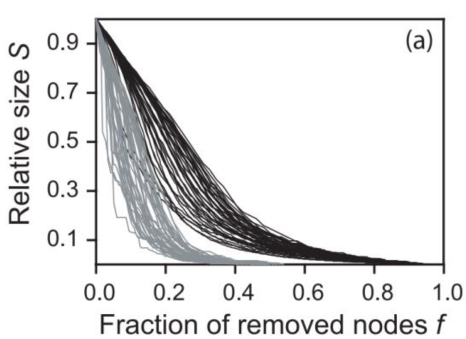
We can differentiate between self-organized and planned cities by assessing the distribution of the information-centrality index. We know that it is distributed exponential in planned cities and as a power law for unplanned cities. This result has been proven by Crucitti et al in their paper.

i. For unplanned, or self-organized cities, $\Pr(C^I > c) \sim c^{\gamma}$

ii. For planned cities, $\Pr(C^I > c) \sim e^{-\frac{c}{s}}$ where s is the parameter.

This is also validated in the paper by Porta et al. (Porta, Crucitti, & Latora, 2006).

This idea is furthered in the paper by Buhl et al where they used the information centrality parameter to talk about network robustness, specifically those of self-organizing cities (Buhl, et al., 2006). For the purposes of objectification, robustness could be considered a similar concept to resilience. In this paper, they find the rate at which the size, S of a graph representing a city reduces as the nodes from it are removed[11]: a) randomly, and b) selectively (decreasing order of the degree of node).



The following result is observed:

Figure 11: Reduction of Relative size with Node Removal

[11] Source: (Buhl, et al., 2006) cf. 5/ Fig. 4a

When each simulation – random and selective – is run for 1000 cycles, the set of grey lines depict the change that happens when nodes are selectively removed and black lines the change when nodes are randomly removed. We see clearly that in self-organizing cities, selective removal causes a larger damage to the integrity of the network than random removal (Buhl, et al., 2006).

We know that a more fragmented network would have reduced connections and therefore, if we were to translate this to the distance of travel between two points, we must get a far larger distance then when removal happens randomly.

Finally, we also see that for self-organizing cities, the random robustness and global efficiency are positively and linearly correlated. We define random robustness as the value of f for which $S = 0.5$.

2.2.2 Network Resilience

A cursory look at the literature available on resilience gives us the view that there is not one unique way in which it is defined. In a paper by Allenby and Fink (Allenby & Fink, 2002) resilience is talked about "capability of a system to maintain its functions and structure in the face of internal and external change and to degrade gracefully when it must". This definition makes it clear that a resilient system might fail to perform to expectations when the operational parameters vary enough so as to not permit it, but it must not be a drastic reduction; the loss of ability must be *graceful* as they say. Yet, one could say that there is not talk of *bouncing back* as the Merriam Webster Dictionary points out in both its definitions which contain the words: capability…to recover, or, adjust easily to change. (Merriam Webster Dictionary, 2016). For this, we look up to the definition by Haimes who wrote: "ability of system to withstand a major disruption within acceptable degradation parameters and to recover with a suitable time and reasonable costs and risks" (Haimes, 2009). This is a much closer definition and one that works for our assessment. It considers Allenby's view of acceptable degradation while also extending it to the recovery aspect.

In order to contextualize this to the idea of resilience against incidents and/or disasters we use the definition by Vugrin who says: "Given the occurrence of a particular disruptive event (or set of events), the resilience of a system to that event (or events) is that system's ability to reduce efficiently both the magnitude and duration of deviation from targeted system performance levels" (Vigrin, Warren, Ehlen, & Camphouse, 2010) We shall use this definition later to contextualize our metric for resilience and apply it to different urban topologies. It contains two important characteristics that were hitherto absent from definitions:

a. That resilience of a system is attached to the nature of the disturbance.
b. That there is some objective metric using which resilience can be measured: the ability to minimize deviation from targeted performance.

While the former is but a statement validating our convictions about the inability of one definition, the latter provides the ground on which to base our scientific endeavor. To that effect, the National Infrastructure Advisory Council through its 2015 report has recommended the Federal Government should standardize the definition of resilience for all to use make the system more transparent (National Infrastructure Advisory Council, 2015).

Within the domain of quantitative resiliency analysis, there can be called three approaches:

a. *Optimization Models*: these are mathematical models which often maximize a resilience function, or minimize time delay. It has been a popular field of research with several models available for as many industries.
b. *Simulation Modelling*: These models use either a discrete event simulation only or combine it with scenario generations. The former was seen in Albores and Shaw and the latter was seen in Carvalho et al, among others. This type of modelling uses simulation methods to create disruptive events and uses a metric, defined ex-ante, to study the response of the network. Sterbenz et al used this for internet networks. Adjetey-Bahun et al used it for a railway network.
c. *Fuzzy logic Modelling*: Here fuzzy linguistic variables are used to set relative importance of resilience parameters.

One of the most comprehensive papers on this is by Hosseini et al. (Hosseini, Barker, & Ramirez-Marquez, 2016) where, in a review paper, they tackle the issue of defining resilience and survey several published definitions to evidence the variety in which it is defined. In the case of this research, it is proposed to have a simulation model to assess resilience. In light of this the following papers found to have been published. We shall review those pertaining to transportation.

We present a review specifically of some of them. In Albores and Shaw, conditions from the implementation of the New Dimension Programme, a disaster management plan from England are modelled using a DES and several experiments are run by varying operational conditions. They concluded that simulation is a reasonable way to model the complex situation and see the response of the network. (Albores & Shaw, 2008)

In Adjetey-Bahun, the authors use a rail transportation network and propose a DES model of the network that, when affected by a *serious perturbation* calculates the total delay of passengers. It does not consider the death/loss if demand following the said perturbation. Then it generates scenarios: a base case when passengers must not decide anything, and two test cases. One of the test cases, allows passengers to change their route if they are waiting beyond a specific time threshold with a given probability. The other test case is the "crisis management situation" where information about perturbation is relayed within a given time to all passengers and those whose paths were on the affected lines shall all take alternative routes. The paper concludes that in the case of the first alternative, the time for recovery of the system does not improve while the delay reduction for passengers is bettered only marginally. In the second alternative, however, the drop in capacity is minimal since passengers choose to not take certain routes at all. Further, the delay reduction is more than the previous case, but as the paper avers, "…the passengers' delay doesn't decrease enough compared to the delay obtained without any management plan…" It concludes that perturbations need to be handled in more ways than merely relaying information (Adjetey-Bahun, Birregah, Chatelet, & Planche, 2014).

We also look at papers written about resilience in a related industry – logistics. In Carvalho et al, they simulate a discrete event simulation of an automobile supply chain (ASC) for a Portuguese manufacturer by modelling time parameters as triangular distributions. Then they generate six scenarios pertaining to the frequency and level of disruption of the network (or the lack of any) and measure two metrics: a) The Lead Time and b) Total Cost. The research finds that for scenarios unaffected by the disturbance, there is a general predictability and uniformity in both the lead times and the total cost. This is a neat result since systematic variation can be accounted for. On the contrary, those scenarios which were affected by the disturbance, there is a randomness with which the total cost varies and the lead time is overshot in all cases after the first ten days. However, the paper generates no contingency measures or means to combat such a disturbance.

CHAPTER 3

METHODOLOGY

Art and science have their meeting point in method.
-Edward G, the First Baron Lytton

3.1 Using Centrality in Transportation Networks

Using centrality as a basis for the analysis of transportation systems is fairly new idea. I discovered this through the paper of Sybil Derrible who uses betweenness centrality calculations on 28 of the world's subway systems to report how they are becomes less "winner takes all" systems and how betweenness was becoming more evenly distributed with size. S/he reported that this parameter was distributed as a power law (Derrible, 2012). This led me to realize that other centrality calculations could be found relevant to the analysis of street networks.

I concluded on Information Centrality as a basis for this research through the work of Estrada and Hatano who mention that Information centrality can be used when "information" must be transmitted between nodes in a network. In our case, vehicular traffic is the "information" we must transfer between nodes and the information centrality is the harmonic mean of the information measure (a proxy) of all the OD links in a network. However, their principal application were social networks. But they also conclude that in Information Centrality, they have found "…the unifying nature of physio-mathematical concepts across the boundaries of many disciplines." (Estrada & Hatano, 2010).

I was finally convinced to the use of this method through the work of Amrit and ter Matt. Their paper described the application of Information Centrality as a means to find the important nodes in a network where information flows as a "walk", not a "path".

Travelling, much like a walk in a graph, is not restricted in terms of which nodes we must pass through. It allows us to choose any node and any number of times. There exists one of these walks which is the shortest route between two points (particularly called a geodesic). Therefore, if a method is relevant to be applied for walks, it is certainly allowed to be applied to geodesics. Their paper shows that information centrality does a fine job in identifying nodes that are important as compared to the other metrics – Freeman closeness, Freeman degree and Bonacich eigenvector (Amrit & ter Matt, 2013). So, this research is essentially an attempt to extend that list to another metric, the SED that does the job like information centrality but allows us to gauge a whole lot more as we shall discuss in further chapters.

3.2 Calculating Sum of Excess Distances

In order to gauge the dependency of parameters, we shall run a correlation analysis. We shall use the Sioux Falls network for a proof of principle. For this network, we first construct an adjacency matrix (AM) and calculate the shortest distance between all points of origin and destination using Dijkstra's algorithm. Then we tabulate this data like an OD matrix and call it the base matrix (BM). Using this as the base matrix, we proceed to remove a node from it. This is done by replacing all the values in the AM under and across that node to 0. A new graph, G' is generated again and the shortest distance algorithm is run again. This generates a new matrix of shortest distances which we shall call the removal matrix for node *i,* RM_i. The process is repeated for all the nodes and so as many RMs are created as there are nodes in a complete network. Using the RM matrices, we calculate the difference matrices, calling these DM_i.

$$BM - RM_i = DM_i$$

So, there are as many difference matrices as there are nodes. Finally, we add the values across all rows and columns for a DM to find the SED value for the node *i*.

$$\sum_{j=i}^{N}\sum_{k=1}^{N} DM_{jk} = SED_i$$

Therefore, as many SED values are calculated as there are nodes in the system.

3.3 Calculating Information Centrality

We have already discussed the mathematical specification of the Information Centrality (IC). In order to calculate it we use the BM and the RMs. We first find the inverse matrices so we can multiply values instead of dividing them. So, we have the IBM (inverse base matrix) and the IRMs (Inverse removal matrices).

$$IRM_{ij} = \frac{1}{RM_{ij}} \forall i, j \in G'$$

We also calculate a distance matrix of nodes by calculating the distance as the crow flies. We shall call this the EM, or a matrix of Euclidean distances.

$$EM(x_1, x_2, y_1, y_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

We then calculate a vector which we call the "Inverse Sum Removed" matrix.

$$ISR_i = \sum_{j=i}^{N}\sum_{k=1}^{N} EM_{jk} * IRM_{jki}$$

$$IBR = \sum_{i=i}^{N}\sum_{j=1}^{N} EM_{ij} * IBM_{ij}$$

Finally, Information Centrality for every node is calculated using the following:

$$IC_i = 1 - \frac{n}{n-2} * \frac{ISR_i}{IBR}$$

3.4 Fitting Distributions

A part of this exposition is to investigate the statistical properties of IC and SED. The hypothesis is, as we have seen, IC is distributed Exponential or Power Law depending on the topology and that SED is distributed Weibull. For this purpose, we shall use several methods. Wherever possible, a maximum likelihood estimation shall be used for estimating the parameters. For the exponential and Weibull distributions, this is used. A goodness of fit test, like a chi square test is used to validate the findings.

For the power law, we use regression by linearizing the CDF of the distribution.

$$\Pr(X \leq x) = 1 - \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

$$\ln(1 - F_i) = (1 - \alpha)\ln x - (1 - \alpha)\ln(x_{min}) \equiv y = \beta x + \epsilon$$

Here, F is coming from plotting points generated by: $F_i = \frac{i-0.3}{n+0.4}$.

If β is the slope of the regression, $\alpha = 1 - \beta$.

For the exponential distribution, the MLE of the rate parameter is given by:

$$\hat{\lambda} = \frac{n}{\sum x_i}$$

For the Weibull distribution, a detailed method for estimating the MLE of the parameters is provided in Appendix C – Optimization Program for MLE of Weibull Distribution. Alternatively, R provides with a **fitdistrplus** package which can help up estimate the MLE parameters as well.

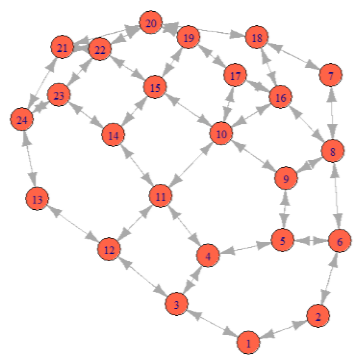
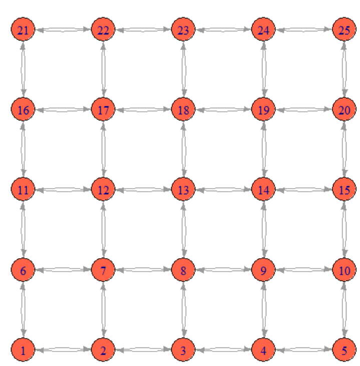
CHAPTER 4

RESULTS AND DISCUSSION

Those who trust chance, must abide by the results of chance.
-Pres. Calvin Coolidge

4.1 Network Representation

From the discussion above it is clear that analyzing our algorithm on the three networks should do the task since they represent most of the cities in the world. So, the first question is, what do they look like? I now present a visual representation of the three networks.

Network Name	**Visualization**
Sioux Falls Network	
Unit25 Network	

Unit100 Network

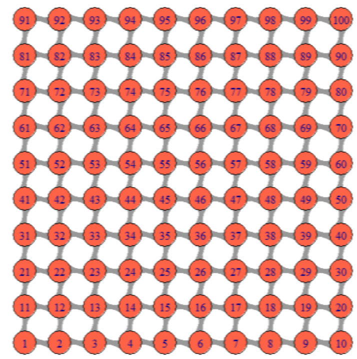
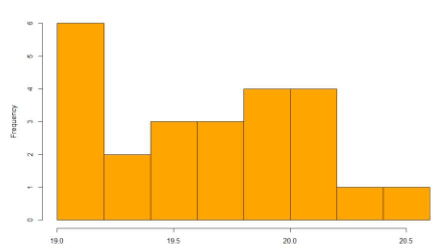


Figure 12: Networks' Visualization

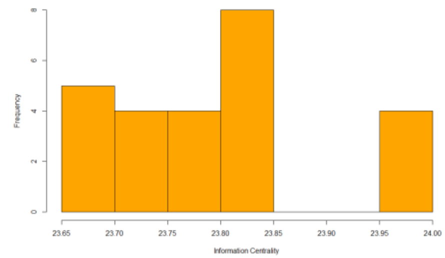
We see that the Sioux Falls Network has 24 nodes and is clearly not a regular grid. Of these 12 are internal nodes and the others are at the boundary. The other two networks are synthetically generated as an extreme case of a regular network. There are two of these to find the differences of scale. The first of these, the Unit25 Network (U25, hereinafter) has been chosen to match with the SFN on scale but not in pattern since they have comparable number of nodes. The other, the Unit100 Network (U100, hereinafter) matches in pattern with the U25 but not in scale.

4.2 Information Centrality Calculations

First we shall analyze the nature of the information centrality. If we plot a histogram of the IC values for the three networks, the following graphs emerge:

Network Name	Distribution of Information Centrality
Sioux Falls Network	

Unit25 Network



Unit100 Network

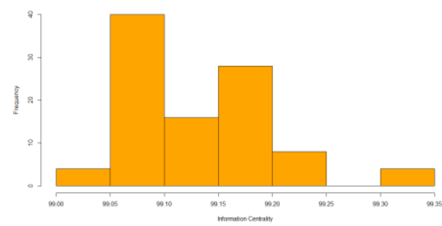


Figure 13: Distribution of Information Centrality

We know from previous knowledge that for the Sioux Falls Network (SFN, hereinafter) this should be distributed as a power law and for the other two, exponentially. Regressing the plotting points versus IC or ln(IC) using the following equations, the following values are generated.

$$\ln(1 - F) = (1 - \alpha) * \ln(IC) - (1 - \alpha) * \ln(IC_{min})$$

$$\ln(1 - F) = -\lambda * IC$$

$$F = (i - 0.3)/(n + 0.25)$$

Network	**R^2 Values**		**Parameter Values**	
	Power Law	*Exponential*	α	λ
Sioux Falls Network	0.87	0.52	40.8	0.05
Unit25 Network	0.83	0.51	197.44	0.041
Unit100 Network	0.945	0.51	1396.68	0.01

Table 2: Fitting Information Centrality

It is striking to see that in all cases presented above, the Information Centrality values fit a Power Law much better than an exponential distribution. This is different from the claims made by the previous papers. A low R^2 value in the exponential fitting is seen because it has been controlled for an intercept.

Another interesting result can be gauged from this result, however. Since the value follows a power law distribution, as could rank the networks based on how "inherently biased" they are. This is an application of Lorenz's curves to our networks. We could say that the sum of all the IC values in the network is akin to the "total information" contained in the network. Then we could plot a graph depicting the cumulative percentage of information (W) versus the cumulative number of nodes that hold it (P). Using this, one could calculate the Gini coefficient of the networks. The following table present the Gini values of the three networks:

Network	Gini (%)
Sioux Falls Network	31.06
Unit25 Network	6.48
Unit100 Network	3.59

Table 3: Gini coefficients of the Networks

4.3 Sum of Excess Distance Calculations

Now we shall present results from the SED calculations. We had hypothesized that its distribution follows a Weibull distribution. A prima facie result that proves this hypothesis comes from plotting the Cullen and Frey graphs for all the cases. These have been presented in Appendix B of this text. We can gauge from all three that the distribution could be Weibull. When the histograms of the values generated are tabulated below.

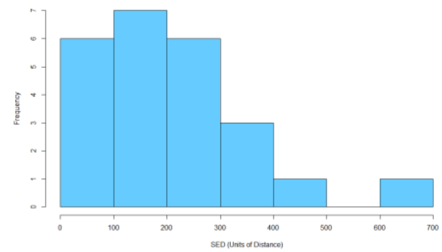
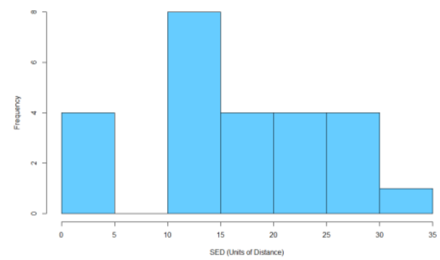
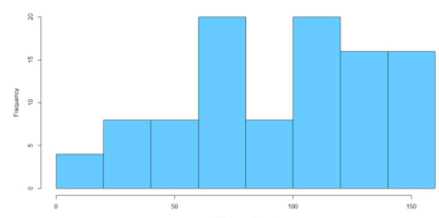
Network Name	Distribution of SED
Sioux Falls Network	
Unit25 Network	
Unit100 Network	

Figure 14: Histogram of the Sum of Excess Distances

The choice of Weibull distribution is valid since it has several form and shapes coming from changes in the shape factor. It's spread is also variable in that it can vary with the scale factor. We choose here to fit the unconstrained Weibull distribution, over the Wight Truncated Weibull Distribution for ease of analysis and because we do not know ex-ante what the end point of the distribution is going to be. In order to estimate the parameters of this Weibull distribution, the Maximum Likelihood Estimator is used. Since the Weibull distribution does not have a closed form, it can either be solved using an optimization algorithm or by using the **fitdist** function in R. Both results are comparable. However, the problem with the optimization is that there are several local optima, which must be discarded before a global optimum is reached. Nevertheless, the optimization model and its results are also presented in Appendix C of this text. The following table captures the results of the **fitdist** function from R.

Network	**Parameter Values**[12]	
	Shape, k	*Scale, λ*
Sioux Falls Network	1.40	217
Unit25 Network	1.03	16
Unit100 Network	1.90	104

Table 4: Parameter values of SED Distribution

From these values now, we can test using a chi-square test whether the distribution follows a Weibull distribution. So, we conduct a chi-square test with the test statistic being:

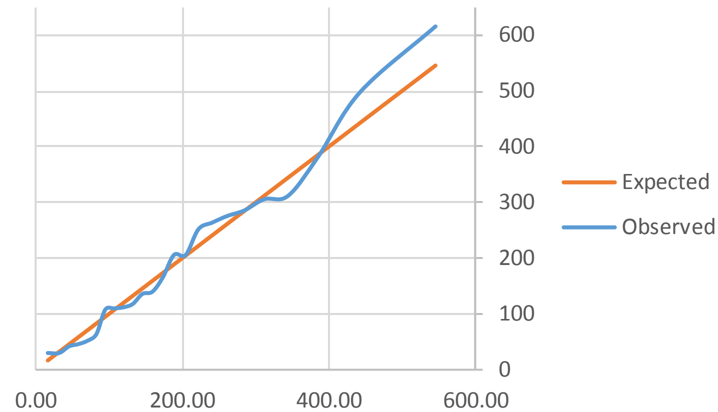
$$\chi^2 = \sum_{i=1}^{n} (E_i - O_i)^2 / E_i \sim \chi^2_{n-3}$$

In all the cases, we find that the distribution does fit the Weibull distribution. We present now the Q-Q plot of the three cases to validate our result. We see that at the extremes, the organic network behaves much better than the regular ones. This is because there are several repetitions in the SED values in the regular networks, U25 and U100, attributable to their regularity. This repetition reduces the count of unique numbers that could scale up with the plotting positions; validated in the fact that the U100 plot fits much better than the U25 plot since the U100 has a larger count of unique values.

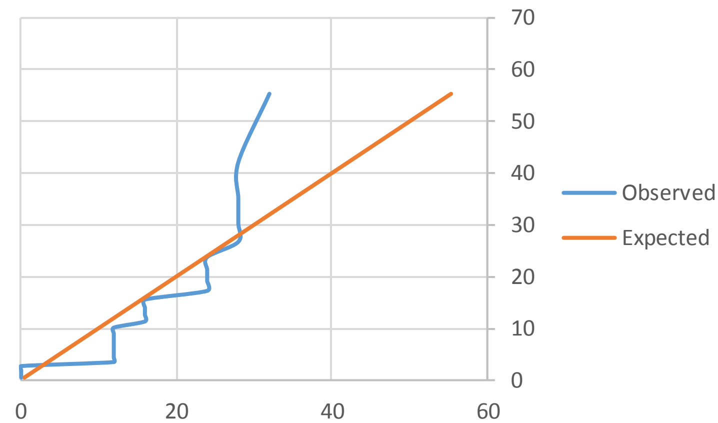
[12] All parameter values are significant with a p-value of the order of 10^{-12}.

Network Name	**Q – Q Plots for Weibull Fit**

Sioux Falls Network



Unit25 Network



Unit100 Network

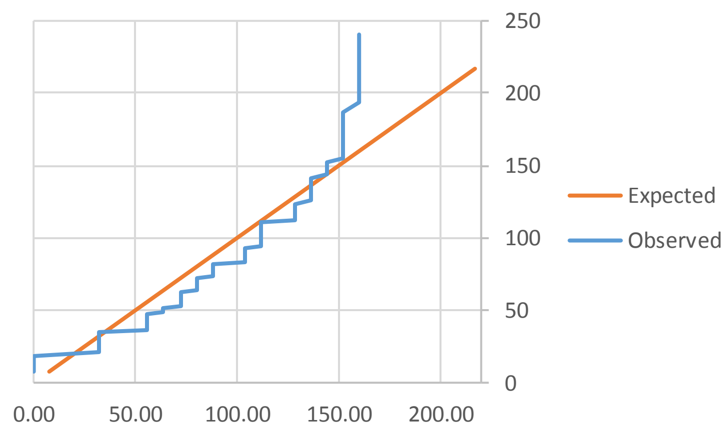


Figure 15: Quantile Plots for SED Weibull Fit

From this it is convincing that the distribution does indeed follow a Weibull distribution. Therefore, we now turn to describing descriptive statistics of this.

Network	Measures of Central Tendencies (distance)		
	Mean	*Median*	*Mode*
Sioux Falls Network	197.78	167.02	88.68
Unit25 Network	15.8	11.21	0.50
Unit100 Network	92.28	85.75	70.18

Table 5: Mean and Mode of SED of all Networks

We shall use values from Table 5 in discussion about the performance of these networks and as metrics to be used for ranking the networks.

Finally, we must prove that the new metric works as a proxy for the old. This can be proven by calculating the correlation between the two values. The following table reports the values of the correlation between IC and SED.

Network	*Correlation, ρ*
Sioux Falls Network	0.650 ± 0.221
Unit25 Network	0.996 ± 0.003
Unit100 Network	0.988 ± 0.005

Table 6: Correlation Values

4.4 Discussion

4.4.1 Network Rankings based on IC and SED

In this section, we shall discuss what all the analyses and results mean for policy and design. First, let us discuss what our results show in the light of the literature. While the literature points to the fact that information centrality for regular networks should follow an exponential distribution, our results show that that is not true in all cases. Under theoretically regular networks, as under organic unplanned networks, IC always follows a power law better than an exponential distribution. We have also seen that a power law allows us to quantify how the total information in the network is distributed – using the 80-20 rule and how much the networks deviate from it. Such a comparison yields a first method to rank networks. We have mentioned Gini percentages in Table 3 and can use that to find out how "unfair" our network is to the vehicles that drive through them. Organizations like Smart Growth America and National Complete Streets Coalition are already pushing for better city streets that are more equitable. Knowing the Gini coefficients of neighborhoods and correlating it with the population's socio-economic situation, one could conclude if certain communities live in areas that inherently unfair with respect to road transportation. Legislators and community action groups can then use this knowledge to make cities and the distribution of populations there more fair and ensure horizontal social equality for all (Smart Growth America, 2016). From our results, we can rank the networks based on their Gini scores thus:

Sioux Falls ***Unit25*** ***Unit100***

Increasing Equity (Reducing Gini values) →

Figure 16: Network Ranking by Equity

One example of this is the suburbanization of American cities. We know that extremely regularly designed suburbs became the order of the day as income levels rose after the World Wars. This was also the phase where ownership of cars became the norm. However, such development was highly unequal racially. The suburbanization this took place in which populations earlier and faster than in communities of color. Work by Schelling[13] has proven this. Therefore, we could conclude that the suburban population made up of predominantly white people tend to live in areas that are inherently more equal than the neighborhoods of the people of color. This must change.

[13] https://www.stat.berkeley.edu/~aldous/157/Papers/Schelling_Seg_Models.pdf

However, more can be gauged from SED values of neighborhoods. First, we can rank networks based on what is the most likely excess distance people must travel in the case of failure of a node. This is known to us from modal values (Table 5) of the distributions they come from. We can rank networks based on this criterion too.

Sioux Falls ***Unit100*** ***Unit25***

"Worst-Case" becoming Better (Reducing Modal SED Values) →

Figure 17: Network Ranking by Worst Case

This ranking system, while they tell us the worst case, does not account for the median of excess distance that vehicles must travel. For these we rank networks using the mean SED values. While the mode tells the SED value that maximizes the pdf, the median tells the SED value that is likely to be seen at least half the time. So, under repeated failures of the same network, the excess distance that vehicles must travel in at least one-half of the cases is reported in Table 3. The ranking is reported as in under:

Sioux Falls ***Unit100*** ***Unit25***

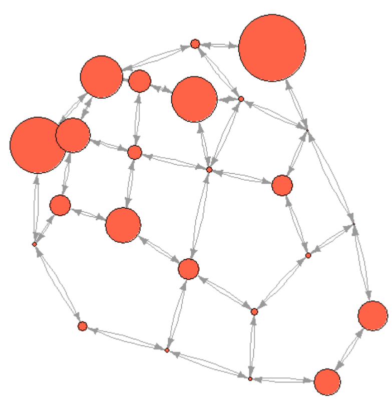
"Worst-Case" becoming Better (Reducing Median SED Values) →

Figure 18: Networks Ranked by Median SED

Organic Networks behave worse than any scale of the organic networks under any of these ranking mechanisms, proving further that they are not good choices for newer cities today. Finally, we define a fourth metric that measures what percentage of nodes are most susceptible to fail. This is known by finding the inverse of the difference between the nodal SED value and the modal SED value of the distribution. For nodes, whose inverse of the difference is greater than the mean is most susceptible to fail. Using this criterion, the following numbers emerge (Table 7) and they can be used to rank the networks such that larger values are less desirable. When seen spatially, we see a clear pattern in the nodes' distribution. The most susceptible nodes fall in the region with the densest nodes for the organic network and in concentric circles for the regular network. This proves to us that under regular grid systems parts of the network behave the same way. Therefore, if a problem is known to have been caused at one location, we should know to employ mitigation efforts at all its complementary locations to as to make the network more resilient. This cannot be said of organic networks where larger susceptibility values are associated with congested regions.

Network	*PMSN*
Sioux Falls Network	62.5
Unit25 Network	64.0
Unit100 Network	52.0

Table 7: PMSN values of Networks

Network Name	**Network Nodes, by Susceptibility to Failure**
Sioux Falls Network	

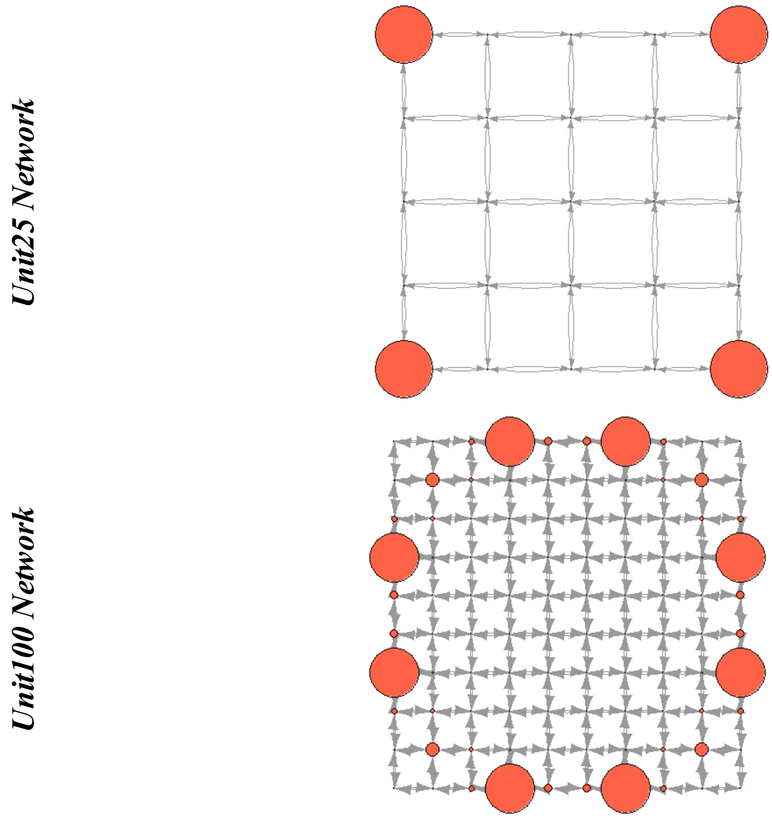


Figure 19: Networks by Node Susceptibility

Besides finding the criteria to rank networks this analysis also presents an interesting finding. When the SED values are mapped on the network to see for spatial patterns, we realize that the spatial distribution of the important nodes – those with higher SED values – varies with the map of the network. While the important nodes happen to be in the periphery of the organic network, for regular networks, they happen to be in the center (Figure 20). This implies that busy city centers planned in the center of a regular network are worse off than when planned in the same location of an organic network. These areas, when not available for traffic circulation – due to events, violence, maintenance, etc. – make the city worse off when the city itself is "unplanned" or, organic. Therefore, it makes sense in a grid network to locate the downtown in one end of the city or distribute commercial activities over regions around several less important nodes. This has implications for location setting of areas within city.

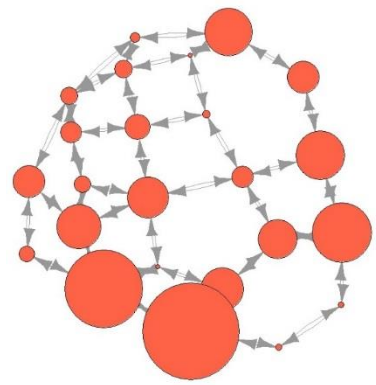
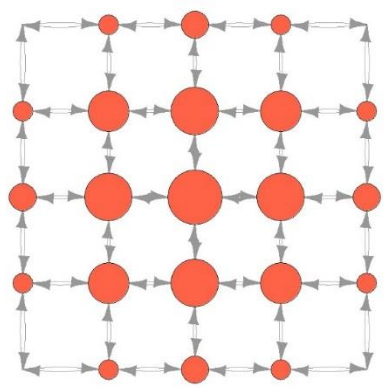
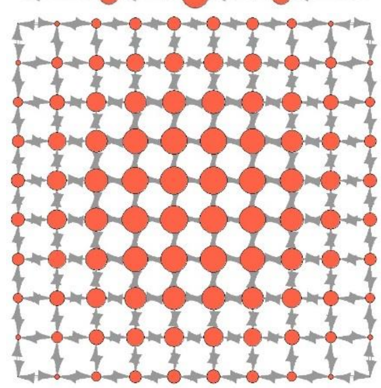
Network Name	Spatial Distribution of SED
Sioux Falls Network	
Unit25 Network	
Unit100 Network	

Figure 20: Spatial Distribution of SED

Another argument that can be made through this analysis is that of an optimally dense network. We see that on the one hand, having more nodes in a network covering a given area gives rise to more number of nodes whose Information Centrality is high. This means that more nodes, when unavailable for access will make the system considerably worse off. Therefore, it can be argued that networks with more nodes in any given area will have a higher cost of maintenance – since these nodes and the edges leading into and out of them must be protected from disruptions.

On the other hand, higher node density will also lead to more possible routes from one point to another. Consequently, the unavailability of one will not make the SED values shoot unreasonably higher – so, the highest SED value (worst case) in should be less in higher node density networks. This is evidenced from that fact that in a regular network the highest value of SED (32 units) is nearly 5 times higher than that in the organic network (6.42 units) of comparable size. This points to the fact that higher node density leads to better resiliency. Therefore, two statements of fact can be made: first, that higher node density leads to higher resiliency measured as SED; and second, that higher node density will also lead to higher costs of maintenance. This is also concluded in the critique by Prof Peter Philips of the Economics Department of the University of Utah titled Why Urban Roads Cost More and Deliver More" (Philips, 2015). While the first is desirable, the second is not. Thus, there exists a trade-off between resiliency and cost of maintenance which can be modelled as a two object Pareto front.

We now turn to the discussion of SED values being distributed Weibull. For all our cases, we generate the Cullen and Frey plots using 100 bootstraps of the sample. These are reported in Appendix B of this document. It is clear from them, that the distribution of SED should lie close to a Weibull distribution (note that Weibull is close to gamma, which is the dotted line underlying the grey beta range). In some cases, though, as in

Figure 25, one sees that the point of observation is close to normal and uniform distribution too. We can eliminate these from the following arguments.

First, the distribution of SED cannot be normal, since SED is bound on the lower side by 0 (a distance, longer than the shortest distance must necessarily be positively away from the initial value). Since the domain of a normal distribution is the entire real line, SED cannot be distributed normal. As for Uniform distribution, we see this proximity more pronouncedly in Figure 25 than in the others. We can attribute it to the fact that there are fewer unique SED values in a UD Network (n = 25). Repetition of values is also symmetric as can be gauged from Figure 24 where nodes lying concentric from the center have the same values of SED. This symmetry is the reason why the observation could be mistaken for a uniform distribution. From all the plots above it is clearly seen that our metric is distributed Weibull.

There is however, another argument to prove that a Weibull distribution would fit best to the metric at hand. We know that Weibull distributions are used to characterize the distributions of the smallest values over several sets of observations, like the breaking strength of a chain of links, for example. In our study, we are dealing with a "minimum" value as well. The SED is, essentially, the *least* extra distance one must travel between two points when the shortest route is not available. Therefore, it would make sense to see that it is distributed Weibull. This finding is in contrast with the distribution of the Information Centrality as reported in Crucitti, Latora and Porta[14]. They show by an extensive (though, by no means exhaustive) list that IC values are distributed exponentially for planned and by a power law for unplanned cities. However, in theirs

[14] (Crucitti, Latora, & Porta, 2006)

or any other publication, one fails to find a theoretical underpinning for this finding. This study too, uses the findings of Crucitti et al to base its analysis, but as has been discussed above proves, albeit qualitatively, that the new metric should be distributed Weibull. This is going beyond the empirical findings of the previous authors and is therefore a better metric to work with.

4.4.2 Environmental Implications of SED

There is another serious implication that is revealed from this SED metric which we can quantify with some ease and that is – when vehicles run extra distances, they also emit more greenhouse gases. Quantifying this make it easy to do a cost-benefit analysis of the network thereby easing choice. So, we shall now proceed to quantify the excess CO_2 emissions due to extra travel length. For this purpose, we shall use emissions from different vehicles as a triangular distribution. This data is provided by the Environmental Protection Agency[15].

$$Emission\ per\ vehicle = Distance * Emissions\ per\ mile\ per\ vehicle$$

For values on the right-hand side of this equation, Distance is the possible SED value and Emissions per mile per vehicle are samples from a triangular distribution modelled from data in Table on Page ES7 of the referenced report. From that data, we know that the highest possible emissions are of a Fiat-Chrysler (402g/mi = 251g/km) and the lowest are of a Mazda (290g/mi = 181g/km) and the modal value is 310g/mi (194g/km)[16]. Using these data, and the fitted distributions of SEDs in the three cases from before, we ran a Monte Carlo Analysis whose results are being presented here.

---

[15] Light-Duty Automotive Technology, Carbon Dioxide Emissions, and Fuel Economy Trends: 1975 to 2016 (US EPA-420-S-16-001, November 2016)

[16] Conversions: 1.6g/mi = 1g/km

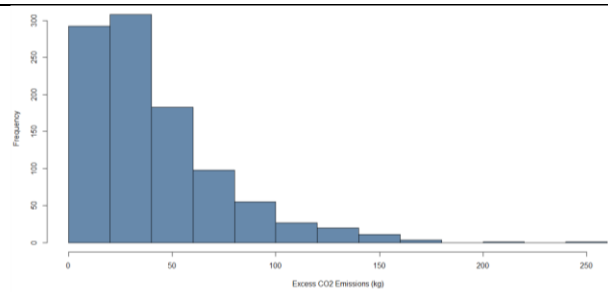
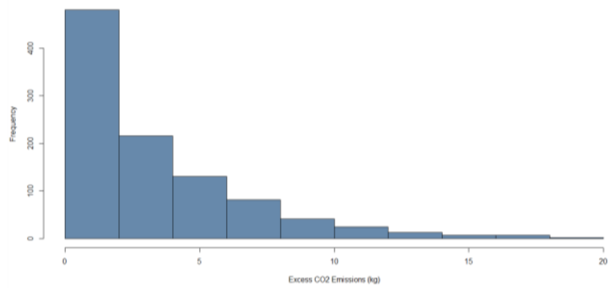
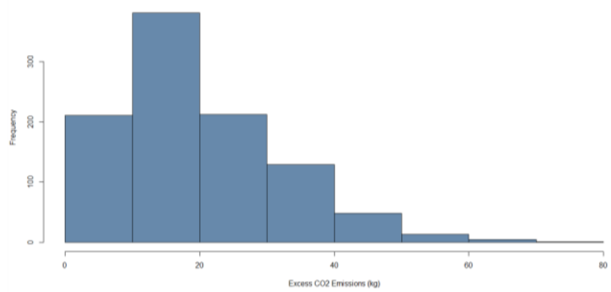
Network Name	Excess Emission of CO_2 per vehicle (n = 1000)	Mean Excess Emissions (kg)
Sioux Falls Network		41.21
Unit25 Network		3.28
Unit100 Network		19.76

Figure 21: Excess CO2 Emissions by Network

The organic network leads to higher emissions that a regular network, even of a larger size. Between the two regular networks, as expected, the smaller one has lesser emissions than the larger one simply because distances scale up as size of the network increases. Nevertheless, it is interesting to note that the increased emissions per vehicle are distributed exponentially. We can now use this excess emission values to rank the networks depending on their environmental impact.

Unit25	***Unit100***	***Sioux Falls***

Worsening Environmental Impact (Increasing CO_2 Numbers)

Figure 22: Networks Ranked by Environmental Impact

CHAPTER 5

FURTHER RESEARCH AND CONCLUSIONS

Eliminate all other factors, and the one which remains must be the truth.
-Sir Arthur Conon Doyle in The Sign of Four

5.1 Further Research

The current analysis focuses on designing and conducting a preliminary investigation into the properties of a new metric to describe the impact of incomplete networks on the shortest distance between any pair of origin and destination. However, this analysis comes with its own set of assumptions. Some of them have been kept for ease of calculations and some for lack of better understanding and model specification.

Uniform Demand Assumption: I believe the biggest assumption is the uniformity of demand. The model assumes that all origins and destinations shall have equal traffic flowing in them thereby eliminating any choice made by the driver to choose an alternative route, even though it might save time.

Binary Road Assumption: The analysis treats road closure as a binary event in that either the road is available, or it is completely unavailable. While we know that this is true and real, a more often seen scenario is when one or more lanes of the road are closed. Any future analysis should use this idea and model the network thereafter. This could be done by using total lane lengths instead of distance to factor in lane closures. We could extend the distance from a set base value by a factor that accounts for how many lanes are closed. As an extension of the same assumption, one could also treat unidirectional closures in the network where road in one direction is closed but the other is not.

Travel Length Indifference: The model assumes that the total vehicle times the excess length it travels is indifferent for society, i.e. whether one vehicle travels x units extra or 1/x vehicles travel 1 unit extra are equal levels of inconvenience. This may not be true and will need to be accounted for in future research.

Coupling Effects: The analysis disregards coupling effects due to failure of two or more nodes at the same time. This is a known phenomenon when the scale of the disruption is large. While the network may have a simple response to a single node failure, the response on two failures could be complex in that it wouldn't be a scaling up. In sparse networks coupling may also lead to disjointing of the networks into two such that there would not exist any path from one to another. These situations are not tackled in this research and should be in the future.

Proving the trade-off hypothesis: The research also only proposes the existence of a trade-off between resiliency and cost. This needs to be investigated further. The presence of such a trade-off shall allow future city and transportation planners to have an ex-ante knowledge of their decisions. Clustering nodes in an area will make it stand better against unavailability of any one of them but will cost more to maintain flow. Distributing them will make it worse off but cost less. Planning neighborhoods based on this knowledge will help in making areas within a city more resilient and accrue monetary benefits.

The Case of Mixed Planning: The analysis above touches on two extreme cases – one of a completely unplanned network and another of a completely planned network. While most neighborhoods in cities across the world can be classified using this, cities as whole might not be. Therefore, we need to extend this analysis, by making the computation more efficient, to whole cities which are generally a mix of planned and unplanned areas – cities like New Delhi, for example are cases in point. While the medieval part of the city is organic and unplanned, the rest of the city has been planned during the early 20th century and revolves around hexagonal intersections.

5.2 Conclusions

Despite the assumptions and shortcomings as discussed above, there are clear conclusions that we can draw from the study. First, we realize that in rigidly regular networks the information centrality follows the power law and not the exponential law as was previously held. From the analysis on information distribution at every node we conclude that the organic networks are inherently more unfair then regular networks of comparable or larger size. Correlating this with socio-economic parameters of the people who reside in them should give us a whole new perspective on thinking about urban sociology.

Another conclusion is that a new metric, the SED does the job satisfactorily in that it correlated significantly with the IC values for all networks considered. This is impressive because now we have a metric to quantify inconvenience that is based not on the assumptions of traffic routing and demand but on simple street design. This can help us make decisions before the plan is implemented for changing street plans is tougher than making policy to reroute traffic.

Further, from SED we conclude that it is distributed Weibull with the reason that it uses the minimum extra distance vehicles must travel in the situation that they are not able to take the first shortest path. This is in keeping with the known applications of Weibull distributions and opens yet another avenue in which Extreme Value distributions are useful. As for the implications of this new metric, we see that several parameters can be designed for ranking networks based on it and this helps in decision making. We could, for example, choose from among the networks if the policy is that the median excess distance cars should have to travel is 40 miles extra. Or, if we were to ask: which network has the most number of susceptible nodes?

This new metric also helps in location setting questions. If one were to ask what the best location of a very critical service, say hospital, should be in a network? This metric could help answer that question: in the center of the network for an organic network and at the periphery for a gridded network, so that even when a node is not available, one need not drive very much extra to reach it.

Finally, SED helps us in better understanding the environmental impacts of road closures. We see that from the various values of excess CO_2 emissions from vehicles having to travel extra distances. This is less for smaller networks but substantial for large ones. In the Sioux Falls networks this has a mean value of 41.21kg per vehicle. This means that if we have a thousand cars having to travel extra on this network, that means extra CO_2 emissions to the tune of 41.2tons. This can have serious implications for the environment. We must therefore, plan neighborhoods where the excess distance is minimized so we can have healthier and less impactful societies.

I hope that further investigation into SED and its behavior and applications shall reveal an even more nuanced analysis as the model is studied for several cities in specific.

Appendix A – Computer Code (in R)

The following is the computer code, written in R used in this thesis

```
#Clearing previous memory
dev.off()
remove(list = ls())

#Driver code
require(igraph)
SF_csv <- read.csv("Sioux Falls Network.csv", sep = ",", header = TRUE)
nmax <- max(SF_csv[,1],SF_csv[,2])

#Creating the adjacency Matrix
SF_Adj <- matrix(nrow = nmax, ncol = nmax)
    SF_Adj[1:24, 1:24] <- 0
    for(i in 1:nrow(SF_csv))
    {
    SF_Adj[SF_csv[i,1],SF_csv[i,2]] <- SF_csv[i,3]
    }

    SF_g <- graph_from_adjacency_matrix(SF_Adj, mode = "directed", weighted = TRUE)

    #Generating and Plotting the graph
    SF_g <- graph_from_adjacency_matrix(SF_Adj, mode = "directed", weighted = TRUE)
    V(SF_g)$color <- "tomato"
    l <- layout.auto(SF_g)
    plot(SF_g, layout=l, edge.arrow.size = 0.2, edge.curved = 0.1, edge.color = "gray60")

    SF_d.base <- distances(SF_g, v = 1:24, to = 1:24, mode = "all", algorithm = "dijkstra")

    library(igraph)

    SF_d.removed <- array(0, c(nmax, nmax, nmax))
```

```
for(i in 1:nmax)
{
local_g <- SF_g
local_Adj <- SF_Adj

local_Adj[i,] <- 0
local_Adj[,i] <- 0

all.nodes <- 1:nmax

local_g <- graph_from_adjacency_matrix(local_Adj, mode = "directed", weighted = TRUE)

#Find Shortest distances
SF_d.removed[,,i] <- distances(local_g, v = all.nodes, to = all.nodes, mode = "all",
algorithm = "dijkstra")
}

#Excess distance arrays
Diff <- array(0, c(nmax, nmax, nmax))
for(i in 1:nmax)
{
Diff[,,i] <- SF_d.removed[,,i] - SF_d.base
}

#Replacing Inf with 0
for(i in 1:nmax)
{
for(j in 1:nmax)
{
for(k in 1:nmax)
{
if((Diff[i,j,k]/4)==Diff[i,j,k])
{
Diff[i,j,k] <- 0
}
}
}
}

#Finding the excess distance over the network
```

```
Excess <- vector(length = nmax)
for(i in 1:nmax)
{
sum = 0
for(j in 1:nmax)
{
for(k in 1:nmax)
{
sum = sum + Diff[k,j,i]
}
}
Excess[i] <- sum
}

cat("Excess distances are:", Excess)
hist(Excess)

#Statistics of Excess
require(fitdistrplus)
hist(Excess)
plot(density(Excess))
descdist(Excess, boot = 100)

#Creating Dummy Code "cat" for Visualisation
E.mean <- mean(Excess)
E.sd <- sd(Excess)
V.cat <- vector(length = nmax)
for(i in 1:nmax)
{
if(Excess[i]>(E.mean + E.sd))
{
V.cat[i] <- 3
}
else if(Excess[i]<(E.mean - E.sd))
{
V.cat[i] <- 1
}
else
{
V.cat[i] <- 2
}
}

#Graph Visualization
```

```
#Distribution by extreme value
l <- layout.auto(SF_g)
V(SF_g)$label <- NA #seq(1:nmax)
colrs <- RColorBrewer::brewer.pal(3, "Reds")
SF_g <- set.vertex.attribute(SF_g, "cat", index = V(SF_g), value = V.cat)
V(SF_g)$color <- colrs[V(SF_g)$cat]
plot(SF_g, layout=l, edge.arrow.size = 0.2, main = "Distribution of SED (Sioux Falls)",
edge.curved = 0.1, edge.color = "gray60", vertex.size = 20)

#Spatial Distribution
V(SF_g)$size <- 0.08*Excess
V(SF_g)$color <- "tomato"
plot(SF_g, layout=l, edge.arrow.size = 0.2, main = "Spatial Distribution (Sioux Falls)",
edge.curved = 0.1, edge.color = "gray60")

for(i in 1:nmax)
{
if(Excess[i] == 0)
{
Excess[i] <- 0.1
}
}
Excess.sort <- sort(Excess)

#Plotting positions
i = seq(1, nmax, 1)
F <- (i-0.3)/(nmax+0.4)

library(fitdistrplus)

#Fitting Weibull Distribution
w.mle <- fitdist(Excess, "weibull", method = "mle")
k = w.mle$estimate[1]
lambda = w.mle$estimate[2]

#Estimating Weibull Expected Values
X_syn_w <- ((-1*log(1-F))^(1/k))*lambda
```

```
#Weibull Tests
chisq.test(X_syn_w, Excess.sort)
ks.test(Excess, X_syn_w, alternative = "two.sided")

#Fitting Exponential Distribution
w.exp <- fitdist(Excess, "exp", method = "mle")
l <- w.exp$estimate

#Estimating Exponential Expected Values
X_syn_e <- -1*log(1-F)/l

#Exponential Tests
chisq.test(X_syn_e, Excess.sort)
ks.test(Excess, X_syn_e, alternative = "two.sided")

#Macrolevel statistics
plot(seq(0, 700, 1), dweibull(seq(0, 700, 1), shape = k, scale = lambda),
main = "pdf of Fitted Distribution", type = "l", xlab = "SED",
ylab = "Pr(SED = s)", col = "orange", lwd = 2.5)
lines(Excess, dweibull(Excess, shape = k, scale = lambda), col = "blue",
type = "p")
h <- hist(Excess, plot = FALSE)
lines(h$mids, h$density, col = "red", type = "l", lwd = 2.5)

W.mode <- lambda*(((k-1)/k)^(1/k))
W.mean <- lambda*gamma((k+1)/k)

#Plotting map for most likely failures
Diff.from.mode <- vector(length = nmax)
for(i in 1:nmax)
{
Diff.from.mode[i] <- abs(Excess[i] -  W.mode)
}

#Percentage of Most Susceptible Nodes
count = 0
for(i in 1:nmax)
```

```
{
if(Diff.from.mode[i]<mean(Diff.from.mode))
{
count = count + 1
}
}
PMSN = 100*count/nmax

heading <- c("Sioux Falls Nodes, By Chances of Failure")
V(SF_g)$size <- 700*(1/Diff.from.mode)
V(SF_g)$color <- "tomato"
plot(SF_g, layout = layout.auto(SF_g), edge.arrow.size = 0.2,
main = heading, edge.curved = 0.1, edge.color = "gray60")

c.index <- vector(length = nmax)
inv.sum.removed <- vector(length = nmax)
d.Euclid <- matrix(nrow = nmax, ncol = nmax)
nodes <- read.csv("SF_Nodes.csv", sep = ",")

inv_d.removed <- 1/SF_d.removed
inv_d.base <- 1/SF_d.base

#The Eucleadian Distance matrix
for(i in 1:24)
{
for(j in 1:24)
{
d.Euclid[i,j] <- ((nodes[i,2] - nodes[j,2])^2 + (nodes[i,3] - nodes[j,3])^2)^0.5
}
}

for(i in 1:nmax)
{
sum = 0
for(j in 1:nmax)
{
for(k in 1:nmax)
{
if(j == k)
{
```

```
sum = sum + 0
}
else if(j!=k)
{
sum = sum + (d.Euclid[k,j]*inv_d.removed[k,j,i])
}
}
}
inv.sum.removed[i] <- sum
}

for(i in 1:nmax)
{
sum = 0
for(j in 1:nmax)
{
if(i == j)
{
sum = sum + 0
}
else if(i != j)
{
sum = sum + (d.Euclid[i,j]*inv_d.base[i,j])
}
}
inv.sum.base <- sum
}


for(i in 1:nmax)
{
c.index[i] <- 1 - ((12/11)*(inv.sum.removed[i]/inv.sum.base))
}

#Index Matrix
c.index.matrix <- matrix(nrow = nmax, ncol = 2)
colnames(c.index.matrix) <- c("Node", "Information Centrality")
c.index.matrix[,1] <- 1:nmax
c.index.matrix[,2] <- round(c.index, digits = 5)

plot(c.index, Excess, log = "y", xlab = "Info. Centrality",
```

```
ylab = "Sum of Excess Distances",
main = "Information Centrality vs Excess distance")

#Test the Correlation
alpha = 0.01
cor.test(Excess, c.index, alternative = "two.sided",
method = "pearson", conf.level = (1-alpha))

c.index <- -1*c.index

#Fitting Information Centrality to a Power Law
lm.power <- lm(log(1-F)~log(sort(c.index)))
summary(lm.power)
alpha.power <- -1*lm.power$coefficients[2]+1

#Fitting Information Centrality to an Exponential Distribution
lm.exp <- lm(log(1-F)~sort(c.index) - 1)
summary(lm.exp)
lambda.exp <- -1*lm.power$coefficients[1]

W <- F^((alpha.power-2)/(alpha.power-1))
plot(F, W, type = "l", xlab = "Cumulative # of Nodes",
ylab = "Cumulative Share of Information", main = "Lorenz Curve for SF Network",
col = "red", lwd = 2.5)
lines(F, F, type = "l", col = "blue", lwd = 2.5)
gini <- 2*sum(W-F)
cat("The Gini coefficient of this network is: ", gini*100, "%")
#End of Main Code
```

##The following piece of code was utilized in calculating the environmental impact from SED. It may be run separately from the rest of the code.

```
remove(list=ls())
dev.off()
library(triangle)
l <- 1000
e <- vector(length = l)
```

```
#For the Sioux Falls Network
for(i in 1:l)
{
r = runif(1)
s <- qweibull(r, shape = 1.40, scale = 217)
p <- qltriangle(r, a = 181, b = 251, c = 194,
logbase = 10)
e[i] <- (s*p)/1000
}
e.sfn <- mean(e)
hist(e, xlab = "Excess CO2 Emissions (kg)", ylab =
"Frequency",
main = "Distribution of Excess CO2 Emissions (SFN)",
col = "#6789AB")
#Distribution Fitting
plot(seq(0, 250, 5), dexp(seq(0, 250, 5), rate =
1/e.sfn), type = "l")
h <- hist(e, plot = FALSE)
lines(h$mids, h$density, type = "l", col = "red")


#For the Unit25 Network
for(i in 1:l)
{
r = runif(1)
s <- qweibull(r, shape = 1.03, scale = 16)
p <- qltriangle(r, a = 181, b = 251, c = 194,
logbase = 10)
e[i] <- (s*p)/1000
}
e.u25 <- mean(e)
hist(e, xlab = "Excess CO2 Emissions (kg)", ylab =
"Frequency",
main = "Distribution of Excess CO2 Emissions
(Unit25)", col = "#6789AB")
#Distribution Fitting
plot(seq(0, 25, 1), dexp(seq(0, 25, 1), rate =
1/e.u25), type = "l")
h <- hist(e, plot = FALSE)
lines(h$mids, h$density, type = "l", col = "red")


#For the Sioux Falls Network
for(i in 1:l)
{
r = runif(1)
```

```
s <- qweibull(r, shape = 1.90, scale = 104)
p <- qltriangle(r, a = 181, b = 251, c = 194,
logbase = 10)
e[i] <- (s*p)/1000
}
e.u100 <- mean(e)
hist(e, xlab = "Excess CO2 Emissions (kg)", ylab =
"Frequency",
main = "Distribution of Excess CO2 Emissions
(Unit100)", col = "#6789AB")
#Distribution Fitting
plot(seq(0, 100, 2), dexp(seq(0, 100, 2), rate =
1/e.u100), type = "l")
h <- hist(e, plot = FALSE)
lines(h$mids, h$density, type = "l", col = "red")
#End of Code
```

Appendix B – Cullen And Frey Plots

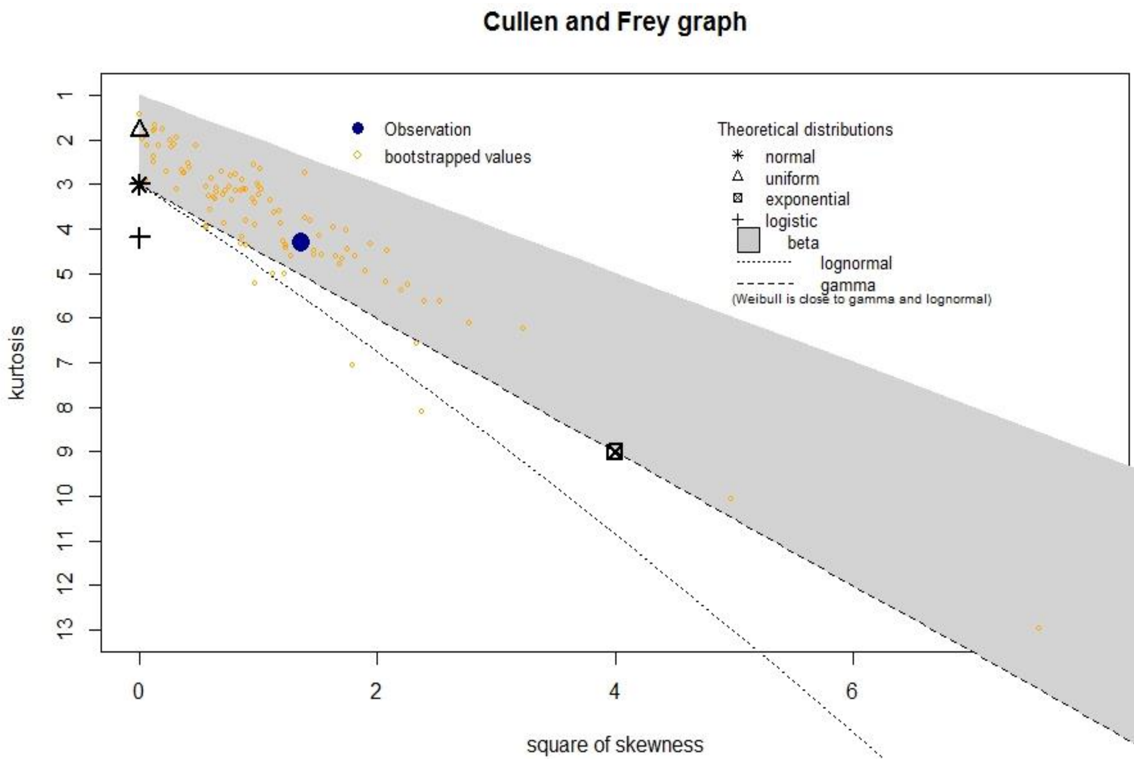


Figure 23: Cullen and Frey Graph for Sioux Falls Network

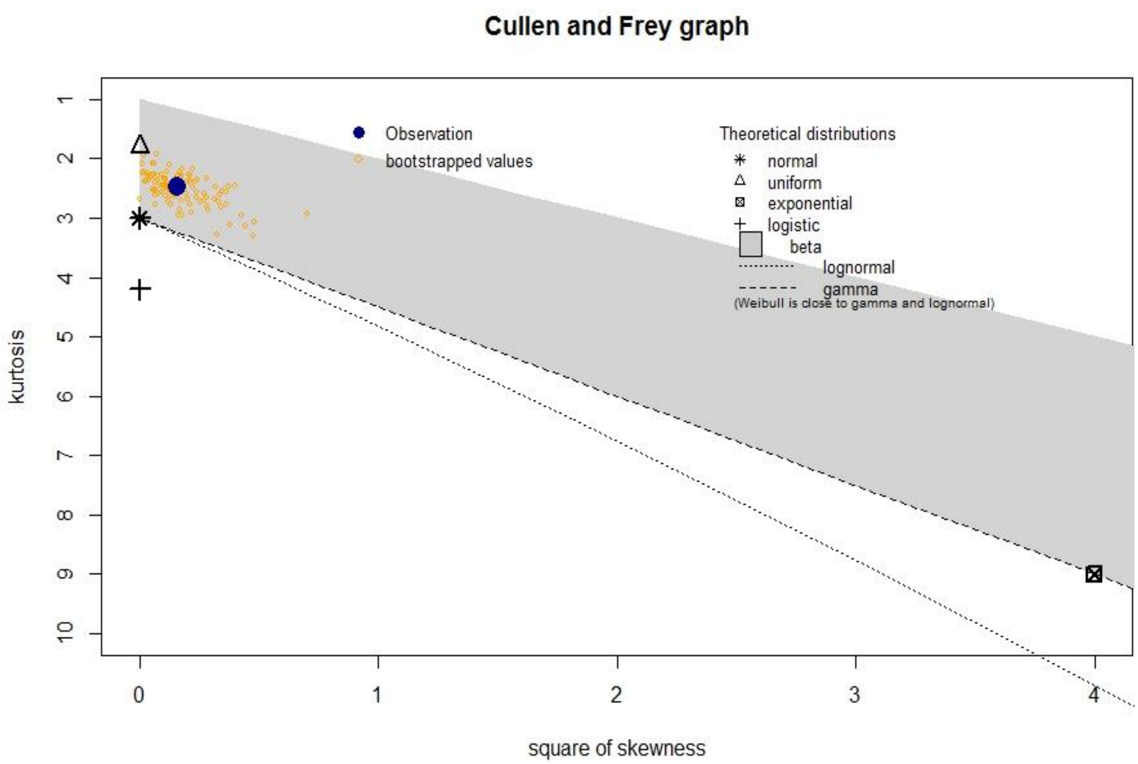


Figure 24: Cullen and Frey Graph for Unit Distance Network (n = 100)

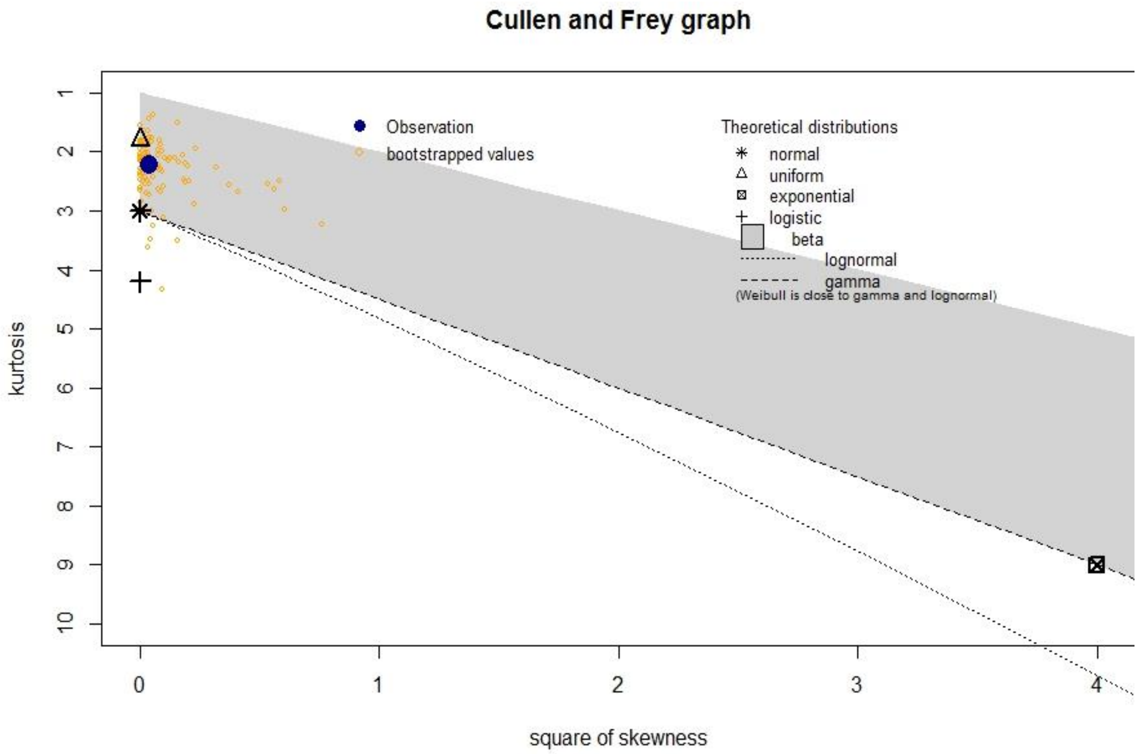


Figure 25: Cullen and Frey Graph for Unit Distance Network (n = 25)

Appendix C – Optimization Program for MLE of Weibull Distribution

In this part, we shall define the problem of finding maximum likelihood estimates of the Weibull Distribution and solve it using a goal seeking method.

For the Weibull Distribution,

$$F(X \leq x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$$

$$f(X = x) = \frac{\mathrm{dF}}{\mathrm{dx}} = \frac{k}{\lambda} * \left(\frac{x}{\lambda}\right)^{k-1} * e^{-\left(\frac{x}{\lambda}\right)^k}$$

The likelihood function, L shall be:

$$L = \prod_{i=1}^{n} f(X = x_i) = \left(\frac{k}{\lambda^k}\right)^n * \prod_{i=1}^{n} x_i^{k-1} * \prod_{i=1}^{n} e^{-\left(\frac{x_i}{\lambda}\right)^k}$$

We shall now write the optimization function:

Objective: max ln(L)

Subject to:

$$\frac{\partial \ln(L)}{\partial \lambda} = 0$$

$$\frac{\partial \ln(L)}{\partial k} = 0$$

The constraints from algebra yield the following:

$$\lambda = \sqrt[k]{\frac{\sum x_i^k}{n}}$$

$$\frac{n}{k} + \sum \ln x_i = n * \ln(\lambda) + \left(\frac{1 - \ln(\lambda)}{\lambda}\right) * \sum x_i^k$$

We now present the solutions to these equations from Excel's GRG Non-Linear Solver.

Network	**Parameter Values**	
	Shape, k	*Scale, λ*
Sioux Falls Network	1.445	220.935
Unit25 Network	1.062	16.242
Unit100 Network	1.433	100.138

Table 8: SED Parameters from Excel Solver

References and Works Cited

The following works have been cited in this thesis (in alphabetical order):

Adjetey-Bahun, K., Birregah, B., Chatelet, E., & Planche, J.-L. (2014). *A simulation-based approach to quantifying resilience indicators in a mass transportation system*. University Park, PA: Pennsylvania State University Press.

Albores, P., & Shaw, D. (2008). Government preparedness: Using simulation to prepare for a terrorist attack. *Computers & Operations Research*, 1924 - 1943.

Allenby, B., & Fink, J. (2002). Social and Ecological Resilience: Towards inherently secure and resilient societies. *Science, 24*(3), 343-364.

Amrit, C., & ter Matt, J. (2013, July). Understanding Information Centrality Metric: A Simulation Approach. Twente, The Netherlands: Working Paper.

Buhl, J., Gautrais, J., Reeves, N., R.V.Sol´e, Valverde, S., Kuntz, P., & Theraulaz, G. (2006). Topological patterns in street networks of self-organized urban settlements. *THE EUROPEAN PHYSICAL JOURNAL B*, 513 - 522.

Carvalho, H., Barroso, A. P., Machado, V. H., Azevedo, S., & Cruz-Machado, V. (2012). Supply chain redesign for resilience using simulation. *Computers & Industrial Engineering*, 329 - 341.

City of Indianapolis vs Edmond, 99-1030 (United States Supreme Court November 28, 2000).

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality in Networks of Urban Streets. *CHAOS*, 16 (015113).

Department of Transportation. (2011, August 15). *Complete Streets*. Retrieved May 26, 2017, from https://www.dot.ny.gov/programs/completestreets

Derrible, S. (2012). Network Centrality of Metro Systems. *PLoS ONE, 7*(7), 1-10. doi:10.1371/journal.pone.004057

Doolittle, N. (2016, April 14). Part of East Avenue to close to all traffic April 26-May 8. *Cornell Chronicle*, pp. http://news.cornell.edu/stories/2016/04/part-east-avenue-close-all-traffic-april-26-may-8.

Estrada, E., & Hatano, N. (2010). Resistance Distance, Information Centrality, Node Vulnerability and Vibrations in Complex Networks. In E. Estrada, M. Fox, D. J. Higham, & G.-L. (. Oppo, *Network Science: Complexity in Nature and Technology* (pp. 13-29). London: Springer-Verlag.

Haimes, Y. (2009). On the Definition of Resilience in Systems. *Risk Analysis, 29*(4), 498-501.

Hillier, B., & Hanson, J. (1984). *The Social Logic of Space.* Cambridge, UK: Cambridge University Press.

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A Review of Definitions and Measures of System Resilience. *Reliablity Engineering and System Safety*, 47-61.

Larson, R. (1981). *Urban Operations Reseach.* Upper Saddle River, NJ: Prentice-Hall.

Latora, V., & Marchiori, M. (2007). A measure of centrality based on the network efficiency. *New Journal of Physics*, 188 - 199.

Merriam Webster Dictionary. (2016, 11 26). Resilience. Retrieved from http://www.merriam-webster.com/dictionary/resilience

National Infrastructure Advisory Council. (2015). *Transportation Sector Resilience Final Report and Recommendations.* Washington DC: NIAC.

NIAC. (2009). *Critical Infrastructure Resilience.* Washington DC: NIAC.

Ognyanova, K. (2015, 06). *Network Visualization with R.* Retrieved 04 20, 2017, from www.kateo.net: http://www.kateto.net/wp-content/uploads/2015/06/Polnet%202015%20Network%20Viz%20Tutorial%20-%20Ognyanova.pdf

Philips, P. (2015). *Why Urban Roads Cost More and Deliver More.* Critique, Salt Lake City.

Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design*, 705 - 725.

Smart Growth America. (2016). *cs-equity.pdf.* Retrieved 06 05, 2017, from old.smartgrowthamerica.org: http://old.smartgrowthamerica.org/documents/cs/factsheets/cs-equity.pdf

The Data Team, The Economist Intelligence Unit. (2016, August 18). *Daily chart | The world's most liveable cities*. Retrieved from www.economist.com: http://www.economist.com/blogs/graphicdetail/2016/08/daily-chart-14

The Economic Intelligence Unit. (2016). *A Summary of the Liveability Ranking and Overview*. London: The Economist.

The Editors of Encyclopaedia Britannica. (2011). Roman Road System. In *Encyclopaedia Britannica.* Encyclopaedia Britannica, Inc. Retrieved 04 27, 2017, from https://www.britannica.com/technology/Roman-road-system

United Nations. (2014). *World Urbanization Prospects.* New York: United Nations.

Vahia, M., & Yadav, N. (2010). Reconstructing the History of Harappan Civilization. *Social Evolution & History*, 27 - 49.

Vigrin, E. D., Warren, D. E., Ehlen, M. A., & Camphouse, R. C. (2010). A Framework for Assessing the Resilience of Infrastructure and Economic Systems. *Sustainable and Resilient Critical Infrastructure Systems*, 77-116.

Wilson, G. A. (2000). *Complex Spatial Systems: The Modeling Foundations of Urban and Regional Analysis.* Upper Saddle River, NJ : Prentice-Hall.