

Tracking Energy and Resource Consumption for Sustainable and Resilient Development

BY

SK NASIR AHMAD

B.S., Bangladesh University of Engineering and Technology, 2009

M.S., University of Illinois at Chicago, Chicago, 2014

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Sybil Derrible, Chair and Advisor

Amid Khodadoust

Krishna Reddy

Thomas Theis

Isabel Cruz, Computer Science

ACKNOWLEDGMENTS

Dedicated to my wonderful parents, Abul Kamal Mumajjad Ahmad and Nazira Begum because of their unwavering support throughout my life.

I would like to express my gratitude to my advisor, Dr. Sybil Derrible, for his support, patience, and encouragement to achieve my goal. His technical and editorial advice was essential for the completion of my dissertation successfully.

Then, I would like to thank Dr. Amid Khodadoust, Dr. Krishna Reddy, Dr. Thomas Theis, and Dr. Isabel Cruz for being in the committee and help to overcome the technical difficulties I have encountered. Moreover, This research was supported, in part, by the National Science Foundation (NSF) CyberSEES Award CCF-1331800, the NSF CAREER Award 1551731, the NSF CPS Award 1646395, the University of Illinois at Chicago Institute for Environmental Science and Policy (IESP) Pre-Doctoral Fellowship, the Dean's Scholar Fellowship by the Dean of the Graduate College of the University of Illinois at Chicago, and by the Department of Civil and Materials Engineering at the University of Illinois at Chicago.

Finally, I would like to thank my wife Jobaida Gulshan Ara Hanif to whom I am utterly grateful for her untiring support, encouragement and patience.

Contribution of Authors

Chapter 1 is the introduction of my dissertation, in which I highlight the scope of this dissertation. **Chapter 2** offers a literature review that places my dissertation question and review preview works in the context of my dissertation. **Chapter 3** represents a traditional statistical analysis of the data that I used in my dissertation. **Chapter 4** represents two published articles (1. Derrible, Ahmad (2015). Network-Based and Binless Frequency Analyses. *PLoS ONE*, 10(11), pp. e0142108. and 2. Ahmad, Derrible, Managi (2018). A network-based frequency analysis of Inclusive Wealth to track sustainable development in world countries. *Journal of Environmental Management*, 218, pp. 348–354.). For the first one, my advisor Dr. Sybil Derrible was the first author and I was the second author, and for the second one I was the primary author; Dr. Sybil Derrible and Dr. Sunshuke Managi are also contributed to the writing of the manuscript published in the *Journal of Environmental Management*. **Chapter 5** represents a published article (Ahmad, Derrible, Eason, Cabezas (2016). Using Fisher information to track stability in multivariate systems. *Royal Society Open Science*, 3, pp. 160582.), for which I was the primary author and major driver of the research. Dr. Sybil Derrible, Dr. Tarsha Eason, and Dr. Heriberto Cabezas contributed to the writing of the manuscript. **Chapter 6** represents a published article (Ahmad, Derrible, Cabezas (2017). Using Fisher information to assess stability in the performance of public transportation systems. *Royal Society Open Science*, 4(4), pp. 160920.), for which I was the primary author and major driver of the research. Dr. Sybil Derrible and Dr. Heriberto Cabezas contributed to the writing of the manuscript. **Chapter 7** represents a published article (Ahmad, Derrible (2018). An information theory based robustness analysis of energy mix in US States. *Energy Policy*, 120, pp. 167–174.), for which I was the primary author and major driver of the research. Dr. Sybil Derrible contributed to the writing of the manuscript

and in the analysis for this research. **Chapter 8** represents my overarching conclusions and the future directions of this field and this research question are discussed.

Table of Contents

ACKNOWLEDGMENTS	ii
Contribution of Authors.....	iii
Table of Contents	v
List of Figures.....	ix
List of Tables	xii
List of Variables	xiii
Summary.....	xv
1. Introduction.....	1
1.1 Traditional Analysis.....	2
1.2 Sustainable Development.....	5
1.3 Networks	6
1.4 Resources	7
1.5 Objective	8
2. Literature Review	12
2.1 Resource Consumption	12
2.2 Inclusive Wealth	16
2.2.1 Human Capital.....	17
2.2.2 Produced Capital.....	17
2.2.3 Natural Capital.....	18
2.3 Sustainability.....	19
2.3.1 The IPAT Equation.....	21
2.3.2 Urban Metabolism	22
2.4 Traditional Statistical Indicators	23
2.4.1 Mean, Mode and Median	23
2.4.2 Outliers	24

2.4.3 Histogram Analysis	25
2.4.4 L-moments.....	25
2.5 Complexity Science	27
2.5.1 Network Science.....	29
2.6 Network Theory and Measures	32
2.6.1 Network	32
2.6.2 Distance in Networks.....	33
2.6.3 Connectivity	33
2.6.4 Giant Cluster.....	34
2.6.5 Social Network	34
2.7 Information Theory	35
2.7.1 Fisher Information	35
3. Tracking Resource Consumption Using Traditional Statistics	37
3.1 Introduction.....	37
3.2 Methodology	37
3.3 Results.....	39
3.4 Discussion	40
4. Tracking Sustainable Development in World Countries Using Network Science	42
.....	
4.1 Introduction.....	42
4.2 Methodology	44
4.2.1 Network Formation.....	44
4.2.2. Orbital Position and Orbital Speed	49
4.2.3. Clustering	51
4.3 Results and Discussion	53
4.4 Conclusion	60
5. Application of Fisher Information to Assess Stability in Multivariate Systems ..	62
5.1 Introduction.....	62

5.2 Background on Fisher Information	63
5.3 Calculation Methodology.....	65
5.4 Interpretation of FI.....	71
5.5 Case Study	72
5.6 Conclusion	74
6. Using Fisher Information to Assess Stability in the Performance of Public Transportation Systems	75
6.1 Introduction.....	75
6.2 Application to Public Transportation Systems.....	78
6.3 Methodology	79
6.4 Results.....	81
6.5 Discussion	84
6.6 Conclusion	89
7. An Information Theory Based Robustness Analysis of Energy Mix in US States 91	
7.1 Introduction.....	91
7.2 Materials and Methods.....	95
7.3 Results and Discussion	97
7.4 Conclusion	107
8. Conclusion	109
8.1 Future Work.....	111
9. References.....	113
10. Appendix A: Orbital Diagrams, Orbital Speed and Heat Maps	125
11. Appendix C: Figures of all US Public Transportation Systems	132
12. Appendix C: Python Scripts	135
12.1 Fisher Information Python Scripts	135
12.1.1 sost.py.....	135
12.1.2 fisher.py	137
12.1.3 smooth.py	142
12.1.4 fisher_main.py	145

12.2 orbital Network-based Frequency Analysis (oNFA) Python Scripts	147
12.2.1 network.py	147
12.2.2 optimum_network.py	150
12.2.3 orbital.py	155
12.2.4 orbital_speed.py	159
12.2.5 oNFA_main.py	161
13. Appendix D: Copyright	163
13.1 Chapter 4	163
13.2 Chapter 5	164
13.3 Chapter 6	165
13.4 Chapter 7	166
14. Curriculum Vitae	167

List of Figures

Figure 1 Triple Bottom Line of Sustainability.....	2
Figure 2 Consumption of Electricity Generation in USA (Data Source: (EIA, 2017)).....	7
Figure 3 Number of Documents for 'Resource Consumption' in Scopus	12
Figure 4 Household Energy Usage in USA in 2009 (Source: EIA)	14
Figure 5 Total Water Withdrawals by Category in USA, 2010.....	15
Figure 6 Number of Documents for 'Sustainability' in Scopus.....	21
Figure 7 Number of Documents for 'Urban Metabolism' in Scopus.....	23
Figure 8 Mean and Mode for a) Uni-modal distributions. b) Multi-modal distributions.	24
Figure 9 Number of Documents for 'Complexity Science" in Scopus.....	27
Figure 10 Number of Documents for 'Network Science" in Scopus	29
Figure 11 Regular Small-world and Random Network (Source: Watts and Strogatz 1998)	30
.....	
Figure 12 Comparison between Regular and Scale Free Network (Source: Barabási and Albert 1999).....	31
Figure 13 Connected versus Disconnected Network A) Connected B) Disconnected.....	34
Figure 14 Histogram Analysis for Human Capital in 2000	38
Figure 15 Change in Arithmetic Mean, Median and Mode of Human, Natural, and Produced Capital from 1990 to 2014	39
Figure 16 Comparison between 2nd L-moment and Standard Deviation of Human, Natural, and Produced Capital from 1990 to 2014	40
Figure 17 Change in L-skewness and L-kurtosis of Human, Natural, and Produced Capital from 1990 to 2014	40

Figure 18 Illustration of the Network-based Frequency Analysis (NFA) method.	46
Figure 19 Impact of ζ on NB Methodology and Network Properties.	48
Figure 20 Orbital Diagram of USA for Human Capital	50
Figure 21 Evolution of the Network-based (N) mode, mean, and median for per-capita human capital, natural capital, and produced capital from 1990 to 2014. (All indicators are reported in constant 2005 U.S. dollars in thousands.)	53
Figure 22 Orbital Diagrams and orbital Speeds of the United States (USA), France, Brazil, and China. The top three rows show the orbital diagrams, where the inner most circles denote the orbital position of the mode at 100 and the outer most circles denote the orbital position of 250. The bottom row shows the orbital speeds since 2000. Orbital speeds between -0.5 and 0.5 are marked with beige, whereas orbital speeds less than 0.5 are marked with red and orbital speeds greater than 0.5 are marked with purple.	54
Figure 23 Number of countries with increasing, decreasing or constant orbital speed. The top row shows the orbital speed since 1990 and the bottom row shows the orbital speed since 2000 for human (first column), natural (second column) and produced (third column)	56
Figure 24 Combinations of Human, natural and produced capital for all the countries for the years: A) 1990, B) 2000, and C) 2014. Combinations 222, 221, 212, 122, 211, 121, 112, and 111 denote groups 1, 2, 3, 4, 5, 6, 7 and 8 respectively (see color bar and table S1 in supplementary section for details about the combinations).	57
Figure 25 Combinations of human, natural and produced capital for all the countries for the years: A) 2020, B) 2030, C) 2040, and D) 2050. Combinations 222, 221, 212, 122, 211, 121, 112, and 111 denote groups 1, 2, 3, 4, 5, 6, 7 and 8 respectively (see color bar and table S1 in supplementary section for details about the combinations).	59

Figure 26 Illustration of the binning process to calculate FI. Note that binning of the points in table 5 resulted in four states.....	68
Figure 27 Probability Distribution.....	69
Figure 28 Amplitude of the Probability Distribution.....	70
Figure 29 Evolution of A) Global-mean Temperature from 1880 to 2015 and B) FI for Global-mean Temperature.	73
Figure 30 Total UPT for eight major UZAs	82
Figure 31 Total VRM for eight major UZAs.....	83
Figure 32 Evolution of FI for eight UZAs with: (a) regime shift with rebound, (b) regime shift with partial rebound, (c) regime shift without	85
Figure 33 FI for eight major UZAs.....	87
Figure 34 Scenario of Fossil Fuel Usage for Electricity Generation in the USA in 2015.	94
Figure 35 Evolution of entropy from 1960 to 2014 for all states in the USA	100
Figure 36 Robustness Ranking in Energy Mix for Electricity Generation for All U.S. States.....	105

List of Tables

Table 1 Global Energy Indexes Evolution between 1973 and 2004 (EIA, 2017)	13
Table 2 List of indicators included in Inclusive Wealth Report 2017	19
Table 3 Combinations and Groups for Clustering	52
Table 4 Number of Countries per Group	58
Table 5 Sample Data for a Time Window	67
Table 6 Patterns in the evolution of FI.....	84
Table 7 Computation of Entropy for California in 2015	95
Table 8 Computation of ΔH for a disruption of 50% in California	97
Table 9 Maximum Change in Entropy for All States in the USA	101
Table 10 Robustness in Energy Mix for Electricity Generation for All U.S. States	103
Table 11 Classification of States Based on Robustness in Energy Mix	107

List of Variables

λ_r = r^{th} Population L-moment

τ_3 = L-skewness

τ_4 = L-kurtosis

A_{ij} = Adjacency Matrix of a Network

d_i = Degree of a Node

ρ = Density of a Network

ζ = Cutoff Percentage

M = Network-based (N) Mode

L_{avg} = Average Shortest Path Length

D = Diameter of a Network

V = Number of Vertices in a Network

V_g = Number of Vertices in the Giant Cluster

p_g = Proportional Number of Vertices in the Giant Cluster

H = Homogeneity Information Index

$O_{i,t}$ = Orbital Position for Node i in Year t

D_i = Orbital Distance

O'_i = Orbital Speed for Node i

HC = Human Capital

NC = Natural Capital

PC = Produced Capital

FI = Fisher Information

VRM = Vehicle Revenue Miles

UPT = Unlinked Passenger Trips

$H(x)$ = Shannon Entropy

ΔH_R = Robustness of the Energy Mix

Summary

The main objective of this dissertation is to analyze trends in energy and resource consumption to measure sustainable and resilient development with specific applications to world countries, the United States (US) states power grid, and US urban public transportation systems. Using human (HC), natural (NC), and produced (PC) capital from Inclusive Wealth as representatives of the triple bottom line of sustainability and inspired by network science, I first introduce a Network-based Frequency Analysis (NFA) method to track sustainable development in world countries from 1990 to 2014. The method compares every country with every other and links them when values are close. The country with the most links becomes the main trend, and the performance of every other country is assessed based on its ‘orbital’ distance from the main trend. Orbital speeds are then calculated to evaluate country-specific dynamic trends. Overall, I find an optimistic trend for HC only, indicating positive impacts of global initiatives aiming towards socio-economic development. In addition, I also find that the relative performance of most countries has not changed significantly in this period, regardless of their gradual development. Furthermore, I develop a technique to cluster countries and project the results to 2050 and find a significant decrease in NC for nearly all countries, suggesting an alarming depletion of natural resources worldwide.

Then, Fisher information (FI), originally developed by Ronald Fisher as a means of measuring the amount of information about an unknown parameter that is present in any observable data, was used to assess stability in multivariate systems. An open source Python script to calculate FI was developed and subsequently applied to analyse stability in the performance of Public transportation systems (PTS) in the 372 US urbanized areas (UZA) reported by the National Transit Database.

Finally, by looking at the 50 US states, Shannon entropy was adapted to measure how state-wide electricity energy mixes have evolved in every US states, and it was then used to assess how robust current energy mixes are to any disruption. I notably observe changes for 26 states between the years 1968 and 1980. From simulating several types of disruptions, I then detect three different classes of states: vulnerable (10 states), moderately robust (17 states), and robust (23 states). Expectedly, some states are particularly vulnerable as they depend predominantly on a single energy source (e.g., West Virginia with 95% coal usage). In contrast, I find seven states (i.e., South Dakota, Illinois, Vermont, Connecticut, Maine, New York, and New Jersey) that have particularly robust energy mixes, all with fossil fuel shares below 50% in 2015.

The main technical contributions of this thesis are the development of the NFA method and the adaptation of Shannon entropy to evaluate robustness of a system, both of which were applied to extract meaningful information to track sustainable and resilient development. Additionally, the development of an open source Python script to compute FI offers another significant contribution of this dissertation. Overall, the methods developed in this dissertation can be applied to any data, irrespective of the scale, to extract meaningful information and gain new insights about a system. The current proliferation of data creates the opportunity to understand our surrounding systems (e.g., urban engineering systems) in a much better way, which is critical to ensure the sustainability and resiliency of any system. Nonetheless, it can be challenging to precisely interpret data and the methods presented in this dissertation possess the potential to tackle those challenges, which is fundamental if we, as a society, aspire to become more sustainable and resilient.

1. Introduction

With the growth of population and advancement in human civilization, the world around us is changing rapidly and so are our trends and patterns in energy and resource consumption. Mankind inherently and fundamentally depends on the Earth's resources for its survival. Providing for all humans, however, is no small feat considering the current population of 7.3 billion people (in 2017), which is further expected to grow and reach 8.5 billion by 2030 (World Bank, 2015). Furthermore, the world is also urbanizing rapidly. The United Nation predicts that by 2050, 66% of the population of the world will live in cities; currently it is at 54% (United Nations, 2017). As a natural consequence of this massive demographic change and urbanization, access to the resources that are vital for the survival of our society become more challenging than ever. Moreover, the rate of depletion of these resources is alarmingly high, partly because of the rapidity at which the world is urbanizing and partly because of the global inability to reduce per capita demand for these resources, resulting in immensely unsustainable trends. This ever-growing demand for resources not only threatens the availability of these resources but it also puts mankind's future security, health and well-being at stake. For a more sustainable world, we need to improve our ability to capture and track trends in energy and resource consumption in order to ensure the balance between people, planet and prosperity, commonly known as the triple bottom line of sustainability (shown in Figure 1). In a similar line of thought, as most failures or crises are highly uncertain by nature and often seem extraordinary, we also need to improve our ability in tracking resiliency for energy and resource consumption. This is all the more important that some initial "small" failures occur in a cascading or escalating form, possibly leading to system collapse (Carlson et al., 2012; Petit et al., 2015).

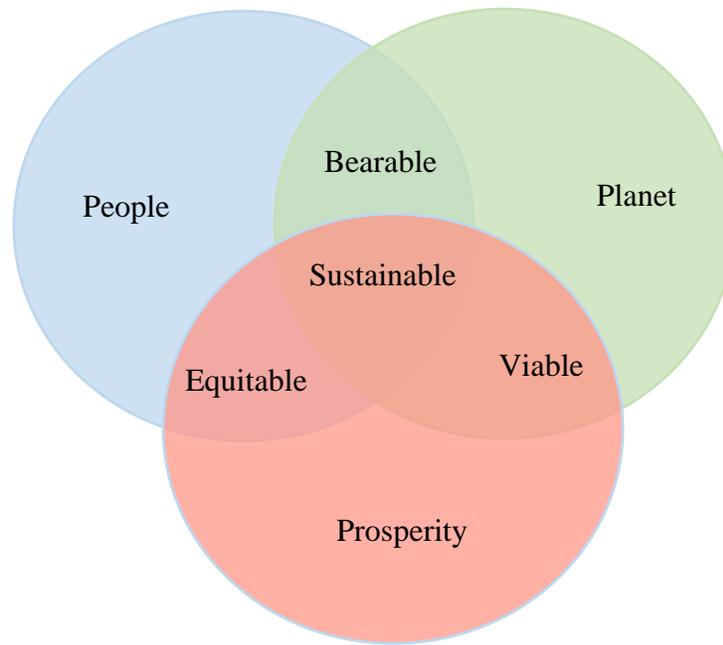


Figure 1 Triple Bottom Line of Sustainability

1.1 Traditional Analysis

Conventionally, traditional statistical metrics (e.g., mean, median, standard deviation) are used to gauge and assess trends and patterns in energy and resource consumption. In this era of advanced computing and decision making, however, coupled with advancement in data collection and data storage technology, enormous amounts of information are generated every day. Traditional metrics can easily fail to provide meaningful insights from this avalanche of information. In particular, errors are frequent in any dataset, and these errors can cause significant biases. In addition, the presence of outliers can also make it challenging to deal with these data. The problems are not new, however, traditional analyses has had these challenges since at least the beginning of statistics (Swetz, 2012), but the scale of the datasets used is so large that semi-automatic processes need to be developed. For example, one of the most common and widely used metrics to gauge any data is the arithmetic mean, which can be highly biased and misleading with the presence of errors and outliers. Moreover, probability distribution functions

are also used extensively to interpret any dataset and find meaningful patterns. Common probability distributions used include normal, lognormal, and logistic distributions, most of which are parametric distributions. As such, they require the calculation of parameters (e.g., mean, median, standard deviation), which, if biased, can wrongly fit a probability distribution function. For instance, the mean is often used to denote the magnitude or location of a distribution, whereas the standard deviation is used to represent the spread of a distribution. Therefore, biases due to the presence of errors and outliers can easily propagate to the interpretation of different probability distribution functions, which can therefore misinform us on actual energy and resource consumption trends.

There are many techniques to fit any data to a distribution from simple regression techniques and method of moments, to method of L-moments, and maximum likelihood to name a few. For each of these techniques, however, an initial assumption of the form is required in order to fit the data. Besides using different traditional statistical tools, frequency analyses using histograms are commonly used to fit those distributions. Histograms are defined by a base and a height, where the base represents some predefined interval and the height represents the frequency of events. In order to plot a histogram, the range of the values needs to be “binned,” which can be achieved by dividing the entire range of values into a sequence of small intervals. Then the frequency of events within those small intervals are counted and used as the height of each bar in a histogram. John Graunt was the first to group the data in this fashion, but no methodical guidelines were given (Scott, 1992). In 1926, Herbert Sturges first suggested some rules to choose the data interval (Sturges, 1926). Since then, different rules (e.g., Rice, Scott,

Freedman-Diaconis) have been developed and used to determine a bin size. One of the most utilized rules to select the bin size is Scott's rule, where the size (h) of the bin is defined as:

$$h = \frac{3.49 \sigma}{n^{1/3}} \quad (1)$$

where, σ is the standard deviation and n is the sample size. Although this equation is extremely simple, we can see that it is based on the standard deviation, and it is therefore prone to outliers as discussed previously. Moreover, the bin size provided is only optimal for normally and randomly distributed data, which is often not the case in nature (or at least not to study energy and resource consumption as we will see). The selection of bin size is therefore highly challenging, as choices of different bin size can yield different interpretations. Moreover, the choice of too large a bin size can fail to capture relevant information, whereas the selection of too small a bin size can lead to exaggeration. In addition, distributions with a multi-modal feature are even more challenging to fit than uni-modal distributions when selecting a proper bin size. Besides this point and as mentioned previously, another critical factor for data analysis is the presence of outliers. Indeed, outliers can easily biased the computation of the statistical metrics, which are used to fit a distribution. To avoid this problem, the outliers are usually removed, which can be based on different formulations. For example, one of the commonly used techniques to identify outliers is the use of the standard interquartile range procedure, which involves using the upper quartile, the lower quartile and the inter quartile range. Furthermore, different techniques to identify outliers can also produce different interpretation, which makes it even difficult to analyze any system with myriad of information. More about these problems are discussed in chapter 3.

1.2 Sustainable Development

The world around us is developing constantly. Because of this development, different human activities are putting pressure on the environment. These environmental stressors are responsible for the alteration of the ecosystem as well as the economy. There is an argument over the exact definition of sustainability but the systems around us, including the ecosystem that are significantly important for human survival, are clearly going through a process of modification. Consequently, a current movement towards higher sustainability and sustainable development is rapidly growing. The systems around us can be simple, complicated, or complex in nature. Moreover, these systems have different characteristics; some systems are substantially brittle and unstable, while some are adaptable. In practice, most systems around us are complex in nature, with myriads of interacting components. According to the Brundtland Commission (Burton, 1987), sustainable development is “development that meets the needs of the present without compromising the ability of future generations to meet their own needs”, which suggests that sustainability is about retaining a condition suitable for social and economic development, without affecting the environment. As mentioned earlier and depicted in Figure 1, the concepts of sustainability applied to integrated systems involves people, planet and prosperity. Hence, to achieve sustainable development, an efficient and proper evaluation and management of human (e.g., social and economic) and natural components (e.g., climate and ecosystems) of systems is required (Karunanithi et al., 2008), as they are significantly complex and integrated. In particular, in this era of Big Data, huge amounts of information about any system are available for comprehensive assessment of that system, and understanding trends quantitatively within that system should therefore now be within reach. Different tools have been applied to assess these complex systems and information theory is one of them. From an information theory viewpoint,

all these complex systems are orderly and well organized dynamic system, but they are also able to undergo abrupt changes in order to shift to a different dynamic regime, commonly referred to as regime shift (e.g., eutrophication of lakes and coastal oceans, regional climate change) (Karunanithi et al., 2008). Many studies have attempted to measure the impact of global trends on sustainability, involving both positive and negative movement of indicators (Kates and Parris, 2003). Therefore, an evaluation and management of the evolution of these changes related to sustainable development is critical. Nonetheless, as numerous indicators are involved with these complex systems, the scrutiny of these systems offers significant challenges.

1.3 Networks

As mentioned earlier, complex systems have interacting components, which means the components of a complex system establish connection with one another and form networks. These networks can have functional flows (e.g., telecommunication network, computer network) or can be free of functional flows akin to social networks. Social networks are significantly complex in nature and myriads of concepts are emerged from their study. Among them, *homophily* is an important concept for this thesis, studied for the analyses of social networks. The concepts of *homophily* emerges from the propensity of human beings to link (e.g., become friends, develop relationships) with similar others. For instance, the existence of common attributes like beliefs, education, and social status have been observed among people involves in homophilic relationships (McPherson et al., 2001). A term similar to homophily is homogamy, which is the marriage between people with alike personalities.

1.4 Resources

The consumption rate of a product should be less than or equal to the production rate of that particular product (Derrible, 2015), which can be inferred from the Brundtland Commission definition on sustainability. Thus, tracking resource (i.e., raw materials, fuel and water) consumption is significantly important to ensure sustainable development. Moreover, resources are consumed for development and maintenance purposes, which also act as a stressor on the ecosystem and the biodiversity. Figure 2 shows amount of energy consumed in the United States (US) for the generation of electricity from 1960 to 2015 (EIA, 2017) and an increasing trend is observed until recently. Furthermore, fair amounts of resources are consumed by residential buildings.

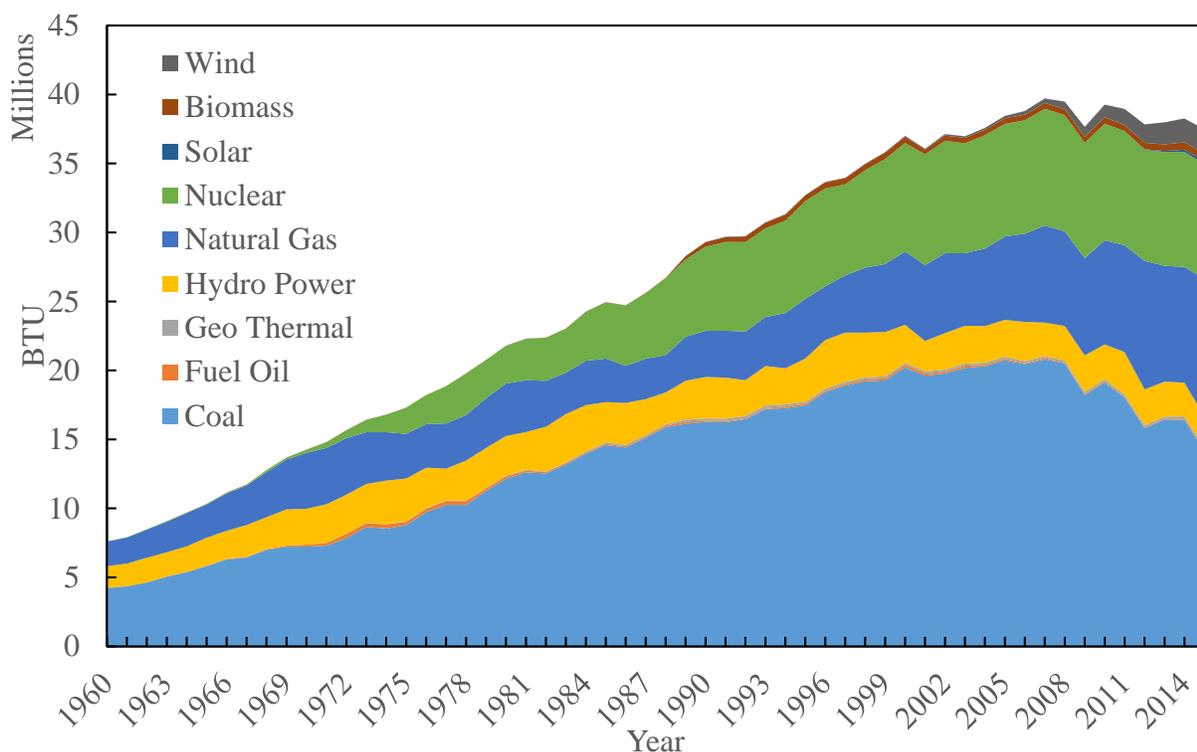


Figure 2 Consumption of Electricity Generation in USA (Data Source: (EIA, 2017))

For instance, in 2014, 41% of total energy was consumed by residential and commercial buildings in the US (EIA, 2017). Although new technologies helped us use more renewable

energy, non-renewable energy remains by far the major source of energy. These scenarios of different resource consumption are similar in other countries due to constant pressure of development. Moreover, emissions of CO₂ and other pollutants are most often associated with these energy and resource consumption. Therefore, the question remains, whether energy consumption, CO₂ emissions, and associated economic development for all the countries possess similar patterns. As energy and resources are consumed for development, the main hypothesis of this thesis is: **‘if two entities (e.g., country, state, city, county) have similar energy and resource consumption patterns, they should also have similar developmental scenario. In turn, if two entities have a similar economic status, they should have a similar environmental performance (i.e., in terms of greenhouse gas emissions)’**. Any dissimilarity can be used to comparatively assess how sustainable an entity is.

1.5 Objective

To test this hypothesis, the main objective of this dissertation is to analyze trends in energy and resource consumption in world countries and US states as a measure of sustainable and resilient development. The main goal is tackled through the following specific objectives:

- **Compute and compare traditional statistics to track energy and resource consumption,**
- **Create a novel and more accurate method to track energy and resource consumption using concepts homophily and network science,**
- **Develop and apply a Network-based Frequency Analysis (NFA) method to track sustainable development in World Countries,**

- **Assess the robustness of the energy mix of US states using Information Theory.**

These objectives are addressed sequentially. Data from the 2014 and 2017 Inclusive Wealth Report (UNU-IHDP and UNEP, 2014; Urban Institute and UNEP, 2017) are used for the analysis. To achieve the specific goals, traditional statistical tools are calculated to find resource consumption patterns. Applying traditional statistical measures will further highlight their shortcomings. First, the arithmetic mean, median, mode, and standard deviation are analyzed as they represent the most common means to get a sense of the data being analyzed. Moreover, to define and explain any probability distribution function, conventional moments are also used. For example, mean is used as the first moment, whereas variance is used as the second moment. We will see that the conventional moments can fail to capture accurate information, especially by its third or higher moment or when the sample size is too small. In contrast, L-moments are similar to conventional moments but estimated from the linear combinations of ordered statistics or L-statistics. Besides computing the first three L-moments, L-skewness and L-kurtosis are also computed for the L-moment analysis and these measures are directly compared with the conventional moments.

In order to address some of the problems associated with the traditional analysis, a novel and more robust methodology has been created that can effectively supplement the traditional method. For this methodology, I use concepts from the emerging field of network science. Network science has evolved intensely for the past 15 years and applied in myriads of fields, often resulting in a better understanding of the complex systems around us that include telecommunication networks, computer networks, biological networks, cognitive and semantic networks, and social networks to name a few (Easley and Kleinberg, 2010; Newman, 2010). The

main concept of the methodology is to compare every value of a dataset with every other by measuring whether they are within a certain range of one another. Then a network is developed based on the concept mentioned earlier and further analyzed. Moreover, the developed network is also assessed through new metrics that are able to provide novel information about the system. The networks created therefore do not possess functional flows; instead, they are conceptually close to social networks. One way to consider the method is that the “peers” of a country are detected when their consumption levels are similar. In the literature, concepts from social network theory have been notably applied to reduce energy consumption by sending Energy Report letters to consumer (Allcott, 2011; Allcott and Mullainathan, 2010). One of the key challenges in developing the network is to find the certain range mentioned above (i.e., what is meant by “similar” consumption), which is to some degree conceptually close to bin width of histogram analysis. An algorithm has been developed to find the optimal range and will be explained later. Optimal networks for various years and for various energy and resources are then analyzed and compared. This approach facilitates the assessment of all the entities (e.g., country, state or city) simultaneously, as they acting as nodes of the network.

Subsequently, information theory has been applied to assess the robustness of energy mix. After its inception, information theory has been widely used in information technology (IT). Here, Shannon entropy will be used to assess robustness in energy mix for all states in the United States (US). Shannon entropy is particularly relevant here because it is able to capture changes due to the combined effect of all energy sources (i.e., energy sources shown in Figure 2). More advanced information theory measures like Fisher Information (Ahmad et al., 2017a, 2016a) were tested but they did not capture any relevant additional information for this particular study. However, the code to calculate Fisher Information has been developed and applied in a simple

case study to evaluate the evolution of FI for the global-mean temperature from 1880 to 2015 (GISTEMP Team, 2016) and to assess stability in the performance of public transportation systems.

The programming language ‘Python’ has been used for all the analysis performed in this dissertation. In particular, I made extensive use of the igraph (Csardi and Nepusz, 2006), NumPy (van der Walt et al., 2011) and Pandas (McKinney, 2010) Python libraries. Moreover, I have developed two Python scripts (available on GitHub) that are directly used in this thesis (<https://github.com/csunlab/nbfreq> , <https://github.com/csunlab/fisher-information>)

2. Literature Review

2.1 Resource Consumption

Resources are critical for human survival. In fact, virtually all the development in this world involves some sort of resource consumption. Resource can be defined from an economist, a physicist, or an ecologist point of view. In economics, any ‘thing’ that enters into the production system can be defined as a resource. Any ‘thing’ irrespective of its usefulness is considered a resource in physics. In ecology, resources are naturally occurring and are able to sustain or benefit any living creature, including humans within an ecosystem. (Göbbling-Reisemann, 2008) defines Resources as “the flows and reservoirs of matter and energy that can sustain or benefit living systems.” Resource management is substantially important to ensure sustainable development. A keyword search for ‘Resource Consumption’ in Scopus resulted 14,127 documents, shown by year in Figure 3.

Resources have been consumed by humans for different household activities,

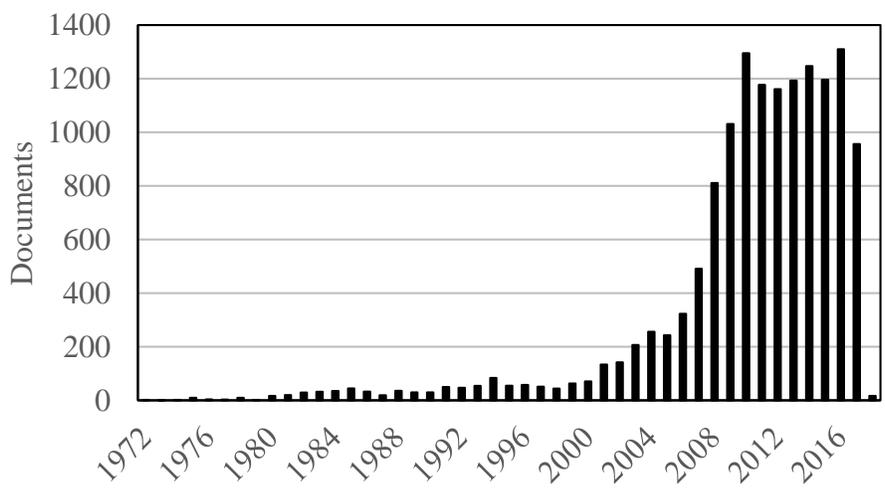


Figure 3 Number of Documents for 'Resource Consumption' in Scopus

development and transportation, initially in the form of raw materials, fuel, and water. The

concern is rising significantly over high energy use, depletion of non-renewable energy sources, and associated adverse environmental impact. According to the International Energy Agency (IEA), energy use has increased by 49% and CO₂ emissions by 43% from 1984 to 2004, with an average yearly increase of 2% and 1% respectively (Pérez-Lombard et al., 2008). Moreover, the average annual rate (3.2%) for developing countries is considerably higher than that of developed countries (1.1%). In absolute terms, China doubled its energy use over the past 20 years with an annual average rate of 3.7% (Pérez-Lombard et al., 2008). Furthermore, several facts are also found during this period, which are: 1) the growth rate of population was lower than that of primary energy consumption, suggesting an increase in per capita energy consumption; 2) CO₂ emissions increased at a lower rate than that of energy consumption; 3) Energy consumption in the form of electricity has increased significantly (Pérez-Lombard et al., 2008). Detailed information on these facts are shown in Table 1 (EIA, 2017).

Table 1 Global Energy Indexes Evolution between 1973 and 2004 (EIA, 2017)

<i>Parameter</i>	<i>1973</i>	<i>2004</i>	<i>Ratio (%)</i>
<i>Population (millions)</i>	3,938	6,352	61.3
<i>GDP (G\$ year 2000)</i>	14,451	35,025	142.4
<i>Per capita income (\$ year 2000)</i>	3,670	5,514	50.2
<i>Primary energy (Mtoe)</i>	6,034	11,059	83.3
<i>Final energy (Mtoe)</i>	4,606	7,644	66
<i>Final energy/primary energy</i>	0.76	0.69	-9.4
<i>Electrical energy (Mtoe)</i>	525 1,374	1,374	161.8
<i>Electrical energy/final energy</i>	0.11	0.18	63.5
<i>Per capita primary energy (toe)</i>	1.53	1.77	15.7
<i>Primary energy intensity (toe/G\$ year 2000)</i>	418	316	-24.4
<i>Final energy intensity (toe/G\$ year 2000)</i>	319	218	-31.5

As mentioned earlier, buildings are the main consumers of energy and the 2009 building energy use for a typical US household is shown in Figure 4. Myriads of studies analyze

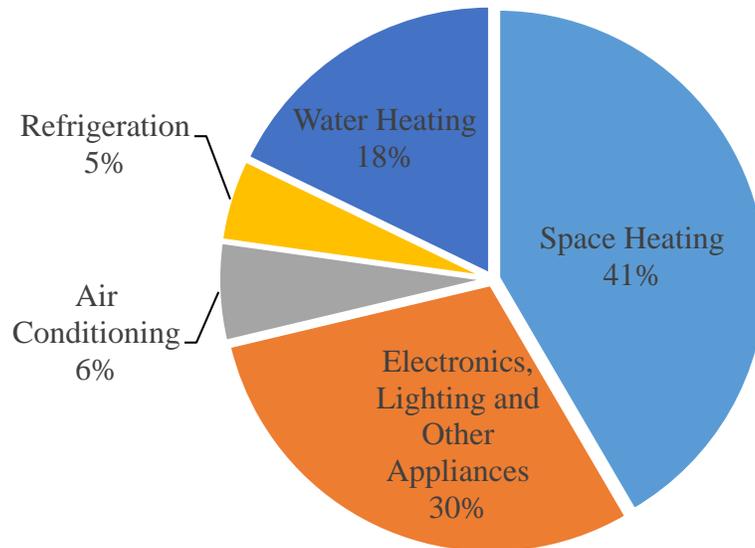


Figure 4 Household Energy Usage in USA in 2009 (Source: EIA)

household energy consumption. The purpose of these studies are to analyze energy consumption patterns from different consumer surveys and to develop models from this information. Moreover, input-output analyses and econometric analyses have also been used to analyze household survey data (Druckman and Jackson, 2008; Park and Heo, 2007; Poortinga et al., 2004; Vringer et al., 2007). Efforts have been made to reduce energy consumption for a sustainable future. For example, the use of energy efficient electronic appliances is considered as a straightforward solution to reduce energy consumption (Bin and Dowlatabadi, 2005; Brandon and Lewis, 1999; Dietz et al., 2013; Fischer, 2008; Wall and Crosbie, 2009). Moreover, efforts have been put into reducing energy consumption by sending Energy Report letters to consumers that showed a comparison of a household's electricity use with its neighbors. This energy conservation program was proven to be significantly cost effective and able to reduce energy consumption by 2% on average (Allcott, 2011; Allcott and Mullainathan, 2010). Furthermore,

the successful implementation of social concepts also encouraged the application of social knowledge to solve different problems related to energy and resource consumption.

Water is another vital resource for human survival and different developmental works. For instance, in the US, approximately 355 billion gallons of water per day (Bgal/d) were withdrawn in 2010. From this amount, 306 Bgal/d of water were withdrawn from freshwater sources and the rest from saline water sources. Major amounts of water were withdrawn (206 Bgal/d) from fresh surface water sources, and fresh groundwater withdrawals amounted to 76 Bgal/d (Maupin et al., 2014). Like other resources, usable water resources are also finite. Therefore, water resource conservation and efficient use of water is important for sustainable development. Similar to energy, the domestic sector is not the predominant consumer of water but large amounts of water are consumed for household activities. In 2010, a total of 13% of the

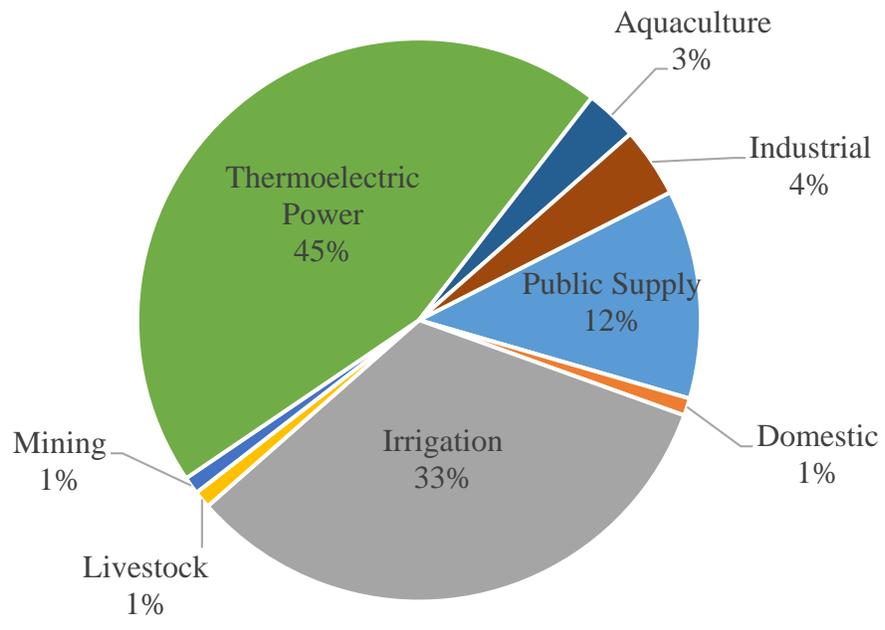


Figure 5 Total Water Withdrawals by Category in USA, 2010

total water was withdrawn for small and medium commercial and domestic use (i.e., public supply and private wells for domestic use), which has been more or less constant since 1985 (Hutson, 2004; Kenny et al., 2009; Maupin et al., 2014; Solley et al., 1998, 1993, 1988). Figure 5 shows water withdrawals by category in the US in 2010 (Maupin et al., 2014). Different studies have been carried out to find water consumption patterns in households, adopting mostly various modeling approaches, like integrated modeling using information collected from different survey; effect on people's motivation to conserve water, and so on and so forth. (Aitken et al., 1994; Berk et al., 1980; Clark and Finley, 2008; Corral-Verdugo et al., 2002; Dandy et al., 1997; Jorgensen et al., 2009).

As stated earlier, with the continuous advancement in information and storage technology, enormous amounts of data are generated every day. These include data on household resource consumption that require advanced techniques in order to extract valuable information from those data. In this era of Big Data, without the help of machines, gathering necessary information from those data is difficult. As a consequence, various techniques using machine learning and agent-based modeling have been developed to track and reduce the energy consumption in households. For example machine learning has been used to learn resource management in a smart environment; use of agent based modeling to model each agent's reaction to behavioral motivation (Fabbriatore et al., 2012; Grassi et al., 2011; Stenudd, 2010).

2.2 Inclusive Wealth

Inclusive Wealth (IW) (Arrow et al., 2012; UNU-IHDP and UNEP, 2014; Urban Institute and UNEP, 2017; World Bank, 2010) comprises a few assets in view of their peripheral commitment to social prosperity. The computation of IW includes the shadow cost of each

capital human (HC), produced (PC), and natural (NC). This price is calculated by the following weighted sum:

$$IW = P_{pc} \times PC + P_{hc} \times HC + P_{nc} \times NC \quad (2)$$

Table 2 includes the indicators of the HC, PC, and NC, which are considered in Inclusive Wealth Report 2017.

2.2.1 Human Capital

Human Capital (HC) is estimated based on a method proposed by Arrow et al. (Arrow et al., 2012) and Klenow and Rodriguez-Clare (Klenow and Rodríguez-Clare, 1997). In computing HC, education data from Barro and Lee (Barro and Lee, 2013) is used, where, education is the average number of years of total schooling per person. Moreover, the present value of the average labor payment per unit of human capital and the expected life's working period are used to calculate the shadow cost of per unit human capital. The general equation for calculating the human capital is as follows:

$$HC = \text{human capital per individual} \times \text{average year of education attainment} \times \text{shadow price per unit of human capital} \quad (3)$$

2.2.2 Produced Capital

The method proposed by Harberger (Harberger, 1988), which, also applied by King and Levine (King and Levine, 1994) and Feenstra et al. (Feenstra et al., 2013) is used to estimate Produced Capital (PC). Perpetual Inventory Method (PIM) is preferred on estimating PC, where, an initial capital estimate is essential. PIM captures the dynamic addition of the produced capital, and it also considers annual investment change. Mathematically, it is calculated by:

$$PC_t = (1 - \text{depreciation rate})^t \times \text{initial capital estimate} + \sum_{j=1}^t \text{investment}_j (1 - \text{depreciation rate})^{t-j} \quad (4)$$

2.2.3 Natural Capital

The summation of the value of agriculture land, forest resources, fossil fuels, minerals, and fishery resources is defined as Natural Capital (NC). Details on estimating NC are available in the 2014 Inclusive Wealth Report. Mathematically, it is defined as:

$$NC = \sum \text{Agricultural Land} + \text{Forest Resources} + \text{Fossil Fuels} + \text{Minerals} + \text{Fishery Resources} \quad (5)$$

Agriculture land values are estimated using their corresponding shadow costs (i.e., net present value (NPV) of future rental flows) and the physical amount of cropland area available every year. Volume of commercially available timber is used for forest resources calculation, where, the total forest area (excluding cultivated forest) is multiplied by timber density per area and percentage of commercially available volume.

The unit rental price and the stocks for each of the 140 countries are multiplied to compute fossil fuels (coal, natural gas, and oil) wealth for the period 1990 to 2014. Sectoral rental rates of different mineral industries obtained from (Narayanan and Walmsley, 2008) are multiplied by their corresponding prices to calculate minerals. Then mineral stocks are multiplied by rental prices to get the total mineral wealth for 1990 to 2014. Fishery resources are roughly estimated from the size of fish stocks. The annual biomass is estimated by using the Schaefer production model proposed by Martell and Froese (Martell and Froese, 2013).

Table 2 List of indicators included in Inclusive Wealth Report 2017

		Natural Capital	Human Capital	Produced Capital
Adjusted IW	IW	Fossil Fuel	Education	Equipment
		Oil		Machineries
		Natural gas		Roads
		Coal		others
		Minerals		
		Bauxite, Nickel, Copper, Phosphate, Gold, Silver, Iron, Tin, Lead and Zinc		
		Forest resources		
	Timber			
	Non-timber forest			
	Factors affecting IW	IW	Agricultural land	
Cropland				
Pastureland				
Fishery				
		Stock		
		(1) Carbon damages		
		(2) Oil capital gains		
		(3) Total factor productivity		

2.3 Sustainability

The most used and quoted definition of sustainability is given by the Brundtland Commission (Burton, 1987) and was mentioned earlier. The principles of sustainability are significantly comprehensive and considerably challenging to implement in practice. The Brundtland Commission's definition of sustainability is indeed substantially broad, including many issues and discussed in different literatures (Carter and Rogers, 2008); for example, by considering the environmental impact of economic activity (Ehrlich and Ehrlich, 1990); ensuring

food security for all (Lal et al., 2002); ensuring basic human needs (Savitz and Weber, 2006); studying the preservation of non-renewable resources (Whiteman and Cooper, 2000). Most of these studies focused on ecological sustainability (e.g., natural environment), rather than explicitly considering social and economic responsibilities (Carter and Rogers, 2008; Daily and Huang, 2001; Hill, 2001; Jennings and Zandbergen, 1995; Sarkis, 2001; Shrivastava, 1995; Starik and Rands, 1995). For instance, according to (Shrivastava, 1995) sustainability is described as the solution for reducing enduring risks of resource depletion. Moreover, it is also regarded as a potential remedy for current instabilities in energy costs, product obligations, and pollution and waste management. Principles of sustainability have been widely used in engineering, considering both the economic and social impact explicitly in addition to the environmental impact (Sikdar, 2003). Furthermore, equal weight tends to be assigned to all the impacts (Gończ et al., 2007). The three components of sustainability, namely people, planet and prosperity, are known as the 'triple bottom line', as depicted in Figure 1. The concept of triple bottom line was originally developed by Elkington (Elkington, 2004, 1998), balancing economic, environmental and social goals from microeconomics point of view (Carter and Rogers, 2008). As sustainability is a broad concept, many issues are also associated with sustainability such as risk management, transparency, strategy, and culture (Carter and Rogers, 2008; Gladwin et al., 1995; Hart, 1995; Henriques and Richardson, 2013; Jennings and Zandbergen, 1995; Sarkis, 2001; Savitz and Weber, 2006; Shrivastava, 1995; Starik and Rands, 1995). Beyond these studies, myriads of studies include to some extent concepts of sustainability. When searching for the keyword 'Sustainability' in Scopus, a total of 105,053 documents were found between 1970 and 2014 and shown in Figure 6.

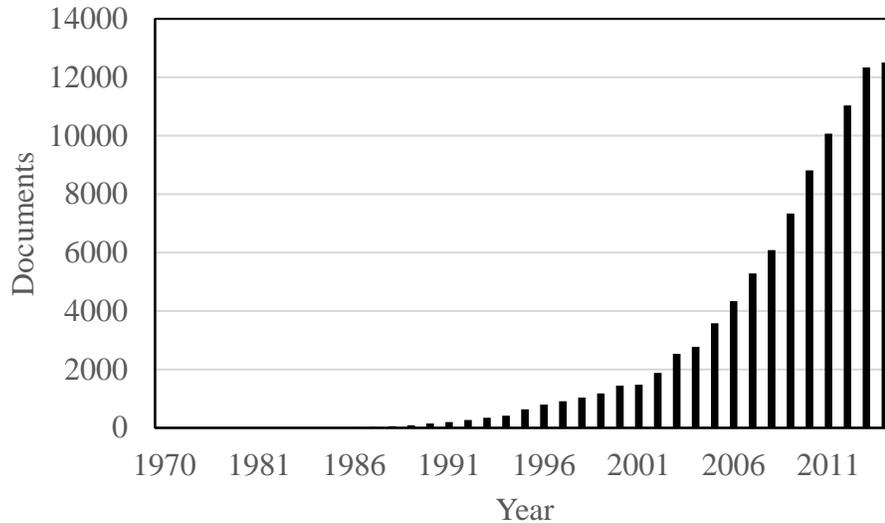


Figure 6 Number of Documents for 'Sustainability' in Scopus

2.3.1 The IPAT Equation

The IPAT equation was first proposed in the 1970s by Paul Ralph Ehrlich and John Paul Holdren, and it was used as a general way to apply concepts of sustainability, measuring the human impact on environment. The equation is defined as:

$$I = P \times A \times T \quad (6)$$

where I is the human impacts on environment, P is the associated population, A is the affluence or level of consumption per person, and T is technology or impact per consumption (Theis and Tomkin, 2011). Although the equation is not mathematically sophisticated, it offers a sound and easy means to analyze sustainability mathematically. As an illustration, with an average water consumption in the US of 100 gallons per capita per day and each gallon of water taking about 4.16 Wh of energy (Derrible, 2015), the energy required to collect, treat and distribute water in a city like Chicago with a population of 2.7 million is:

$$I_{\text{water}} = 2.7 \times 10^6 \times 100 \times 4.16 = 1.12 \times 10^9 \text{ Wh} = 1.12 \text{ GWh per day}$$

2.3.2 *Urban Metabolism*

Urban metabolism (UM) is inspired from the metabolism of organisms where cities alter resources like raw materials, fuel, and water into built environment, human biomass and waste (Decker et al., 2003). According to (Engel-Yan et al., 2007), urban metabolism is defined as “the sum total of the technical and socio-economic process that occur in cities, resulting in growth, production of energy and elimination of waste.” With rapid urbanization and growing concern over climate change, urban metabolism has become a key concept for a sustainable city. The UM framework was first applied by (Wolman, 1965) to a hypothetical US city of one million people. Moreover, the UM of several cities around the world was measured and compared (Kennedy et al., 2011). From a technical perspective, UM is similar material flow analysis (MFA) that has also been applied to cities (Bader and Baccini, 1996; Brunner and Rechberger, 2004), where MFA is defined as a method of measuring flows and stocks of materials in a system. UM therefore applies a MFA type of approach but looking at energy and resources, and different mathematical models have been developed and successfully used to study urban metabolism (Baccini, 1997; Fung and Kennedy, 2005; Hendriks et al., 2000; Kennedy et al., 2015). To

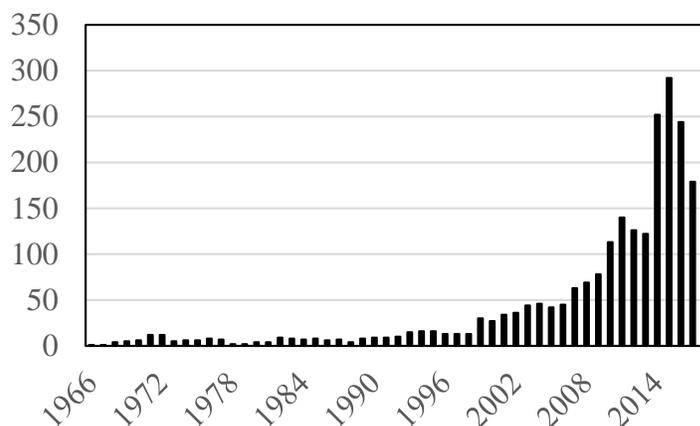


Figure 7 Number of Documents for 'Urban Metabolism' in Scopus

achieve a sustainable city, information about its urban metabolism is vital. An increasing trend in research on urban metabolism is also found while searching for the keyword in Scopus and shown in Figure 7.

2.4 Traditional Statistical Indicators

2.4.1 Mean, Mode and Median

Among seemingly countless statistical metrics, the most widely used one is arguably the arithmetic mean, commonly referred to as mean or average. The term arithmetic mean is used to differentiate it from other means, i.e., the geometric and harmonic means. The arithmetic mean is used in almost all fields including economics, engineering, science, and sociology to name a few. Besides computing the mean of a distribution or data, the arithmetic mean is also used as a common measure of central tendencies, where central tendency refers to a central or typical value of a probability distribution. In fact, the mean is also commonly known as the center or location of the distribution. The mean is widely used in different fields, but it is not a significantly robust metric, as it is highly biased in the presence of outliers. Instead, the *median*

is also used to represent the center or location of a distribution, especially for skewed distributions. Moreover, another widely used metric for traditional analysis is the *mode*, which is

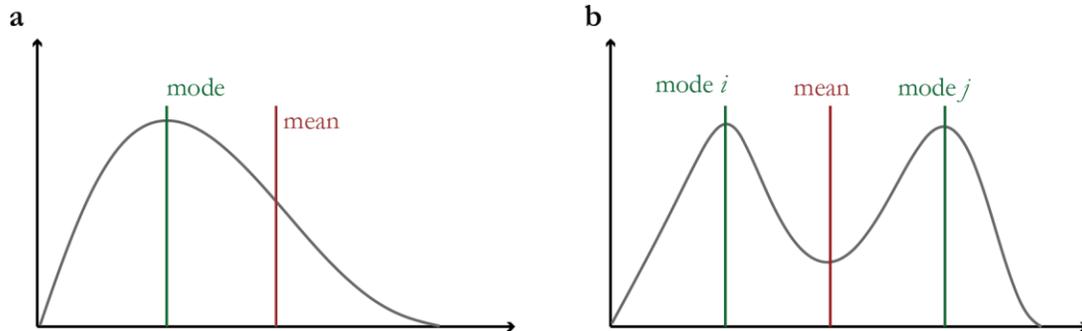


Figure 8 Mean and Mode for a) Uni-modal distributions. b) Multi-modal distributions.

the value with the highest frequency. For a symmetric distribution (e.g., normal distribution), the mean, median and mode are similar, but for skewed distributions (e.g., log normal distribution) they are different. Figure 8 shows the mean and mode for uni-modal and multi-modal distributions.

2.4.2 Outliers

Outliers or extreme observations are common in most datasets. These outliers can easily bias expected results. In practice, outliers are sometimes discarded for simplification and different rules exist to do so. For example, the upper quartile, lower quartile and interquartile range have been used to identify outliers using the following equation

$$\text{Outlier} \begin{cases} \text{if the value is lower than } Q_1 - 1.5IQR \\ \text{if the value is greater than } Q_3 + 1.5IQR \end{cases} \quad (7)$$

where Q_1 is the lower quartile, Q_3 is the upper quartile and IQR is the interquartile range. This method, however, can also discard real information, and it is therefore not robust.

2.4.3 Histogram Analysis

Histograms are commonly used for the graphical representation of a distribution and it was first introduced by Karl Pearson (Pearson, 1895). To plot a histogram, the range is first calculated, which is the difference between maxima and minima that is divided into separate bins. The number of bins depends directly on the bin size and as mentioned earlier in the introduction, selecting a bin size is highly challenging.

Mathematically, histograms involve discrete data. Histograms are also used to define a probability distribution function (PDF):

$$PDF : f(x) = \frac{P(x_i - \frac{dx}{2} < x \leq x_i + \frac{dx}{2})}{dx} \quad (8)$$

where, $P(x_i - dx/2 < x \leq x_i + dx/2)$ represents the probability of variable x to lie in the given range. Analogous to the selection of bin size, the selection of range is also significantly challenging. Finally, the integration of the PDF between 0 to 1 gives the cumulative distribution function (CDF).

2.4.4 L-moments

In statistics, moments are used to define a distribution. For instance, the zeroth moment represents the total mass (e.g., for a probability distribution, the zeroth moment will be the total probability or one), whereas the first moment represents the location of the distribution, and the mean is widely used as the first moment for different distributions. Moreover, the second moment signifies the variance and third moment represents skewness. As mentioned in the introduction, the conventional moments often fail to capture expected outcomes, especially for

small samples. Moreover, the third or higher moment also fail to capture the anticipated results.

Order statistics have been used by (Hosking, 1990) to compute linear moments also known as the

L-moments defined as:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r} \quad (9)$$

where λ_r is the the r^{th} population L-moment for a random variable X and $X_{r-k:r}$ denotes the $r-k^{\text{th}}$ order statistic ($r-k^{\text{th}}$ smallest value) in an independent sample of size n from the distribution of X and E denotes expected value. The first four moments are simply defined as follows

$$\lambda_1 = EX \quad (10)$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) \quad (11)$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \quad (12)$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} + X_{1:4}) \quad (13)$$

Besides defining all the moments, Hosking also defined L-skewness and L-kurtosis which are defined as

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (14)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad (15)$$

where, τ_3 is the L-skewness and τ_4 is the L-kurtosis. Moreover L-scale is defined as the ratio between λ_2 and λ_1 .

2.5 Complexity Science

Although ‘Complexity Science’ or the ‘Science of Complexity’ has become popular in the recent past, it has a long history and can be traced back at least to the works of (Simon, 1965). Despite its rapid emergence, however, there is still no generally accepted definition for complexity. (Chu et al., 2003) defined ‘Complexity Science’ as the interaction of physics, mathematics, biology, economics, engineering and computer science, with a goal to avoid unrealistic models, which are developed based on some simplifications or idealization. This process of simplification was termed as ‘spherical cows’ by Bak, who was one of the pioneers to formulate the general principles of complex system by explaining ‘self-organized criticality’ (Bak, 2013; Bak et al., 1988; Bak and Creutz, 1994). ‘Spherical cows’ is a comical metaphor for oversimplified mathematical model, where a simple sphere represents as cow. A keyword search for ‘Complexity Science’ in Scopus resulted 47,963 documents and the trend is shown in Figure 9.

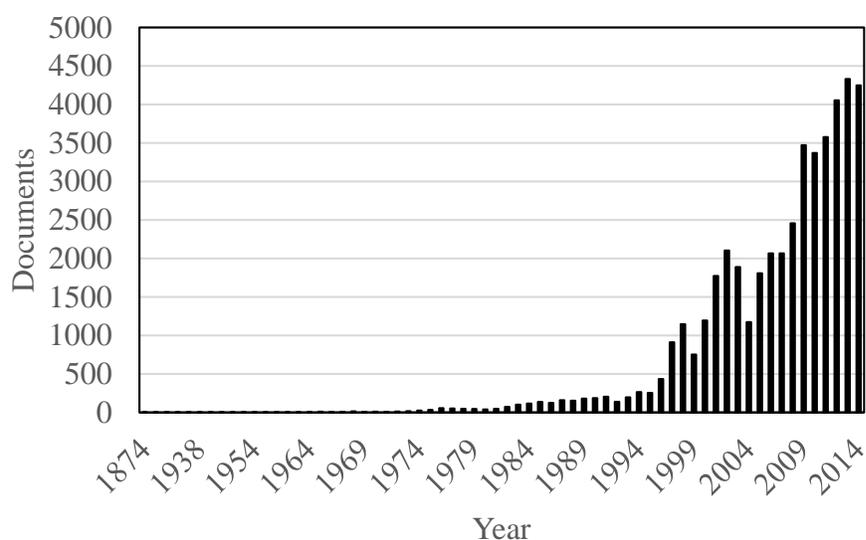


Figure 9 Number of Documents for 'Complexity Science' in Scopus

One of the predominant methods used in complexity science is agent-based modeling (ABM) (Casti, 1997; Epstein, 1999). ABM has been used widely to develop different models, including in John Holland's ECHO model (Holland, 1999, 1995; Hrabar et al., 1997; Smith and Bedau, 2000), and for the simulation of social systems (Epstein and Axtell, 1996) and artificial stock market (Arthur et al., 1996); (Chu et al., 2003). Although Bak was one of the pioneers, his theory was repeatedly criticized (Newman, 1996; Newman and Sneppen, 1996). Later John Holland introduced the idea of complex adaptive system (CAS). He observed that, although many complex systems are superficially different from one another, they share hidden similarities. According to (Holland, 2014):

“Complex systems exhibit several kinds of telltale behavior, which includes, self-organization into patterns, as occurs with flocks of birds or schools of fish chaotic behavior, where small changes in initial conditions ('the flapping of a butterfly's wings in Argentina') produce large changes ('a hurricane in the Caribbean'); 'fat-tailed' behavior, where rare events (e.g., mass extinctions and market crashes) occur much more often than would be predicted by a normal (bell-curve) distribution adaptive interaction, where interacting agents (as in markets or the Prisoner's Dilemma) modify their strategies in diverse ways as experience accumulates.”

2.5.1 Network Science

Network science is an important sub-science of complexity science and it has been widely used to study complex networks like social networks, computer networks, infrastructure networks, and communication networks. Theories from different fields, which include graph theory from mathematics, data mining and information visualization from computer science, and statistical mechanics from physics, have been used to develop and enrich network science. The concept of graph theory, and in turn network science, was first laid by Leonhard Euler in 1736 in his famous paper ‘Seven Bridges of Königsberg’ (Shields, 2012). Since its birth, many scientists around the world have contributed to making it a science of its own. Searching for the keyword ‘Network Science’ in Scopus revealed 139,964 documents (Figure 10), which is more than the number of documents found for the keyword ‘Complexity Science.’

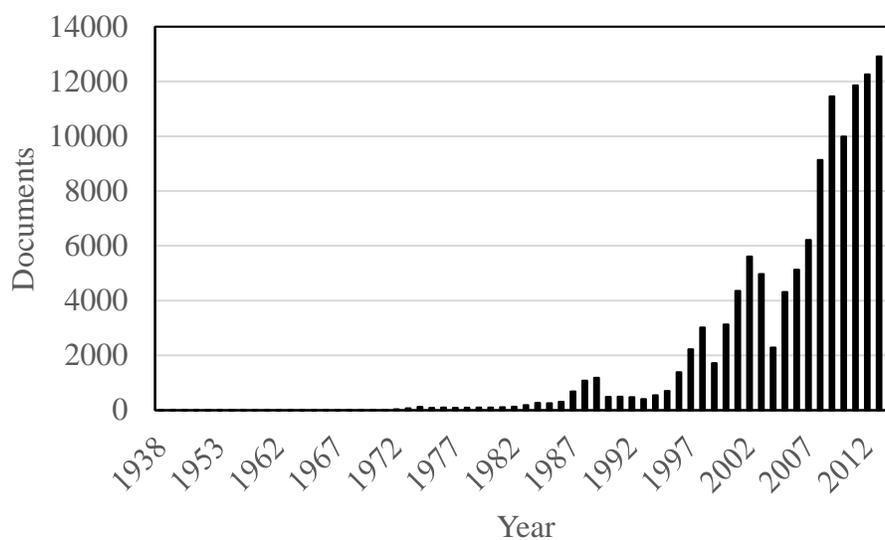


Figure 10 Number of Documents for 'Network Science' in Scopus

Paul Erdős and Alfréd Rényi were the first to apply probabilistic theory in network science by developing the notion of random graphs, better known as Erdős–Rényi model (Erdős and Rényi, 1960, 1959). In Erdős–Rényi (ER) random graph model, any node has a certainty probability to be connected to another node. Later, Duncan Watts and Steven Strogatz identified small-world networks as a class of random graphs and described the small-world problem mathematically in 1998 (Watts and Strogatz, 1998). In a small-world network, most nodes are not directly connected to each other but are accessible through a small number of hops or steps. Watts and Strogatz develop a new graph model, now known as the Watts and Strogatz model. Graphs generated using this model have two main properties: 1) a small average shortest path length and (ii) a large clustering coefficient. From left to right, Figure 11 shows a regular graph, a small-world, and a random network (Watts and Strogatz, 1998).

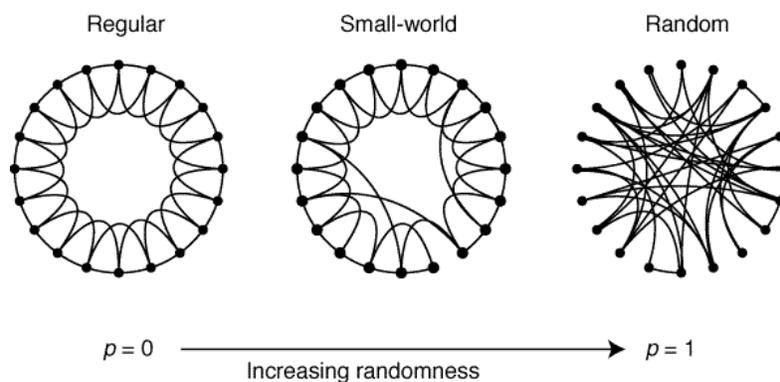


Figure 11 Regular Small-world and Random Network (Source: Watts and Strogatz, 1998)

Moreover, the realm of network science rapidly emerged after the work of Albert-Barabási, who developed the concept of scale free network, where the degree distribution of a scale free network follows a power law. The degree of a node is its number of connections and the probability distribution of these degrees over the network is known as the degree distribution. The basic properties of the Albert-Barabási model: 1) the degree distribution of the developed network follows a power law, and 2) the nodes are connected based on preferential attachment,

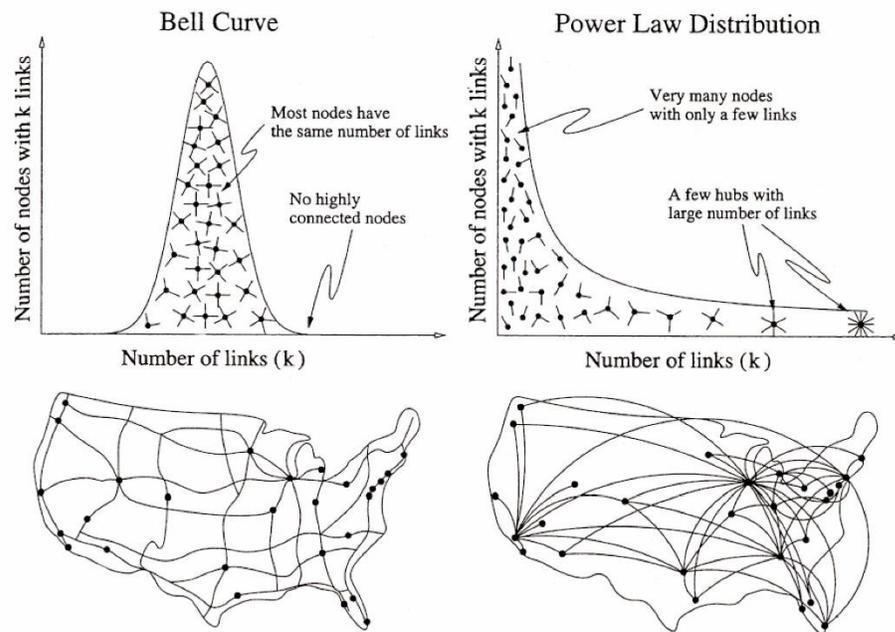


Figure 12 Comparison between Regular and Scale Free Network (Source: Barabási and Albert 1999)

where preferential attachment refers to a process, where new nodes will have a higher propensity to be connected to a node with a high degree (Barabási and Albert, 1999). Put differently, preferential attachment is analogous to terms like “cumulative advantage” and “the rich get richer.” Figure 12 shows the comparison between a “regular” and a scale free network, where the degree distribution of a regular network follows a normal distribution and that of a scale free network follows a power law distribution.

2.6 Network Theory and Measures

2.6.1 Network

A network or a graph G is a collection of vertices/nodes N joined by edges/links L ; $G = \{N, L\}$. Based on the “direction” and number of links, there can be different types of network. For example, if a direction is provided for the links, then the network becomes directed, otherwise it is undirected. Moreover, multiple links can exist between nodes (e.g., multigraph) or nodes can be connected themselves (e.g., self-loop). Furthermore, a network can be weighted, when links of that networks are weighted. In addition, networks are commonly represented in the form of a matrix, which has a binary format (except for multigraphs and weighted graphs), and known as adjacency matrix, where:

$$A_{ij} \begin{cases} 1 & \text{If there is link between } n_i \text{ and } n_j \\ 0 & \text{Oterwise} \end{cases} \quad (16)$$

where, n_i and n_j are two nodes. From an adjacency matrix, several metrics can be computed. Among them, the most common one is the degree of a node, where the degree d_i of a node i is the total number of links of that node (i.e., number of connections) and is defined as:

$$d_i = \sum_{j=1}^N A_{ij} \quad (17)$$

where, N is the total number of nodes. The study of the distribution of degrees is common in network analysis, i.e., how many nodes have a degree 1, 2, 3, and so on. Different degree distributions are observed for different networks. For instance, the degree distribution of the networks generated by the Erdős–Rényi random graph model will have a normal distribution (Erdős and Rényi, 1960, 1959), whereas networks from Albert-Barabási model follow a power

law (Barabási and Albert, 1999). Analyzing the distribution of a property in networks can reveal noteworthy insights about their properties.

2.6.2 Distance in Networks

Another important feature of networks relates to distance. In graph theory, distance refers to geodesic distance between two nodes, which is the number of edges required to connect those two nodes in a shortest path. Accordingly, the average shortest path length is defined as the average of the shortest path lengths of that network. On the other hand, the largest of all these shortest path lengths is defined as the diameter of the network. For this work, to relate these measures with conventional statistical measures, the average shortest path length can provide an idea about the “spread” of the network, while diameter can be used as the “range” of that network.

2.6.3 Connectivity

An important measure in network theory is related to the concept of connectivity. Among different measures of connectivity, *density* (ρ) calculates the number of links in a network divided by the total number of potential links in this network, which can be calculated as $\frac{1}{2} N(N-1)$ for an undirected network. Density therefore takes the form:

$$\rho = \frac{2L}{N(N-1)} \quad (18)$$

where L is the number of links and N is the number of nodes.

2.6.4 Giant Cluster

Another common feature in any network is the existence of sub-network. Although all the nodes in a network can be connected to one another, it is not necessary. Hence, there are two kinds of networks, i.e., connected and disconnected, which are depicted in Figure 13.

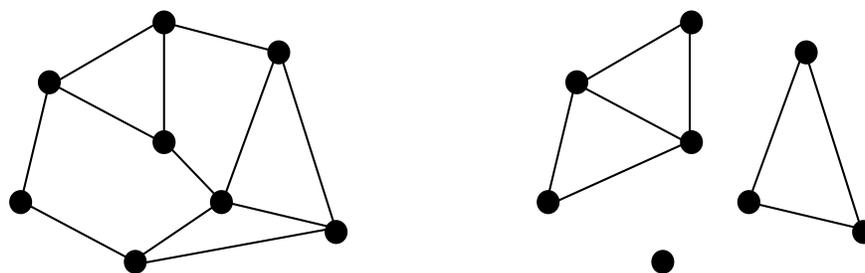


Figure 13 Connected versus Disconnected Network A) Connected B) Disconnected

In most disconnected networks, as they grow, i.e., as they accumulate more nodes, one sub-network tends to absorb most of these new nodes. This particular sub-network refers to as *giant cluster*. Different properties of a network including the ones mentioned above can be found in (Newman, 2010)).

2.6.5 Social Network

Social networks are typically considered to be complex networks, where each actor of a society is a node and they are connected to others via their interactions. Instead of analyzing individual actors and their activities, the analysis of social networks focuses on all the nodes (Wasserman and Galaskiewicz, 1994). Being significantly complex in nature, social networks possesses interesting properties, and theories like ‘Six Degrees of Separation’ has been used to

explain this complex nature (Barabási, 2002). An important concept of social network is homophily, which was explained earlier as the propensity of people to connect with people that shares similar characteristics (e.g., income, religion).

2.7 Information Theory

Information theory deals with the quantification of information. Rooted in statistics, information theory resides between computer science, mathematics, physics and engineering, and it has been widely applied from cryptology to ecosystem dynamics (Frieden and Gatenby, 2007). Claude Shannon is regarded as the pioneer in information theory, as he developed the concept of information theory to process signal, compress data and store data (Shannon, 2001). Since its inception, it has been applied to statistical inference, natural language processing, and cryptography to name a few. Entropy is a key measure of information, and it was notably applied by Theil (Theil, 1967) to measure inequality in a distribution. Nonetheless, it originates from statistical mechanics developed by Boltzmann (Boltzmann, 1872) and popularized in information theory by Shannon (Shannon, 2001). Mathematically it is defined as:

$$H(x) = - \sum p(x_i) \log (p(x_i)) \quad (19)$$

where $H(x)$ is the entropy, and $p(x_i)$ is the probability of event x_i , Entropy is bounded by 0 and ∞ .

2.7.1 Fisher Information

Fisher information (FI) was first developed by statistician R.A Fisher (1922), and it offers a measure of indeterminacy. In other words, FI can measure the amount of information of an unknown parameter that is present in observable data. Mathematically, Fisher Information (FI), $I(\theta)$, is defined as (Karunanithi et al., 2008):

$$I(\theta) = \int \frac{dX}{p_0(X|\theta)} \left[\frac{\partial p_0(X|\theta)}{\partial \theta} \right]^2 \quad (20)$$

where $P_0(X|\theta)$ is the probability density of obtaining a particular value of X in the presence of θ . Fisher information can be also used as a measure of dynamic order. As it is able to analyze dynamic regime, it can be used to assess the stability of any system, measuring whether the system is stable or losing its functionality or even moving towards a regime shift. FI has been used in a variety of applications from deriving fundamental laws of thermodynamics (Frieden, 1998) to assessing dynamic order in real and model systems (Brian D. Fath et al., 2003; Fath and Cabezas, 2004; Karunanithi et al., 2008; A.L. Mayer et al., 2006; Pawlowski et al., 2005) to sustainable environmental management and resilience (Cabezas and Fath, 2002; Eason and Cabezas, 2012; Eason and Garmestani, 2012; González-Mejía et al., 2014; Ingwersen et al., 2014; Mayer et al., 2007; Shastri et al., 2008; US EPA, 2010). However, there has been limited use in the engineering arena. As FI is a central measure used in this dissertation, the methodology will be detailed later in Chapter 5.

3. Tracking Resource Consumption Using Traditional Statistics

3.1 Introduction

Traditional statistics offer many benefits and are often enough to fully investigate a data set. It is also considerably helpful for preliminary exploration more substantial data, but insights may be misleading as we will see. From the collected data on human, produced and natural capital, we can first calculate several traditional statistical measures. The data contains yearly information for all the countries in the world from 1990 to 2014.

For this study, the arithmetic mean, the median and the mode are computed first. Moreover, concepts of order statistics are implemented to compute L-moments. Additionally, L-skewness and L-kurtosis, as defined by (Hosking, 1990), are also calculated.

3.2 Methodology

The arithmetic mean is calculated by taking the mean of all values given at the country level. For the median, the values in the data are sorted in ascending order and then the midpoint value is taken as the median of the data. To compute the mode, I first performed a traditional histogram analysis. As mentioned earlier, histogram analyses depend on bin size. In this case, Scott's rule was applied to determine the bin size, which is equal to $3.49 \sigma / n^{1/3}$ with a standard deviation of σ and sample size of n . The mode is then calculated by taking the midpoint of the bar with the highest frequency from the histogram. For example, Figure 14 illustrates the histogram analysis for human capital in 2000, where the mode is identified between 0 and \$427169.72 and the mode is therefore selected as \$213584.86, i.e., the midpoint of the range. In the same year, the mean is \$171175.76, and the median is \$37453.35

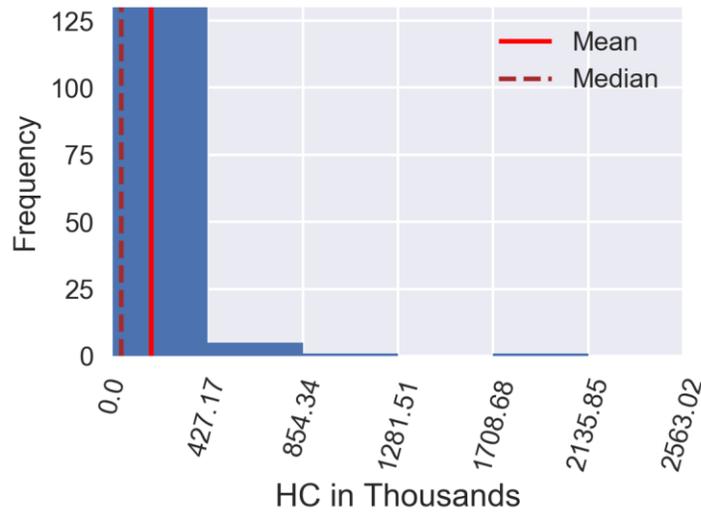


Figure 14 Histogram Analysis for Human Capital in 2000

After computing the traditional statistical metrics, the L-moments, L-skewness and L-kurtosis were computed based on the definitions provided in chapter 2. The first four moments are calculated using equations 21-24 (Wang, 1996) defined as:

$$\lambda_1 = \frac{1}{n} \sum_{i=1}^n x_i \quad (21)$$

$$\lambda_2 = \frac{1}{2} \frac{1}{n} \sum_{i=1}^n ({}^{i-1}C_1 - {}^{n-1}C_1) x_i \quad (22)$$

$$\lambda_3 = \frac{1}{3} \frac{1}{n} \sum_{i=1}^n ({}^{i-1}C_2 - 2{}^{i-1}C_1 {}^{n-1}C_1 + {}^{n-1}C_2) x_i \quad (23)$$

$$\lambda_4 = \frac{1}{4} \frac{1}{n} \sum_{i=1}^n ({}^{i-1}C_3 - 3{}^{i-1}C_2 {}^{n-1}C_1 + 3{}^{i-1}C_1 {}^{n-1}C_2 - {}^{n-1}C_3) x_i \quad (24)$$

where x_i is the sample value from any distribution $\{x_1, x_2, \dots, x_n\}$, where values ranked in ascending order. ${}^m C_k$ is simply the combinatorial expression:

$${}^m C_k = \frac{m!}{k!(m-k)!} \quad (25)$$

where m is the total number of items, and k is the number of combinations of k -distinct elements; and $k < m$ and ${}^m C_k$ is zero when $k \geq m$. Results from equation 21-24 were used to compute L-

skewness, which is the ratio between λ_3 and λ_2 . Moreover, L-kurtosis was also computed by taking the ratio between λ_4 and λ_2 .

3.3 Results

The arithmetic mean, mode, and median are computed for human, produced, and natural capital for each year and the results are shown in Figure 15. Moreover, Figure 16 shows the comparison between 2nd L-moment and conventional 2nd moment (standard deviation) for all the indicators mentioned above, whereas Figure 17 depicts the change in L-skewness and L-kurtosis.

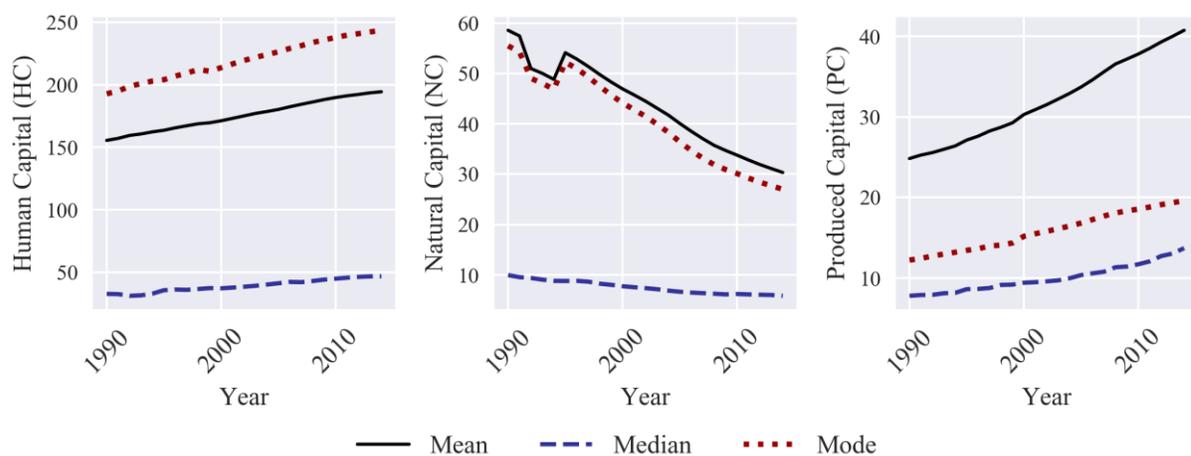


Figure 15 Change in Arithmetic Mean, Median and Mode of Human, Natural, and Produced Capital from 1990 to 2014

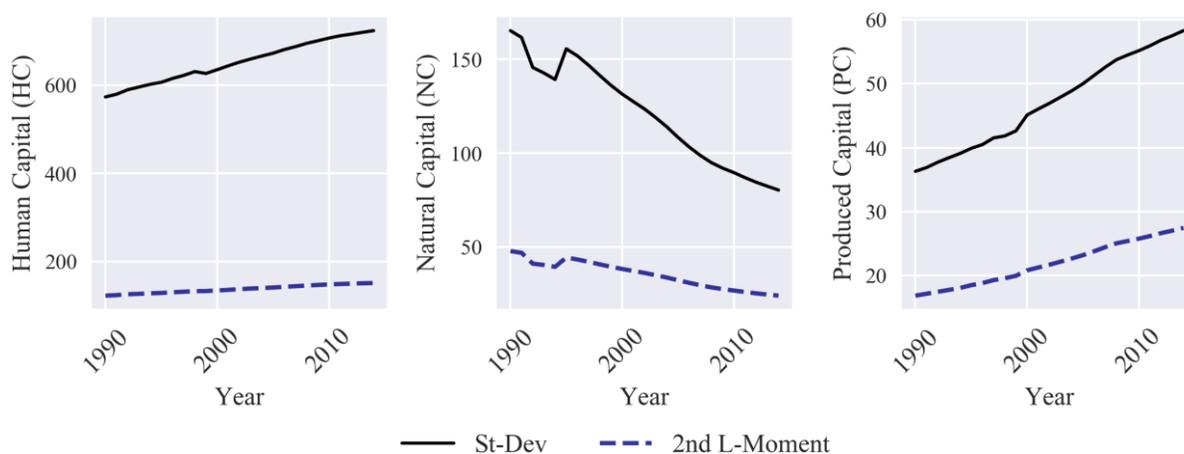


Figure 16 Comparison between 2nd L-moment and Standard Deviation of Human, Natural, and Produced Capital from 1990 to 2014

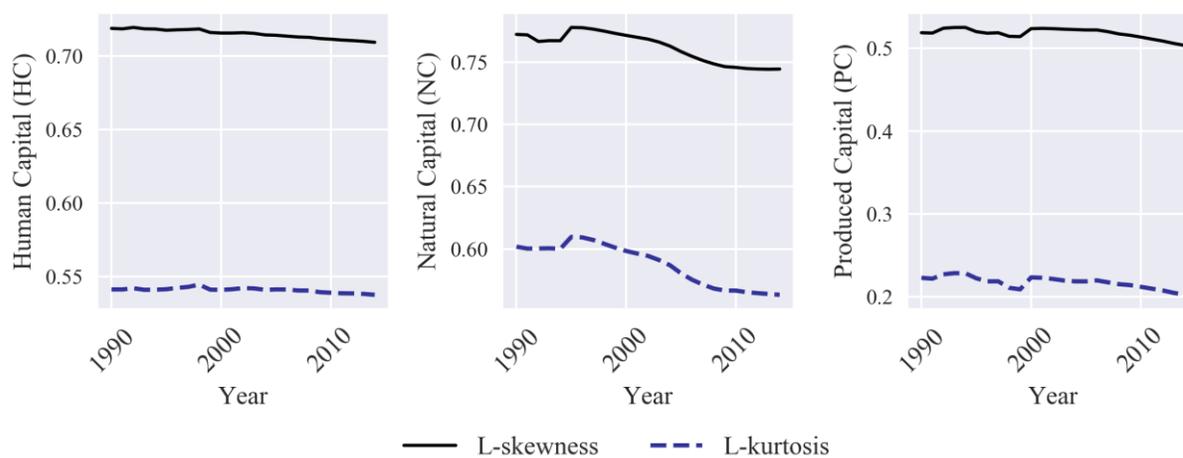


Figure 17 Change in L-skewness and L-kurtosis of Human, Natural, and Produced Capital from 1990 to 2014

3.4 Discussion

From Figure 15, we can see that the arithmetic mean, the median and the mode follow overall similar trends for all the indicators (i.e., human, natural, and produced capital). In particular, 139 countries for HC, 120 countries for produced capital, and only 13 countries for

NC have higher capital value in 2014 than in 1990. Thus, except for natural capital, we observe a gradual increasing trend for both human and produced capital. Moreover, from Figure 15 we can observe that for all three capitals the mode was always greater than the median indicating that the distribution for HC, NC, and PC are skewed towards left from 1990 to 2014. Moreover, as all of the data used for this analysis contain per capita information, the arithmetic mean computed here is thus a mean of the mean. Therefore, the outcomes carry mean-of-the-mean-type errors.

Analogous trends for the mode and the mean compared to the median were found for all three capitals, where the mode follows the trend of the mean but the median remained stable from 1990 to 2010.

Furthermore, we can compare the standard deviation with the 2nd L-moment illustrated in Figure 16. From the figure we can observe that both curves follow similar pattern and the 2nd L-moment is lower than the standard deviation. Additionally, from Figure 17, we note that the L-skewness remains positive from 1990 to 2010, indicating a positively skewed or skewed-towards-the-left distribution, which corroborates with the earlier interpretations from Figure 15.

From these results, we can infer that they provide useful and valuable information, but as all these traditional indicators follow similar trends like the arithmetic mean, thus, we are unable to find anything new about the data. Therefore, we use network science to develop a novel method to supplement traditional statistics and extract more information from any data, which will be discussed in details in the next chapter.

4. Tracking Sustainable Development in World Countries Using Network Science^{1,2}

4.1 Introduction

The world has changed dramatically since the end of the 20th century, and the pace of change shows no sign of slowing down. The global societal aspiration for development seems to systematically lead to the consumption of more resources, which not only puts additional pressure on the world's natural resources but it also jeopardizes the environment and plays a key role in altering the climate. According to several studies (Meadows et al., 1972; Rockström et al., 2009; Wackernagel et al., 2002), the depletion of resources, climate change, and the degradation of the environment bear clear signs of unsustainable development. Moreover, studies have also shown that standards of living can be on the rise even as stresses on the environment are increasing (DeFries, 2014; Johnson, 2000; Raudsepp-Hearne et al., 2010). Therefore, tracking progress (or lack of it) in sustainable development is critical (Bettencourt and Kaur, 2011; Clark and Dickson, 2003; Kates, 2001; Levin and Clark, 2010; Parris and Kates, 2003). The Gross Domestic Product (GDP), Human Development Index (HDI), and ecological footprint have long been used to track human development (Rees, 1992; UNDP, 2016; Wackernagel and Rees, 1996), but many have shown that these indices can fail to capture whether a development is sustainable or not (Dasgupta et al., 2015; Polasky et al., 2015). By contrast, the Inclusive Wealth (IW) index, also called Comprehensive Wealth or Genuine Wealth, was designed purposefully to track sustainable development (Arrow et al., 2004; UNU-IHDP and UNEP, 2014; World Bank,

¹ Part of this chapter is published in PLoS ONE. Citation: Derrible, Ahmad (2015). Network-Based and Binless Frequency Analyses. PLoS ONE, 10(11), pp. e0142108.

² Part of this chapter is published in the Journal of Environmental Management. Citation: Ahmad, Derrible, Managi (2018). A network-based frequency analysis of Inclusive Wealth to track sustainable development in world countries. Journal of Environmental Management, 218, pp. 348–354.

2010). IW is defined as the sum of three capitals (defined in monetary terms) representing the triple bottom line of sustainability: human (i.e., society), produced (i.e., economy), and natural capital (i.e., environment); see supplementary information for details on what variables are included in the three capitals.

The main goal of this study is to capture and evaluate the main sustainable development trends and trajectories of world countries relative to every other using IW data from 1990 to 2014 (UNU-IHDP and UNEP, 2014; Urban Institute and UNEP, 2017). Traditionally, the arithmetic mean or median are used to capture the general trend, but both can be easily biased by the presence of extreme values in a dataset. For this study, we prefer to adopt a Network-based Frequency Analysis (NFA) approach (Derrible and Ahmad, 2015), which is not affected by extreme values and which is able to capture general trends in a more robust way. Moreover, NFA allows us to measure the evolution of an entity relative to the evolution of other entities. More specifically, the objectives of this study are to: a) create a method to estimate the main human, natural, and produced capital trends in the world using concepts from network science; b) assess the relative performance of every country from the general trends captured; c) cluster countries based on their performance in all the three capitals; and d) project current trends up to 2050 to depict the potential future performance of every country. For this work, we use data from the 2014 and 2017 Inclusive Wealth Report (UNU-IHDP and UNEP, 2014; Urban Institute and UNEP, 2017). We selected human, natural, and produced capital for all the countries for our analysis.

One of the main contributions of this work is the deployment of the NFA method, whose results can offer a significant contribution to current knowledge and complement traditional statistical approaches. Briefly, among 140 analyzed countries, 110 showed an increasing trend

for human capital, whereas only 4 countries for natural capital and 6 countries for produced capital showed an increasing trend from 2000 to 2014. Moreover, based on the combined performance of human, natural, and produced capital, only 3 countries show an optimistic trend, whereas 18 countries show a decreasing trend, and the 119 remaining countries have not changed significantly from 2000 to 2014.

4.2 Methodology

4.2.1 Network Formation

The main goal of the methodology is to track general trends in a dataset and to track the performance of individual entities within the dataset. Traditional frequency analyses techniques can fail to capture real trends in a dataset (Derrible and Ahmad, 2015). To overcome these problems, inspired from concepts of *homophily* in network science, we analyze multivariate temporal data by transforming a dataset into a network. The main concept behind it is to compare values of a distribution with one another by measuring whether they are similar or not (i.e., whether they lie within a certain range of one another). In other words, while traditional histograms have fixed ranges, this methodology adapts the ranges around individual values. Although not rooted in the general field of Network Science (Newman, 2010), the method essentially compares to forming a network. From a Machine Learning perspective, the foundation of the methodology is closest to kernel density estimation (KDE) (Parzen, 1962; Rosenblatt, 1956), where the critical challenge is to select the optimum bandwidth. The method works as follows. First, we convert each dataset to a network by comparing the performance of all countries with every other. Formally, we connect two nodes together when their values are within a certain range, ζ , of each other. A network is analytically represented by an adjacency

matrix, A_{ij} , where the cells take a value of 1 if nodes i and j are connected and 0 otherwise. For this study, adjacency matrices are defined as:

$$A_{il} = \begin{cases} 1 & \text{if } (x_i - \zeta) \leq x_j \leq (x_i + \zeta) \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

where, x_i represents the value of node I , x_j represents the value of node j , and ζ is called the cutoff percentage. For instance, x_i could be the human capital per capita of the USA, and x_j could be the human capital per capita of China.

The degree d_i of a node i is the total number of links of that node (i.e., number of connections) and is defined as:

$$d_i = \sum_{j=1}^N A_{ij} \quad (27)$$

Importantly for this work, the value of the node with the highest degree (i.e., value with the most connections) becomes the Network-based (N) mode M of the distribution, since it is the value that is the most ‘similar’ to other values, and therefore most representative of an entire dataset:

$$M = V(\max(d_i)) \quad (28)$$

The N mode acts as the most representative value for the network, especially for a uni-modal distribution. Other modes can be determined by plotting the degree against value, for a multi-modal distribution. Furthermore, since a network is generated here, a plethora of metrics can be computed. Of relevance, the average shortest-path length, which is defined as the average of all shortest-path lengths, L_{avg} , between every pair of nodes, becomes a measure of spread of a distribution akin to the standard deviation. The diameter of the network, D , which is the largest of all shortest-path lengths, gives us a different measure of range.

To illustrate the method, Figure 18 shows a random sampling of 10 values from the normal distribution with a mean of 5 and standard deviation of 2, where the 10 values are {6.90,

2.83, 3.49, 4.85, 5.33, 5.65, 3.69, 2.58, 5.74, 7.28}. The left-hand side applies a traditional binning process, where we chose a bin width of 1 and the right-hand side applies the proposed method with $\zeta = 0.5$, therefore representing a range of 1 as well. We can see that in this particular case, the network-based approach better captures the properties of the simulated distribution than the traditional histogram. Moreover, the traditional histogram places the mode of the distribution in a range 5 to 6, as opposed to a more desirable discrete value, compared to the network based methodology which gives 5.33 as the mode.

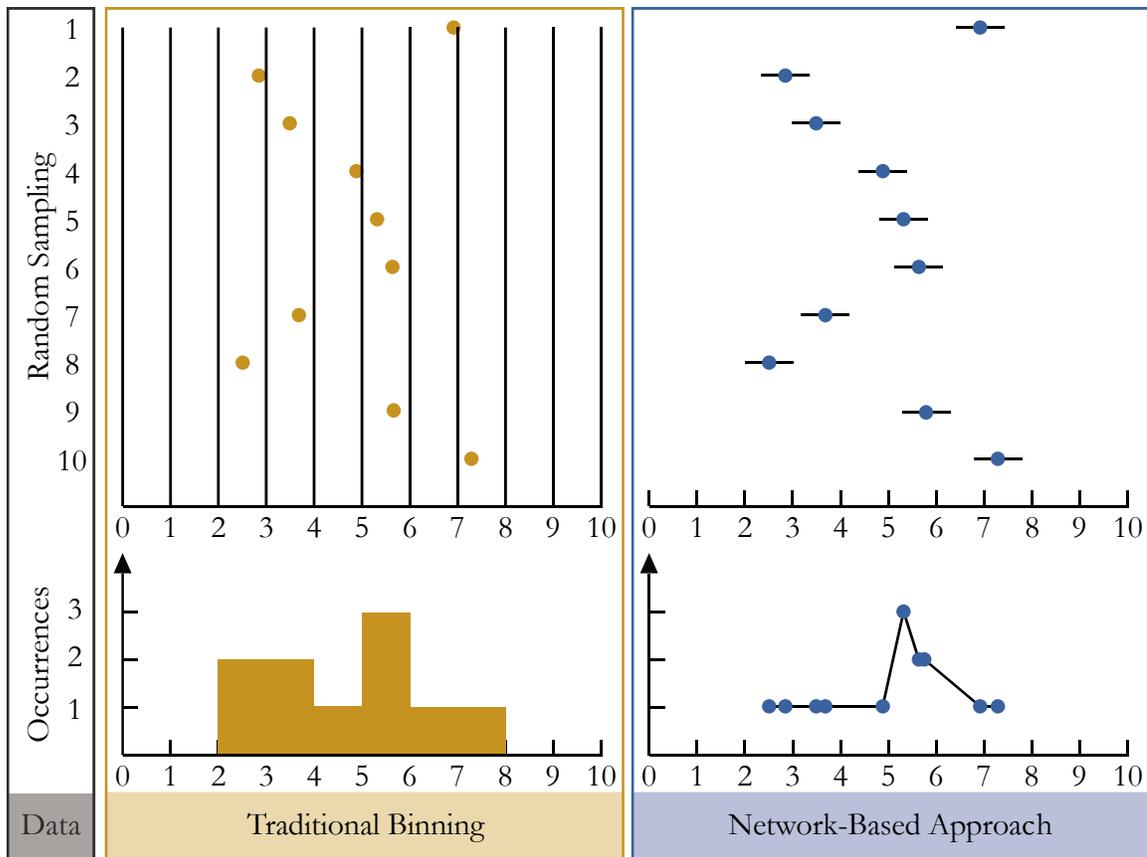


Figure 18 Illustration of the Network-based Frequency Analysis (NFA) method.

From equation 26, the selection of ζ can have significant impacts on the results. A ζ too small may omit important connections, while a ζ too large artificially inflates the main mode of the distribution. The main objective is therefore to select a ζ for which the network properties

have become stable. Because we are not fitting a distribution to the values, we cannot seek a method to minimize the mean squared error as is usually the case in statistics. Instead, we propose to look at the evolution of the number of nodes in the largest cluster. As we slowly increase the value of ζ , clusters (akin to communities) form and a larger / giant cluster rapidly emerges containing V_g nodes. We define the proportional number of vertices in the giant cluster as $p_g = V_g / V$. After a certain value of ζ , the giant cluster admits few or no new nodes, and p_g does not change, which suggests it has reached stability. Moreover as an additional condition, we ensure that the giant cluster contains at least 60% vertices or the p_g is greater than or equal to 0.60 for this analysis. As an example, Figure 19 e shows the increase of p_g with ζ for a 100 random normally distributed, $N(100,20)$, values.

Figure 19 a-c shows scatter plots of individual values vs. their degrees for three different ζ . We can see that a small ζ looks more like noise and it takes the shape of a normal distribution as we increase ζ , after which the entire distribution starts to inflate. In this particular case, p_g becomes stable at the point for $\zeta = 10$ (Fig. 19 e), where all nodes are part of the giant cluster, i.e., $p_g = 1$, which suggests there are no outliers in the distribution, as expected. Stability in p_g remains a user-defined notion at this point. For this work, we say that the network is optimal when p_g remains identical for several increases of ζ ; the exact number of increases depends on the mag magnitude of each increase set. For large increases, fewer consecutive increases of ζ need to be identical.

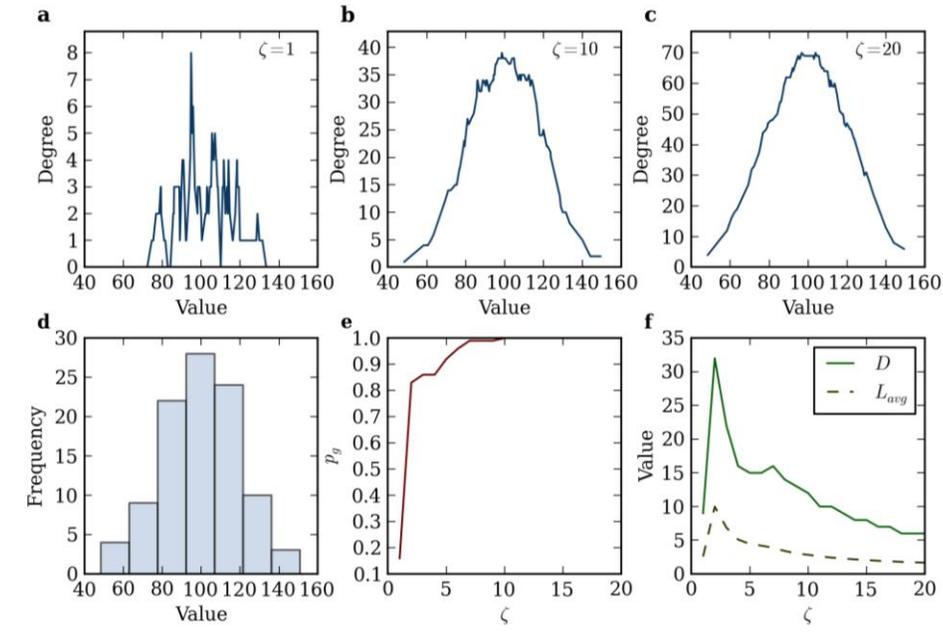


Figure 19 Impact of ζ on NB Methodology and Network Properties.

One means to further validate visually ζ is to analyze the evolution of the diameter D and average path length L_{avg} of the network created. Figure 19 f shows D and L_{avg} for the normal distribution generated. We can see that for significantly small values of ζ , both D and L_{avg} tend to be small, since few nodes are connected. By then increasing ζ , D and L_{avg} also increase up to a maximum, connecting more nodes that had no previous connections, after which they tend to decrease, connecting nodes that were already connected. We find that the constructed networks tend to reach a stable point at or shortly after D and L_{avg} have reached a unique peak for unimodal distributions, and at or shortly after the last of multiple peaks of D and L_{avg} for multimodal distributions.

The selected value ζ_s carries much information. In general, a higher value suggests the distribution has a large spread and is therefore more “heterogeneous”, and a low value suggests the distribution is more “homogeneous”. We can therefore define a homogeneity information

index by dividing ζ_s by the mode M : $H = \zeta_s/M$. For instance, in Figure 19, the network becomes stable for $\zeta = 10$, the highest degree is 39 for the node with value 98.5387, and therefore the homogeneity index of the distribution is $M = 10/98.5387 = 0.1015$. Although H is bounded by $[0, \infty]$, it is most often within the range $[0,1]$, since for values larger than 1, $\zeta_s > M$, and all values below M are connected.

4.2.2. Orbital Position and Orbital Speed

Then, using the optimal network measured, a technique is developed to track the discrete relative position of that node compared to the mode of the distribution (e.g., how far or close that node is from the mode). The concept of distance from network science is used to do so.

Moreover, we can account for direction in the network, whether one step away is above or below the mode. To avoid negative values, the mode can then be set at any positive constant (e.g., 100). Here in this network-based methodology a single step is equal to the cutoff ζ , therefore the discrete relative position of a node i from the mode is defined as orbital position $O_{i,t}$ and expressed mathematically as:

$$O_{i,t} = \frac{x_{i,t} - M_t}{\zeta_t} + C \quad (29)$$

where, $O_{i,t}$ is the orbital position for node i in year t , $x_{i,t}$ is the corresponding value for node i in year t , M_t is the mode in year t , ζ_t is the cutoff in year t and C is any positive constant (in this work, for convenience, we place the mode at orbit 100, and therefore $C = 100$). Moreover, in line with the notion of orbits, O_i necessarily has to be an integer. This particular equation offers one of the pivotal contributions of this thesis. The performance of a country is therefore assessed based on its distance from the mode as opposed to being based on a fixed value. In concrete terms, it is conceptually close to tracking the evolution of human capital in real terms in contrast

to measuring it in nominal terms. Despite an increase in consumption, an orbital position can still decrease if the general trend (i.e., the mode) witnessed a larger increase.

The term orbital position refers to the location of electrons around a nucleus. Electrons orbit around the nucleus and absorb or emit energy as they change orbits. Similarly, a node (i.e., country) can also change its orbital position over time. For example, if a country is able to increase its natural capital, then that country will move outwards, from an inner orbit to an outer orbit or vice versa based on natural capital. Orbital positions of every node are plotted over time in polar format to generate the orbital diagram for each country, where the timeline proceeds in the counter clockwise direction. Figure 20 shows the orbital diagram of the USA for human capital from 1990 to 2010. For better visualization, orbital positions over 300 are not shown in this dissertation. In the figure, the first dotted ring closest to the center represents orbit 100, i.e., the mode.

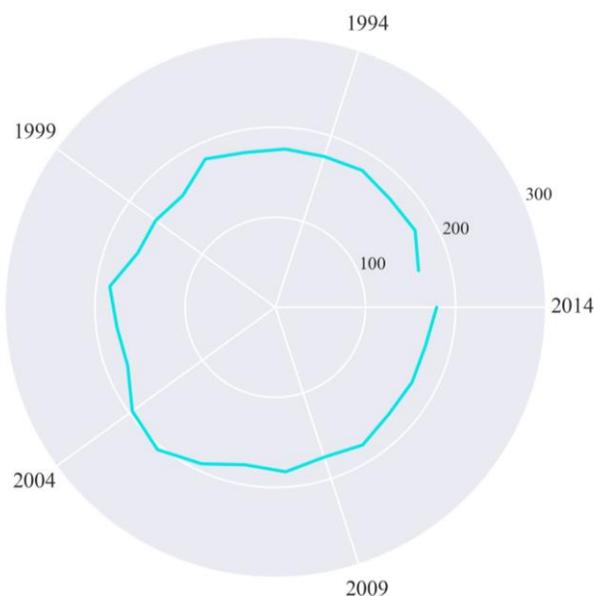


Figure 20 Orbital Diagram of USA for Human Capital

Moreover, the change in orbital positions from the previous year is also computed by taking the difference between orbital positions in two consecutive years. We can then define a

total ‘orbital distance,’ D_i , for each country by taking the summation of those differences in orbital positions for the entire time period. For a timeline from 1 to T , D_i is expressed as:

$$D_i = \sum_2^T O_{i,t} - O_{i,t-1} \quad (30)$$

where, $O_{i,t}$ is the orbital position for node i in year t and where, $O_{i,t-1}$ is the orbital position for node i in year $t-1$.

The speed at which the countries’ orbital positions are changing is calculated by dividing D_i from equation 6 of any country by its number of available data points and is expressed as:

$$O'_i = \frac{D_i}{T_i} = \frac{\sum_2^T O_{i,t} - O_{i,t-1}}{T_i} \quad (31)$$

where, O'_i is the orbital speed for node i and T_i is the number of available data points within the time line of 1 (i.e., 1990) to T (i.e., 2010) for node i . For example, for human capital, the orbital distance for Belgium is 18 and since this distance travelled over a period of 14 years (from 2000 to 2014), the orbital speed of Belgium for human capital is $18/14 = 1.29$.

4.2.3. Clustering

We selected human, natural, and produced capital for all the countries for our analysis. Moreover, as we observe the presence of a giant cluster (i.e., sub-network containing most countries) for each dataset, where all countries (i.e., nodes) inside the giant cluster are directly or indirectly connected to one another. Therefore, for each capital and each year, a country can be either a part of the giant cluster or remain outside of the giant cluster, thus yielding two distinct groups for each dataset. As we have three datasets (i.e., human, natural and produced) for each year, we have $2^3 = 8$ possible combinations per year, and therefore we can group all countries

into 8 different clusters. To find the combination for each country, we assign a value of ‘2’ if a country is in the giant cluster and a value of ‘1’ if it is not. Moreover, we denote those eight groups from 1 to 8. Table SI 1.2 shows all the combinations and group name for the clustering.

From Table 3, we can observe that countries in group 8 have a combination of ‘111’, meaning all of their capital values (i.e., human, natural, and produced) are significantly higher from the rest of the world and thus, located outside the giant cluster. Moreover, countries in groups 7,6, and 5 have two capital values outside the giant cluster and countries in groups 4,3, and 1 have only one capital values outside the giant cluster. Finally, countries in group 1 have all their capital values within the giant cluster.

Table 3 Combinations and Groups for Clustering

Human	Natural	Produced	Combination	Group	Example for 1990
1	1	1	111	8	USA
1	1	2	112	7	Sudan
1	2	1	121	6	France
2	1	1	211	5	Qatar
1	2	2	122	4	South Korea
2	1	2	212	3	Brazil
2	2	1	221	2	Bahrain
2	2	2	222	1	China

Subsequently, using orbital speeds and assuming that on average future trends will follow current ones, we can project the potential future orbital position of a country:

$$O_{i,t} = O_{i,t_0} + O'_i \times (t - t_0) \quad (32)$$

where $O_{i,t}$ is the orbital position for the country i in the year t and O'_i is the orbital speed for the country i . The year 2014 is used as t_0 . Then, we find the maximum orbital position of the country within the giant cluster for HC, NC, and PC from 1990 to 2014 and use those values as cutoffs to decide whether a country will be in the giant cluster or not in the future.

4.3 Results and Discussion

Figure 21 presents three ways of representing the main trends in Human Capital (HC), Natural Capital (NC), and Produced Capital (PC) of all the countries in the world. The two standard representations are the arithmetic mean and median. To these two, we add the N mode that captures the main trend from the Network-based Frequency Analysis (NFA); essentially, the N mode arguably better represents the general trend since it is not skewed by asymmetrical distributions unlike the mean and the median.

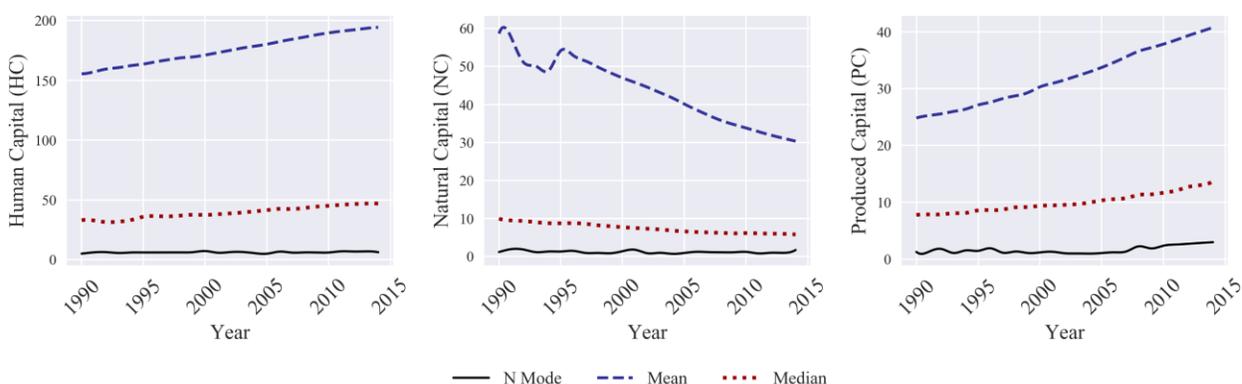


Figure 21 Evolution of the Network-based (N) mode, mean, and median for per-capita human capital, natural capital, and produced capital from 1990 to 2014. (All indicators are reported in constant 2005 U.S. dollars in thousands.)

An increasing global trend is detected in both the mean and the median values for HC and PC, while NC is decreasing (Figure 21). In contrast, the N mode stayed relatively stable for HC and NC, and it has been showing signs of increase for PC since 2009. Moreover, we find on average that 78% of HC, 83% of NC, and 72% of PC are below the corresponding mean, suggesting a large right-tail in the frequency distribution of the three capitals. Furthermore, we observe that among the 140 analyzed countries, 139 countries for HC, 13 countries for NC, and 120 countries for PC have higher capitals in the year 2014 than in the year 1990. Overall, when

summing the three variables to compute the IW, only 90 countries have higher values in 2014 compared to 1990, mainly because 127 countries see a significant decrease in NC.

Figure 22 shows the orbital diagrams and orbital speeds for the United States (USA), France, Brazil, and China for human, natural, and produced capital from 1990 to 2014. Orbital diagrams and orbital speeds for all countries are provided in the Appendix A (Figure S1, S2, S3, S4, and S5)

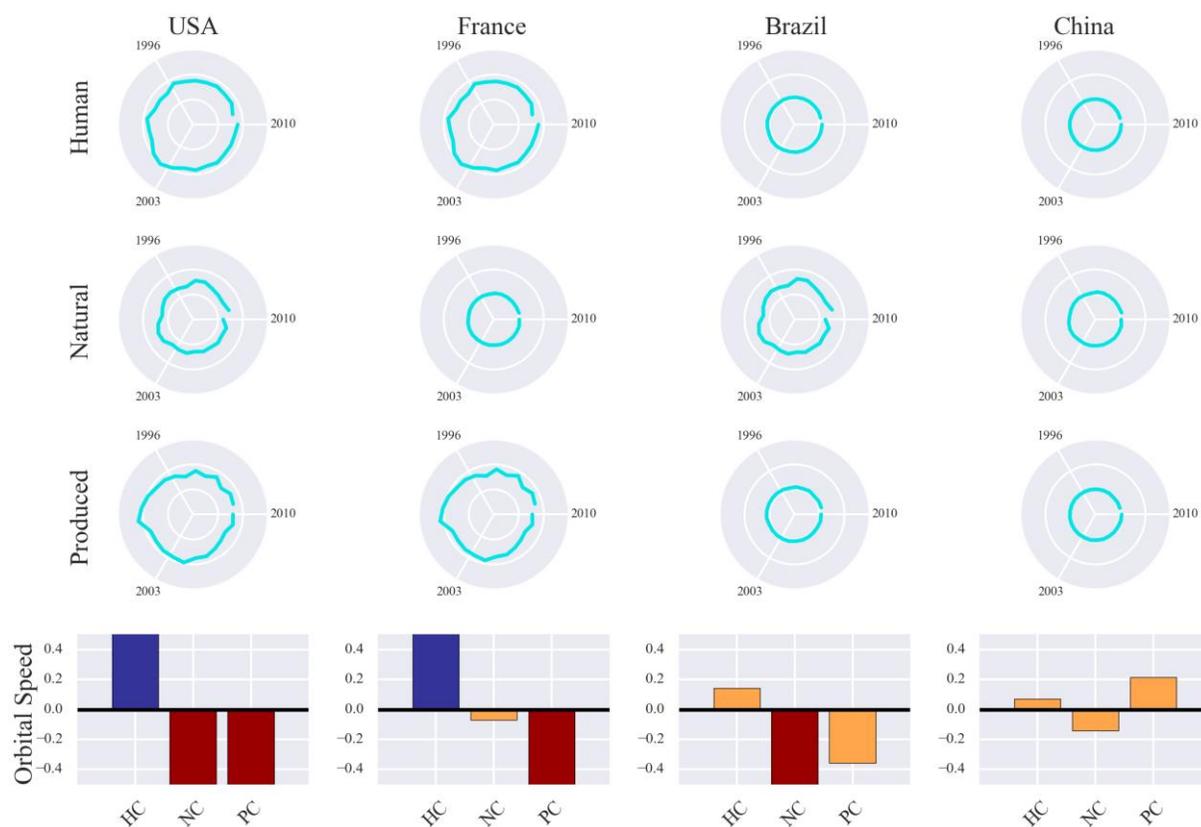


Figure 22 Orbital Diagrams and orbital Speeds of the United States (USA), France, Brazil, and China. The top three rows show the orbital diagrams, where the inner most circles denote the orbital position of the mode at 100 and the outer most circles denote the orbital position of 250. The bottom row shows the orbital speeds since 2000. Orbital speeds between -0.5 and 0.5 are marked with beige, whereas orbital speeds less than -0.5 are marked with red and orbital speeds greater than 0.5 are marked with purple.

Overall, we find mixed results when analyzing each country's performance relative to the N mode. For instance, from Figure 22, we can see that the orbital positions of the USA for all capitals are consistently located away from the N mode (i.e., the inner circle). Additionally, except for HC, we observe a decreasing trend for the two other capitals, indicating that the USA is depleting its natural and produced capitals. Only a few countries have a scenario similar to the USA. In contrast, the orbital positions of all capitals for most countries (i.e., about 70%) are close to the N mode, akin to China that barely changed its orbital positions during the analyzed period. Indeed, despite China's increase in HC and PC, per capita values remain nearly constant and close to the N mode from 1990 to 2014. France's orbital positions are close to the N mode only for NC, with a slightly negative orbital speed; the orbital positions of HC and PC are away from the N mode with a negative orbital speed for PC and a positive orbital speed for HC. Therefore, similar to the USA, France is also depleting its natural and produced capitals, but the depletion rate for the natural capital is significantly lower (i.e., 0.07 geodesic distance per year) than that of the USA (i.e., 0.71 geodesic distance per year). Conversely, in the case of Brazil, except for NC, the orbital positions of the other two capitals remain close to N mode. We also observe a decreasing trend for NC and PC, whereas HC remains marginally positive from 1990 to 2014 for Brazil.

Orbital speed captures whether a country's orbital positions are increasing, decreasing, or remaining constant. We compute the orbital speeds of all countries from 1990 to 2014, and from years 2000 to 2014 to capture more recent dynamics. Figure 23 shows the frequency analysis of all countries based on whether they are increasing, decreasing, or remaining constant.

From Figure 23, we can see that for HC, few countries have decreasing orbital speeds. In fact, orbital speeds largely increase for most countries. Considering the N mode has stayed

nearly constant, this indicates a global gain of human capital, which is desirable. On the contrary, the scenario for the natural and produced capitals are reversed as 129 and 87 countries since 1990, and 120 and 108 countries since 2000 have decreasing orbital speeds for NC and PC. Moreover, from Figure 21, we found decreasing trends in both the mean and the median values for NC. Looking at trends in orbital speeds, we find decreasing trends for both NC and PC for most countries. This observation notably suggests that despite a global increase in PC, 108 countries failed to keep pace with the global trend since 2000. In contrast, both the traditional and NFA

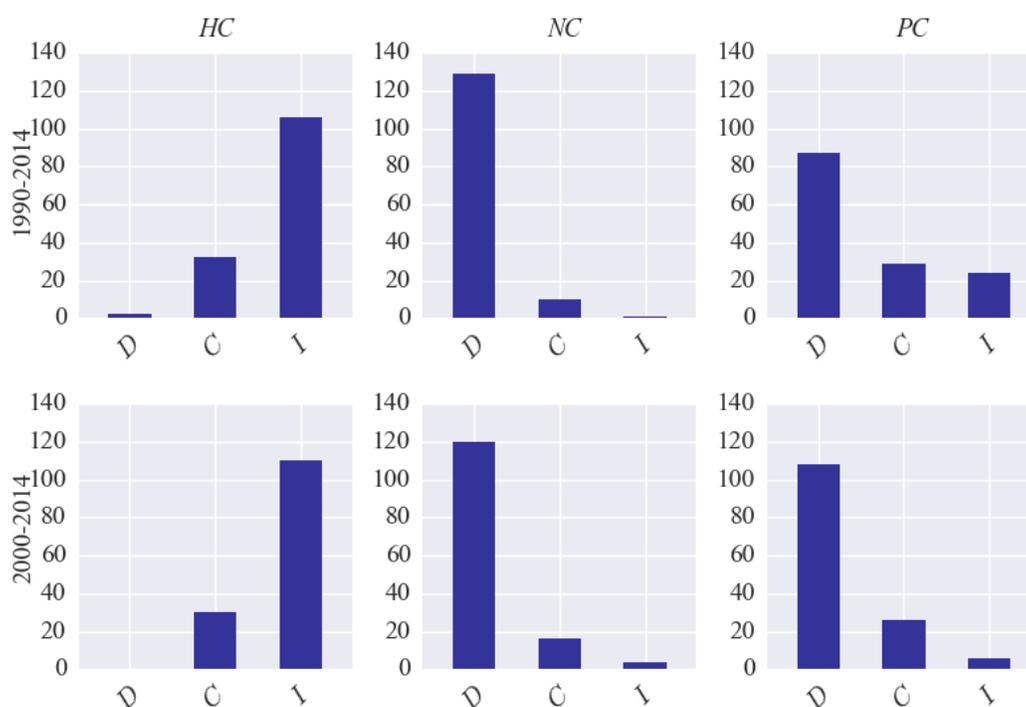


Figure 23 Number of countries with increasing, decreasing or constant orbital speed. The top row shows the orbital speed since 1990 and the bottom row shows the orbital speed since 2000 for human (first column), natural (second column) and produced (third column)

approaches corroborate the progressive trend in HC, likely thanks to major global initiatives that aim to significantly encourage socio-economic development in developing countries like

‘Agenda 21’ (United Nations Division for Sustainable Development, 1992) and the Millennium Development Goals (UNDP, 2000).

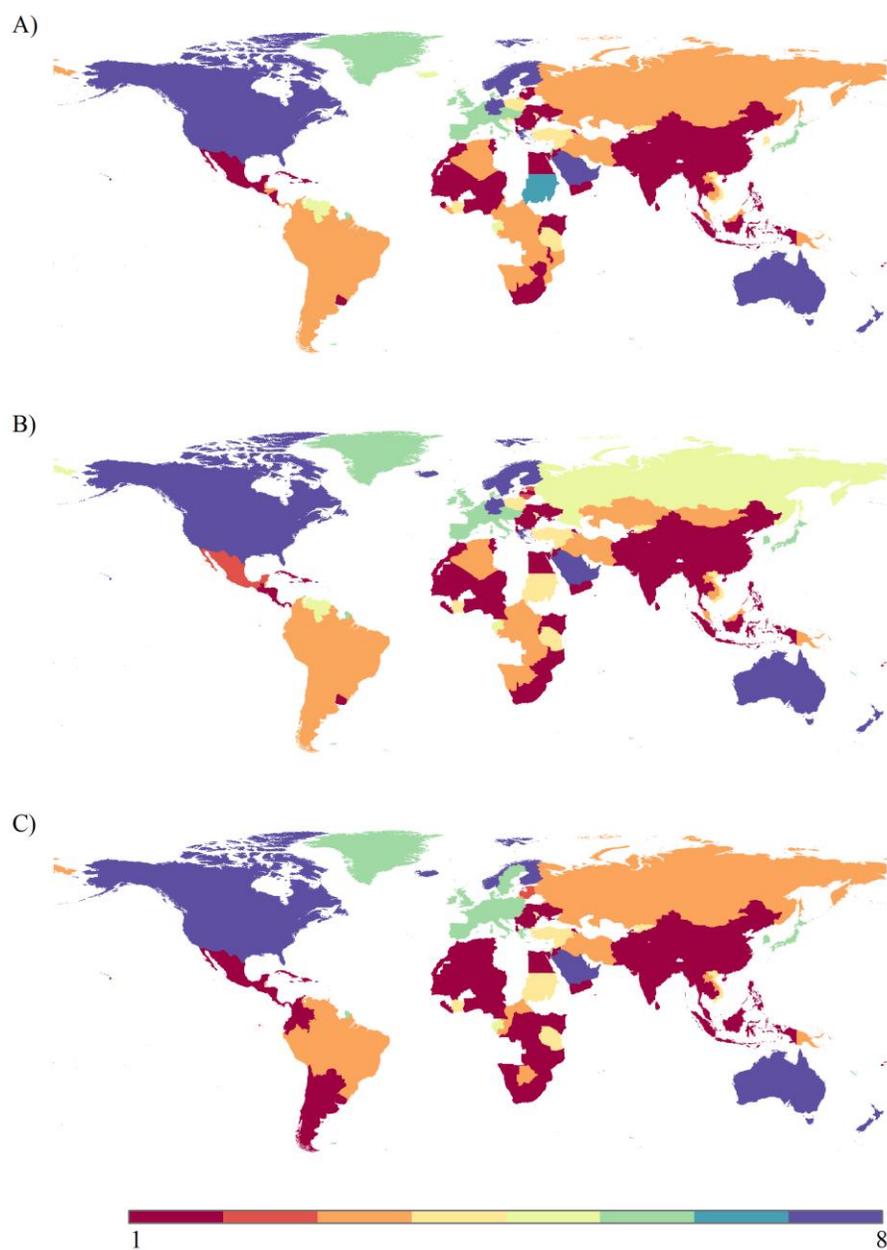


Figure 24 Combinations of Human, natural and produced capital for all the countries for the years: A) 1990, B) 2000, and C) 2014. Combinations 222, 221, 212, 122, 211, 121, 112, and 111 denote groups 1, 2, 3, 4, 5, 6, 7 and 8 respectively (see color bar and table S1 in supplementary section for details about the combinations).

Figure 24 shows the clusters of all countries for the years 1990, 2000, and 2014. For example, in 2014, France was outside the giant cluster for HC, inside the giant cluster for NC, and outside the giant cluster for PC, it is, therefore, part of the cluster with a combination of ‘121’ and group of 6, as we assign a value of ‘2’ if a country is in the giant cluster and a value of ‘1’ if it is not. Among the eight possible clusters (i.e., two distinct groups for HC, NC, and PC, thus $2^3=8$), four of them are much more common: 1) countries with high HC, NC, and PC (e.g., the USA); 2) countries with high HC and PC but low NC (e.g., France); 3) countries with low HC and PC but high NC (e.g., Brazil); and 4) countries with low HC, NC, and PC (e.g., China). Overall, about a quarter of the world countries maintained higher values (i.e., values outside the giant cluster) for at least two capitals from 1990 to 2014. These countries mostly include developed countries including the USA, Australia, and European countries. Among the remaining 75% of world countries, we find that most countries’ capitals for all three capitals remain within the giant cluster. Moreover, we find that the number of countries per group did not change significantly from 1990 to 2014. Table 4 shows the number of countries per group for the years 1990, 2000, and 2014. A heatmap tracking the cluster of every country every year is available in the Appendix A (Figure S6, and S7).

Table 4 Number of Countries per Group

Group Description	Group	1990	2000	2014
High human, natural and produced capital	8 (111)	11	14	11
High human and natural but scarce produced capital	7 (112)	1	0	0
High human and produced but scarce natural capital	6 (121)	20	22	27
Scarce human but high natural and produced capital	5 (211)	6	4	1
High human but scarce natural and produced capital	4 (122)	10	9	8
Scarce human and produced capital but high natural capital	3 (212)	31	25	16
Scarce human and natural capital but high produced capital	2 (221)	2	6	6
Scarce human, natural and produced capital	1 (222)	59	60	71

Then, using equation 32, we project orbital positions for all the countries for the years 2020, 2030, 2040, and 2050. Subsequently, we use the maximum orbital positions of the country within the giant cluster for HC, NC, and PC from 1990 to 2014 (i.e., 119 for NC and 118 for both HC and PC) to decide whether a country will be in the giant cluster in 2020, 2030, 2040, and 2050. For example, the orbital position for Russia's NC was 158 in the year 2014, but it will decrease to 117 in the year 2050, which is below the cutoff of NC (i.e., 118). Thus, although Russia was outside the giant cluster in 2014, it is projected to become part of the giant cluster by the year 2050 if it follows the current decreasing trend for NC. Moreover, we also find the projected cluster for each country in the future. As we found an increasing trend for HC and

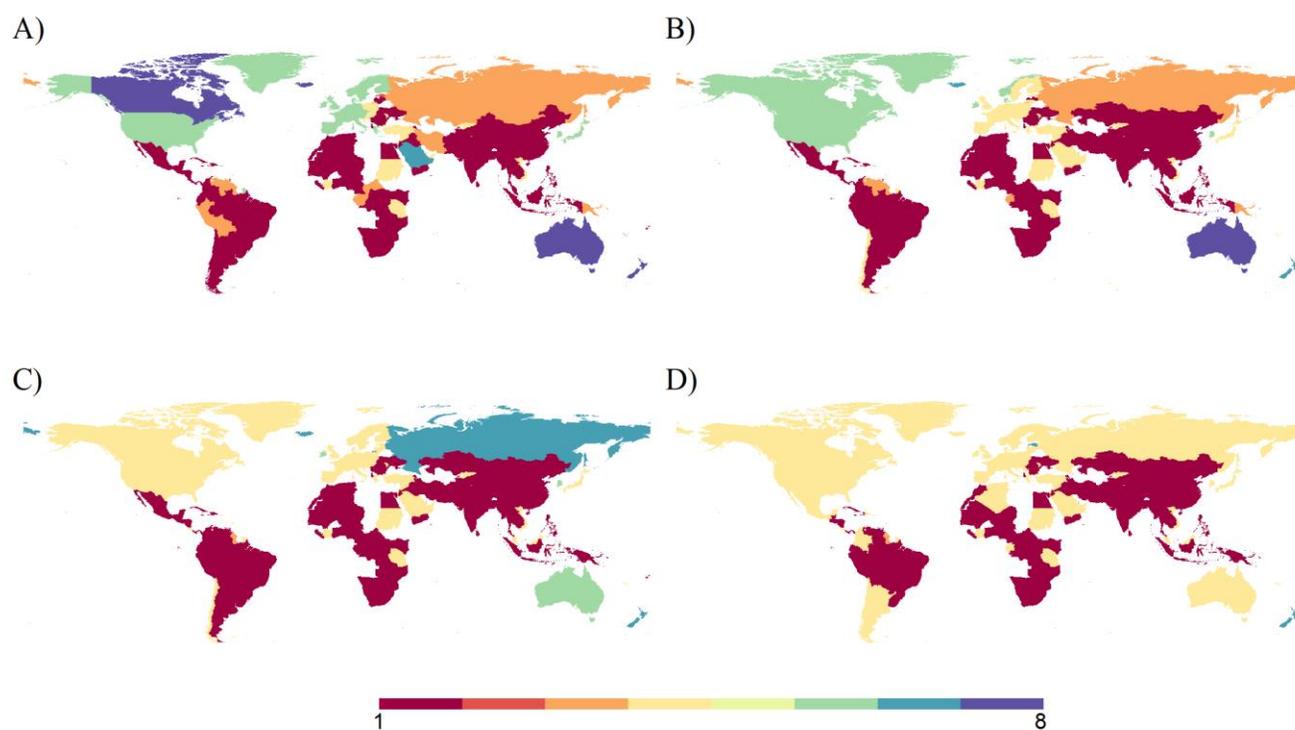


Figure 25 Combinations of human, natural and produced capital for all the countries for the years: A) 2020, B) 2030, C) 2040, and D) 2050. Combinations 222, 221, 212, 122, 211, 121, 112, and 111 denote groups 1, 2, 3, 4, 5, 6, 7 and 8 respectively (see color bar and table S1 in supplementary section for details about the combinations).

decreasing trends for NC and PC for most countries, most of them are projected to move or try to

move away from the giant cluster for HC while simultaneously trying to come closer or enter the giant cluster for NC and PC by the year 2050. As a consequence, by the year 2050, we expect two predominant cluster combinations: 1) 77 countries within the giant cluster for HC, NC, and PC; and 2) 59 countries outside the giant cluster for HC but within the giant cluster for NC and PC. Figure 25 shows the projected clusters for all the countries for the years 2020, 2030, 2040, and 2050. Moreover, we also find that countries with low HC and PC but with high NC (e.g., Brazil, Iran, Congo) will be affected more significantly and fail to keep their high NC, thus quickly joining the cluster with low HC, NC, and PC.

4.4 Conclusion

As the world has entered the Anthropocene (Crutzen, 2002; Steffen et al., 2007), tracking and assessing sustainable development in every country is paramount, if we, as a society, aspire to ensure a safe operating space for humanity (Rockström et al., 2009). For this research, we studied global trends in the evolution of human, natural, and produced capital by adopting a network science approach, which has successfully been applied in other fields (Ahmad and Derrible, 2015; Newman, 2010; Xu et al., 2011). This technique enables us to evaluate the evolution of individual country compared to the main trend by measuring the orbital positions; conventional techniques generally assess evolution in absolute terms only (e.g., GDP growth). To visualize these orbital positions, we generated “orbital diagrams” and we calculated “orbital speeds” to evaluate the evolving patterns (i.e., increasing, decreasing, or stable) of every country. Moreover, based on whether a country is connected to the main trend or not for all three capitals, we classify all countries per year and track their evolution from 1990 to 2014. The method created here can also be applied in a variety of cases; for example to track the performance of

individual cities over time. We find an encouraging optimistic trend for human capital, likely thanks to major global initiatives that aim to significantly encourage socio-economic development in developing countries like ‘Agenda 21’ (United Nations Division for Sustainable Development, 1992) and the Millennium Development Goals (UNDP, 2000), which also resulted in a positive trend for IW for 60% of the world countries. Meanwhile, the trends for natural and produced capital are rapidly decreasing and require immediate attention for a more robust sustainable development. Moreover, among 140 analyzed countries, 119 countries remain in the same group in 2014 as they were in 2000, although 90% of these countries’ GDP increased in the same period (Dasgupta et al., 2015; UNU-IHDP and UNEP, 2014), suggesting that the relative performance of most of the countries did not change much in this period, regardless of their gradual development, which is alarming as a decrease in produced and natural capitals along with an increase in GDP also suggests that most of the countries are consuming resources to increase their GDP, thus exhibiting features of unsustainable development. This result is alarming as decrease in produced and natural capital along with increase in GDP also suggests that most of the countries are consuming resources to increase their GDP, thus exhibiting features of unsustainable development. Moreover, by projecting the current trends in the future, we find that the number of countries with low human, natural, and produced capital will rise from 71 in 2014 to 77, and countries with higher values (i.e., values outside the giant cluster) for at least two capitals will decrease from 39 to 3 by 2050, if the current trends hold true. Furthermore, countries with low human and produced capitals but high natural capital (e.g., Brazil, Congo, Iran) will fail to sustain their high natural capital. Greater and more consistent effort should therefore be expanded to sustain natural capital as well as to increase human and produced capital in the years to come rather than focusing solely on the GDP growth.

5. Application of Fisher Information to Assess Stability in Multivariate Systems ¹

5.1 Introduction

As we pass through the twenty-first century, a massive advancement of information technology and rise of big data affects all the sectors around us, presenting substantial opportunities (Cottrill and Derrible 2015, Derrible 2016b, 2016a, Peiravian *et al* 2014, Derrible *et al* 2010) . There is a long history of using information by humans, but in modern days the process has become more advanced and robust (Derrible and Ahmad, 2015; Karduni et al., 2016). Efforts have been made to gain more information not only about observable phenomena but also about latent parameters inherent in system data. Increasing computing power and data availability, coupled with powerful data mining techniques have facilitated the growth and development of a plethora of approaches to discern and capture patterns in system behavior. Many of these concepts originate from information theory. Rooted in statistics, information theory resides between computer science, mathematics, physics and engineering, and has been widely applied from cryptology to ecosystem dynamics (Frieden and Gatenby, 2007; González-Mejía et al., 2014). Fisher information (FI), a key method in information theory, offers great promise for data mining applications. It was developed by Ronald Fisher (Fisher, 1922) as a means of measuring the amount of information about an unknown parameter that can be obtained by observations. Since then, it has been adapted into a means of monitoring system variables to assess patterns and evaluate stability in system dynamics (Karunanithi et al., 2008). FI has been used in a variety of applications from deriving fundamental laws of thermodynamics (Frieden,

¹ Published in Royal Society Open Science. Citation: Ahmad, Derrible, Eason, Cabezas (2016). Using Fisher information to track stability in multivariate systems. Royal Society Open Science, 3, pp. 160582.

1998) to assessing dynamic order in real and model systems (Karunanithi *et al* 2008, Fath *et al* 2003b, Fath and Cabezas 2004, Pawlowski *et al* 2005, Mayer *et al* 2006a) to sustainable environmental management and resilience (Mayer *et al* 2006b, Cabezas and Fath 2002, US EPA 2010, Ingwersen *et al* 2014, Mayer *et al* 2007, Eason and Garmestani 2012, Eason *et al* 2014, Spanbauer *et al* 2014, Eason and Cabezas 2012, Gonzalez-Mejia *et al* 2012, Gonzalez-Mejía *et al* 2012, Eason *et al* 2016, Gonzalez Mejia 2011, Gonzalez-Mejia *et al* 2015). In this chapter, I first present a brief overview of FI theory, describe the main calculation algorithm for FI and provide a simple computational example. The FI algorithm was previously coded in MATLAB, and the deployed applications are accessible by contacting the code developer directly (US EPA, 2010). Here we present an open source Python code for FI calculations, which is freely available. Furthermore, to demonstrate the use of FI for relatively complex systems, we calculate FI to analyze the evolution of the global-mean temperature from 1880 to 2015.

5.2 Background on Fisher Information

Fisher information was first developed by statistician R.A Fisher (Fisher, 1922), as a measure of indeterminacy. In other words, it can be used to measure the amount of information about an unknown parameter, θ that is present in observable data. Mathematically, Fisher Information, $I(\theta)$, is defined as (Karunanithi *et al.*, 2008):

$$I(\theta) = \int \frac{dX}{p_0(X|\theta)} \left[\frac{\partial p_0(X|\theta)}{\partial \theta} \right]^2 \quad (33)$$

where, $p_0(X|\theta)$ is the probability density of obtaining a particular value of X in the presence of θ .

In practice, it is essentially impossible to use Equation 33 because the computation of the derivative of the $(\partial p_0(X|\theta)/\partial \theta)$ component is required which depends on the numeric value of the unknown parameter θ . Through numerous derivation steps, (A.L. Mayer *et al.*, 2006) adapted

this equation for application to real systems based on the probability of observing various states of the system $p(s)$, such that:

$$I = \int \frac{ds}{p(s)} \left[\frac{dp(s)}{ds} \right]^2 \quad (34)$$

Equation 34 is the foundational form of Fisher Information used in this work. The equation is further simplified by eliminating the complication of handling a small $p(s)$ in the denominator. To overcome this problem $p(s)$ is replaced by its amplitude ($q^2(s) = p(s)$), which after some manipulation gives (A.L. Mayer et al., 2006):

$$I = 4 \int \left[\frac{dq(s)}{ds} \right]^2 ds \quad (35)$$

Karunanithi et al. (Karunanithi et al., 2008) further simplified this equation by assuming discrete steps so that $dq \approx \Delta q = q_i - q_{i+1}$ and $ds \approx \Delta s = s_i - s_{i+1}$. For sequential steps, $s_i - s_{i+1} = 1$, Equation 3 is written as:

$$FI \approx 4 \sum_{i=1}^m [q_i - q_{i+1}]^2 \quad (36)$$

where, m is the number of states. A state is defined as a condition of the system determined by specifying a value for each of the variables that characterize its behavior (Karunanithi et al., 2008). Equation 36 is used to compute Fisher information (henceforth denoted as FI) numerically for systems characterized by multiple discrete data. The following section will discuss the step-by-step procedure to compute FI. Complete details on the method and related derivations may be found in (Cabezas and Fath, 2002; Brian D. Fath et al., 2003; Karunanithi et al., 2008; Mayer et al., 2007).

5.3 Calculation Methodology

Evaluating changes in the probability of detecting different states of a system over time is the foundation of computing FI. Hence, information about a system's condition or state over time is required. A system can first be defined by n measurable variables (y_i), which are able to characterize the system and its state at any point in time (Karunanithi et al., 2008). The selection of variables is crucial, and effort should be put into selecting variables that are not only pertinent to a system but that also capture critical properties of a system. Each data point v_i at time t_j , $(v_{i,j})$, representing the entire system in a phase space is defined by the set of variables $v_{i,j} = \{y_1(t_j), y_2(t_j), \dots, y_n(t_j)\}$, to aid in categorizing the system into discrete states. In practice, to measure stability, we note that small fluctuations in a variable do not systematically translate into a regime change. Moreover, some inherent or small measurement error also frequently occurs. We define these fluctuations and small errors as measurement uncertainty which represents random variation in our system data.

Numerically, a parameter Δy_i is defined as measurement uncertainty such that, if:

$$|y_i(t_j) - y_i(t_k)| \leq \Delta y_i \quad (37)$$

is true for all variables y_i at time t_j and t_k then the two points are indistinguishable, and they are consequently "binned" together in the same state. In other words, if a system is defined by n measurable variables then a state is exemplified as a n dimensional hyper-rectangular box, where each side represents the uncertainty for each variable. Here, this set of Δy_i defines the size of state for the system.

Usually, unless reported with the data, the measurement of uncertainty is unknown.

Hence, (Karunanithi et al., 2008) recommend choosing a relatively stable time period in each

time series, and then computing the standard deviation (SD) of each variable Y with population mean \mathcal{G} and using Chebyshev's inequality, defined by:

$$P(|Y - \mathcal{G}| < k.SD) \geq (1 - \frac{1}{k^2}) \quad (38)$$

Equation 38 indicates that for any form of a probability distribution, “the proportion of the observations falling within k standard deviations of the [population] mean (\mathcal{G}) is at least $1 - 1/k^2$ ” (Lapin, 1980). Thus, Δy_i is chosen as $\pm k \times SD$. To ensure at least 75% of the data would occur within the level of uncertainty, a k of 2 can be selected as $1 - 1/2^2 = 0.75$ (Eason and Cabezas, 2012).

In other words, for one variable, two points can be considered to belong to the same state and be indistinguishable, if they vary within this defined level of uncertainty for this variable. Overall, this means that the state of a system is represented by all the points that are “binned” within a range of uncertainty (Karunanithi et al., 2008).

As mentioned earlier, the goal of FI is to capture dynamic behavior in terms of the probability of observing various states of a system. To move through the data, the time period is divided into time windows composed of several time steps (e.g., eight consecutive years), and one measure of FI is calculated for the time window which we attribute to the last time step of the window so that only past data are used in the computation (Eason et al., 2016). The time window is then moved by a defined number of steps. Two parameters are therefore used to define the moving window, which are the size of the window and the increment of the window. Both of these parameters are expressed in terms of time steps and depending on the data, time steps could be any units of time like years, months, days, weeks etc. These two parameters are used to move through the data such that the size of the window is greater than the amount of movement for each window in order to capture behavior that may extend beyond the boundary of

the window (Eason and Cabezas, 2012). The numerical example below illustrates this point. Then, the probability densities $p(s)$ and eventually FI for each window are computed. The size of the window depends on the amount of data available, but from empirical tests, a window size of at least eight-time steps has been recommended (Cabezas and Fath, 2002). Further details on the computation algorithm can be found in the US EPA report published in 2010 (US EPA, 2010).

After determining the parameters for the integration window (window size, window increment and size of state), the binning process can be initiated (Eason and Cabezas, 2012; Fisher, 1922). To begin, the first point of the time window is selected as the center of the first state and a hyper-rectangle, whose sides are defined by Δy_i for each variable of that system, is placed around that point. The points that lie within the hyper-rectangle are binned together. Then, the next unbinned point in the window is taken as the center of the next hyper-rectangle and similar points found within that hyper-rectangle are binned together. This process continues until all the points in the time window are binned or placed in different states.

Table 5 Sample Data for a Time Window

Time Step	Y_1	Y_2
1	0.6	1.5
2	2	1.5
3	0.3	1
4	3.5	4.8
5	0.95	2
6	3.1	4
7	2.4	1.8
8	2.7	2.1
ΔY	0.5	1

Following the approach presented by US EPA (US EPA, 2010), Figure 26 shows the binning process for a system, which is defined by two variables with size of state of 0.5 and 1 respectively (data shown in Table 5). In Figure 26, we can observe that, first, point 1 is chosen as

the center of the hyper-rectangle and point 1, 3 and 5 binned together to form state 1. Then, point 2, which is the next unbinned point taken as the center of a new hyper-rectangle and point 2 and 7 binned together to form state 2. After that, next unbinned point 4 is taken as the center of a new hyper-rectangle and point 4 and 6 binned together to form state 3. Finally, the only remaining unbinned point 8 forms state 4. Although point 7 and 8 are within the size of state to be binned together, but they are not, as point 7 is binned with point 2 earlier. Eight time steps are defined in each window and result in one measure of FI, which is plotted at the end of the window. For example, time steps 1 to 8 could represent data from year 2001 to 2008. For this example, we assign the value of FI to time step 8 (e.g., 2008). The next time window will go from time 2 to 9 (e.g., 2002 to 2009), followed by time step 3 to 10, etc.

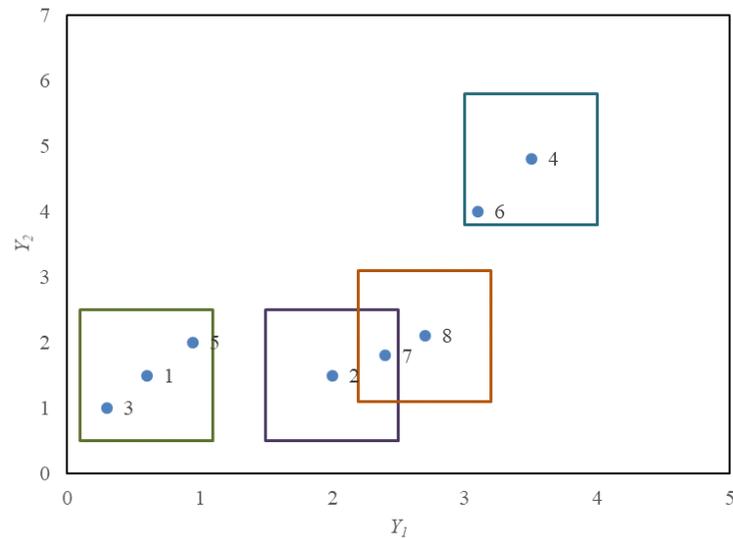


Figure 26 Illustration of the binning process to calculate FI. Note that binning of the points in table 5 resulted in four states.

When all the points are binned together, then probability distribution (p_i) for each window is estimated by using Equation 39 (A.L. Mayer et al., 2006):

$$p_i = \frac{\text{Number of points in state}}{\text{Total number of point in window}} \quad (39)$$

The probability distribution for the sample data in Table 5 is shown in Figure 27. Then the amplitude, q ($q_i = \sqrt{p_i}$) and FI for each window is calculated by using Equation 36, where the initial and final q_i is set as zero. Figure 27 and 28 display the $p(s)$ and $q(s)$ for each state based on the sample data in Table 5. The FI for the sample data using Equation 4 is: $4 \times [(0-0.61)^2 + (0.61-.5)^2 + (0.5-0.5)^2 + (0.5-0.35)^2 + (0.35-0)^2] = 4 \times (0.375 + 0.13 + 0 + 0.21 + 0.125) = 4 \times 0.534 = 2.136$.

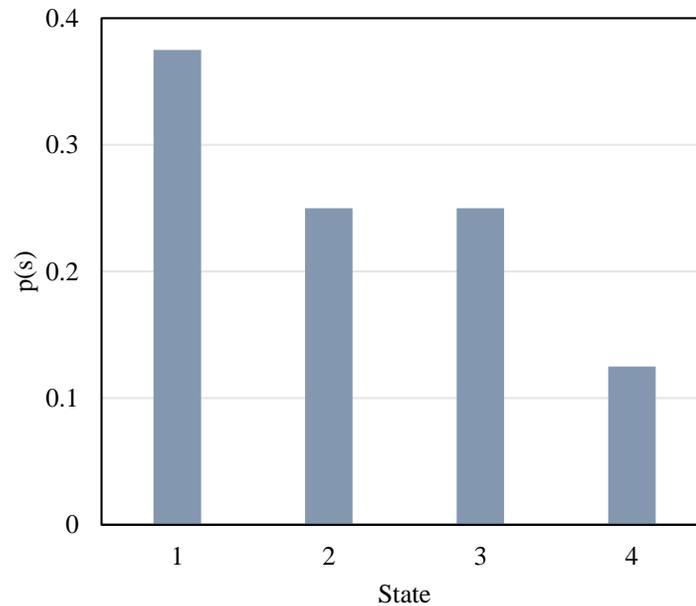


Figure 27 Probability Distribution

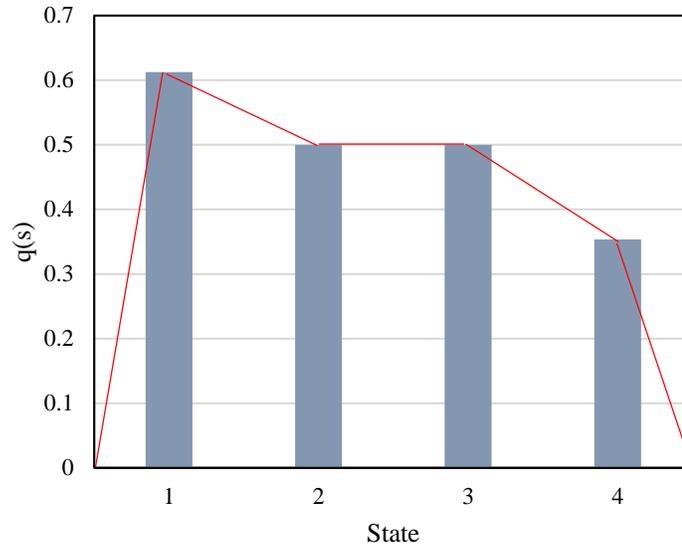


Figure 28 Amplitude of the Probability Distribution

In practice, system variables fluctuate such that not all the variables meet the size of state criteria. Therefore, a new parameter called tightening level (TL) is introduced to adjust the binning criteria. The tightening level (TL) adjusts the binning criteria such that a point can be declared to be within a given hyper-rectangle (a particular state of the system) when a certain percentage of the variables meet the size of states criteria (Fisher, 1922). For example, if a system is characterized by 100 variables and 95 of the variables indicate that a particular point fits within the state being evaluated, then the two points would be binned together at a 95% tightening level. There are no specific criteria for setting the tightening level, hence we take the average of all the TLs from strict (TL = 100%) to the lowest TL in which more than one state is observed in a window (Fisher, 1922). Moreover, to focus on the trends in dynamic order and not fluctuations, we may report a smoothed FI by averaging neighboring FI values (Eason and Cabezas, 2012). For example, if a time step of 3 is chosen for the block average, then 3 consecutive FI values (e.g., FI_1 , FI_2 and FI_3) are averaged, and that average value acts as the representative for all the three consecutive FI values. The averaging is essentially a high

frequency filter. Note the number of years including the averaging has to be smaller to capture more accurate trends in dynamic order.

5.4 Interpretation of FI

The Sustainable Regimes Hypothesis was developed to provide a construct for interpreting FI (Cabezas and Fath, 2002; B.D. Fath et al., 2003; Karunanithi et al., 2008).

- A system is considered to be in an orderly dynamic regime when a non-zero FI remains nearly constant over time (i.e., $d\langle FI \rangle / dt \approx 0$).
- A steady decrease in FI indicates that the system is losing its order, functionality, stability, and the patterns are breaking down. This declining trend may provide warning of an impending regime shift (Eason et al., 2014) or even a catastrophe, but the index alone will not pinpoint any particular indicator that contributes to the shift. Potential drivers may be identified post hoc using approaches (e.g., Spearman Rank Order correlations (Eason and Cabezas, 2012; Gonzalez-Mejía et al., 2012)).
- A steady increase in FI indicates that the system is becoming more organized/stable.
- A sharp decrease in FI indicates a regime shift and the intensity of the shift is related to the depth of the drop of FI (Karunanithi et al., 2008).

Further, researchers have noted that the actual FI value is not as important as the ability of the system to remain within a desired regime. Accordingly, when comparing different regimes, note that a stable system regime has a relatively high and stable mean FI (μFI) and low standard deviation in FI (σFI) (Eason and Garmestani, 2012; US EPA, 2010) than others.

Researchers have studied the behavior of FI in the neighborhood of a tipping point (Gonzalez-Mejia et al., 2015). While most systems tend to exhibit declining FI as a warning of

impending transitions (Eason et al., 2014; Eason and Garmestani, 2012), a number of theoretical scenarios have been explored to model expected behavior under different conditions (Eason et al., 2014). From this study, it is clear that the behavior of FI depends heavily on the trends in the variables as the system approaches a tipping point.

5.5 Case Study

To illustrate the utility of FI for assessing system stability, we assessed the evolution of FI as global-mean temperature changed from 1880 to 2015. The data was collected from the National Aeronautics and Space Administration (NASA), Goddard Institute for Space Studies (GISTEMP Team, 2016). The data included monthly global temperature anomalies in 0.01 degrees Celsius from the base period of 1950 to 1980 (i.e., how average monthly temperatures diverge from average temperatures recorded from 1950 to 1980). In order to assess how the average temperatures evolved over time, we organized the time series data such that each month represents one system variable and end up with 12 variables describing global temperature anomalies from January to December for each time step (year). The rationale is that if the climate is stable, the temperature of each of the months would be about the same every year regardless of seasonal variations. For this analysis, the window size of 40 years is chosen, as changes in climate tend to transpire rapidly over only a few decades (e.g., 40 years) (Adams et al., 1999; Audrey L. Mayer et al., 2006; Taylor, 1999). Using the approach described previously, the size of state is calculated and found to be 35.96, 39.21, 34.14, 31.17, 32.94, 24.68, 24.14, 31.28, 25.87, 36.43, 27.25 and 40.81 respectively for the twelve variables (i.e., global temperature anomalies from January to December) used for the analysis. The Python

scripts supplied at <https://github.com/csunlab/fisher-information> (Accessed 10/7/2016) were used to compute FI for this study.

Figure 29A shows the evolution of the global-mean temperature from 1880 to 2015 and Figure 29B provides the FI for the corresponding data. A window size of 40 is chosen, a moving window increment of 1 is used and FI is assigned at the end of the window; therefore, the first value reported is for 1919, representing 1880 to 1919. From the Figure, we observe a significant change in FI from 1978 with continual decrease since then, suggesting a rapid change in the global-mean temperature. Moreover, the FI for the period of 1919 to 1978 was more stable with an average (μFI) of 5.09 and standard deviation (σFI) of 0.89, than the period of 1979 to 2015 ($\mu FI = 4.04$, $\sigma FI = 1.32$). The decline in FI from 1979 to 2015 represents a 62.62% change indicating significant variation in global temperature patterns.

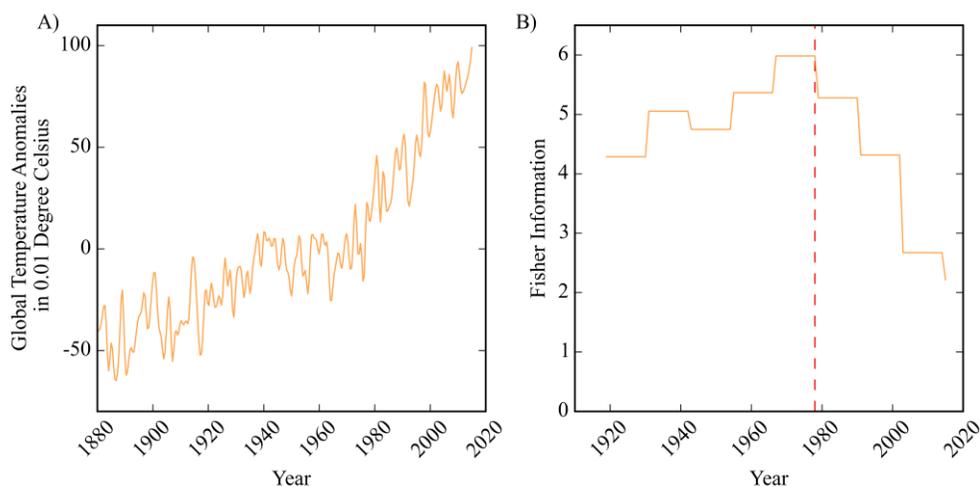


Figure 29 Evolution of A) Global-mean Temperature from 1880 to 2015 and B) FI for Global-mean Temperature.

Naturally, these analyses are not sufficient to fully capture how the global climate is performing, however the change in the FI trajectory during the late 1970s corresponds with the period in which our global societal demand (ecological footprint) also began to surpass the global biocapacity to supply that demand (Global Footprint Network, 2010). Moreover, the later

part of the twentieth century is also noted for major anthropogenic global environmental impacts (Steffen et al., 2007) and studies identify this period as the base of a new Anthropocene epoch (Lewis and Maslin, 2015).

5.6 Conclusion

The main objective of this chapter was to present FI as a useful method for data mining applications by demonstrating its utility in assessing patterns in complex system data (Ahmad and Derrible, 2015). FI has been applied to a variety of systems but the transferability and utility of the method have been hindered by the algorithm development in Matlab with deployed applications available only by contacting the code developer directly. The creation of an open access Python script offers significant opportunities for the general scientific community to facilitate the calculation of FI for any multivariate data. The assessment of global temperature provides a simple case study and suggests that it appears to have destabilized in the latter part of the twentieth century (i.e., since 1978) which corresponds to increasing ecological demand, declining biocapacity, and the initial stages of the new Anthropocene. This case study afforded the ability to demonstrate the power of the index and shows how FI can provide information about trends in complex system behavior. This effort showcased FI as a viable tool for mining data. By providing public access to the Python script for FI, we hope to expand use of the method to a broader audience who may be interested in methods for detecting hidden trends and identifying signals useful for system evaluation and management.

6. Using Fisher Information to Assess Stability in the Performance of Public Transportation Systems ¹

6.1 Introduction

As cities are expanding and becoming increasingly integrated (S. Derrible, 2016; Sybil Derrible, 2016), the future of public transportation systems (PTS) is bright. In 2014, a record of 10.8 billion trips were made by public transportation in the USA (APTA, 2015). Moreover, with constant urbanization and a strengthening of urban cores (Delmelle et al., 2014), the role that PTS will have to play in cities in the future can only increase. This is desirable for several reasons. To start, PTS can significantly help reduce traffic congestion that has a significant toll on urban economies every year. They also tend to be more sustainable than private vehicles from the viewpoint of greenhouse gas emissions (Derrible et al., 2010). Nonetheless, they are complex, and evaluating their performance over time presents significant challenges, especially as they depend on many different variables that continuously evolve over time (Grant et al., 2011). The performances of PTS are typically evaluated by looking at a range of metrics that can vary from agency to agency but that are rarely combined to get an overall performance assessment measure of a system. For example, the Chicago Transit Authority (CTA) focuses on six core areas of service: ridership, schedule, efficiency, cleanliness, safety and courteousness (CTA, 2017). In the scientific literature, many studies have developed their own performance metrics to evaluate PTS (Bertini and El-Geneidy, 2003; Derrible and Kennedy, 2009; Henderson et al., 1991; Kittelson et al., 2003; Liao and Liu, 2010; Pratt and Lomax, 1996), notably

¹ Published in Royal Society Open Science. Citation: Ahmad, Derrible, Cabezas (2017). Using Fisher information to assess stability in the performance of public transportation systems. Royal Society Open Science, 4(4), pp. 160920.

wrestling with issues of scale (e.g., state versus regional versus city level) (Baird and Stammer Jr., 2000; Bertini and El-Geneidy, 2003; Cramer et al., 2009; Grant et al., 2011).

Recently, significant advances in data science and information science have enabled the development of new and powerful techniques to analyze urban data (Ahmad and Derrible, 2015; Cottrill and Derrible, 2015; Derrible and Ahmad, 2015; Karduni et al., 2016; Kermanshah and Derrible, 2016, 2017; Peiravian et al., 2014). Within this general context, in this chapter, we show how Fisher information (FI) can be used to combine relevant metrics into one performance measure. FI can specifically be used to measure the ‘stability’ of a system. To compute FI, we use a Python script mentioned in the previous chapter (Ahmad et al., 2016a). On the one hand, PTS consist of many modes, from bus to heavy rail, that depend on many factors. On the other hand, seasonal fluctuations in ridership are common in almost all transit systems, making the analysis of monthly data difficult.

In information theory, complex systems are considered to be dynamic, orderly and well organized, but they also have the potential to undergo abrupt changes that can dramatically alter their performance. These changes are commonly referred to as regime shifts, e.g., eutrophication of lakes and coastal oceans and regional climate change (Karunanithi et al., 2008). Regime shifts happen in PTS as well, from the introduction of a new transit mode to sudden and substantial changes in ridership. FI is a key method developed by Fisher (Fisher, 1922) that offers a means of measuring the amount of information about an unknown parameter (e.g., performance) based on current observations, and it has been used to assess dynamic order in real and model systems (Brian D. Fath et al., 2003; Fath and Cabezas, 2004; Karunanithi et al., 2008; Audrey L. Mayer et al., 2006; Pawlowski et al., 2005). Moreover, FI is particularly able to combine many variables to assess the overall performance and stability of a system.

FI can be significantly useful to transit planners for three main reasons. First, FI provides an effective measure of overall performance of a PTS, and in particular, it is able to detect early warning signs that may lead to regime shifts. Second, it provides practical information to transit planners on which other PTS is going or has gone through similar situations. Third, because FI is not sensitive to differences in scale of the input variables, it can be used to categorize PTS across an entire region or a country (e.g., overall assessment of transit across the USA).

The main objective of this study is to measure the stability, order and regime shifts in PTS over time in all urbanized areas (UZAs) of the USA. More specifically, this chapter aims to:

- Compute FI for all PTS in the 372 UZAs reported by the National Transit Database (NTD),
- Analyze and interpret results in FI to detect patterns in the evolution of PTS, and
- Categorize PTS based on the interpretation from the computed FI.

To achieve these goals, monthly public transit data from 2002 to 2016 were collected from the NTD (NTD, 2016). In particular, we use unlinked passenger trips (UPTs) and vehicle revenue miles (VRM) for all modes reported in the NTD. Using these data, we can then compute FI for all US transit systems, and we can assess the overall pattern in the 14-year period. Overall, we find the presence of eight different patterns. Specifically, we detect regime shifts in 254 PTS, and we observe decreasing FI trends for 308 PTS, which may lead to regime shifts (Eason et al., 2014). Additionally, we find increasing FI trends for 136 PTS. Full details on FI are not provided here, but the reader is referred to the report of the US Environmental Protection Agency (US EPA, 2010) and Ahmad et al 2016 that clearly explain how FI is calculated step by step.

6.2 Application to Public Transportation Systems

PTS are commonly considered as the most sustainable motorized transportation systems, and they have been present in the USA for more than a century. With the significant change in technology, economy and socio-political environment, PTS have also changed substantially over time and have undergone several regime shifts. For a history of PTS, the reader is referred to Vuchic (Vuchic, 2005). By looking at the historical data, FI can therefore be used to track the changes experienced by PTS over time, which can help better understand patterns in ridership for instance. More specifically, we look into the FI for all UZAs in the USA as reported by the NTD for (i) rail, (ii) bus, (iii) others (i.e., modes which are neither rail nor bus) and (iv) all (i.e., overall performance). The NTD defines two major categories of PTS modes: rail and non-rail.

Moreover, among non-rail modes, the bus is by far the most predominant. The modes in the rail category used for this analysis are commuter rail (CR), heavy rail (HR), hybrid rail (YR), light rail (LR) and monorail/automated guideway (MG). The modes in the bus category used are commuter bus (CB), bus (MB), bus rapid transit (RB) and trolleybus (TB). Finally, other modes are also present, that we refer to as ‘others’, and they include: Alaska railroad (AR), cable car (CC), inclined plane (IP), street car rail (SR), demand response (DR), demand response—taxi (DT), aerial tramway (TR), ferryboat (FB), jitney (JT), publico (PB) and vanpool (VP) (NTD, 2016).

In this work, PTS data for all US public transportation authorities have been collected from the NTD that reports monthly data from January 2002 for four main types of data: UPTs, VRM, vehicle revenue hours and vehicles operated in maximum service (peak vehicles).

Because the three later are heavily correlated, for our analysis, we solely use UPT and VRM.

UPT is defined as the number of passengers who board public transportation vehicles, and VRM

is defined as the miles that are travelled by vehicles while in revenue service (NTD, 2016).

Another way to consider these datasets is that UPT offers an indicator of the demand for transit, while VRM offers an indicator of the supply. Moreover, we used all data points from January 2002 to December 2016.

6.3 Methodology

First, as already mentioned, the data for each PTS in each UZA are divided into four different systems: (i) rail, (ii) bus, (iii) other and (iv) all. Second, the total UPT and VRM for all the UZAs are analyzed to get an idea of overall public transit patterns in the USA. As a UZA can host several transit agencies that can have multiple transit modes, each system is defined by its total number of transit modes across all agencies. As an example, for the rail mode in the Chicago UZA, there are three different rail modes run by three different transit agencies: the CTA, Metra, and the Northern Indiana Commuter Transportation District. Information for both UPT and VRM for these three modes is available. As we collect two variables for each system (UPT and VRM), the Chicago rail mode is represented by six variables (i.e., two variables per mode). Overall, the Chicago has a total of 32 transit modes, giving us 64 variables.

Subsequently, using the procedure described in (Ahmad et al., 2016a; US EPA, 2010), FI for all UZAs present in the NTD was computed. The window size selected to measure FI was 12 for 12 calendar months since the NTD reports monthly data. Moreover, we choose a window increment of 1. In other words, we first compute the FI for January 2002 to December 2002, we then compute the FI for February 2002 to January 2003, and so on until December 2016. We

then calculate the yearly average to assess yearly performance; note that this two-step process allows us to account for seasonal variations in our calculations while outputting a yearly result¹.

On a technical note, FI requires a ‘size of state’ defined by the size of the hyper-rectangle used (Karunanithi et al., 2008), where any point outside of this hyper-rectangle forms a new state. In this work, we calculated the standard deviation of all windows and identified the smallest value as an indicator of the stable period. We then used Chebyshev’s inequality (Lapin, 1980) that multiplies the standard deviation by two to obtain the size of state for that particular variable. More information can be found in (Ahmad et al., 2016a; US EPA, 2010) and in the chapter 5.

As mentioned earlier, FI can increase or decrease or remain stable over a period of time. It can also undergo a sharp decrease, suggesting the occurrence of a regime shift. In this article, we define a drop in FI of 3 or greater as a regime shift. Moreover, in order to find the presence of an increasing or decreasing trend, we applied a Mann–Kendall non-parametric test (Kendall and Gibbons, 1990; Mann, 1945) with a confidence level of 95%. This nonparametric test yields three different outputs for the overall pattern: (i) increasing, (ii) decreasing and (iii) no detectable pattern. Moreover, we also detected how much the system is able to rebound after a regime shift or a decrease. The detection of a rebound of more than 75% of an original value is defined as ‘full rebound’, whereas a rebound in between 25% and 75% of the original value is defined as a ‘partial rebound’, and failure of rebounding to at least 25% of the original value is defined as ‘no rebound’ in this analysis. If the Mann–Kendall non-parametric test failed to detect any pattern and no regime shift is spotted, then the evolution of those FI is also classified as ‘no

¹ This is akin to the problem of taking the mean of the squares versus the square of the mean.

pattern/others'. Finally, we performed a frequency analysis to find the number of UZAs that belong to the different patterns in the evolution of the FI.

6.4 Results

The total UPT and VRM from 2002 to 2016 for eight UZAs are shown in Figures 30 and 31. As defined in the NTD, the UZAs include New York–Newark, NY–NJ–CT; Washington, DC–VA–MD; Chicago, IL–IN; Boston, MA–NH–RI; San Francisco–Oakland, CA; Philadelphia, PA–NJ–DE–MD; Atlanta, GA; and Sacramento, CA. Moreover, because the data are relatively noisy, we only show yearly averages (despite the fact that we use the original monthly data for the computation of FI).

Using these data, the FI of rail, bus, other and all were calculated for all 372 UZAs reported by NTD. Overall, we observe eight different patterns in the evolution of FI based on the Sustainable Regimes Hypothesis mentioned earlier. Specifically, we are able to detect patterns for 698 out of 1146 different PTS; the other PTS either do not exist (e.g., Milwaukee, WI, does not have a rail system), or no detectable pattern can be found. The eight patterns are listed in Table 6. Moreover, Table 6 also shows the frequency analysis for all the four categories (i.e., bus, rail, other and all). We can notably observe 254 PTS with regime shifts. We also detect 308 PTS with decreasing FI, and 136 PTS with increasing FI. Figure 32 shows the FI for eight UZAs representing the eight different patterns observed. Moreover, the FI for the eight major UZAs used in Figures 30 and 31 are shown in Figure 33. The same results for all PTS can be found in (Ahmad et al., 2017b).

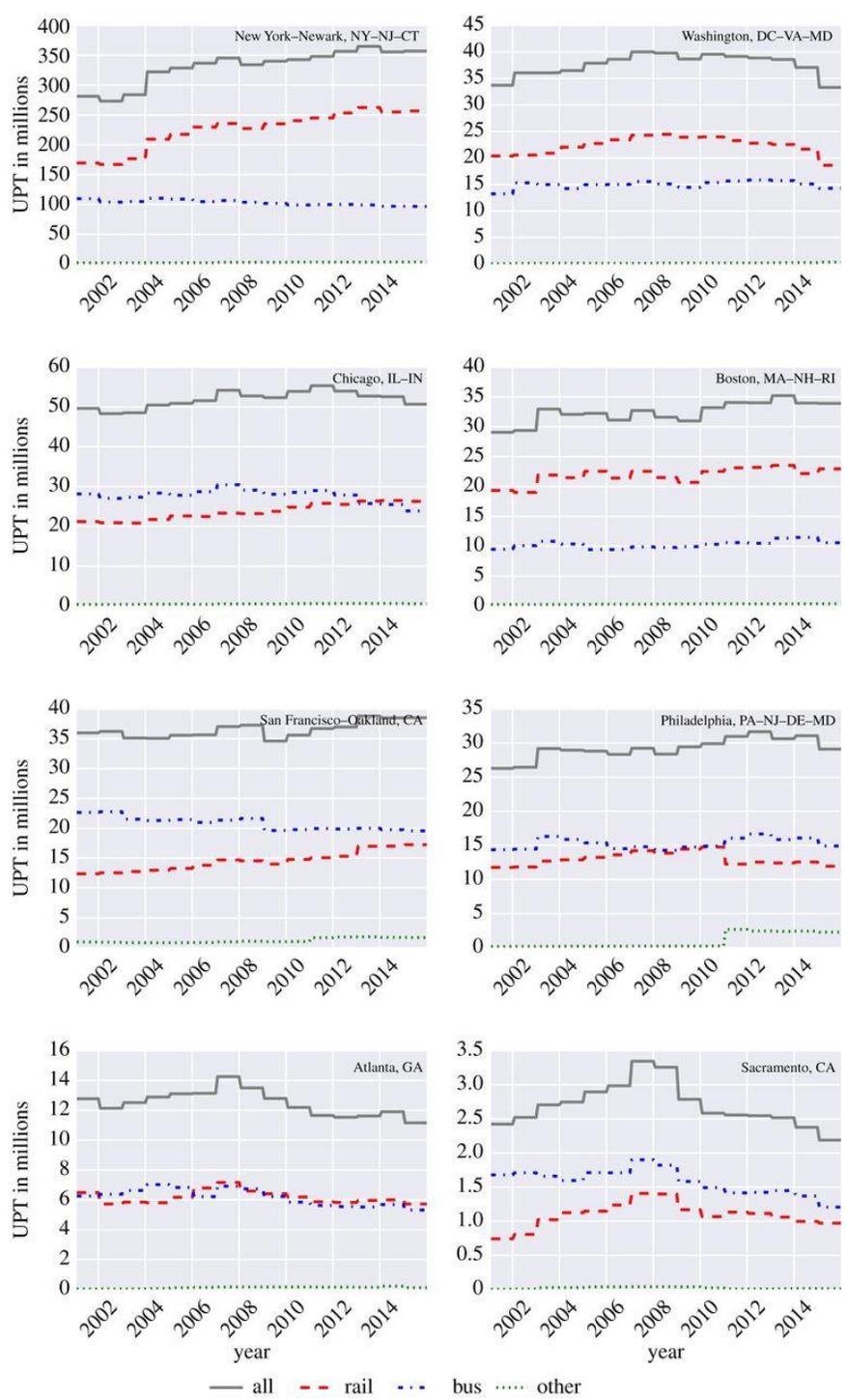


Figure 30 Total UPT for eight major USZAs

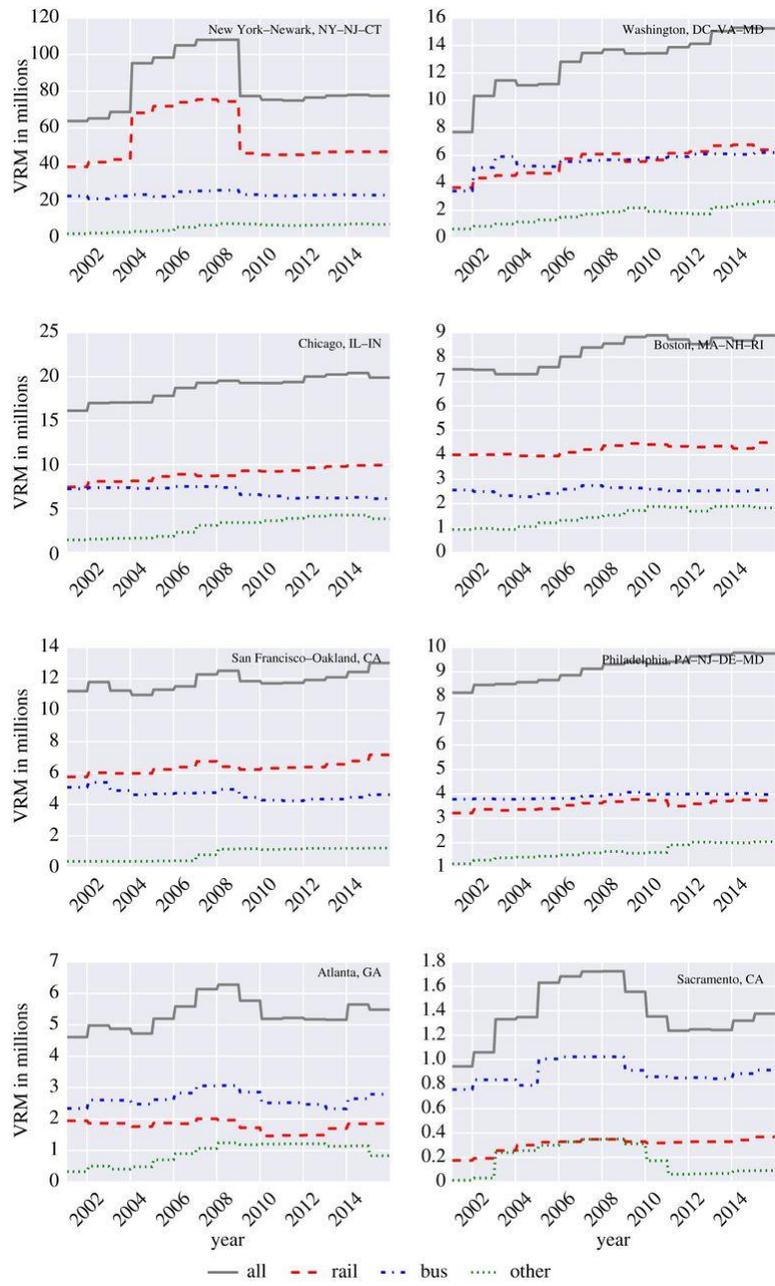
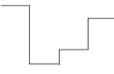
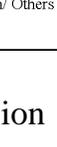
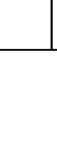
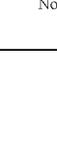
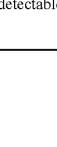
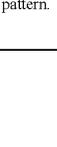
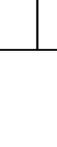
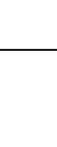


Figure 31 Total VRM for eight major USZAs.

Table 6 Patterns in the evolution of FI

Pattern	Properties	Illustration	Example	Frequency Analysis			
				Bus	Rail	Other	All
Regime Shift with Rebound	Drop in FI of 3 or greater and a rebound of 75% or greater from the minimum FI.		Bus- Richmond, VA Rail- Little Rock, AR Other- Utica, NY All- Utica, NY	36	1	36	34
Regime Shift with Partial Rebound	Drop in FI of 3 or greater and a rebound of 25% to 75% from the minimum FI.		Bus- Ithaca, NY Rail- Sacramento, CA Other- Green Bay, WI All- Corvallis, OR	25	5	24	18
Regime Shift without Rebound	Drop in FI of 3 or greater without any rebound from the minimum FI.		Bus- Salem, OR Rail- Portland, ME Other- Mount Vernon, WA All- Burlington, VT	27	6	20	22
Decrease with Rebound	Gradual decrease in FI with a rebound of 75% or greater from the minimum FI.		Bus-Lancaster-Palmdale, CA Other-Dover-Rochester, NH-ME	1	0	1	0
Decrease with Partial Rebound	Gradual decrease in FI with a rebound of 25% to 75% from the minimum FI.		Bus- Seattle, WA Rail- Portland, OR-WA Other-Medford, OR All-New York-Newark, NY-NJ-CT	44	8	50	59
Decrease without Rebound	Gradual decrease in FI without any rebound.		Bus-Boston, MA-NH-RI Rail-Chicago, IL-IN Other- Eugene, OR All-Yakima, WA	40	10	38	57
Increase	Gradual increase in FI.		Bus- New Haven, CT Rail-Memphis, TN-MS-AR Other-Fairbanks, AK All-Rochester, NY	45	5	45	41
No Pattern/ Others	No detectable pattern.		Bus-Buffalo, NY Rail- Springfield, MA-CT Other-Portland, ME All-Raleigh, NC	154	337	158	141

6.5 Discussion

From Figure 30, we can see that the total UPT for the eight major cities remained stable from 2002 to 2016, except for some minor fluctuations. In New York, the total UPT for bus and other remained flat, whereas the UPT for rail and all increased slightly in 2004 and they have continued to increase with a mild slope since 2004; this suggests that the overall changes in ridership in New York depend chiefly on the rail modes. In Washington, DC, the total UPT for all the modes remained mostly uniform throughout the period, but the ridership for rail and all followed a particularly similar pattern, suggesting a dominance of the rail mode for overall ridership patterns. The evolution of the UPT in Chicago remained stable for all four modes, and the overall ridership pattern is analogous to the pattern in the bus mode, indicating that the bus may be the dominant public transportation mode in Chicago. For Boston, and akin to

Washington and New York, the rail mode is dominant. In San Francisco–Oakland, like other UZAs, the ridership patterns were stable, except for a slight decrease in the years 2003 and 2009. Moreover, we also see that bus is the major public transportation mode as it follows the trends of the overall ridership pattern. Like Chicago, in Philadelphia, the total UPT trends for all the modes were stable, and the overall pattern was similar to the bus. For Atlanta, a small increase is observed until 2007, which is followed by a gradual decrease. Furthermore, between the years 2002 and 2006, the total UPT for the bus was slightly higher than that of rail, and the ridership for the remaining years remained nearly equal. Finally, for Sacramento, the UPT for all systems increased gradually until 2009 and it then decreased steadily until 2016.

Looking at VRM (Figure 31), we see that except for New York, Washington, San Diego and Atlanta, the VRM of the other UZAs were stable from 2002 to 2016. New York experienced

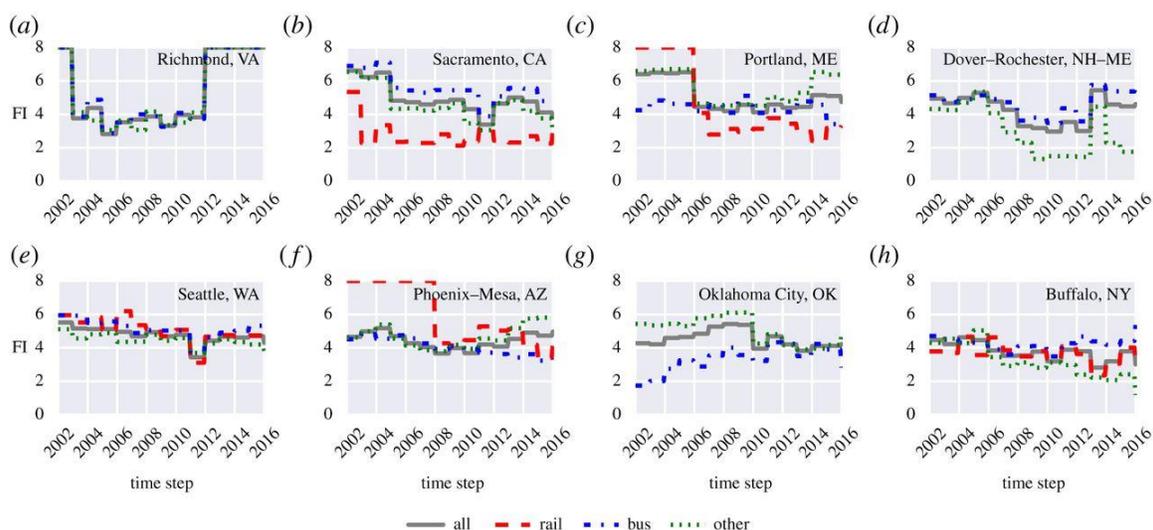


Figure 32 Evolution of FI for eight UZAs with: (a) regime shift with rebound, (b) regime shift with partial rebound, (c) regime shift without rebound, (d) decrease with rebound, (e) decrease with partial rebound, (f) decrease without rebound, (g) increase and (h) other.

a jump in VRM in 2004, which then dropped in 2009, and remained around the 40 million mark until 2016. For Washington, DC, the VRM increased gradually from 2002 to 2016. Although rail ridership in Washington was higher than that of the bus, revenue miles for both the modes were

almost similar. In Atlanta, we can observe that the VRM increased from 2004 to 2008, followed by a decrease until 2010, and then remained almost constant until 2016. Finally, the VRM patterns for Sacramento, CA, are similar to the patterns found for the UPT, except for the rail system that remained relatively stable after an increasing period between 2002 and 2005.

From Figure 30, we can identify that except for New York, Washington and Boston, the bus attracts most of the riders, which is also reflected in the higher bus VRM. Nonetheless, and despite the fact that the bus is dominant in Chicago and San Francisco, the rail VRM were higher from 2002 to 2016. In Philadelphia and Atlanta, the bus VRM were higher than all other modes.

Focusing in FI results, from Table 6, we can observe that the pattern ‘no pattern/others’ dominates for all categories. The majority of the UZAs for rail, bus, others and all fall in this category, which suggests that most PTS are either significantly stable or perpetually looking for stability. Among the other PTS, we can see that a significant number follow an increasing trend, including the bus systems in San Jose, New Orleans, Oklahoma City and Milwaukee. Moreover, from the PTS that experienced a regime shift, few were able to rebound completely or partially.

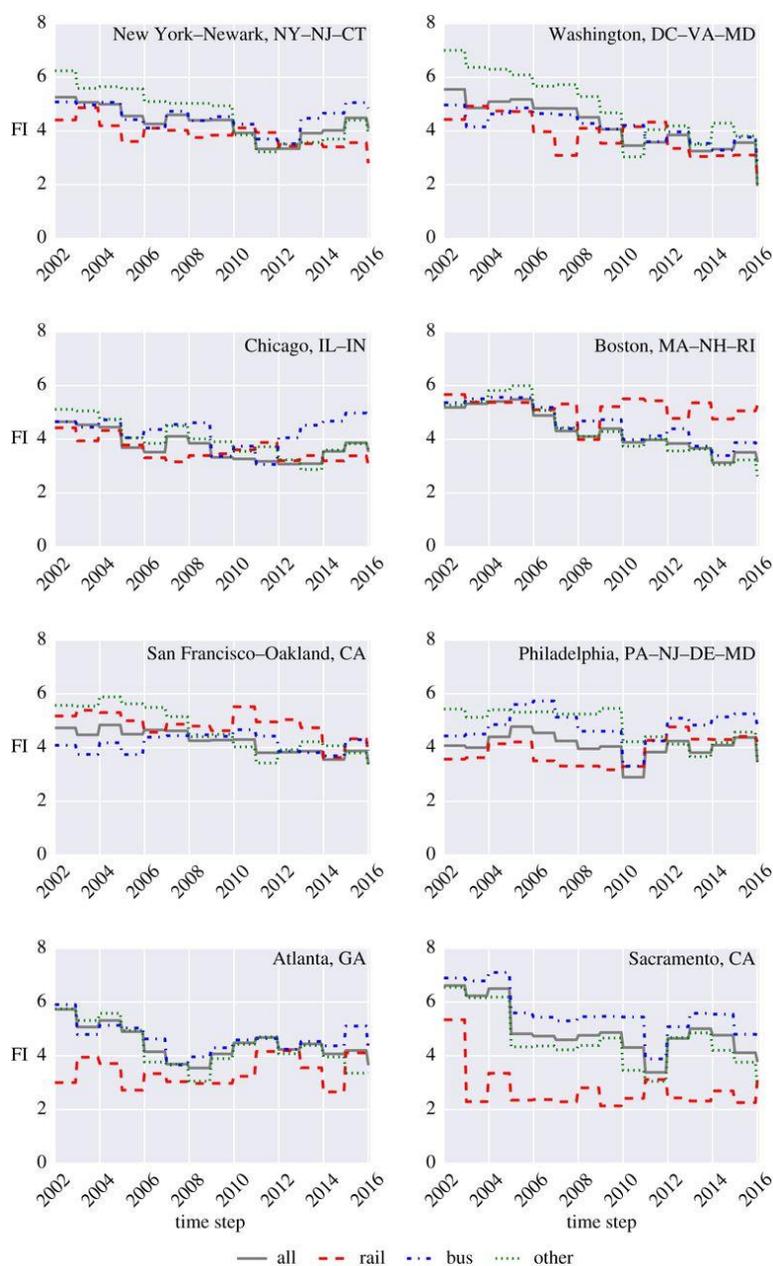


Figure 33 FI for eight major UZAs.

Regime shifts are mostly observed in small UZAs because of the introduction or termination of a new mode, Sacramento offering a notable exception. In 2003, both the UPT and the VRM increase gradually, and yet the rail mode underwent a regime shift. A similar situation occurred in 2009, but no regime shift was detected. Moreover, the prime reason for rebound after a regime shift is also due to the introduction or termination of a new transit mode. In the case of a

decreasing pattern (e.g., rail systems in Chicago and Washington), few PTS were able to rebound to their earlier state, and nearly an equal number of PTS have rebounded partially or failed to rebound after experiencing a decrease in FI. These results give somewhat of a cause for concern as decreasing patterns may exhibit a warning for an upcoming regime shift since uncertainty in the system is increasing (see Sustainable Regimes Hypothesis in §5.4).

Figure 32 shows examples for the eight different patterns observed overall. A regime shift with rebound was found in the bus system of Richmond, VA. The rail system of Sacramento experienced a regime shift with partial rebound, whereas the rail system in Portland, ME, also experienced a regime shift but it failed to rebound. A decreasing trend in FI is found in Dover–Rochester, NH–ME, as well as in Seattle, WA, and Phoenix–Mesa, AZ. However, in Dover–Rochester, NH–ME, all systems managed to rebound completely, while the Seattle, WA, transit system rebounded partially, and the Phoenix -Mesa, AZ, bus system failed to rebound altogether. Moreover, the Oklahoma City, OK, bus system shows an increasing trend in FI, and no detectable pattern was found for the PTS of Buffalo, NY.

From Figure 33, we can observe that in New York, the rail, other and all modes have a decreasing trend in FI, and while the rail system was unable to rebound, the others and all categories only rebounded partially. Moreover, no detectable pattern was found for the bus system in New York. In Washington, all modes have a decreasing trend in FI, and only the other system was able to rebound partially; the other modes failed to rebound. For Chicago, no detectable patterns were found for the bus system, whereas the rail and all systems show a decreasing pattern without any rebound, and the other system shows a decreasing pattern with partial rebound. From analyzing the FI for Boston, a decreasing pattern without any rebound is observed, except for the rail mode, for which no detectable patterns were found. Rail, other and

all show a decreasing pattern without any rebound in San Francisco, but no detectable pattern was identified for the bus mode. For Philadelphia, a decreasing pattern with partial rebound is observed for the others and no detectable pattern was found for the remaining three systems. No detectable patterns were found for the rail and bus systems in Atlanta, whereas decreasing trends with partial rebound are observed for the remaining two. Finally, for Sacramento, a regime shift was found for the rail system in 2003, essentially capturing the increasing period for both the UPT and the VRM of the rail system.

Overall, this analysis showed no negative regime shifts; we mostly observed regime shifts due to the introduction of a new mode. While this is desirable, we also noticed that several modes seem to be on a decreasing trend, which can lead to a regime shift or dysfunction if nothing is done (e.g., bus system in Boston and rail in Chicago). An analysis of the trend in FI can therefore help transit agencies identify whether their systems are currently maintaining or losing stability, leading to possible measures to improve performance.

6.6 Conclusion

The main objective of this chapter was to measure the stability, order and regime shifts of the PTS of all US UZAs by using concepts of FI. To achieve this goal, a Python code (Ahmad et al., 2016b) was used to compute FI for all the PTS. In particular, UPT and VRM datasets from the NTD were collected and used for this analysis. FI for the bus, rail, other and all modes were computed for all 372 UZAs. We notably found the presence of eight different patterns, and the majority of the systems belong to the final category (i.e., no pattern/others). This suggests that most PTS are searching for a stable state. Among the remaining PTS, a significant number experienced a decrease in FI (i.e., 308 PTS), which is a cause for concern. By contrast, we also

find optimistic trends for 136 UZAs with an increasing FI. Furthermore, several regime shifts were detected for different UZAs. A regime shift can either be positive, for example, for the case of introducing a new service (e.g., in San Diego), but it can also be negative, for example, when a needed service is terminated. Moreover, besides providing FI for eight UZAs showcasing the eight different patterns, FI for eight major UZAs were provided. Overall, PTS offer myriads of benefits, but they are also complex by nature. Considering PTS have a bright future ensuring their success is critical. This is particularly important as the era of autonomous vehicle is rapidly approaching and will likely result in a decrease in operation costs for PTS. FI offers a means to combine multiple variables of a complex system to determine its overall stability, which can prove to be valuable in practice.

7. An Information Theory Based Robustness Analysis of Energy Mix in US States ¹

7.1 Introduction

Fossil fuels have traditionally been the primary energy source for electricity generation in the United States (US). Before 1970, fossil fuels were used to generate about 80% of the electricity. This proportion dropped to 71% by 1978. Two economic recessions (1973 and 1980), partly triggered by the two oil crises in the 1970s (Hamilton, 2011; Kerr, 1998; Schurr, 1983), resulted in a negative growth in the US gross domestic product (GDP) and a high unemployment rate (“Bureau of Labor Statistics,” 2018). In 2015, after more than three decades, fossil fuels still accounted for 65% of all energy sources for electricity generation.

The events of the 1970s therefore had a substantial impact on the energy mix for electricity generation in a relatively short period of time—we quantify how much of an impact in this study. In particular, uncertainties linked with the availability of fossil fuels, the availability of cheaper fuels, accompanied by significant technological innovation, drove electric utility companies to diversify their energy source (Dooley, 1998; Estanqueiro, 2010; Margolis and Kammen, 1999; Weijermars et al., 2012). It is therefore not unreasonable to image that similar events could occur in the future. More generally, we need to recognize that states that rely on a single or few energy sources can be inherently more vulnerable to external shocks. This applies both to fossil fuel and non-fossil fuel energy sources; for example, fossil fuels can be vulnerable to shortages of supply (e.g., because of a war or economic crisis) and non-fossil fuel sources can be vulnerable to climate change (e.g., rainfall patterns for hydroelectric power). At the same

¹ Published in Energy Policy. Citation: Ahmad, Derrible (2018). An information theory based robustness analysis of energy mix in US States. Energy Policy, 120, pp. 167–174.

time, severe external shocks and crises are highly uncertain by nature; some even occur in a cascading or escalating form (Petit et al., 2015). This uncertainty remains a vital challenge for robustness or vulnerability analysis (Ben-Haim, 2012; A. Kermanshah and Derrible, 2017; Park et al., 2013), and various models have been developed to understand uncertainty in any system (Dempster, 1967; Edwards, 1969; Ferson, 2002; Klir, 2005; Miranda, 2008; Shafer and others, 1976; Wald, 1945; Walley, 1996; Zadeh, 1965). In addition, and partly because of this uncertainty, the preferred approach in robustness analysis is often to analyze the worst case or extreme scenarios (Ben-Haim, 2012; Wald, 1945). For example, electric power networks have been analyzed comprehensively to assess their vulnerability to any type of failure (Cao et al., 2006; Carreras et al., 2001; Ding and Han, 2008, 2006; Mei et al., 2011), often using network science or graph models (Arianos et al., 2009; Chassin and Posse, 2005; Holmgren, 2006; Pagani and Aiello, 2013; Sun, 2005). In this article, we analyze the robustness of the electricity grid by focusing on the proportion of energy sources used for electricity generation, and we start from the premise that a more diverse energy mix is generally desirable from a robustness viewpoint. More precisely, as a state with a more diverse energy mix is more robust than a state relying on a single energy source, the main goal of this study is to numerically measure the diversity of each U.S. state energy mix. Towards that end, we define a measure of robustness using Shannon entropy and analyze the impact of several extreme energy mix scenarios for all US states.

According to Energy Information Administration (EIA), in 2016, 4.08 trillion kilowatt-hours (kWh) of electricity were generated at utility-scale facilities in the US; an electric utility-scale facility is defined as “a corporation, person, agency, authority, or other legal entity or instrumentality aligned with distribution facilities for delivery of electric energy for use primarily by the public” (EIA, 2017). This electricity was generated from nine sources of energy: 1) coal,

2) natural gas, 3) distillate fuel oil and kerosene-type jet fuel, 4) nuclear, 5) hydropower, 6) geothermal, 7) photovoltaic and solar thermal, 8) wind, and 9) wood and waste (biomass). In 2015, about 65% of the electricity was generated from fossil fuels (sources 1-3), 22% from nuclear energy (source 4), and 13% from renewable energy sources (sources 5-9). These proportions vary greatly from state to state, however. On the one hand, some states (e.g., Utah, Missouri) depend predominantly on fossil fuel sources, and on the other hand, some states (e.g., Vermont) depend almost entirely on renewable energy sources. Figure 34 shows a map of the 50 US states grouped based on their fossil fuel usage for electricity generation in 2015, where we find that only 5 states used less than 25%, 7 states used 25% to 50%, 20 states used 50% to 75%, and 18 states used more than 75% fossil fuel to generate their electricity. In particular, 12 states used less than 50% fossil fuel sources to generate their electricity: South Dakota, Illinois, Vermont, Connecticut, South Carolina, Maine, New Hampshire, Washington, Oregon, New York, New Jersey, and Idaho. We note, however, that these 12 states used to rely mostly on fossil fuel sources to generate electricity, but they went through significant changes, either regime shifts or steady transitions, that we study in this article.

indicating the significant impact of major events in the 1970s. Finally, in the robustness analysis, we find that 10 US states are particularly vulnerable as they depend on a single fossil fuel source for electricity generation. In contrast, we also detect 7 US states whose energy mixes are well distributed while having less than 50% fossil fuels.

7.2 Materials and Methods

In this study, we use Shannon entropy to capture changes in energy grid mix. Shannon entropy was developed by Claude Shannon (2001), and it captures the diversity or homogeneity of a system; i.e., a higher diversity translates into higher entropy. It is defined as:

$$H(x) = -\sum p(x_i) \cdot \log_2(p(x_i)) \quad (40)$$

where $H(x)$ is the Shannon entropy and $p(x_i)$ is the probability of event x_i . Although $H(x)$ is bounded by $[0, \infty]$, entropy is maximum when all events have the same probability. We use the natural logarithm for this study. As we found nine major sources for electricity generation, we have nine possible events per year for each state. To compute entropy, we replace $p(x_i)$ in equation 1 with the shares of the nine main sources of energy for electricity generation. Table 7 shows the computation of entropy for California in 2015, which is 1.53. With nine equal probabilities (i.e., $p(x) = 1/9$), the maximum possible entropy is 2.20.

Table 7 Computation of Entropy for California in 2015

Energy Source	Usage in %	$p(x_i)$	$p(x_i) \cdot \log(p(x_i))$
Nuclear	12	0.12	-0.25
Hydropower	8	0.08	-0.20
Geothermal	7	0.07	-0.19
Solar	9	0.09	-0.22
Wind	7	0.07	-0.19
Biomass	5	0.05	-0.15
Natural Gas	52	0.52	-0.34
Total	100		$H = 1.53$

Shannon entropy is particularly relevant here because it is able to capture changes due to the combined effect of all energy sources. More advanced information theory measures like Fisher Information (Ahmad et al., 2017a, 2016a) were tested but they did not capture any relevant additional information for this particular study.

Subsequently, we again use Shannon entropy to evaluate the robustness of a US state to a disruption in one of its energy sources. For this, we use the year 2015 as the base year (since it is the last year for which data were available) and we reduce the share of each energy source by 5%, 10%, 25%, 50%, 75% and 100% to calculate the new probability:

$$p(x_{i,j,\Delta p}) = \begin{cases} \text{if } i=j, \frac{p(x_i) - p(x_j) * \Delta p}{1 - p(x_j) * \Delta p} \\ \text{else, } \frac{p(x_i)}{1 - p(x_j) * \Delta p} \end{cases} \quad (41)$$

where $p(x_{i,j,\Delta p})$ is the new probability for each source j (e.g., natural gas) for $\Delta p = \{0.05, 0.10, 0.25, 0.50, 0.75, 1\}$. For example in California, when, $i = j = \text{natural gas}$ and $\Delta p = 0.5$, we get $p(x_{\text{natural gas}, \text{natural gas}, 0.05}) = (0.52 - 0.52 \times 0.5) / (1 - 0.52 \times 0.5) = 0.35$. Similarly, when $i \neq j$ (e.g., $i = \text{nuclear}$ and $j = \text{natural gas}$), we get $p(x_{\text{nuclear}, \text{natural gas}, 0.05}) = 0.12 / (1 - 0.52 \times 0.5) = 0.16$. The new probabilities (i.e., $p(x_{i,j,\Delta p})$) and equation 1 are then used to compute the new entropy $H(x_{\Delta p, j})$ and equation 3 is used to compute the change in entropy:

$$\Delta H_{\Delta p, j} = H(x_{\Delta p, j}) - H(x) \quad (42)$$

where $\Delta H_{\Delta p, j}$ is the difference between $H(x)$ and $H(x_{\Delta p, j})$. As we have nine energy sources for electricity generation, we also have nine different $\Delta H_{\Delta p, j}$ for each Δp . Finally, we sum all $\Delta H_{\Delta p, j}$ for each Δp to compute how robust an electricity grid is if any of the energy sources are disrupted by 5%, 10%, 25%, 50%, 75%, and 100%. We denote the sum of all ΔH across any Δp as ΔH_R , which quantitatively measures the robustness of the energy mix. Table 8 shows the computation of ΔH_R for a disruption of 50% in California (i.e., $\Delta p=0.5$), where the second

column shows the entropy of California in 2015, the third column shows the entropy for new probabilities (i.e., $p(x_{i,j,\Delta p})$), and the last column shows the difference between second and third columns.

As a more distributed energy mix yields higher entropy, the new entropy (i.e. $H(x_{\Delta p, j})$) is higher than the current one ($H(x)$) if a state's energy mix is not well distributed; and thus ΔH_R will be positive. Finally, to compare the ΔH_R from all states with an "ideal" scenario, we compute a ΔH_R with nine equal probabilities (i.e., assuming $p(x) = 1/9$) and denote it as ΔH_{R^*} .

Table 8 Computation of ΔH for a disruption of 50% in California

Energy Source	$H(x)$	$H(x_{\Delta p, j})$	$\Delta H_{\Delta p, j}$
Nuclear	1.53	1.48	-0.055
Hydropower	1.53	1.48	-0.053
Geothermal	1.53	1.48	-0.051
Solar	1.53	1.48	-0.055
Wind	1.53	1.48	-0.051
Biomass	1.53	1.49	-0.045
Natural Gas	1.53	1.78	0.252
ΔH_R			-0.058

7.3 Results and Discussion

Figure 35 shows the evolution of entropy from 1960 to 2015 for all US states. We can notably detect regime shifts (i.e., sharp change in entropy) and steady transitions (i.e., steady change in entropy) for several states. For example, we detect regime shift in Louisiana from 1979 to 1985. The entropy of Louisiana was 0 in 1979 (100% natural gas) and it had increased to 0.82 by 1985 (50% natural gas). Moreover, Louisiana used 6% nuclear energy to generate its electricity in 1985, which had increased to 23% by 2015, thus exhibiting a regime shift. Similar scenarios are also observed for Florida, where the entropy increased from 0.93 (0% nuclear

energy) in 1972 to 1.17 (13% nuclear energy) in 1973; for Illinois, where the entropy increased from 0.41 (1% nuclear energy) in 1969 to 0.84 (23% nuclear energy) in 1974; for Missouri, where the entropy increased from 0.23 (2% nuclear energy) in 1984 to 0.60 (14% nuclear energy) in 1985; for Mississippi, where the entropy increased from 0.72 (1% nuclear energy) in 1984 to 1.03 (34% nuclear energy) in 1987; for New Hampshire, where the entropy increased from 0.59 (0% nuclear energy) in 1988 to 1.30 (40% nuclear energy) in 1990; for New Mexico, where the entropy increased from 0.19 (2% coal and 97% natural gas) in 1962 to 0.73 (41% coal and 58% natural gas) in 1966; for New York, where the entropy increased from 1.12 (36% nuclear energy and 0% natural gas) in 1978 to 1.40 (21% nuclear energy and 21% natural gas) in 1982; and for Pennsylvania, where the entropy increased from 0.16 (0% nuclear energy) in 1973 to 0.63 (16% nuclear energy) in 1975, where we observe an increase in entropy during the regime shift.

On the contrary, we also find states with entropies that decrease. For instance, Utah's entropy decreased from 1.05 in 1968 to 0.42 in 1980, as Utah's usage of coal increased from 40% to 89%, and its usage of hydropower decreased from 40% to 7% from 1968 to 1980. Similar scenarios are observed for Colorado, where the entropy decreased from 0.79 (78% coal and 12% natural gas) in 1980 to 0.38 (90% coal and 2% natural gas) in 1985; for Georgia, where the entropy decreased from 0.94 (60% coal and 15% natural gas) in 1962 to 0.48 (81% coal and 0% natural gas) in 1966; for Massachusetts, where the entropy decreased from 1.45 (39% nuclear energy and 19% natural gas) in 1972 to 0.37 (92% nuclear energy and 2% natural gas) in 1978; for South Dakota, where the entropy decreased from 0.95 (60% hydropower and 23% natural gas) in 1962 to 0.39 (90% hydropower and 4% natural gas) in 1972, and for Tennessee, where the entropy decreased from 0.74 (15% nuclear energy and 74% natural gas) in 1985 to 0.33 (0%

nuclear energy and 90% natural gas) in 1986. Moreover, we also find no or insignificant changes for Kentucky, West Virginia, and Indiana, as these states still depend predominantly on fossil fuel sources. For instance, Kentucky used 85% coal for its electricity generation in 1960, and this proportion had increased to 94% by 2015. Similarly, West Virginia used 98% coal for its electricity generation in 1970, and this proportion had found 95% in 2015, whereas Indiana used 93% coal for its electricity generation in 1970, and this proportion was 82% in 2015. Table 9 shows the maximum increase and decrease in entropy between two consecutive years for each state along with the observed year. While analyzing Table 9, we find that 26 states had a maximum increase of 0 to 0.25, 22 states had a maximum increase of 0.25 to 0.50, and only Rhode Island and Massachusetts had a maximum increase beyond 0.50 in entropy from 1960 to 2015. In contrast, we observe that 39 states had a maximum decrease of 0 to 0.25, 10 states had a maximum decrease of 0.25 to 0.50, and only one state (Massachusetts) had a maximum decrease beyond 0.50 from 1960 to 2015. Additionally, considering the absolute change (i.e., increase or decrease), we find that 28 states out of 50 states had a change of 0.25 or above, suggesting that more than 56% states went through a regime shift in energy mix during the analyzed period. Furthermore, 26 states had their maximum change (i.e., increase or decrease) in between 1968 to 1980, suggesting that most states were affected by the crises of the 1970s.

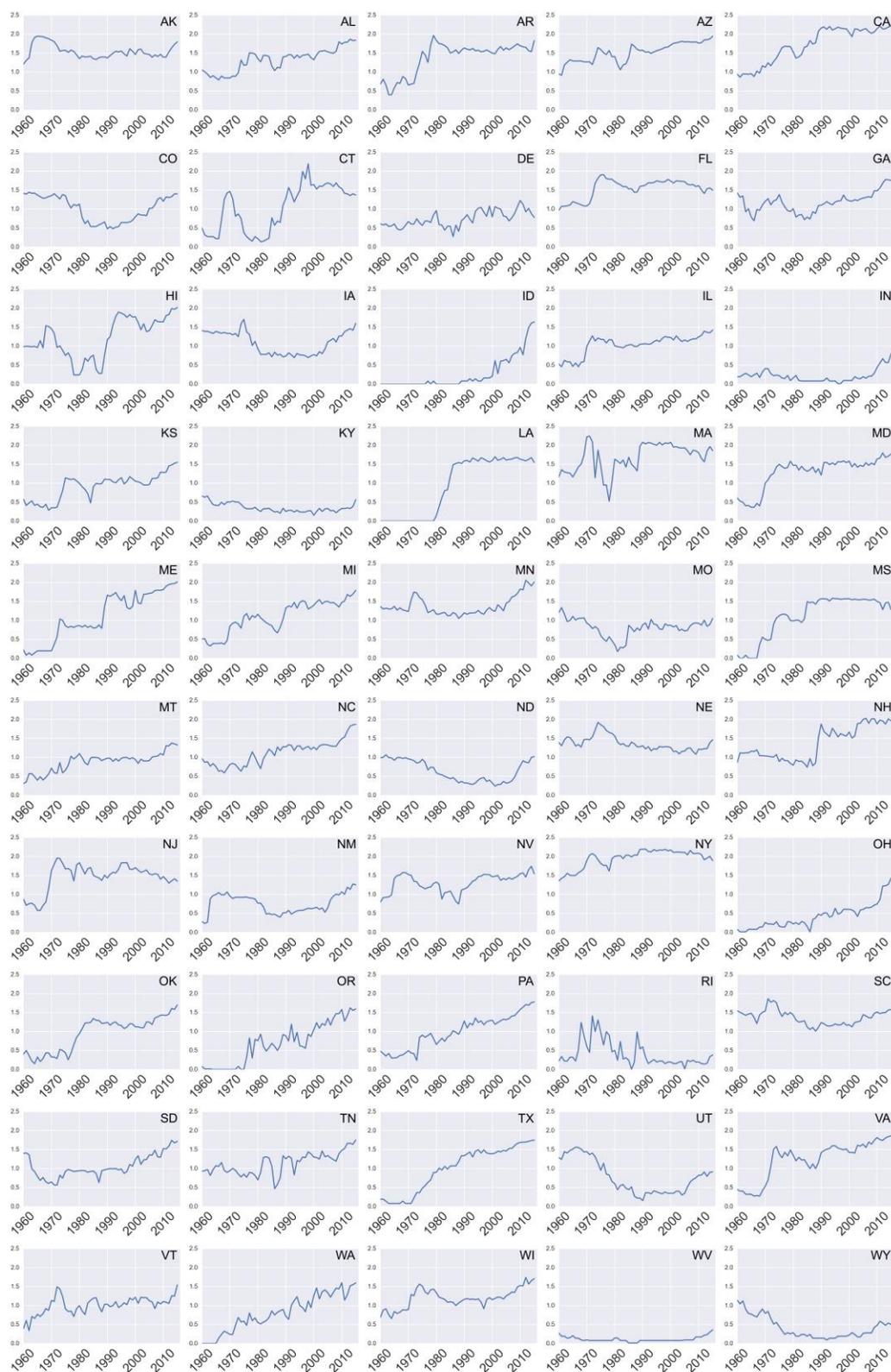


Figure 35 Evolution of entropy from 1960 to 2014 for all states in the USA

Table 9 Maximum Change in Entropy for All States in the USA

State	Maximum Change		State	Maximum Change	
	Increase, Year	Decrease, Year		Increase, Year	Decrease, Year
Alaska	0.28, 1962	-0.1, 1972	Montana	0.2, 1972	-0.19, 1973
Alabama	0.23, 1973	-0.18, 1984	North Carolina	0.18, 1981	-0.12, 1979
Arkansas	0.32, 1977	-0.18, 1962	North Dakota	0.11, 2008	-0.14, 1976
Arizona	0.24, 1985	-0.15, 1980	Nebraska	0.14, 1973	-0.11, 1980
California	0.21, 1988	-0.11, 1980	New Hampshire	0.46, 1988	-0.14, 1990
Colorado	0.13, 2004	-0.25, 1980	New Jersey	0.33, 1969	-0.19, 1977
Connecticut	0.37, 1967	-0.39, 1998	New Mexico	0.43, 1962	-0.13, 1982
Delaware	0.21, 1993	-0.26, 1980	Nevada	0.3, 1964	-0.27, 1981
Florida	0.24, 1972	-0.08, 2010	New York	0.23, 1978	-0.11, 1977
Georgia	0.21, 1966	-0.29, 1962	Ohio	0.25, 2011	-0.11, 1995
Hawaii	0.4, 1967	-0.31, 1977	Oklahoma	0.17, 1972	-0.13, 1975
Iowa	0.24, 1973	-0.23, 1975	Oregon	0.35, 1976	-0.37, 1977
Idaho	0.3, 2011	-0.24, 2001	Pennsylvania	0.42, 1973	-0.13, 1990
Illinois	0.28, 1969	-0.11, 1979	Rhode Island	0.67, 1971	-0.32, 1982
Indiana	0.1, 2009	-0.09, 1978	South Carolina	0.23, 1970	-0.16, 1974
Kansas	0.3, 1984	-0.17, 1983	South Dakota	0.21, 1987	-0.26, 1962
Kentucky	0.09, 1988	-0.09, 1962	Tennessee	0.4, 1988	-0.41, 1985
Louisiana	0.25, 1984	-0.06, 2001	Texas	0.13, 1978	-0.09, 1992
Massachusetts	0.51, 1973	-0.66, 1972	Utah	0.14, 1990	-0.2, 1976
Maryland	0.25, 1969	-0.12, 1989	Virginia	0.31, 1972	-0.14, 1974
Maine	0.4, 1988	-0.23, 2000	Vermont	0.27, 1971	-0.23, 1974
Michigan	0.26, 1969	-0.13, 1997	Washington	0.26, 1991	-0.32, 2010
Minnesota	0.21, 1970	-0.14, 1976	Wisconsin	0.28, 1970	-0.12, 1996
Missouri	0.37, 1984	-0.14, 1962	West Virginia	0.06, 2008	-0.06, 1960
Mississippi	0.31, 1984	-0.13, 2011	Wyoming	0.1, 2008	-0.15, 1962

Table 10 shows the values ΔH_R for a disruption of 5%, 10%, 25%, 50%, and 75% for all states. From Table 10 we observe that the magnitude of a ΔH_R increased exponentially with an increase in disruption. Thus, the magnitude of a ΔH_R for a disruption of 75% is significantly higher than that of a disruption of 5%. To be able to visualize the results, despite these differences in scales, we rank all states for each disruption from 1 (most robust) to 50 (most vulnerable). In particular, we find that the most robust state with a minimum ΔH_R is Alabama for a disruption of 75% (ranked as 1) and the most vulnerable state with a maximum ΔH_R is West Virginia for a disruption of 75% (ranked as 50). Figure 36 shows changes in ranking for robustness (i.e., ΔH_R) for disruptions of 5%, 10%, 25%, 50%, and 75% for all states. Moreover, from Figure 36, we find 13 states with a consistent ranking of 15 or less, 14 states with a ranking of 35 and higher, and 17 states with a ranking between 15 to 35 for all disruptions. From these results, we can classify all US states into three different classes using the ΔH_R value for a disruption of 75%. The classes are:

1. *Vulnerable*: positive ΔH_R , depending predominantly on a single energy source (e.g., Kentucky, West Virginia).
2. *Moderately Robust*: negative ΔH_R but greater than -0.28 or the difference between ΔH_R and ΔH_{R^*} is greater than 0.12, depending on a dominant source for more than 50% of energy grid mix (e.g., California, Massachusetts).
3. *Robust*: negative ΔH_R and less than -0.28 or the difference between ΔH_R and ΔH_{R^*} is less than 0.12, depending on five to six energy sources out of a maximum of nine, but does not have a dominant energy source (e.g., Maine, New York, Pennsylvania).

Table 10 Robustness in Energy Mix for Electricity Generation for All U.S. States

State	5%	10%	25%	50%	75%
Alabama	-0.00086	-0.00356	-0.02514	-0.12708	-0.384
Georgia	-0.00093	-0.00362	-0.02469	-0.12432	-0.37838
Maine	-0.00105	-0.00392	-0.02568	-0.12541	-0.37194
Arizona	-0.00098	-0.00369	-0.02446	-0.12113	-0.36468
Oklahoma	-0.00067	-0.00296	-0.02195	-0.11517	-0.36206
North Carolina	-0.00077	-0.00321	-0.02287	-0.11723	-0.36086
Virginia	-0.00091	-0.00347	-0.02329	-0.11697	-0.3576
Pennsylvania	-0.00073	-0.00305	-0.02182	-0.11298	-0.35274
Arkansas	-0.00073	-0.00306	-0.02175	-0.1117	-0.34643
New York	-0.00071	-0.00297	-0.02121	-0.10976	-0.34259
Tennessee	-0.0007	-0.00292	-0.02089	-0.10836	-0.34077
Texas	-0.00069	-0.00289	-0.02063	-0.10708	-0.33734
Louisiana	-0.00065	-0.00271	-0.01943	-0.10186	-0.32689
Hawaii	-0.00049	-0.00245	-0.01932	-0.10323	-0.32668
Vermont	-0.0006	-0.00252	-0.01823	-0.09696	-0.31665
New Jersey	-0.00052	-0.00218	-0.01624	-0.09115	-0.31539
Maryland	-0.00072	-0.00274	-0.01875	-0.09785	-0.31497
South Dakota	-0.00073	-0.00282	-0.01921	-0.09821	-0.31049
Illinois	-0.00052	-0.0022	-0.01627	-0.08997	-0.30712
Minnesota	-0.00064	-0.00266	-0.01888	-0.09695	-0.30347
Alaska	-0.00058	-0.00242	-0.01732	-0.0907	-0.29193
Michigan	-0.00054	-0.00228	-0.01644	-0.08728	-0.28531
Connecticut	-0.00043	-0.00188	-0.01422	-0.08045	-0.28211
Iowa	-0.00047	-0.00196	-0.01434	-0.07849	-0.26747
Kansas	-0.00052	-0.00207	-0.01454	-0.07718	-0.25824

State	5%	10%	25%	50%	75%
New Hampshire	-0.00032	-0.00172	-0.01399	-0.07676	-0.25471
Montana	-0.0004	-0.00168	-0.01246	-0.07044	-0.25211
Florida	-0.0004	-0.00169	-0.01228	-0.06701	-0.23279
South Carolina	-0.00045	-0.00178	-0.0125	-0.06744	-0.23223
Idaho	-0.00037	-0.00155	-0.01125	-0.06135	-0.21385
Oregon	-0.00041	-0.00161	-0.01124	-0.0609	-0.21317
Mississippi	-0.00034	-0.00142	-0.01038	-0.05783	-0.20997
Nebraska	-0.0003	-0.00126	-0.00941	-0.054	-0.20017
California	-0.00037	-0.00154	-0.01098	-0.05805	-0.19531
Ohio	-0.00028	-0.00117	-0.00864	-0.04876	-0.18125
Wisconsin	-0.00035	-0.00127	-0.00839	-0.04566	-0.16718
New Mexico	-0.00015	-0.00077	-0.00648	-0.03993	-0.16116
Colorado	-0.00022	-0.00093	-0.007	-0.0405	-0.15679
Massachusetts	-0.00018	-0.00075	-0.00569	-0.03392	-0.13506
Washington	0.000129	0.000423	0.002344	0.007212	-0.0122
Nevada	0.000126	0.000518	0.003483	0.013913	0.009444
North Dakota	0.00016	0.000667	0.004671	0.02166	0.039422
Utah	0.000383	0.001208	0.006905	0.03099	0.069777
Indiana	0.000406	0.001224	0.006739	0.030336	0.070742
Missouri	0.000256	0.001066	0.007567	0.037381	0.093025
Delaware	3.01E-05	0.000639	0.006848	0.039523	0.11629
Wyoming	0.000304	0.00128	0.00942	0.052093	0.177554
Kentucky	0.000333	0.001399	0.010277	0.056482	0.190004
Rhode Island	0.000305	0.001286	0.009516	0.053847	0.198236
West Virginia	0.000338	0.001425	0.010555	0.060239	0.229654

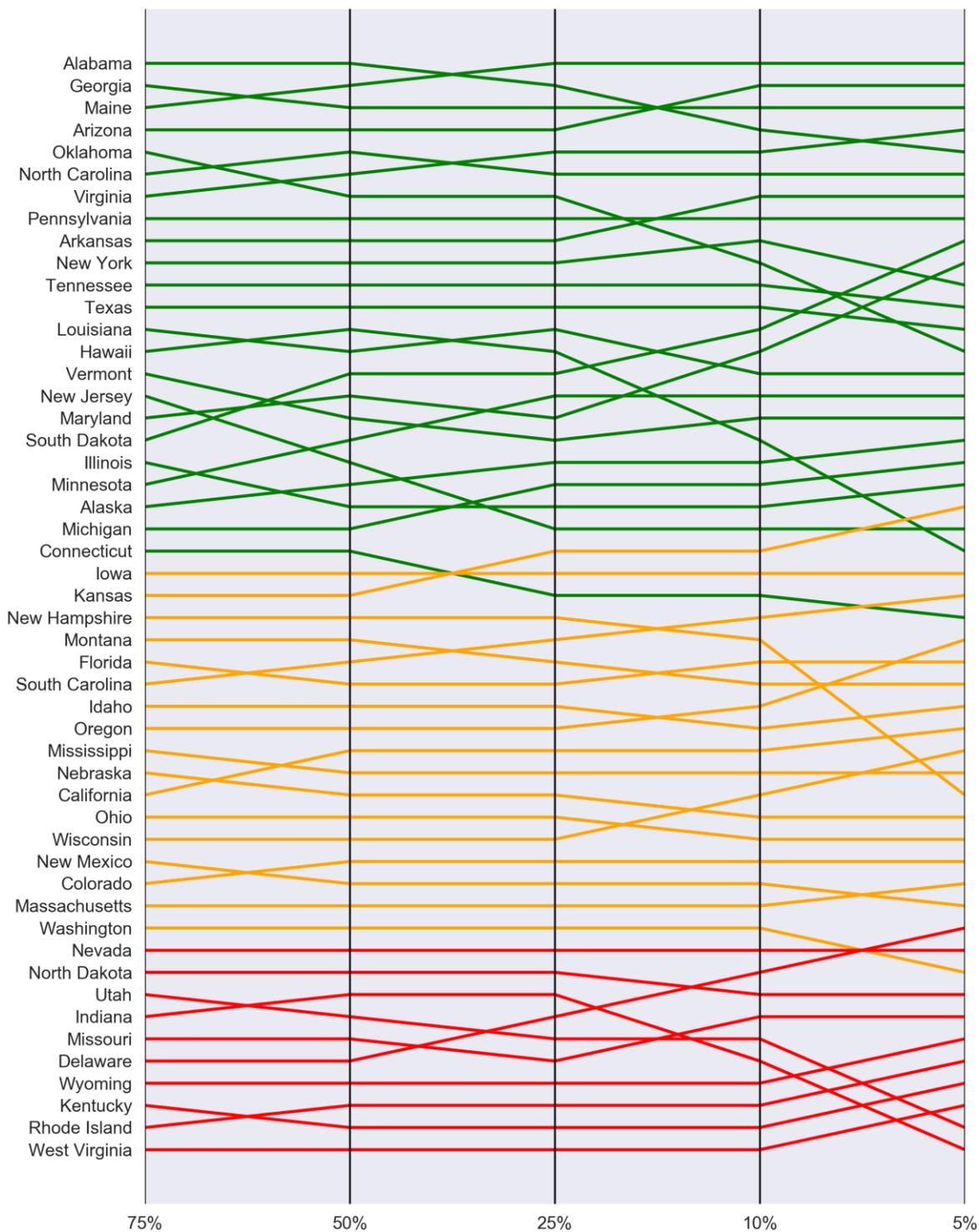


Figure 36 Robustness Ranking in Energy Mix for Electricity Generation for All U.S. States

Table 11 show the three classes of states, where we find 10 states as vulnerable (i.e. $\Delta H_R > 0$), 17 states as moderately robust (i.e. $-0.28 < \Delta H_R < 0$) and the remaining 23 states as robust (i.e. $\Delta H_R < -0.28$). From Figure 36 we can also observe that while using a high percentage of fossil fuels (see Figure 34), some states are classified as robust thanks to their diverse energy mix. For example, Texas used 78% fossil fuels to generate its electricity in 2015, but the share of coal is 34% and natural gas is 44%. We find that 13 states (i.e., Georgia, Arkansas, Pennsylvania, Oklahoma, Tennessee, Minnesota, Virginia, Louisiana, Maryland, Arizona, Michigan, North Carolina, and Alabama) achieve a ΔH_R of less than -0.28 although their fossil fuel usage is greater than 50%. In contrast, despite using a high percentage of non-fossil fuel sources, some states are detected as moderately robust as their energy mix is not well distributed. For example, Washington state used only 16% fossil fuels in 2015, but 68% of its energy mix comes from hydroelectric power. Finally, we detect 7 states (i.e., South Dakota, Illinois, Vermont, Connecticut, Maine, New York, and New Jersey) with a fossil fuel share of less than 50% and $\Delta H_R < -0.28$.

Table 11 Classification of States Based on Robustness in Energy Mix

Robustness	States	Count
Vulnerable	Kentucky, Missouri, West Virginia, Indiana, Nevada, North Dakota, Delaware, Rhode Island, Utah, and Wyoming	10
Moderately Robust	Nebraska, New Mexico, California, Ohio, Wisconsin, Kansas, Colorado, Massachusetts, Florida, South Carolina, New Hampshire, Washington, Oregon, Mississippi, Montana, Idaho, and Iowa	17
Robust	South Dakota, Georgia, Arkansas, Pennsylvania, Oklahoma, Tennessee, Illinois, Vermont, Minnesota, Virginia, Louisiana, Connecticut, Maryland, Maine, New York, Arizona, New Jersey, Texas, Michigan, Alabama, North Carolina, Hawaii, and Alaska	23

7.4 Conclusion

Traditionally, the United States have relied heavily on fossil fuels for electricity generation, although several crises have had substantial impacts on the energy mix used. As the global reserves of fossil fuels are depleting rapidly (Hoel and Kverndokk, 1996; Höök and Tang, 2013; Mohr et al., 2015; Shafiee and Topal, 2009) and as climate change and weather patterns in general introduce uncertainties in how much renewable energy can be used for electricity generation, tracking and assessing the robustness of the energy mix in US states for electricity generation is increasingly important. In this study, we measured the evolution of the energy mix of the 50 US states using Shannon entropy. In particular, we showed that sudden and significant changes in the energy mix are not rare as the US was affected by the economic and energy crises of the 1970s, and as a result, most states had to significantly change their energy mix. Following the same logic, we can posit that similar crises and events can happen in the future. The rise of

new technologies such as the smart grid and microgrids are also likely to have a substantial impact. Moreover, we also evaluated the robustness of the 50 US states under multiple energy source disruption scenarios, and we found 10 states as vulnerable, 17 states as moderately robust, and the remaining 23 states as robust. However, among the 23 robust states, we detect only 7 states (i.e., South Dakota, Illinois, Vermont, Connecticut, Maine, New York, and New Jersey) with a fossil fuel share of less than 50%. Naturally, this analysis omits many important variables to be able to fully assess the robustness of the 50 US states electricity grid. For starter, we do not include operating elements such as flexible versus intermittent electricity generation, power plant age and efficiency. We also do not account for electricity transmission and distribution (that are often the cause of blackouts). Moreover, to classify the states, we only consider robustness in energy mix, and we do not take into account elements of sustainability and environmental emissions that possess feedback effects on the robustness of the electricity grid. Nonetheless, this study shows that many states have gone through regime shifts in how they generate electricity suggests, and having a more diverse energy mix can improve the robustness of the electricity grid. In fact, in that respect, we also observed that some states are performing much better than others and they should therefore be better able to withstand several external shocks in the supply of their energy sources.

8. Conclusion

The main goal of this dissertation was to analyze the evolution of energy and resource consumption trends in the world and its effect on environment and economy. To achieve this goal, I used per capita human, natural, and produced capital for all the countries in the world from 1990 to 2014. Data was collected from the data from the 2014 and 2017 Inclusive Wealth Report (UNU-IHDP and UNEP, 2014; Urban Institute and UNEP, 2017). Detecting current trends in energy and resource consumption is vital if we aspire to become more sustainable and develop effective policies. Traditional statistical indicators are extremely useful, but they can fail to capture patterns of the complex systems around us. I first pointed their limitations by calculating several measures and noticing contradictory and biased results, especially owing to the presence of outliers or errors in the data sets.

I then formulated a new methodology, inspired from the concepts of homophily in network science. Essentially, by comparing every single value of a dataset with every other, the methodology links two values when they are similar to each other. Technically close to the construction of a network, the node with the highest degree is assimilated to the mode of the distribution. Moreover, I use the proportional number of nodes in the largest cluster to determine the cutoff for the optimal network. Thanks to its nature, outliers have little to no impact on the metrics calculated. This feature is particularly important since current datasets often have significant outliers that are hard to detect manually (e.g., from sensor data that generate millions of points). Moreover, the method can be applied to data sets of any size. Moreover, borrowing the semantics from nuclear physics and using the notion of ‘distance’ in networks, an attempt was made to find the discrete relative position of any country from the mode, which is defined as ‘orbital position.’ Then the orbital positions of each country were plotted over time in a polar

format to generate orbital diagrams for each country and each indicator analyzed (i.e., energy consumption, CO₂ emission and GDP). Finally, the total “travelled” distances for all the countries were computed by taking the summation of the changes in orbital positions from the previous year and subsequently orbital speeds are computed by dividing the travelled distances by number of available data points within the timeline. From the results, I find an encouraging and optimistic trend for human capital, likely thanks to major global initiatives that aim to significantly encourage socio-economic development in developing countries like ‘Agenda 21’ (United Nations Division for Sustainable Development, 1992) and the Millennium Development Goals (UNDP, 2000). Meanwhile, the trends for natural and produced capital are rapidly decreasing and require immediate attention. Moreover, projections of current trends to 2050 further demonstrate the decline of natural resources. Orbital diagrams, orbital speeds, and heat maps depicting current and projected trends for all the countries for all the indicators analyzed are provided in the appendix section.

Furthermore, to assess the stability and sustainability of a system, the concept of Fisher Information was adapted and used. An open source Python library has been developed to compute FI and is shown in the appendix. Then, first, the code was applied to a simple case study to evaluate the evolution of global-mean temperature from 1880 to 2015. After that, the code was applied to assess the stability and sustainability of the public transit systems of the urbanized areas (UZA) in the US. For this analysis data was collected from the National Transit Database for unlinked passenger trips and vehicle revenue miles. A total of 372 UZAs were analyzed and eight distinct patterns were uncovered: 1) regime shift with rebound, 2) regime shift with partial rebound 3) regime shift without rebound 4) decrease with rebound 5) decrease with partial rebound 6) decrease without rebound 7) increase and 8) no pattern/others are found

for the evolution of fisher information. Most of UZAs failed to show any definite pattern and fall into the last category (i.e., no pattern/others).

Finally, I applied Shannon entropy to track the evolution of energy mix in US states and found that most of the states are affected by the oil crises in 1970s. Moreover, I also use Shannon entropy to evaluate the robustness of US states energy mix based on the hypothesis that a more diverse energy mix is generally desirable from a robustness viewpoint. From the robustness analysis results, I found three different classes of states, which are 1) vulnerable, 2) moderately robust, and 3) robust. Among the robust states only 7 states are found relatively environmentally friendly as they use less than 50% fossil fuel to generate their electricity.

Overall, this work presented in this dissertation provided a new means to better understand how energy and resources are being consumed in the world. Most importantly, it provided a new methodology to study trends in the evolution of any data over time, forming the main contribution of my work. Moreover, the application of information theory to analyze the evolution and robustness of energy mix is novel and possesses potential to extract more information in the general goal to live in a more sustainable and resilient world.

8.1 Future Work

As mentioned earlier, a novel network based methodology has been developed, which provides significant potential to supplement traditional analysis. The methodology is currently able to handle a single indicator only (e.g., human capital). It can therefore be further developed to increase its ability and handle multiple indicators at once by concatenating networks from individual variables to generate a multi-layer network. Besides, I only applied the NFA method

to global scale with only 140 nodes (i.e., countries), but it can be applied to much larger scale with thousands of nodes (e.g., census tracts).

Furthermore, Fisher Information was computed for public transit systems in the US, providing significant information about those systems. As Fisher Information is able to handle multiple indicators simultaneously, it can thus be applied to assess the stability or sustainability of any system.

Finally, the robustness analysis in chapter 7 omits many important variables to be able to fully assess the robustness of the 50 US states electricity grid. For starter, we do not include operating elements such as flexible versus intermittent electricity generation, power plant age and efficiency. We also do not account for electricity transmission and distribution (that are often the cause of blackouts). Moreover, to classify the states, we only consider robustness in energy mix, and we do not take into account elements of sustainability and environmental emissions that possess feedback effects on the robustness of the electricity grid. Thus, all these can be incorporated for an comprehensive robustness analysis.

9. References

- Adams, J., Maslin, M., Thomas, E., 1999. Sudden climate transitions during the Quaternary. *Prog. Phys. Geogr.* 23, 1–36.
- Ahmad, N., Derrible, S., 2015. Evolution of Public Supply Water Withdrawal in the USA: A Network Approach: Water Withdrawal in the USA: A Network Approach. *J. Ind. Ecol.* 19, 321–330. <https://doi.org/10.1111/jiec.12266>
- Ahmad, N., Derrible, S., Cabezas, H., 2017a. Using Fisher information to assess stability in the performance of public transportation systems. *R. Soc. Open Sci.* 4, 160920. <https://doi.org/10.1098/rsos.160920>
- Ahmad, N., Derrible, S., Cabezas, H., 2017b. Using Fisher Information to Assess Stability in the Performance of Public Transportation Systems [WWW Document]. URL https://figshare.com/projects/Using_Fisher_Information_to_Assess_Stability_in_the_Performance_of_Public_Transportation_Systems/20041 (accessed 4.17.17).
- Ahmad, N., Derrible, S., Eason, T., Cabezas, H., 2016a. Using Fisher information to track stability in multivariate systems. *R. Soc. Open Sci.* 3, 160582. <https://doi.org/10.1098/rsos.160582>
- Ahmad, N., Derrible, S., Eason, T., Cabezas, H., 2016b. Fisher Information [WWW Document]. URL <https://github.com/csunlab/fisher-information> (accessed 9.28.15).
- Aitken, C.K., McMahon, T.A., Wearing, A.J., Finlayson, B.L., 1994. Residential Water Use: Predicting and Reducing Consumption1. *J. Appl. Soc. Psychol.* 24, 136–158. <https://doi.org/10.1111/j.1559-1816.1994.tb00562.x>
- Allcott, H., 2011. Social norms and energy conservation. *J. Public Econ.* 95, 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>
- Allcott, H., Mullainathan, S., 2010. Behavior and energy policy. *Science* 327, 1204–1205. <https://doi.org/10.1126/science.1180775>
- APTA, 2015. Record 10.8 Billion Trips Taken On U.S. Public Transportation In 2014 [WWW Document]. URL http://www.apta.com/mediacenter/pressreleases/2015/Pages/150309_Ridership.aspx (accessed 10.22.15).
- Arianos, S., Bompard, E., Carbone, A., Xue, F., 2009. Power grid vulnerability: A complex network approach. *Chaos Interdiscip. J. Nonlinear Sci.* 19, 013119. <https://doi.org/10.1063/1.3077229>
- Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., Levin, S., Mäler, K.-G., Schneider, S., Starrett, D., others, 2004. Are we consuming too much? *J. Econ. Perspect.* 18, 147–172.
- Arrow, K.J., Dasgupta, P., Goulder, L.H., Mumford, K.J., Oleson, K., 2012. Sustainability and the measurement of wealth. *Environ. Dev. Econ.* 17, 317–353.
- Arthur, W.B., Holland, J.H., LeBaron, B., Palmer, R.G., Tayler, P., 1996. Asset pricing under endogenous expectations in an artificial stock market. Available SSRN 2252.
- Baccini, P., 1997. A city's metabolism: Towards the sustainable development of urban systems. *J. Urban Technol.* 4, 27–39. <https://doi.org/10.1080/10630739708724555>
- Bader, B.P.-P., Baccini, P., 1996. *Regionaler Stoffhaushalt; Erfassung, Bewertung und Steuerung.* Berlin7 Heidelb.

- Baird, M., Stammer Jr., R., 2000. Measuring the Performance of State Transportation Agencies: Three Perspectives. *Transp. Res. Rec. J. Transp. Res. Board* 1729, 26–34. <https://doi.org/10.3141/1729-04>
- Bak, P., 2013. *How nature works: the science of self-organized criticality*. Springer Science & Business Media.
- Bak, P., Creutz, M., 1994. Fractals and self-organized criticality, in: *Fractals in Science*. Springer, pp. 27–48.
- Bak, P., Tang, C., Wiesenfeld, K., 1988. Self-organized criticality. *Phys. Rev. A* 38, 364–374.
- Barabási, A.-L., 2002. *Linked: the new science of networks*. Perseus Books Group New York.
- Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *science* 286, 509–512.
- Barro, R.J., Lee, J.W., 2013. A new data set of educational attainment in the world, 1950–2010. *J. Dev. Econ.* 104, 184–198. <https://doi.org/10.1016/j.jdeveco.2012.10.001>
- Ben-Haim, Y., 2012. Why Risk Analysis is Difficult, and Some Thoughts on How to Proceed: **Perspective**. *Risk Anal.* 32, 1638–1646. <https://doi.org/10.1111/j.1539-6924.2012.01859.x>
- Berk, R.A., Cooley, T.F., LaCivita, C.J., Parker, S., Sredl, K., Brewer, M., 1980. Reducing consumption in periods of acute scarcity: The case of water. *Soc. Sci. Res.* 9, 99–120. [https://doi.org/10.1016/0049-089X\(80\)90001-0](https://doi.org/10.1016/0049-089X(80)90001-0)
- Bertini, R., El-Geneidy, A., 2003. Generating Transit Performance Measures with Archived Data. *Transp. Res. Rec. J. Transp. Res. Board* 1841, 109–119. <https://doi.org/10.3141/1841-12>
- Bettencourt, L.M.A., Kaur, J., 2011. Evolution and structure of sustainability science. *Proc. Natl. Acad. Sci.* 108, 19540–19545. <https://doi.org/10.1073/pnas.1102712108>
- Bin, S., Dowlatabadi, H., 2005. Consumer lifestyle approach to US energy use and the related CO2 emissions. *Energy Policy* 33, 197–208. [https://doi.org/10.1016/S0301-4215\(03\)00210-6](https://doi.org/10.1016/S0301-4215(03)00210-6)
- Boltzmann, L., 1872. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, *Sitzungs. Akad. Wiss. Wein* 66 (1872), 275–370; English: Further Studies on the Thermal Equilibrium of Gas Molecules. *Kinet. Theory* 2, 88–174.
- Brandon, G., Lewis, A., 1999. Reducing household energy consumption: A qualitative and quantitative field study. *J. Environ. Psychol.* 19, 75–85.
- Brunner, P.H., Rechberger, H., 2004. Practical handbook of material flow analysis. *Int. J. Life Cycle Assess.* 9, 337–338.
- Bureau of Labor Statistics [WWW Document], 2018. URL <https://data.bls.gov/pdq/SurveyOutputServlet> (accessed 1.18.18).
- Burton, I., 1987. Report on reports: Our common future: The world commission on environment and development. *Environ. Sci. Policy Sustain. Dev.* 29, 25–29.
- Cabezas, H., Fath, B.D., 2002. Towards a theory of sustainable systems. *Fluid Phase Equilibria* 194, 3–14.
- Cao, Y., Chen, X., Sun, K., 2006. Identification of vulnerable lines in power grid based on complex network theory [J]. *Electr. Power Autom. Equip.* 12, 1–5.
- Carlson, L., Bassett, G., Buehring, W., Collins, M., Folga, S., Haffenden, B., Petit, F., Phillips, J., Verner, D., Whitfield, R., 2012. *Resilience: theory and applications*. Argonne National Laboratory. Argonne, Illinois, USA.

- Carreras, B., Newman, D., Dobson, I., Poole, A., 2001. Evidence for self-organized criticality in electric power system blackouts, in: 34th Hawaii International Conference on System Sciences, Maui, Hawaii.
- Carter, C.R., Rogers, D.S., 2008. A framework of sustainable supply chain management: moving toward new theory. *Int. J. Phys. Distrib. Logist. Manag.* 38, 360–387.
- Casti, J.L., 1997. *Would-be worlds: How simulation is changing the frontiers of science*. N. Y.
- Chassin, D.P., Posse, C., 2005. Evaluating North American electric grid reliability using the Barabási–Albert network model. *Phys. Stat. Mech. Its Appl.* 355, 667–677.
<https://doi.org/10.1016/j.physa.2005.02.051>
- Chu, D., Strand, R., Fjelland, R., 2003. Theories of complexity. *Complexity* 8, 19–30.
<https://doi.org/10.1002/cplx.10059>
- Clark, W.A., Finley, J.C., 2008. Household water conservation challenges in Blagoevgrad, Bulgaria: a descriptive study. *Water Int.* 33, 175–188.
<https://doi.org/10.1080/02508060802023264>
- Clark, W.C., Dickson, N.M., 2003. Sustainability science: The emerging research program. *Proc. Natl. Acad. Sci.* 100, 8059–8061. <https://doi.org/10.1073/pnas.1231333100>
- Corral-Verdugo, V., FRÍAS-ARMENTA, M., PÉREZ-URIAS, F., ORDUÑA-CABRERA, V., Espinoza-Gallego, N., 2002. Residential Water Consumption, Motivation for Conserving Water and the Continuing Tragedy of the Commons. *Environ. Manage.* 30, 527–535. <https://doi.org/10.1007/s00267-002-2599-5>
- Cottrill, C.D., Derrible, S., 2015. Leveraging Big Data for the Development of Transport Sustainability Indicators. *J. Urban Technol.* 22, 45–64.
<https://doi.org/10.1080/10630732.2014.942094>
- Cramer, A., Cucarese, J., Tran, M., Lu, A., Reddy, A., 2009. Performance Measurements on Mass Transit: Case Study of New York City Transit Authority. *Transp. Res. Rec. J. Transp. Res. Board* 2111, 125–138. <https://doi.org/10.3141/2111-15>
- Crutzen, P.J., 2002. Geology of mankind. *Nature* 415, 23–23. <https://doi.org/10.1038/415023a>
- Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal Complex Syst.* 1695.
- CTA, 2017. CTA Performance Metrics & Reports [WWW Document]. Transit Chic. URL <http://www.transitchicago.com/performance/default.aspx> (accessed 11.10.16).
- Daily, B.F., Huang, S., 2001. Achieving sustainability through attention to human resource factors in environmental management. *Int. J. Oper. Prod. Manag.* 21, 1539–1552.
<https://doi.org/10.1108/01443570110410892>
- Dandy, G., Nguyen, T., Davies, C., 1997. Estimating Residential Water Demand in the Presence of Free Allowances. *Land Econ.* 73, 125. <https://doi.org/10.2307/3147082>
- Dasgupta, P., Duraiappah, A., Managi, S., Barbier, E., Collins, R., Fraumeni, B., Gundimeda, H., Liu, G., Mumford, K., 2015. How to measure sustainable progress. *Science* 350, 748–748.
- Decker, E.H., Elliott, S., Smith, F.A., Blake, D.R., Rowland, F.S., 2003. Energy and material flow through the urban ecosystem.
- DeFries, R., 2014. *The Big Ratchet: How Humanity Thrives in the Face of Natural Crisis*. Basic Books.
- Delmelle, E., Zhou, Y., Thill, J.-C., 2014. Densification without Growth Management? Evidence from Local Land Development and Housing Trends in Charlotte, North Carolina, USA. *Sustainability* 6, 3975–3990. <https://doi.org/10.3390/su6063975>

- Dempster, A.P., 1967. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Stat.* 38, 325–339. <https://doi.org/10.1214/aoms/1177698950>
- Derrible, S., 2016. Urban infrastructure is not a tree: Integrating and decentralizing urban infrastructure systems. *Environ. Plan. B Plan. Des.* <https://doi.org/10.1177/0265813516647063>
- Derrible, Sybil, 2016. Complexity in future cities: the rise of networked infrastructure. *Int. J. Urban Sci.* 1–19. <https://doi.org/10.1080/12265934.2016.1233075>
- Derrible, S., 2015. Class Notes of “Cities and Sustainable Infrastructure”.
- Derrible, S., Ahmad, N., 2015. Network-Based and Binless Frequency Analyses. *PLOS ONE* 10, e0142108. <https://doi.org/10.1371/journal.pone.0142108>
- Derrible, S., Kennedy, C., 2009. Network Analysis of World Subway Systems Using Updated Graph Theory. *Transp. Res. Rec. J. Transp. Res. Board* 2112, 17–25. <https://doi.org/10.3141/2112-03>
- Derrible, S., Saneinejad, S., Sugar, L., Kennedy, C., 2010. Macroscopic model of greenhouse gas emissions for municipalities. *Transp. Res. Rec. J. Transp. Res. Board* 174–181.
- Dietz, T., Stern, P.C., Weber, E.U., 2013. Reducing carbon-based energy consumption through changes in household behavior. *Daedalus* 142, 78–89. https://doi.org/10.1162/DAED_a_00186
- Ding, M., Han, P., 2008. Vulnerability assessment to small-world power grid based on weighted topological model. *Proc.-Chin. Soc. Electr. Eng.* 28, 20.
- Ding, M., Han, P., 2006. Small-world topological model based vulnerability assessment algorithm for large-scale power grid [J]. *Autom. Electr. Power Syst.* 8, 002.
- Dooley, J.J., 1998. Energy R&D in the United States (No. PNNL-12188, 15001470). <https://doi.org/10.2172/15001470>
- Druckman, A., Jackson, T., 2008. Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. *Energy Policy* 36, 3177–3192. <https://doi.org/10.1016/j.enpol.2008.03.021>
- Easley, D., Kleinberg, J., 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.
- Eason, T., Cabezas, H., 2012. Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado. *J. Environ. Manage.* 94, 41–49. <https://doi.org/10.1016/j.jenvman.2011.08.003>
- Eason, T., Garmestani, A.S., 2012. Cross-scale dynamics of a regional urban system through time. *Reg. Dev.* 36, 55–77.
- Eason, T., Garmestani, A.S., Cabezas, H., 2014. Managing for resilience: early detection of regime shifts in complex systems. *Clean Technol. Environ. Policy* 16, 773–783. <https://doi.org/10.1007/s10098-013-0687-2>
- Eason, T., Garmestani, A.S., Stow, C.A., Rojo, C., Alvarez-Cobelas, M., Cabezas, H., 2016. Managing for resilience: an information theory-based approach to assessing ecosystems. *J. Appl. Ecol.* n/a-n/a. <https://doi.org/10.1111/1365-2664.12597>
- Edwards, P., 1969. *The encyclopedia of philosophy*.
- Ehrlich, P.R., Ehrlich, A.H., 1990. *The population explosion*. Simon and Schuster New York.
- EIA, 2017. U.S. Energy Information Administration [WWW Document]. US Energy Inf. Adm. URL <https://www.eia.gov/state/seds/seds-data-complete.php?sid=US> (accessed 4.20.17).
- Elkington, J., 2004. Enter the triple bottom line. *Triple Bottom Line Does It Add Up* 11, 1–16.

- Elkington, J., 1998. *Cannibals with Forks: The Triple Bottom Line of the 21st Century*. Oxford, Royaume-Uni, Capstone.
- Engel-Yan, J., Cuddihy, J., Kennedy, C., 2007. The changing metabolism of cities. *J. Ind. Ecol.* 11, 43–59.
- Epstein, J.M., 1999. Agent-based computational models and generative social science. *Gener. Soc. Sci. Stud. Agent-Based Comput. Model.* 4, 4–46.
- Epstein, J.M., Axtell, R., 1996. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.
- Erdős, P., Rényi, A., 1960. On the evolution of random graphs. *Publ Math Inst Hung. Acad Sci* 5, 17–61.
- Erdős, P., Rényi, A., 1959. On random graphs I. *Publ Math Debr.* 6, 290–297.
- Estanqueiro, A., 2010. The future energy mix paradigm: How to embed large amounts of wind generation while preserving the robustness and quality of the power systems?, in: *Wind Power*. InTech.
- Fabbricatore, C., Boley, H., Karduck, A.P., 2012. Machine learning for resource management in smart environments. *IEEE Int. Conf. Digit. Ecosyst. Technol.* <https://doi.org/10.1109/DEST.2012.6227910>
- Fath, B.D., Cabezas, H., 2004. Exergy and Fisher information as ecological indices. *Ecol. Model.* 174, 25–35.
- Fath, Brian D., Cabezas, H., Pawlowski, C.W., 2003. Regime changes in ecological systems: an information theory approach. *J. Theor. Biol.* 222, 517–530.
- Fath, B.D., Cabezas, H., Pawlowski, C.W., 2003. Regime changes in ecological systems: An information theory approach. *J. Theor. Biol.* 222, 517–530. [https://doi.org/10.1016/S0022-5193\(03\)00067-5](https://doi.org/10.1016/S0022-5193(03)00067-5)
- Feenstra, R.C., Inklaar, R., Timmer, M.P., 2013. Penn world table version 8.0. *Gener. Penn World Table Available Download Www Ggdc Netpwt*.
- Person, S., 2002. *RAMAS Risk Calc 4.0 software: risk assessment with uncertain numbers*. CRC press.
- Fischer, C., 2008. Feedback on household electricity consumption: A tool for saving energy? *Energy Effic.* 1, 79–104. <https://doi.org/10.1007/s12053-008-9009-7>
- Fisher, R.A., 1922. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 222, 309–368. <https://doi.org/10.1098/rsta.1922.0009>
- Frieden, B.R., Gatenby, R.A., 2007. *Exploratory Data Analysis Using Fisher Information*.
- Frieden, R., 1998. *Physics from Fisher information: a unification*. Cambridge University Press.
- Fung, M., Kennedy, C.A., 2005. An integrated macroeconomic model for assessing urban sustainability. *Environ. Plan. B* 32, 639.
- GISTEMP Team, 2016. NASA GISS surface temperature (GISTEMP) analysis. *Trends Compend. Data Glob. Change*.
- Gladwin, T.N., Kennelly, J.J., Krause, T.-S., 1995. Shifting Paradigms for Sustainable Development: Implications for Management Theory and Research. *Acad. Manage. Rev.* 20, 874. <https://doi.org/10.2307/258959>
- Global Footprint Network, 2010. *Ecological footprint atlas 2010*. Retrieved May 25, 2014.
- Goñcz, E., Skirke, U., Kleizen, H., Barber, M., 2007. Increasing the rate of sustainable change: a call for a redefinition of the concept and the model for its implementation. *J. Clean. Prod.* 15, 525–537. <https://doi.org/10.1016/j.jclepro.2006.05.018>

- Gonzalez Mejia, A.M., 2011. Fisher Information-Sustainability Analysis of Several US Metropolitan Statistical Areas. University of Cincinnati.
- Gonzalez-Mejia, A.M., Eason, T., Cabezas, H., Suidan, M.T., 2012. Computing and interpreting Fisher Information as a metric of sustainability: Regime changes in the United States air quality. *Clean Technol. Environ. Policy* 14, 775–788. <https://doi.org/10.1007/s10098-011-0445-2>
- González-Mejía, A.M., Eason, T.N., Cabezas, H., Suidan, M.T., 2014. Social and economic sustainability of urban systems: comparative analysis of metropolitan statistical areas in Ohio, USA. *Sustain. Sci.* 9, 217–228. <https://doi.org/10.1007/s11625-013-0227-3>
- Gonzalez-Mejía, A.M., Eason, T.N., Cabezas, H., Suidan, M.T., 2012. Assessing Sustainability in Real Urban Systems: The Greater Cincinnati Metropolitan Area in Ohio, Kentucky, and Indiana. *Environ. Sci. Technol.* 46, 9620–9629. <https://doi.org/10.1021/es3007904>
- Gonzalez-Mejia, A.M., Vance, L., Eason, T., Cabezas, H., 2015. Recent developments in the application of Fisher information to sustainable environmental management, in: Klemes, J.J. (Ed.), *Assessing and Measuring Environmental Impact and Sustainability*. Butterworth-Heinemann, pp. 25–72.
- Gößling-Reisemann, S., 2008. What Is Resource Consumption and How Can It Be Measured?: Theoretical Considerations. *J. Ind. Ecol.* 12, 10–25. <https://doi.org/10.1111/j.1530-9290.2008.00012.x>
- Grant, M., Plaskon, T., Trainor, S., Suter, S., 2011. State DOT Public Transportation Performance Measures: State of the Practice and Future Needs. Transportation Research Board, Washington, DC.
- Grassi, M., Nucci, M., Piazza, F., 2011. Towards an ontology framework for intelligent smart home management and energy saving. Presented at the Proceedings - ISIE 2011: 2011 IEEE International Symposium on Industrial Electronics, pp. 1753–1758. <https://doi.org/10.1109/ISIE.2011.5984327>
- Hamilton, J.D., 2011. Historical oil shocks. National Bureau of Economic Research.
- Harberger, A.C., 1988. Perspectives on capital and technology in less-developed countries. *Estud. Econ. Chile*.
- Hart, S., 1995. A Natural-Resource-Based View of the Firm. *Acad. Manage. Rev.* 20, 986–1014. <https://doi.org/10.2307/258963>
- Henderson, G., Kwong, P., Adkins, H., 1991. Regularity indices for evaluating transit performance.
- Hendriks, C., Obernosterer, R., Müller, D., Kytzia, S., Baccini, P., Brunner, P.H., 2000. Material flow analysis: a tool to support environmental policy decision making. Case-studies on the city of Vienna and the Swiss lowlands. *Local Environ.* 5, 311–328.
- Henriques, A., Richardson, J., 2013. *The triple bottom line: Does it all add up*. Routledge.
- Hill, M.R., 2001. Sustainability, greenhouse gas emissions and international operations management. *Int. J. Oper. Prod. Manag.* 21, 1503–1520. <https://doi.org/10.1108/EUM0000000006292>
- Hoel, M., Kverndokk, S., 1996. Depletion of fossil fuels and the impacts of global warming. *Resour. Energy Econ.* 18, 115–136.
- Holland, J.H., 2014. *Complexity: A very short introduction*. Oxford University Press.
- Holland, J.H., 1999. Echoing emergence: Objectives, rough definitions, and speculations for echo-class models, in: *Complexity*. Perseus Books, pp. 309–342.
- Holland, J.H., 1995. *Hidden order: How adaptation builds complexity*. Basic Books.

- Holmgren, A. A., 2006. Using graph models to analyze the vulnerability of electric power networks. *Risk Anal.* 26, 955–969.
- Höök, M., Tang, X., 2013. Depletion of fossil fuels and anthropogenic climate change—A review. *Energy Policy* 52, 797–809.
- Hosking, J.R., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B Methodol.* 105–124.
- Hraber, P.T., Jones, T., Forrest, S., 1997. The ecology of Echo. *Artif. Life* 3, 165–190.
- Hutson, S.S., 2004. Estimated use of water in the United States in 2000. Geological Survey (USGS).
- Ingwersen, W., Cabezas, H., Weisbrod, A.V., Eason, T., Demeke, B., Ma, X., Hawkins, T.R., Lee, S.-J., Bare, J.C., Ceja, M., 2014. Integrated Metrics for Improving the Life Cycle Approach to Assessing Product System. *Sustainability* 6, 1386–1413. <https://doi.org/10.3390/su6031386>
- Jennings, P.D., Zandbergen, P.A., 1995. Ecologically Sustainable Organizations: An Institutional Approach. *Acad. Manage. Rev.* 20, 1015. <https://doi.org/10.2307/258964>
- Johnson, D.G., 2000. Population, Food, and Knowledge. *Am. Econ. Rev.* 90, 1–14. <https://doi.org/10.1257/aer.90.1.1>
- Jorgensen, B., Graymore, M., O’Toole, K., 2009. Household water use behavior: An integrated model. *J. Environ. Manage.* 91, 227–236. <https://doi.org/10.1016/j.jenvman.2009.08.009>
- Karduni, A., Kermanshah, A., Derrible, S., 2016. A protocol to convert spatial polyline data to network formats and applications to world urban road networks. *Sci. Data* 3, 160046. <https://doi.org/10.1038/sdata.2016.46>
- Karunanithi, A.T., Cabezas, H., Frieden, B.R., Pawlowski, C.W., 2008. Detection and assessment of ecosystem regime shifts from fisher information. *Ecol. Soc.* 13.
- Kates, R.W., 2001. ENVIRONMENT AND DEVELOPMENT: Sustainability Science. *Science* 292, 641–642. <https://doi.org/10.1126/science.1059386>
- Kates, R.W., Parris, T.M., 2003. Long-term trends and a sustainability transition. *Proc. Natl. Acad. Sci.* 100, 8062–8067.
- Kendall, M.G., Gibbons, J.D., 1990. Rank correlation methods, 5th ed. ed. E. Arnold ; Oxford University Press, London : New York, NY.
- Kennedy, C., Pincetl, S., Bunje, P., 2011. The study of urban metabolism and its applications to urban planning and design. *Environ. Pollut.* 159, 1965–1973.
- Kennedy, C.A., Stewart, I., Facchini, A., Cersosimo, I., Mele, R., Chen, B., Uda, M., Kansal, A., Chiu, A., Kim, K.-G., Dubeux, C., La Rovere, E.L., Cunha, B., Pincetl, S., Keirstead, J., Barles, S., Pusaka, S., Gunawan, J., Adegbile, M., Nazariha, M., Hoque, S., Marcotullio, P.J., Otharan, F.G., Genena, T., Ibrahim, N., Farooqui, R., Cervantes, G., Sahin, A.D., 2015. Energy and material flows of megacities. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5985–5990. <https://doi.org/10.1073/pnas.1504315112>
- Kenny, J.F., Barber, N.L., Hutson, S.S., Linsey, K.S., Lovelace, J.K., Maupin, M.A., 2009. Estimated use of water in the United States in 2005. US Geological Survey Reston, VA.
- Kermanshah, A., Derrible, S., 2017. Robustness of road systems to extreme flooding: using elements of GIS, travel demand, and network science. *Nat. Hazards* 86, 151–164. <https://doi.org/10.1007/s11069-016-2678-1>
- Kermanshah, A., Derrible, S., 2016. A geographical and multi-criteria vulnerability assessment of transportation networks against extreme earthquakes. *Reliab. Eng. Syst. Saf.* 153, 39–49. <https://doi.org/10.1016/j.res.2016.04.007>

- Kermanshah, Derrible, 2017. Robustness of road systems to extreme flooding: using elements of GIS, travel demand, and network science. *Nat. Hazards* 86, 151–164.
<https://doi.org/10.1007/s11069-016-2678-1>
- Kerr, R.A., 1998. GEOLOGY: The Next Oil Crisis Looms Large--and Perhaps Close. *Science* 281, 1128–1131. <https://doi.org/10.1126/science.281.5380.1128>
- King, R.G., Levine, R., 1994. Capital fundamentalism, economic development, and economic growth. *Carnegie-Rochester Conf. Ser. Public Policy* 40, 259–292.
[https://doi.org/10.1016/0167-2231\(94\)90011-6](https://doi.org/10.1016/0167-2231(94)90011-6)
- Kittelsohn, Associates, et al., 2003. TCRP Report 88: A Guidebook for Developing a Transit Performance- Measurement System. Transit Cooperative Research Program of the National Academies, Washington, DC.
- Klenow, P.J., Rodríguez-Clare, A., 1997. The Neoclassical Revival in Growth Economics: Has It Gone Too Far? *NBER Macroecon. Annu.* 12, 73–103. <https://doi.org/10.1086/654324>
- Klir, G.J., 2005. *Uncertainty and information: foundations of generalized information theory.* John Wiley & Sons.
- Lal, R., Hansen, D.O., Uphoff, N., 2002. *Food security and environmental quality in the developing world.* CRC Press.
- Lapin, L.L., 1980. *Statistics: meaning and method.* Harcourt Brace Jovanovich New York.
- Levin, S., Clark, W., 2010. *Toward a science of sustainability: report from toward a science of sustainability conference.* Princet. Environ. Inst. Princet.
- Lewis, S.L., Maslin, M.A., 2015. Defining the Anthropocene. *Nature* 519, 171–180.
<https://doi.org/10.1038/nature14258>
- Liao, C.-F., Liu, H., 2010. Development of Data-Processing Framework for Transit Performance Analysis. *Transp. Res. Rec. J. Transp. Res. Board* 2143, 34–43.
<https://doi.org/10.3141/2143-05>
- Mann, H.B., 1945. Nonparametric tests against trend. *Econom. J. Econom. Soc.* 245–259.
- Margolis, R.M., Kammen, D.M., 1999. Evidence of Under-investment in Energy R&D in the United States and the Impact of Federal Policy. *Energy Policy* 27, 575–584.
- Martell, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience. *Fish Fish.* 14, 504–514. <https://doi.org/10.1111/j.1467-2979.2012.00485.x>
- Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., Linsey, K.S., 2014. *Estimated Use of Water in the United States in 2010 (Circular).*
- Mayer, A.L., Pawlowski, C., Fath, B.D., Cabezas, H., 2007. Applications of Fisher Information to the management of sustainable environmental systems, in: *Exploratory Data Analysis Using Fisher Information.* Springer, pp. 217–244.
- Mayer, A.L., Pawlowski, C.W., Cabezas, H., 2006. Fisher Information and dynamic regime changes in ecological systems. *Ecol. Model.* 195, 72–82.
<https://doi.org/10.1016/j.ecolmodel.2005.11.011>
- Mayer, Audrey L., Pawlowski, C.W., Cabezas, H., 2006. Fisher Information and dynamic regime changes in ecological systems. *Ecol. Model.* 195, 72–82.
<https://doi.org/10.1016/j.ecolmodel.2005.11.011>
- McKinney, W., 2010. Data structures for statistical computing in Python, in: *Proceedings of the 9th.* pp. 51–56.
- McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* 415–444.

- Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.W., 1972. The limits to growth. N. Y. 102.
- Mei, S., Zhang, X., Cao, M., 2011. Power Grid Complexity. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-16211-4>
- Miranda, E., 2008. A survey of the theory of coherent lower previsions. *Int. J. Approx. Reason.* 48, 628–658. <https://doi.org/10.1016/j.ijar.2007.12.001>
- Mohr, S.H., Wang, J., Ellem, G., Ward, J., Giurco, D., 2015. Projection of world fossil fuels by country. *Fuel* 141, 120–135. <https://doi.org/10.1016/j.fuel.2014.10.030>
- Narayanan, B., Walmsley, T., 2008. Global trade, assistance, and production: the GTAP 8 data base, center for global trade analysis. West Lafayette Purdue Univ.
- Newman, M., 2010. *Networks: An Introduction*, 1 edition. ed. Oxford University Press, USA.
- Newman, M.E.J., 1996. Self-organized criticality, evolution and the fossil extinction record. *Proc. R. Soc. Lond. B Biol. Sci.* 263, 1605–1610.
- Newman, M.E.J., Sneppen, K., 1996. Avalanches, scaling, and coherent noise. *Phys. Rev. E* 54, 6226.
- NTD, 2016. National Transit Database [WWW Document]. URL <http://www.ntdprogram.gov/ntdprogram/> (accessed 7.2.15).
- Pagani, G.A., Aiello, M., 2013. The Power Grid as a complex network: A survey. *Phys. Stat. Mech. Its Appl.* 392, 2688–2700. <https://doi.org/10.1016/j.physa.2013.01.023>
- Park, H.-C., Heo, E., 2007. The direct and indirect household energy requirements in the Republic of Korea from 1980 to 2000—An input–output analysis. *Energy Policy* 35, 2839–2851. <https://doi.org/10.1016/j.enpol.2006.10.002>
- Park, J., Seager, T.P., Rao, P.S.C., Convertino, M., Linkov, I., 2013. Integrating Risk and Resilience Approaches to Catastrophe Management in Engineering Systems: Perspective. *Risk Anal.* 33, 356–367. <https://doi.org/10.1111/j.1539-6924.2012.01885.x>
- Parris, T.M., Kates, R.W., 2003. Characterizing and measuring sustainable development. *Annu. Rev. Environ. Resour.* 28, 559–586. <https://doi.org/10.1146/annurev.energy.28.050302.105551>
- Parzen, E., 1962. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* 33, 1065–1076. <https://doi.org/10.1214/aoms/1177704472>
- Pawlowski, C., Fath, B.D., Mayer, A.L., Cabezas, H., 2005. Towards a sustainability index using information theory. *Energy* 30, 1221–1231. <https://doi.org/10.1016/j.energy.2004.02.008>
- Pearson, K., 1895. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans. R. Soc. Lond. A* 343–414.
- Peiravian, F., Derrible, S., Ijaz, F., 2014. Development and application of the Pedestrian Environment Index (PEI). *J. Transp. Geogr.* 39, 73–84.
- Pérez-Lombard, L., Ortiz, J., Pout, C., 2008. A review on buildings energy consumption information. *Energy Build.* 40, 394–398. <https://doi.org/10.1016/j.enbuild.2007.03.007>
- Petit, F., Verner, D., Brannegan, D., Buehring, W., Dickinson, D., Guziel, K., Haffenden, R., Phillips, J., Peerenboom, J., 2015. Analysis of Critical Infrastructure Dependencies and Interdependencies (No. ANL/GSS--15/4, 1184636). <https://doi.org/10.2172/1184636>
- Polasky, S., Bryant, B., Hawthorne, P., Johnson, J., Keeler, B., Pennington, D., 2015. Inclusive wealth as a metric of sustainable development. *Annu. Rev. Environ. Resour.* 40, 445–466.

- Poortinga, W., Steg, L., Vlek, C., 2004. Values, Environmental Concern, and Environmental Behavior: A Study into Household Energy Use. *Environ. Behav.* 36, 70–93. <https://doi.org/10.1177/0013916503251466>
- Pratt, R., Lomax, T., 1996. Performance Measures for Multimodal Transportation Systems. *Transp. Res. Rec. J. Transp. Res. Board* 1518, 85–93. <https://doi.org/10.3141/1518-15>
- Raudsepp-Hearne, C., Peterson, G.D., Tengö, M., Bennett, E.M., Holland, T., Benessaiah, K., MacDonald, G.K., Pfeifer, L., 2010. Untangling the Environmentalist's Paradox: Why Is Human Well-being Increasing as Ecosystem Services Degrade? *BioScience* 60, 576–589. <https://doi.org/10.1525/bio.2010.60.8.4>
- Rees, W.E., 1992. Ecological footprints and appropriated carrying capacity: what urban economics leaves out. *Environ. Urban.* 4, 121–130.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. *Nature* 461, 472–475. <https://doi.org/10.1038/461472a>
- Rosenblatt, M., 1956. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* 27, 832–837. <https://doi.org/10.1214/aoms/1177728190>
- Sarkis, J., 2001. Manufacturing's role in corporate environmental sustainability - Concerns for the new millennium. *Int. J. Oper. Prod. Manag.* 21, 666–686. <https://doi.org/10.1108/01443570110390390>
- Savitz, A.W., Weber, K., 2006. *The triple bottom line*. San Franc. Jossey-Boss.
- Schurr, S.H., 1983. Energy efficiency and economic efficiency: an historical perspective. *Energy Product. Econ. Growth Schurr Sonebulum Wood Eds Oelgeschlager Gunn Hain Publ.*
- Scott, D.W., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics. Wiley.
- Shafer, G., others, 1976. *A mathematical theory of evidence*. Princeton university press Princeton.
- Shafiee, S., Topal, E., 2009. When will fossil fuel reserves be diminished? *Energy Policy* 37, 181–189. <https://doi.org/10.1016/j.enpol.2008.08.016>
- Shannon, C.E., 2001. *A mathematical theory of communication*. ACM SIGMOBILE Mob. Comput. Commun. Rev. 5, 3–55.
- Shastri, Y., Diwekar, U., Cabezas, H., 2008. Optimal control theory for sustainable environmental management. *Environ. Sci. Technol.* 42, 5322–5328.
- Shields, R., 2012. Cultural topology: The seven bridges of Königsburg, 1736. *Theory Cult. Soc.* 29, 43–57.
- Shrivastava, P., 1995. The role of corporations in achieving ecological sustainability. *Acad. Manage. Rev.* 20, 936–960.
- Sikdar, S.K., 2003. Sustainable development and sustainability metrics. *AIChE J.* 49, 1928–1932. <https://doi.org/10.1002/aic.690490802>
- Simon, H.A., 1965. The architecture of complexity. *Gen. Syst.* 10, 63–76.
- Smith, R.M., Bedau, M.A., 2000. Is Echo a complex adaptive system? *Evol. Comput.* 8, 419–442.
- Solley, W.B., Merk, C.F., Pierce, R.R., 1988. *Estimated Use of Water in the United States in 1985 (No. CIR-1004)*. United States Geological Survey.

- Solley, W.B., Pierce, R.R., Perlman, H.A., 1998. Estimated use of water in the United States in 1995 (No. CIR-1200). United States Geological Survey.
- Solley, W.B., Pierce, R.R., Perlman, H.A., 1993. Estimated use of water in the United States in 1990 (No. CIR-1081). United States Geological Survey.
- Spanbauer, T.L., Allen, C.R., Angeler, D.G., Eason, T., Fritz, S.C., Garmestani, A.S., Nash, K.L., Stone, J.R., 2014. Prolonged instability prior to a regime shift. *PloS ONE*.
- Starik, M., Rands, G.P., 1995. Weaving an Integrated Web: Multilevel and Multisystem Perspectives of Ecologically Sustainable Organizations. *Acad. Manage. Rev.* 20, 908. <https://doi.org/10.2307/258960>
- Steffen, W., Crutzen, P.J., McNeill, J.R., 2007. The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. *AMBIO J. Hum. Environ.* 36, 614–621. [https://doi.org/10.1579/0044-7447\(2007\)36\[614:TAAHNO\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2007)36[614:TAAHNO]2.0.CO;2)
- Stenudd, S., 2010. Using machine learning in the adaptive control of a smart environment. *vtffi*.
- Sturges, H.A., 1926. The choice of a class interval. *J. Am. Stat. Assoc.* 21, 65–66.
- Sun, K., 2005. Complex networks theory: A new method of research in power grid, in: *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES. IEEE*, pp. 1–6.
- Swetz, F.J., 2012. *The Search for Certainty: A Journey Through the History of Mathematics, 1800-2000*. Courier Corporation.
- Taylor, K., 1999. Rapid climate change. *Am. Sci.* 87, 320–327.
- Theil, H., 1967. *Economics and information theory*. North-Holland Amsterdam.
- Theis, T., Tomkin, J. (Eds.), 2011. *Sustainability: A Comprehensive Foundation*. OpenStax CNX.
- UNDP, 2016. | Human Development Reports [WWW Document]. URL <http://hdr.undp.org/en/2016-report> (accessed 2.20.18).
- UNDP, 2000. Millennium Development Goals [WWW Document]. URL http://www.undp.org/content/undp/en/home/sdgoverview/mdg_goals.html (accessed 1.26.16).
- United Nations, 2017. World’s population increasingly urban with more than half living in urban areas | UN DESA | United Nations Department of Economic and Social Affairs.
- United Nations Division for Sustainable Development, 1992. Agenda 21 [WWW Document]. URL (accessed 10.5.16).
- UNU-IHDP, UNEP, 2014. *Inclusive Wealth Report 2014*. Cambridge University Press.
- Urban Institute, UNEP, 2017. *Inclusive Wealth Report 2017, Working Draft*.
- US EPA, 2010. San Luis Basin sustainability metrics project: A methodology for evaluating regional sustainability. US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Washington, DC.
- van der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* 13, 22–30. <https://doi.org/10.1109/MCSE.2011.37>
- Vringer, K., Aalbers, T., Blok, K., 2007. Household energy requirement and value patterns. *Energy Policy* 35, 553–566. <https://doi.org/10.1016/j.enpol.2005.12.025>
- Vuchic, V.R., 2005. *Urban transit: operations, planning and economics*. J. Wiley & Sons, Hoboken, N.J.
- Wackernagel, M., Rees, W., 1996. Our Ecological Footprint. *Green Teach.* 45, 5–14.

- Wackernagel, M., Schulz, N.B., Deumling, D., Linares, A.C., Jenkins, M., Kapos, V., Monfreda, C., Loh, J., Myers, N., Norgaard, R., Randers, J., 2002. Tracking the ecological overshoot of the human economy. *Proc. Natl. Acad. Sci.* 99, 9266–9271.
<https://doi.org/10.1073/pnas.142033699>
- Wald, A., 1945. Statistical decision functions which minimize the maximum risk. *Ann. Math.* 265–280.
- Wall, R., Crosbie, T., 2009. Potential for reducing electricity demand for lighting in households: An exploratory socio-technical study. *Energy Policy* 37, 1021–1031.
<https://doi.org/10.1016/j.enpol.2008.10.045>
- Walley, P., 1996. Measures of uncertainty in expert systems. *Artif. Intell.* 83, 1–58.
[https://doi.org/10.1016/0004-3702\(95\)00009-7](https://doi.org/10.1016/0004-3702(95)00009-7)
- Wang, Q.J., 1996. Direct sample estimators of L moments. *Water Resour. Res.* 32, 3617–3619.
- Wasserman, S., Galaskiewicz, J., 1994. *Advances in social network analysis: Research in the social and behavioral sciences.* Sage Publications.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. *nature* 393, 440–442.
- Weijermars, R., Taylor, P., Bahn, O., Das, S.R., Wei, Y.-M., 2012. Review of models and actors in energy mix optimization – can leader visions and decisions align with optimum model strategies for our future energy systems? *Energy Strategy Rev.* 1, 5–18.
<https://doi.org/10.1016/j.esr.2011.10.001>
- Whiteman, G., Cooper, W.H., 2000. Ecological embeddedness. *Acad. Manage. J.* 43, 1265–1282.
- Wolman, A., 1965. The metabolism of cities. *Sci. Am.* 213, 179–190.
- World Bank, 2015. *The World Population Prospects: 2015 Revision (Report).*
- World Bank, 2010. *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium.* The World Bank.
- Xu, M., Allenby, B.R., Crittenden, J.C., 2011. Interconnectedness and Resilience of the U.S. Economy. *Adv. Complex Syst.* 14, 649–672.
<https://doi.org/10.1142/S0219525911003335>
- Zadeh, L.A., 1965. Fuzzy sets. *Inf. Control* 8, 338–353.

10. Appendix A: Orbital Diagrams, Orbital Speed and Heat Maps

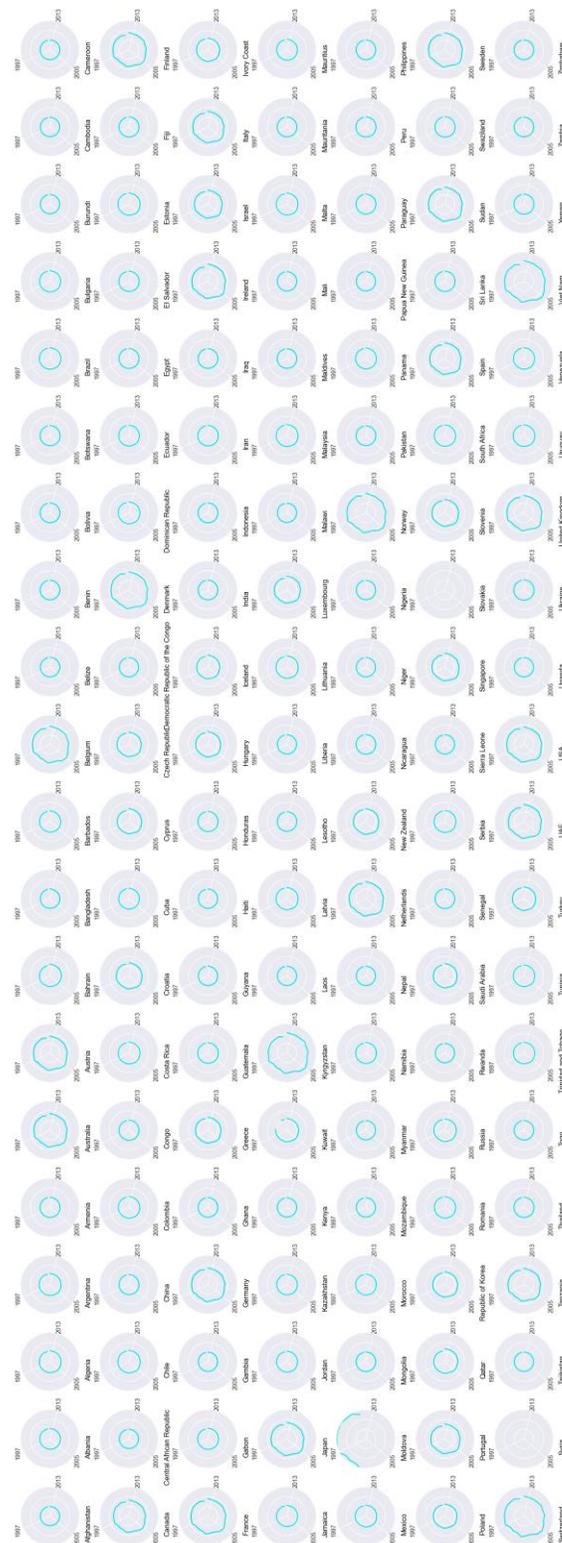


Figure S1 Orbital Diagrams for HC

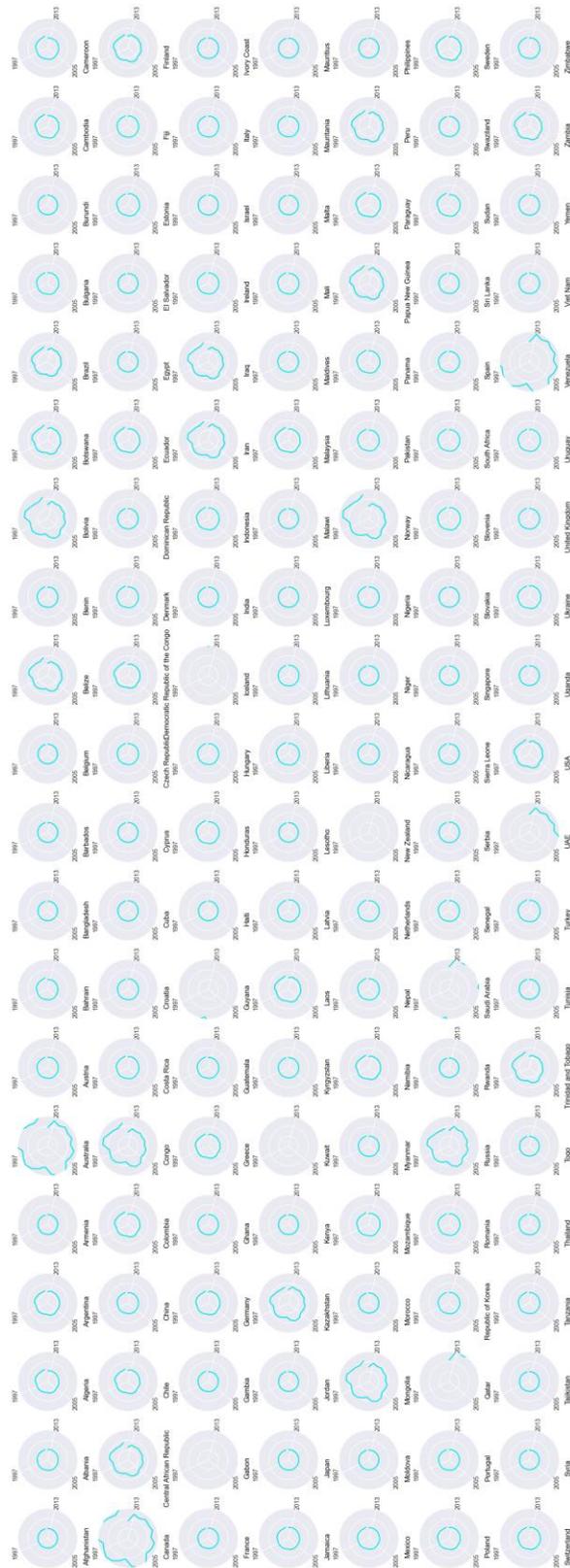


Figure S2 Orbital Diagrams for NC

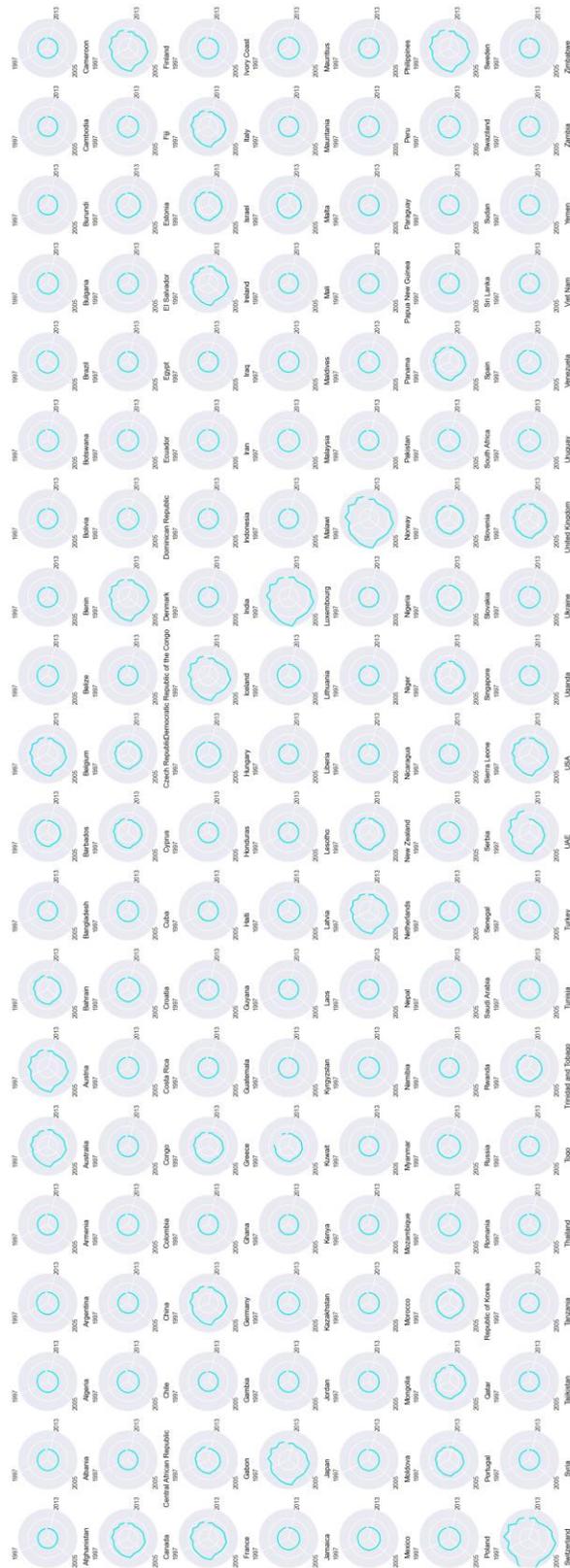


Figure S3 Orbital Diagrams for PC

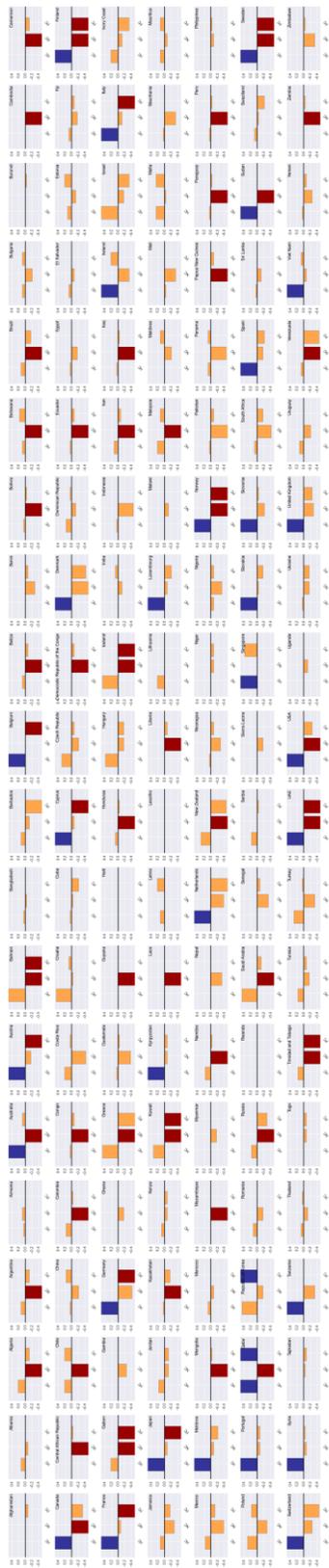


Figure S4 Orbital Speeds from 1990 to 2014

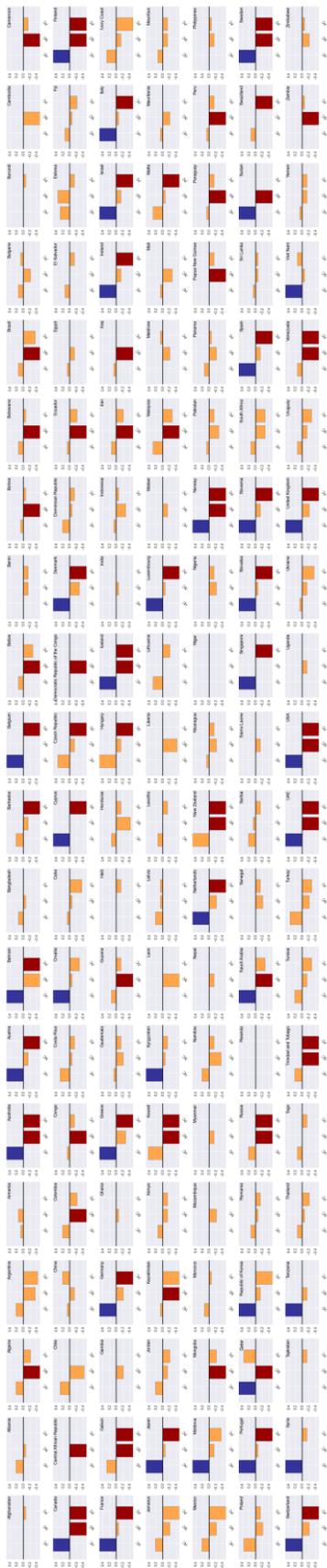


Figure S5 Orbital Speeds from 2000 to 2014

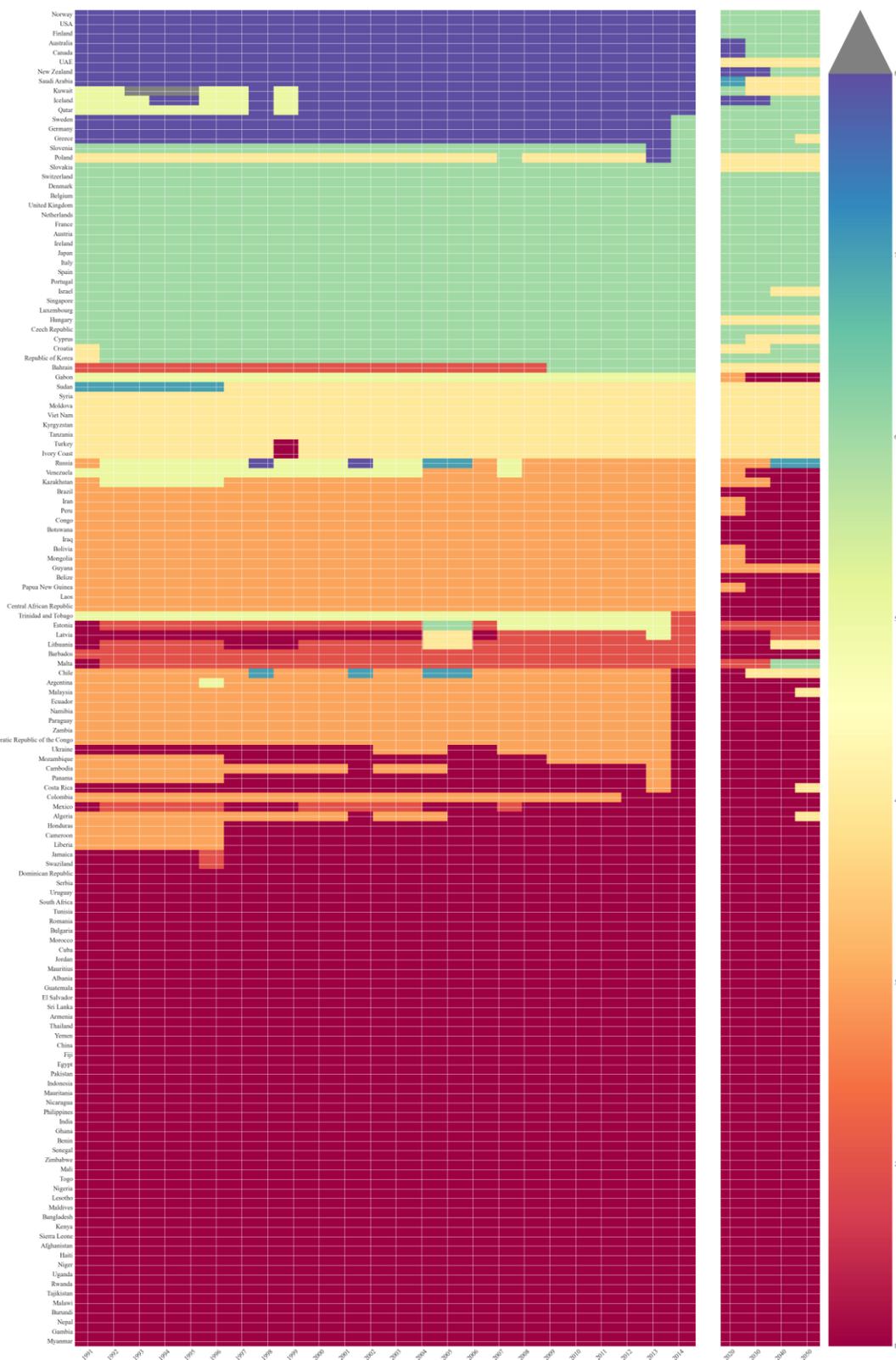


Figure S6 Evolution of Groups from 1990 to 2014 and Projections from 2020 to 2050 Using Orbital Speeds from 1990 to 2014

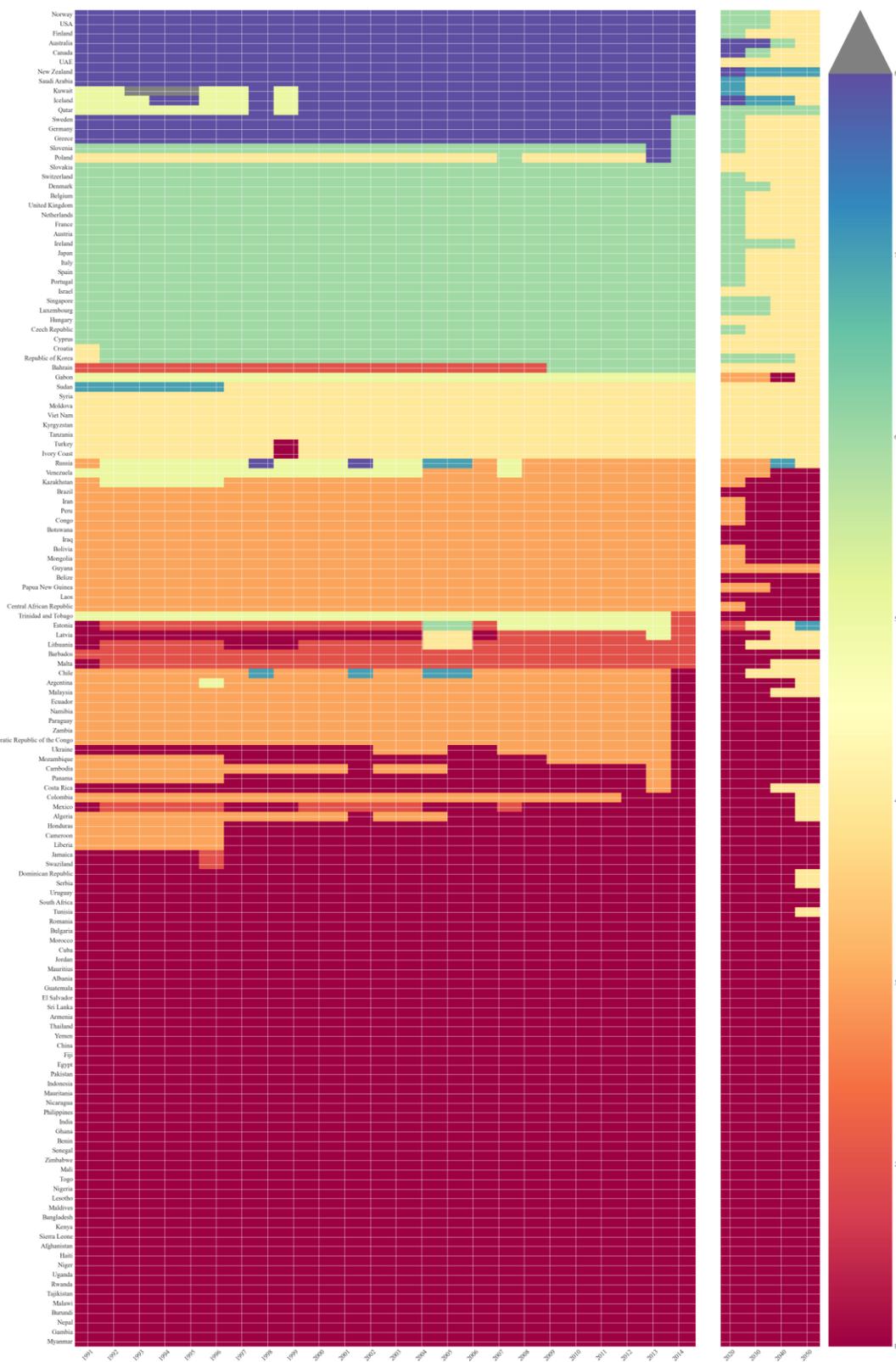
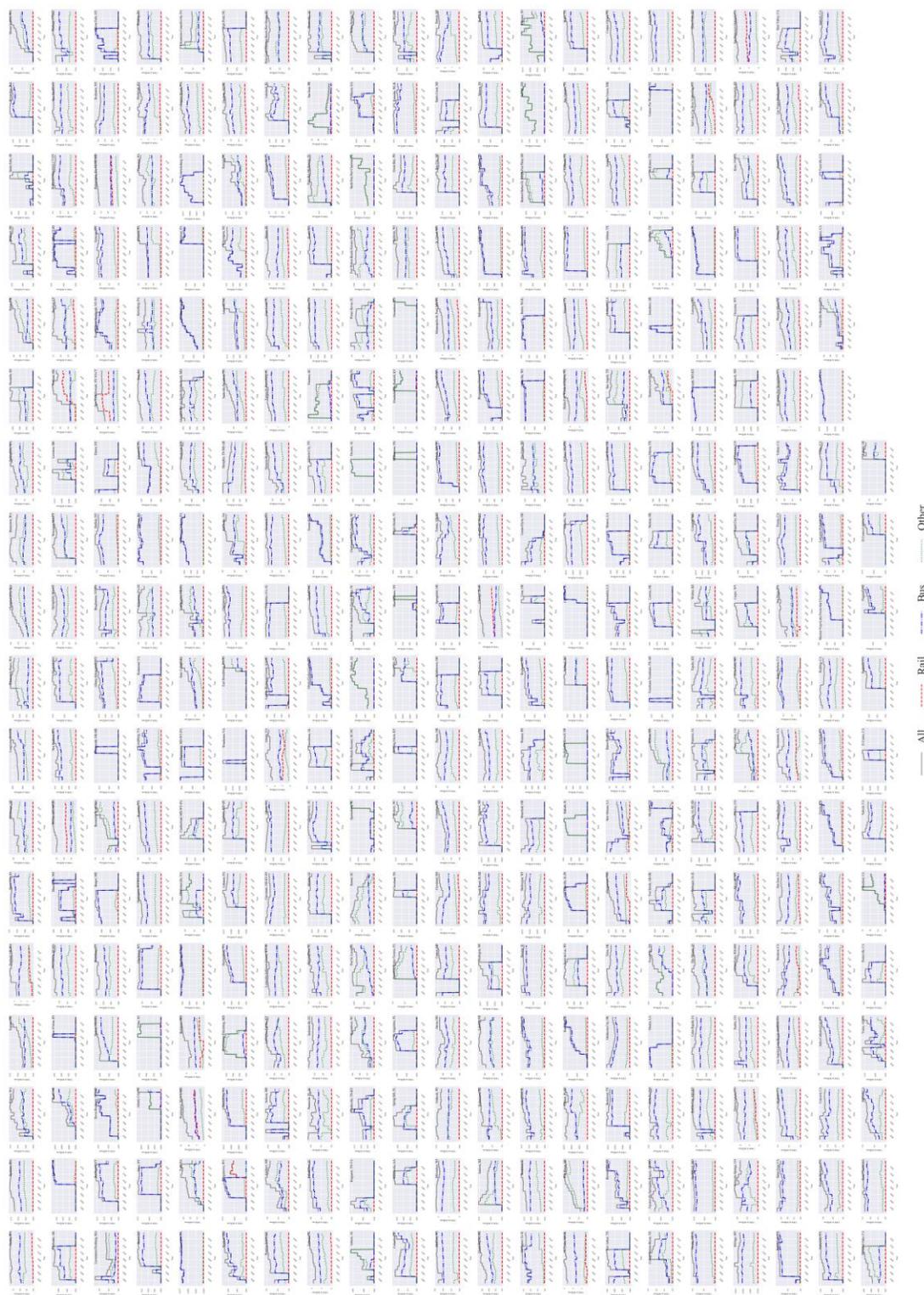


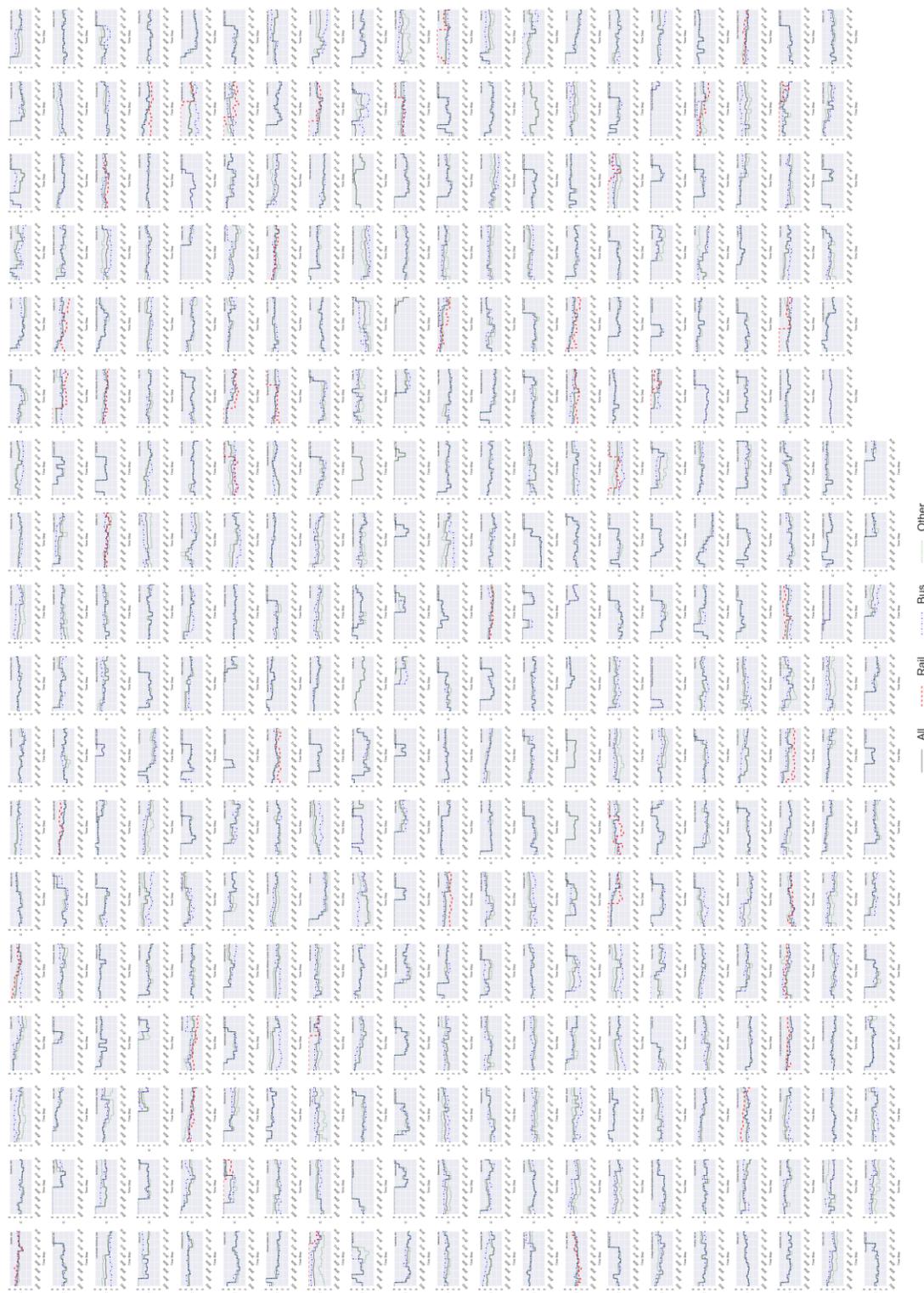
Figure S7 Evolution of Groups from 1990 to 2014 and Projections from 2020 to 2050 Using Orbital Speeds from 2000 to 2014

11. Appendix B: Figures of all US Public Transportation Systems





Total Vehicle Revenue Miles for All UZAs.



— AI -.- Rail -.- Bus -.- Other

Fisher Information for AI UZAs.

12. Appendix C: Python Scripts

12.1 Fisher Information Python Scripts

Four python scripts (sost.py, fisher.py, smooth.py, fisher_main.py) are written, which are needed to compute fisher information and provided below.

12.1.1 *sost.py*

```
import csv

import pandas as pd

import numpy as np

def SOST(f_name,s_for_sd):

    out=open(f_name+'.csv','rb')

    data=csv.reader(out)

    Data=[]

    for row in data:

        Data.append(row)

    out.close()

    Data_num=[]

    for row in Data:

        temp=[]

        for i in range(1,len(row)):

            if row[i]=="":

                temp.append(0)

            else:

                temp.append(float(row[i]))
```

```
Data_num.append(temp)
df=pd.DataFrame(Data_num)
sos=[]
for j in range(len(df.columns)):
    sos_temp=[]
    for i in df.index:
        A=list(df[j][i:i+s_for_sd])
        A_1=[float(i) for i in A if i!=0 ]
        if len(A_1)==s_for_sd:
            sos_temp.append(np.std(A_1,ddof=1))
    if len(sos_temp)==0:
        sos.append(0)
    else:
        sos.append(min(sos_temp)*2)
df_sos=pd.DataFrame(sos)
df_sos=df_sos.transpose()
df_sos.to_csv('{}_sost.csv'.format(f_name),index=False,header=False)
```

12.1.2 fisher.py

```
import csv

import pandas as pd

import math

import matplotlib.pyplot as plt

def FI(f_name,step,step_1):

    out=open(f_name+'.csv','rb')

    data=csv.reader(out)

    Data=[]

    for row in data:

        Data.append(row)

    out.close()

    Data_num=[]

    Time=[]

    for row in Data:

        Time.append(row[0])

        temp=[]

        for i in range(1,len(row)):

            if row[i]=="":

                temp.append(0)

            else:

                temp.append(float(row[i]))

        Data_num.append(temp)
```

```
out=open('{ }_sost.csv'.format(f_name),'rb')

data=csv.reader(out)

Data=[]

for row in data:

    Data.append(row)

out.close()

sost=[]

for i in Data[0]:

    sost.append(eval(i))

FI_final=[]

k_init=[]

for i in range(0,len(Data_num),step_1):

    Data_win=Data_num[i:i+step]

    win_number=i

    if len(Data_win)==step:

        Bin=[]

        for m in range(len( Data_win)):

            Bin_temp=[]

            for n in range(len( Data_win)):

                if m==n:

                    Bin_temp.append('I')

            else:

                Bin_temp_1=[]
```

```

for k in range(len(Data_win[n])):
    if (abs(Data_win[m][k]-Data_win[n][k]))<=sost[k]:
        Bin_temp_1.append(1)
    else:
        Bin_temp_1.append(0)
    Bin_temp.append(sum(Bin_temp_1))
Bin.append(Bin_temp)
FI=[]
for tl in range(1,101):
    tl1=len(sost)*float(tl)/100
    Bin_1=[]
    Bin_2=[]
    for j in range(len(Bin)):
        if j not in Bin_2:
            Bin_1_temp=[j]
            for i in range(len(Bin[j])):
                if Bin[j][i]!='T' and Bin[j][i]>=tl1 and i not in Bin_2:
                    Bin_1_temp.append(i)
            Bin_1.append(Bin_1_temp)
            Bin_2.extend(Bin_1_temp)
prob=[0]
for i in Bin_1:
    prob.append(float(len(i))/len(Bin_2))

```

```

prob.append(0)

prob_q=[]

for i in prob:

    prob_q.append(math.sqrt(i))

FI_temp=0

for i in range(len(prob_q)-1):

    FI_temp+=(prob_q[i]-prob_q[i+1])**2

FI_temp=4*FI_temp

FI.append(FI_temp)

for i in range(len(FI)):

    if FI[i]!=8.0:

        k_init.append(FI.index(FI[i]))

        break

FI_final.append(FI)

if len(k_init)==0:

    k_init.append(0)

for i in range(0,len(FI_final)):

FI_final[i].append(float(sum(FI_final[i][min(k_init):len(FI_final[i])])/len(FI_final[i][min(k_init
):len(FI_final[i])]))

    FI_final[i].append(Time[(i*step_1+step)-1])

out=open("FI.csv","wb")

new=csv.writer(out)

for i in FI_final:

```

```
new.writerow(i)
```

```
out.close()
```

```
plt.plot(range(step,len(FI_final)+step),[i[-2] for i in FI_final ],
```

```
'b',label='FI')
```

```
plt.ylim(0,8.5)
```

```
plt.ylabel('Fisher Information')
```

```
plt.xlabel('Time')
```

```
plt.tight_layout()
```

12.1.3 smooth.py

```
import csv

import matplotlib.pyplot as plt

def FI_smooth(f_name,step,step_win,xtick_step):

    out=open('FI.csv','rb')

    data=csv.reader(out)

    Data=[]

    for row in data:

        Data.append(row)

    out.close()

    FI=[]

    time=[]

    for row in Data:

        FI.append(eval(row[-2]))

        time.append(row[-1])

    FI_smth=[]

    for i in range(step,len(FI)+step,step):

        for j in range(i-step,i):

            FI_smth.append(float(sum(FI[i-step:i])/len(FI[i-step:i])))

    FI_smth=FI_smth[0:len(FI)]

    plt.plot(range(step_win,len(FI_smth)+step_win),FI_smth,'r',label='Smoothed')

    plt.xlabel('Time Step')

    plt.ylabel('Fisher Information')
```

```

if xtick_step!='def':
    plt.xticks(range(step_win,len(FI_smth)+step_win,xtick_step),
               [time[i] for i in range(0,len(FI_smth),xtick_step)],rotation=75)
else:
    plt.xticks(range(step_win,len(FI_smth)+step_win,3),
               [time[i] for i in range(0,len(FI_smth),3)],rotation=75)

plt.legend()

plt.tight_layout()

plt.savefig(f_name+'FI_'+'.pdf')

plt.savefig(f_name+'FI_'+'.png',dpi=1000)

plt.close('all')

out=open('FI.csv','rb')

data=csv.reader(out)

Data=[]

for row in data:
    Data.append(row)

out.close()

for i in range(len(Data)):
    Data[i].append(FI_smth[i])

out=open("%s_FI.csv"%(f_name),"wb")

new=csv.writer(out)

new.writerow(['Time_Step','FI','Smooth_FI'])

```

```
for i in Data:  
    data_temp=[i[-2],i[-3],i[-1]]  
    new.writerow(data_temp)  
out.close()
```

12.1.4 *fisher_main.py*

```

import datetime

import os

Time=datetime.datetime.now()

f_name=raw_input('enter file name-')

w_size=int(input('enter window size-'))

w_incre=int(input('enter window increment-'))

sm_step=int(input('enter step for block average for smoothing of the FI-'))

X_tick=raw_input('Provide step for xticks(Y)-')

if X_tick.upper()=='Y':

    xtick_step=int(input('enter step for xticks-'))

else:

    xtick_step='def'

def main(f_name,w_size,w_incre,xtick_step):

    if raw_input("Want to use default size of state? enter Y

    otherwise enter N and provide a .csv file named 'file name'_sost.csv-")=='Y':

        from sost import SOST

        SOST(f_name,w_size)

    from fisher import FI

    FI(f_name,w_size,w_incre)

    from smooth import FI_smooth

    FI_smooth(f_name,sm_step,w_size,xtick_step)

main(f_name,w_size,w_incre,xtick_step)

```

```
os.remove('FI.csv')
```

```
print 'Total time taken-',datetime.datetime.now()-Time
```

12.2 orbital Network-based Frequency Analysis (oNFA) Python Scripts

12.2.1 *network.py*

```

from igraph import*

import pandas as pd

import numpy as np

import os

def network_complete(filename):

    df=pd.read_csv('%s.csv'%(filename))# read the csv file using pandas

    df=df.replace(np.NaN,-5,regex=True)# take care of the empty cells

    def network(year,cutoff):

        df1=df[['Country','%s'%(year)]]

        df1=df1[df1['%s'%(year)]!=-5]

        g=Graph()# create the graph using igraph

        value=list(df1['%s'%(year)])

        country=list(df1['Country'])

        g.add_vertices(len(value))

        g.vs["value"]=value #store all the values as node attributes.

        # compute cutoof as a percentage of the median

        cutoff_1=np.median(value)*float(cutoff)/100.0

        # create the edge list. Two nodes are connected when the

        #values are within the cutoff

        edge=[]

```

```

for i in range(len(value)):
    for k in range(len(value)):
        if (g.vs["value"][i]-cutoff_1)<=g.vs["value"][k]<=(g.vs["value"][i]+cutoff_1)
and i!=k :
        edge.append((i,k))

g.add_edges(edge) # add the edges to the graph

g.simplify(multiple=True,loops=True,combine_edges=None)

c1=g.clusters(2)

c3=c1.giant() # get the giant cluster

# Store name, value, degree and orbital position of each node in the graph

final=[]

for i in range(0,len(value)):

    final.append([i,country[i],g.vs["value"][i],g.degree(i),

                c1.membership[i]])

df2=pd.DataFrame(final).sort_values(by=[2],ascending=False)

df2.columns=['ID','Country','Value','Degree','cluster']

df_mode=df2[df2['Degree']==max(df2['Degree'])]

df_mode_1=list(df_mode.iloc[-1])

# return all parameters for each generated network

return [cutoff,cutoff_1,g.vcount(),g.diameter(), g.density(),df_mode_1[1],

        df_mode_1[2],df_mode_1[0],g.average_path_length(),c3.density(),

        float(c3.vcount())/g.vcount(),float(cutoff_1)/float(df_mode_1[2])]

```

```

os.mkdir('%s_Result'%(filename))

mypath=os.path.join(os.getcwd(),'%s_Result'%(filename))

for year in range(1990,2015,1):

    print ('starting %s'%(year))

    parameters=[]

# generate 100 networks for each column using cutodd from 1% of

# the median to 100% of the median

for cutoff in range(1,101):

    parameters.append(network(year,cutoff))

df_param=pd.DataFrame(parameters)

df_param.columns=['Cutoff_per','Cutoff','Nodes','Dia','Density',

                  'Country','Mode','Mode_ID','Avg_shtst','G_density','G_ratio',

                  'H_index']

stdev_list=[]

for i in range(len(df_param['G_ratio'])-5):

    stdev_list.append('%0.2f'%(np.std(df_param['G_ratio'].iloc[i:i+5],

                                     ddof=1)))

stdev_list.extend([0]*(len(df_param['G_ratio'])-len(stdev_list)))

df_param['stdev']=stdev_list

# save all parameters for 100 networks for each column in a csv file

df_param.to_csv(os.path.join(mypath,

                              '%s_parameters_%s.csv'%(filename,year)),index=False)

```

12.2.2 *optimum_network.py*

```

import csv

from igraph import*

import pandas as pd

import numpy as np

import os

import math

from operator import itemgetter

import matplotlib.pyplot as plt

def network_optimum(filename,year):

    df=pd.read_csv('%s.csv'%(filename)) # read the csv file using pandas

    df=df.replace(np.NaN,-5,regex=True) # take care of the empty cells

    mypath=os.path.join(os.getcwd(),'%s_Result'%(filename)) # Access the folder path,

```

where, all results are stored

```

df1=df[['Country','%s'%(year)]]

df1=df1[df1['%s'%(year)]!=-5]

g=Graph() # create the graph using igraph

value=list(df1['%s'%(year)])

country=list(df1['Country'])

g.add_vertices(len(value))

g.vs["value"]=value #store all the values as node attributes.

df_co=pd.read_csv(os.path.join(mypath,

```

```

's_parameters_%.csv'%(filename,year))

df_co=df_co[df_co['stdev']==0 ]

# find the optimum cutoff for the optimum network, where five consecutive
# cutoffs yield minimal/zero change in the size of the giant cluster and also
# ensure a certain percentage (i.e., 60%) of nodes in the giant cluster

df_co=df_co[df_co['G_ratio']> 0.60]

cutoff_1=float(list(df_co.iloc[0])[1]) # get the optimum cutoff

mode=float(list(df_co.iloc[0])[6]) # get the N mode

mode_id=int(list(df_co.iloc[0])[7]) # get the ID of the N mode

# create the edge list. Two nodes are connected when the
#values are within the cutoff

edge=[]

for i in range(len(value)):

    for k in range(len(value)):

        if (g.vs["value"][i]-cutoff_1)<=g.vs["value"][k]<=(g.vs["value"][i]+cutoff_1) and
i!=k :

            edge.append((i,k))

g.add_edges(edge) # add the edges to the graph

g.simplify(multiple=True,loops=True,combine_edges=None)

c1=g.clusters(2)

c3=c1.giant() # get the giant cluster

```

```

# Store name, value, degree and orbital position of each node in the graph

final=[]

for i in range(0,len(value)):

    if g.vs["value"][i]>mode:

        final.append([i,country[i],g.vs["value"][i],g.degree(i),

                    c1.membership[i],

                    g.shortest_paths(mode_id,i)[0][0],

                    100+math.ceil((g.vs["value"][i]-mode)/cutoff_1)])

    else:

        final.append([i,country[i],g.vs["value"][i],g.degree(i),

                    c1.membership[i],

                    g.shortest_paths(mode_id,i)[0][0],

                    100+math.floor((g.vs["value"][i]-mode)/cutoff_1)])

final=sorted(final,key=itemgetter(2),reverse=True)

df_optimum=pd.DataFrame(final)

df_optimum.columns=['ID','Country','Value','Degree',

                    'cluster','sht','Rank']

# Store all results in a csv file

try:

    out=open(os.path.join(mypath,

                    '%s_optimum_%s.csv'%(filename,year)), 'wb')

    new=csv.writer(out)

    new.writerow(df_co.columns)

```

```

new.writerow(df_co.iloc[0])

new.writerow(['edge',g.ecount()])

new.writerow(['Giant_cluster',c1.membership[mode_id],
              'size',c3.vcount()])

new.writerow(['ID','Country','Value','Degree',
              'cluster','sht','Rank'])

for row in final:
    new.writerow(row)

new.writerow(['Cluster analysis'])

for i in range(len(c1)):
    new.writerow([i,[country[j] for j in c1[i]],
                  len([country[j] for j in c1[i]])])

out.close()

except:

out=open(os.path.join(mypath,
                      '%s_optimum_%s.csv'%(filename,year)), 'w')

new=csv.writer(out)

new.writerow(df_co.columns)

new.writerow(df_co.iloc[0])

new.writerow(['edge',g.ecount()])

new.writerow(['Giant_cluster',c1.membership[mode_id],
              'size',c3.vcount()])

new.writerow(['ID','Country','Value','Degree',

```

```

'cluster','sht','Rank'])

for row in final:

    new.writerow(row)

new.writerow(['Cluster analysis'])

for i in range(len(c1)):

    new.writerow([i,[country[j] for j in c1[i]],

                    len([country[j] for j in c1[i]])])

out.close()

# plot the degree distribution

plt.plot(df_optimum['Value'],df_optimum['Degree'])

plt.xscale('log') # transform the x axis to log scale

plt.xlim(xmin=1000)

plt.ylim(0,50)

plt.annotate('Year-%s\nCutoff-%.2f\nMode-%.2f\nCountry-%s\nH_index-

%.2f'%(year,cutoff_1,mode,

list(df_co.iloc[0])[5],float(list(df_co.iloc[0])[-2])),xy=(0.97, 0.96),

xycoords='axes fraction', fontsize=12,

            horizontalalignment='right', verticalalignment='top')

plt.savefig(os.path.join(mypath,'%s_optimum_%s.pdf'%(filename,year)))

plt.savefig(os.path.join(mypath,'%s_optimum_%s.png'%(filename,year)))

plt.close('all')

```

12.2.3 orbital.py

```
import pandas as pd

import os

import numpy as np

import matplotlib.pyplot as plt

def orbital_diagram(filename):

    mypath=os.path.join(os.getcwd(),'%s_Result'%(filename))

    # Read the first year data

    df_orbit=pd.read_csv(os.path.join(mypath,

                                     '%s_optimum_%s.csv'%(filename,1990)),

                        index_col='Cutoff_per')

    end_point=list(df_orbit.index).index('Cluster analysis')

    col=df_orbit.columns[:6]

    df_orbit=df_orbit[col]

    header=df_orbit.iloc[3]

    df_orbit=df_orbit.iloc[4:end_point]

    df_orbit.columns=header

    df_orbit=df_orbit[['Country','Rank']]

    df_orbit.columns=['Country','%s'%(1990)]
```

```
# read the other years data and then merge it with the first year
for year in range(1991,2015):

    df_temp=pd.read_csv(os.path.join(mypath,
        '%s_optimum_%s.csv'%(filename,year)),
        index_col='Cutoff_per')
    end_point=list(df_temp.index).index('Cluster analysis')
    col=df_temp.columns[:6]
    df_temp=df_temp[col]
    header=df_temp.iloc[3]
    df_temp=df_temp.iloc[4:end_point]
    df_temp.columns=header
    df_temp=df_temp[['Country','Rank']]
    df_temp.columns=['Country','%s'%(year)]
    df_orbit=pd.merge(df_orbit,df_temp,how='outer',on='Country')
    df_orbit.sort_values(by=['Country'],ascending=True,inplace=True)
    df_orbit.to_csv(os.path.join(mypath,
        '%s_orbital_position.csv'%(filename)),
        index=False)
```

```

# Create the rank clock for all the countries

df_rank=pd.read_csv(os.path.join(mypath,
                                '%s_orbital_position.csv'%(filename)),
                    index_col='Country')

theta=[2*np.pi*i/len(df_rank.columns) for i in range(1,
                                len(df_rank.columns)+1)]

fig=plt.figure(figsize=(36.4,15))

for i in range(len(df_rank.index)):

    ax=fig.add_subplot(8,20,i+1,polar=True)

    ax.plot(theta,df_rank.iloc[i],color='#00e6e6')

    ax.set_thetagrids([360.0*j/25 for j in range(8,25,8)],
                      range(1990+7,2015,8)) # theta frids at a time interval of 8

    ax.tick_params(axis='y',pad=1.2)

    ax.set_ylim(0,300)

    plt.rgrids((100,200,300))

    ax.set_yticklabels([])

    ax.text(0.35,-0.15,
            df_rank.index[i],
            verticalalignment='center', horizontalalignment='center',
            transform=ax.transAxes,fontsize=12,color='black')

fig.tight_layout()

```

```
fig.text(0.5,0.07,'Orbital Diagrams for %s'%(filename),fontsize=30,  
        horizontalalignment='center',color='#660000')  
fig.savefig('%s_Orbital_Diagram.png'%(filename))  
fig.savefig('%s_Orbital_Diagram.pdf'%(filename))  
plt.close('all')
```

12.2.4 orbital_speed.py

```

import os

import pandas as pd

import numpy as np

def orbital_speed(filename):

    mypath=os.path.join(os.getcwd(),'%s_Result'%(filename))

    df=pd.read_csv(os.path.join(mypath,

    '%s_orbital_position.csv'%(filename)),index_col='Country')

    df1=pd.read_csv(os.path.join(mypath,

    '%s_orbital_position.csv'%(filename)),index_col='Country')

    df1=df1.replace(np.nan,-5,regex=True)

    def o_speed(x):

        x_sum=0

        for i in range(1,len(x)):

            if x[i]!=-5 and x[i-1]!=-5:

                x_sum+= x[i]-x[i-1]

        return x_sum

    df1['speed_total']=df1.apply(o_speed,axis=1)

    df1['avg_speed']=float(0)

    avg_speed=[]

    for i in range(len(df1.index)):

        avg_speed.append('% .2f'%(float(df1['speed_total'][i])/float(df.iloc[i].count()-1)))

```

```
df1['avg_speed']=avg_speed
```

```
df1.to_csv(os.path.join(mypath,'%s_Orbital_Speed.csv'%(filename)))
```

12.2.5 oNFA_main.py

```
import csv

from igraph import*

import pandas as pd

import numpy as np

import os

import datetime

import seaborn

Time=datetime.datetime.now()

try:

    f_name=raw_input('enter filename-')

except:

    f_name=input('enter filename-')

df_main=pd.read_csv('%s.csv'%(f_name))

start_yr=int(df_main.columns[1])

end_yr=int(df_main.columns[-1])+1

print ('Analyzing %s'%(f_name))

import network

network.network_complete(f_name)

import optimum_network

for year in range(start_yr,end_yr,1):

    optimum_network.network_optimum(f_name,year)
```

```
import orbital  
  
orbital.orbital_diagram(f_name)  
  
import orbital_speed  
  
orbital_speed.orbital_speed(f_name)  
  
print ('Finished Analyzing %s'%(f_name))  
  
print ('Total time taken-%s'%(datetime.datetime.now()-Time))
```

13. Appendix D: Copyright

13.1 Chapter 4

<p>Citation: Derrible S, Ahmad N (2015) Network-Based and Binless Frequency Analyses. PLoS ONE 10(11): e0142108. https://doi.org/10.1371/journal.pone.0142108</p>
<p>Editor: Irene Sendiña-Nadal, Universidad Rey Juan Carlos, SPAIN</p>
<p>Received: June 1, 2015; Accepted: October 16, 2015; Published: November 3, 2015</p>
<p>Copyright: © 2015 Derrible, Ahmad. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited</p>
<p>Data Availability: Relevant data are either within or fully referenced (and available for free) in the paper and its Supporting Information file.</p>
<p>Funding: This research was supported, in part, by NSF Award CCF-1331800, by the University of Illinois at Chicago Institute for Environmental Science and Policy (IESP) Pre-Doctoral Fellowship, and by the Department of Civil and Materials Engineering at the University of Illinois at Chicago.</p>
<p>Competing interests: The authors have declared that no competing interests exist.</p>

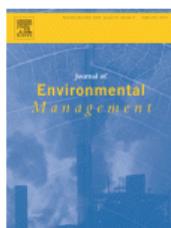


RightsLink®

Home

Create Account

Help



Title: A network-based frequency analysis of Inclusive Wealth to track sustainable development in world countries

Author: Nasir Ahmad, Sybil Derrible, Shunsuke Managi

Publication: Journal of Environmental Management

Publisher: Elsevier

Date: 15 July 2018

© 2018 Elsevier Ltd. All rights reserved.

LOGIN

If you're a [copyright.com](#) user, you can login to RightsLink using your [copyright.com](#) credentials. Already a RightsLink user or want to [learn more?](#)

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>



Using Fisher information to track stability in multivariate systems

Nasir Ahmad, Sybil Derrible, Tarsha Eason, Heriberto Cabezas

Published 9 November 2016. DOI: [10.1098/rsos.160582](https://doi.org/10.1098/rsos.160582)

ARTICLE INFORMATION

DOI <https://doi.org/10.1098/rsos.160582>

Published By The Royal Society

Online ISSN 2054-5703

History Received August 8, 2016

Accepted October 12, 2016

Published online November 9, 2016.

Copyright & Usage © 2016 The Authors.

Published by the Royal Society under the terms of the Creative Commons Attribution License

<http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

KEYWORDS

Fisher information, data mining, big data, information science

AUTHOR INFORMATION

Nasir Ahmad, Sybil Derrible, Tarsha Eason, Heriberto Cabezas



Using Fisher information to assess stability in the performance of public transportation systems

Nasir Ahmad, Sybil Derrible, Heriberto Cabezas

Published 26 April 2017. DOI: [10.1098/rsos.160920](https://doi.org/10.1098/rsos.160920)

ARTICLE INFORMATION

DOI <https://doi.org/10.1098/rsos.160920>

Published By The Royal Society

Online ISSN 2054-5703

History Received November 14, 2016

Accepted March 28, 2017

Published online April 26, 2017.

Copyright & Usage © 2017 The Authors.

Published by the Royal Society under the terms of the Creative Commons Attribution License

<http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

AUTHOR INFORMATION

Nasir Ahmad, Sybil Derrible, Heriberto Cabezas

KEYWORDS

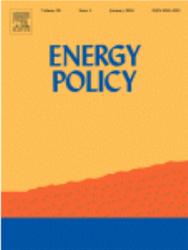
public transportation, transit performance, Fisher information, stability, regime shift

13.4 Chapter 7









Title: An information theory based robustness analysis of energy mix in US States

Author: Nasir Ahmad, Sybil Derrible

Publication: Energy Policy

Publisher: Elsevier

Date: September 2018

© 2018 Elsevier Ltd. All rights reserved.

LOGIN

If you're a **copyright.com** user, you can login to RightsLink using your copyright.com credentials.

Already a **RightsLink** user or want to [learn more?](#)

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

14. Curriculum Vitae

SK Nasir Ahmad

Ph.D. Candidate at the
University of Illinois at Chicago

1255 W Bryn Mawr Ave, APT-101, Chicago, IL-60660

Tel: +1-774-334-1873

sahmad38@uic.edu , nasir1873@yahoo.com

LinkedIn: www.linkedin.com/in/nasir-ahmad-ba72b282

Research Gate:

https://www.researchgate.net/profile/Nasir_Ahmad16

Google Scholar:

<https://scholar.google.com/citations?user=8ZPeXs0AAAAJ&hl=en&oi=ao>

1 Education and Skills

Education

UNIVERSITY OF ILLINOIS AT CHICAGO <u>Ph.D. Candidate</u>	2012-Current
UNIVERSITY OF ILLINOIS AT CHICAGO <u>Master of Science in Civil Engineering</u>	2012-2014
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY (BUET) <u>B.Sc. in Civil Engineering</u>	2004-2009

Languages

Bangla: native

English: fluent

IT Skills

Python (including scikit-learn, keras, igrph, networkx, numpy, scipy, pandas, matplotlib libraries, seaborn), JavaScript, MATLAB, LaTeX, ArcGIS, R, gephi, Apple, Microsoft Windows, MS Office (Word, Excel, PowerPoint, Publisher)

2 Professional Experience

UNIVERSITY OF ILLINOIS AT CHICAGO <u>Research Assistant</u>	2012-Current
DEPARTMENT OF PUBLIC HEALTH ENGINEERING <u>Assistant Engineer</u>	Jan, 2012 - July, 2012
BANGLADESH HOUSE BUILDING FINANCE CORPORATION <u>Assistant Engineer</u>	Oct, 2010 - Dec, 2011

3 Publications

Referred publications

- 1) **Ahmad**, Derrible (2018). An information theory based robustness analysis of energy mix in US States. *Energy Policy*, 120, pp. 167–174.
- 2) **Ahmad**, Derrible, Managi. A Network-Based Frequency Approach to Representing the Inclusive Wealth of World Countries. *Journal of Environmental Management*, 218, pp. 348–354.
- 3) Tennant, **Ahmad**, Derrible (2017). Complexity Analysis in the Sport of Boxing. *Journal of Complex Networks*.
- 4) **Ahmad**, Derrible, Cabezas (2017). Using Fisher information to assess stability in the performance of public transportation systems. *Royal Society Open Science*, 4(4).

- 5) Wang, Koç, Derrible, **Ahmad**, Pino, Kooij (2017). Multi-criteria robustness analysis of metro networks. *Physica A: Statistical Mechanics and its Applications*, 474, pp. 19–31.
- 6) **Ahmad**, Derrible, Eason, Cabezas (2016). Using Fisher information to track stability in multivariate systems. *Royal Society Open Science*, 3, pp. 160582.
- 7) Derrible, **Ahmad** (2015). Network-Based and Binless Frequency Analyses. *PLoS ONE*, 10(11), pp. e0142108.
- 8) **Ahmad**, Derrible (2015). Evolution of Public Supply Water Withdrawal in the USA: A Network Approach. *Journal of Industrial Ecology*, 19(2), pp. 321–330.

Conference Proceedings

- 9) Wang, Koc, Derrible, **Ahmad**, Pino, Kooij (2015). Quantifying the Robustness of Metro Networks. In 6th INSTR - The Value of Reliability, Robustness and Resilience, Nara, Japan.

Conference Abstracts & Technical Reports

- 10) **Ahmad**, Derrible (2016). A Network-Based Approach to Track Social, Economic, and Environmental Trends in World Countries. In 2016 International Symposium on Sustainable Systems and Technology (ISSST), Phoenix, AZ.
- 11) Derrible, **Ahmad**, Theis (2015). Using Fisher Information to Measure the Sustainability of Urban Systems. In 2015 International Symposium on Sustainable Systems and Technology (ISSST), Dearborn, MI.
- 12) **Ahmad**, Derrible, Kermanshah (2015). Towards Sustainable Development: A Network Approach. In 8th International Conference in Industrial Ecology, Surrey, UK.
- 13) Derrible, **Ahmad**, Theis (2015). Capturing the Sustainability of Urban Systems using Fisher Information: Application to Public Transport. In 8th International Conference in Industrial Ecology, Surrey, UK.
- 14) **Ahmad**, Derrible (2015). Analyzing Multivariate Time Series Data using a Network-Based Approach. In 2015 Network Frontier Workshop, Evanston, IL.

- 15) Kermanshah, **Ahmad**, Berkelhammer, Derrible (2015). Resilience in Complex Systems: A GIS Analysis of Extreme Events Impact on Infrastructure Networks. In International Society for Industrial Ecology, Surrey, UK.
- 16) Peiravian, **Ahmad**, Kermanshah, Derrible (2014). Questioning Box-Counting Method as a Tool for Fractal Characterization of Physical Networks. In NetSci 2014 International School and Conference on Network Science, Berkeley, CA.
- 17) **Ahmad**, Kermanshah, Peiravian, Derrible (2014). Network Science: A Potential Tool for Analyzing Water Consumption in the USA. In NetSci 2014 International School and Conference on Network Science, Berkeley, CA.
- 18) Kermanshah, Peiravian, **Ahmad**, Derrible (2014). Investigating Transportation Network Resilience to Extreme Events. In NetSci 2014 International School and Conference on Network Science, Berkeley, CA.
- 19) **Ahmad**, Derrible (2014). Evolution of Water Consumption in the USA: A Network Approach. In 2014 Illinois Water, Urbana, IL.

4 Teaching Experience

- **Cities and Sustainable Infrastructure (CME 440)** **Fall 2016**
Graduate and 4th year undergraduate course – 51 students

Teaching Assistant
- **Water Resources Engineering (CME 311)** **Spring 2016**
Undergraduate course – 46 students

Teaching Assistant
- **Cities and Sustainable Infrastructure (CME 494)** **Spring 2015**
Graduate and 4th year undergraduate course – 36 students

Teaching Assistant
- **Senior Design I (CME 396)** **Fall 2013**
Undergraduate course – 58 students

Teaching Assistant

- **Hydraulic Design (CME 403)** **Spring 2013**

Undergraduate course – 42 students

Teaching Assistant

- **Design of Prestressed Concrete Structures (CME 410)** **Fall 2012**

Graduate and 4th year undergraduate course – 43 students

Teaching Assistant

5 Honors and Awards

- **Dean's Scholar Fellowship** **July, 2016**

Dean of the Graduate College, University of Illinois at Chicago

The Dean's Scholar Fellowship is a one-year, non-renewable award presented by the Dean of the Graduate College in recognition of a student's scholarly achievement.

The award is intended to provide the most distinguished, advanced-level graduate students with a period of time dedicated solely to the completion of their programs.

- **SIEYP Scholarship** **May, 2015**

International Society for Industrial Ecology

For the 2015 ISIE Conference and Symposium on Industrial Ecology for Young Professionals in Surrey, UK.

- **IESP Predoctoral Fellowship** **July, 2014**

The Institute for Environmental Science and Policy, University of Illinois at

Chicago

- **NTU Winter School Fellowship**

February, 2014

Nanyang Technological University, Singapore

For the Workshop on Introduction to Complexity and Complexity Science at

Nanyang Technological University, Singapore

6 Membership

Member, Complex and Sustainable Urban Networks (CSUN)

Member, International Society for Industrial Ecology (ISIE)