

ANALYTICAL METHODS FOR COMPUTING THE RESILIENCE,
RECOVERY, AND TRANSFORMATION OF COMMUNITIES AND THEIR
CONSTITUENT SYSTEMS IN THE AGE OF BIG DATA

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Benjamin A. Rachunok

In Partial Fulfillment of the
Requirements for the Degree

of

Doctor of Philosophy

December 2020

Purdue University

West Lafayette, Indiana

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL**

Dr. Roshanak Nateghi, Chair

School of Industrial Engineering

Dr. P. Suresh Rao

Lyles School of Civil Engineering

Dr. Daniel DeLaurentis

School of Aeronautics and Astronautics

Dr. Christopher Quinn

Department of Computer Science, Iowa State University

Approved by:

Dr. Abhijit Deshmukh

Head, School of Industrial Engineering

ACKNOWLEDGMENTS

My experience has been that a Ph.D. is truly an instance of “it takes a village.” I sat down to write this acknowledgement and it was a great experience to think through all the people who contributed to getting to this point whether they knew it or not. I owe everyone in this acknowledgement this an enormous debt of gratitude. First, I have to give my deepest appreciation and thanks to my Ph.D. advisor Roshi. I can’t imagine having a better experience in graduate school, so thank you immensely Roshi for all the support, guidance, encouragement, and opportunities to grow. Similarly, I appreciate all the time support and encouragement my committee has given me as well; no question this work is immeasurably improved by their input. Thanks to my mom Sue and brother Charlie for always supporting me and asking “how research is going” even when they know it will ignite an hour of meaningless ranting on my part. Of course thanks to Caitlin for being so supportive in all the non-work times and helping me brainstorm titles, pick color palettes, pick me up from work in the rain, and always being ready to blow off steam at AJs when necessary.

A huge thanks to LASCI lab members past and present: Ellen, Aiyshwariya, Minsoo, Debs, Renee, Jackson, Kendrick, Sayanti and Negin for being a perpetual source of new ideas and support. Thanks also to Moya who was a great undergraduate researcher and contributed so much to these research ideas. The same goes for all my IE compatriots: Jackie, Ben, Bryan, Viplove, James (and Kyle), Oscar, Vivia, Mustafa, Esmail, Anna, Marissa, Naveen, Edgar, Ye, Brijesh. Thank you all for making Grissom such a warm, fun, and supportive place to be. Similarly, thanks to all the IE faculty and Anita, Leza, Elizabeth, Morgan, Lindsay, Abhi, Pat, Dave and everyone who makes the IE department such a great place to be. Also thanks to all of my mentors, friends, and collaborators from other universities and labs for setting a research standard of excellence. A Ph.D. wouldn’t be possible if you went insane

during it, so I owe my friends for keeping my afloat. I'm so thankful to have found the strangest group of humans on earth in Tyler, Thomas, Marcus, Michael and my son Mark, without whom I would have gotten much more sleep.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	x
ABSTRACT	xiii
1 INTRODUCTION	1
1.1 Climate Change and Communities	1
1.2 Introduction to Resilience Theory	3
1.3 Urban & Community Resilience	4
1.3.1 Measurement	6
1.3.2 Transformation in Communitites	6
1.3.3 Social Media and Community Resilience	7
1.4 Engineering Systems Resilience	10
1.4.1 System Representations	10
1.4.2 Disruption Impacts in Engineering Systems	14
1.4.3 Interdependence	15
1.4.4 Transformation in Engineered Systems	17
1.5 Summary of Included Papers	17
1.5.1 The Sensitivity of Electric Power Infrastructure Resilience to the Spatial Distribution of Disaster Impacts	17
1.5.2 Interdependent Infrastructure System Risk Resilience to Nat- ural Hazards	18
1.5.3 Twitter and Disasters: A Social Resilience Fingerprint	19
1.5.4 Improving Operational Measures of Community Resilience	20
2 THE SENSITIVITY OF ELECTRIC POWER INFRASTRUCTURE RE- SILIENCE TO THE SPATIAL DISTRIBUTION OF DISASTER IMPACTS	21
2.1 Introduction	21
2.2 Background	22
2.3 Methods	24
2.3.1 Electric Power Network	25
2.3.2 Disruption generation algorithms	26
2.3.3 Performance metric calculation	28
2.3.4 Simulation Methodology	29
2.4 Results	30
2.4.1 Static measures of impact	30

	Page
2.4.2	Dynamic measures of impact 35
2.5	Conclusion 37
3	INTERDEPENDENT INFRASTRUCTURE SYSTEM RISK & RESILIENCE TO NATURAL HAZARDS 39
3.1	Introduction 39
3.2	Methods and Data 40
3.2.1	Methods 40
3.2.2	Failures 41
3.2.3	Failure Generation Methods 42
3.2.4	Recovery 43
3.2.5	Data and parameters 44
3.3	Results 45
3.3.1	Spatial Differences in Initial Disruption 45
3.3.2	Changes in recovery of systems 46
3.4	Conclusion 47
4	TWITTER AND DISASTERS: A SOCIAL RESILIENCE FINGERPRINT . 49
4.1	Introduction 49
4.2	Background 51
4.2.1	Community resilience 51
4.2.2	Twitter 52
4.3	Data and Methods 55
4.3.1	Tweet Acquisition 55
4.3.2	Data processing 57
4.3.3	Social Resilience Fingerprint 58
4.4	Results 60
4.4.1	Event Similarity 60
4.4.2	Critical Components of Community Resilience 66
4.4.3	Posterior Analysis 66
4.5	Discussion 68
4.6	Conclusion and Future Research 70
5	IMPROVING OPERATIONAL MEASURES OF COMMUNITY RESILIENCE 72
5.1	Introduction 72
5.2	Contrastive Community Networks 75
5.2.1	Community Risk Factors 75
5.2.2	System Resilience: Restored Access to Electricity 76
5.2.3	Engineering Resilience Model 77
5.2.4	Self Organized Maps 79
5.2.5	Contrastive Community Networks 81
5.3	Results 84
5.3.1	Quantifying Community Transformation 84
5.3.2	Transformation vs. Degradation 85

	Page
5.3.3 Engineering Resilience and Transformation	86
5.3.4 Triggering Transformation	89
5.3.5 Resilience Traps	89
5.3.6 Improved Operationalization of Community Resilience	91
REFERENCES	95
A APPENDIX FOR CHAPTER 1	111
B APPENDIX FOR CHAPTER 2	116
C APPENDIX FOR CHAPTER 3	118
D APPENDIX FOR CHAPTER 4	122
VITA	136

LIST OF TABLES

Table	Page
2.1 Summary statistics for the distribution of efficiency for respective failure modes with the percentage of optimal network efficiency listed in parentheses. Failure fraction represents the fraction of the network which was induced as failed in each iteration. Results presented here are for failures in 60% of the network. Complete results are presented in Appendix Table A.1 and A.2.	33
2.2 P-values for two-sample, two tailed, Kolmogorov-Smirnov tests between the efficiency and LCC of given initial failure methods and failure fraction. Results at the 0.6 failure fraction are presented in this article. Results use a significance level of $\alpha = 0.05$. Values of zero listed with one significant digit indicate $p < 1.11022e - 16$; this cutoff is the numerical precision of the machine used for computations.	34
2.3 Summary statistics for the distribution of largest connected component (LCC) for respective failure modes with percentage of the optimal value listed in parentheses. Results presented here are for failures in 60% of the network. Complete results are presented in Appendix Table A.3 and A.4. .	34
2.4 Summary statistics for the distribution of percent of county customers without power in a static analysis. All numbers represent the fraction of the total population of the county without power.	35
3.1 Comparison of initial disruption types. P-values taken from 2-sample Kolmogorov-Smirnov tests. Values of 2.2×10^{-16} represent a p-value smaller than the numerical precision of R	46
4.1 Resilience components, their description, and community elements from that category	53
4.2 Size, date, and type of twitter events analyzed. Ordered by starting date of event.	56
4.3 Closest Events. The Pearson correlation is calculated between all pairs of events, with the closest match listed. The correlations above 80% have been highlighted in bold.	64
5.1 CCN importance and engineering resilience importance.	88
5.2 Engineering resilience importance for non-transformation risk factors . . .	92

Table	Page
A.1 Summary statistics for the distribution of efficiency for respective failure modes. Failure fraction represents the fraction of the network which was induced as failed in each iteration. Results presented in the body of the work represent a failure fraction of 0.6	112
A.2 Summary statistics for the distribution of efficiency for respective failure modes scaled by the efficiency when the network is fully repaired. In this table, a value of 100 is the same performance metric seen at a fully repaired system.	113
A.3 Summary statistics for the distribution of largest connected component (LCC) for respective failure modes.	114
A.4 Summary statistics for the distribution of largest connected component (LCC) for respective failure modes, scaled by the total LCC when the network is fully repaired. In this table, a value of 100 is the same performance metric seen at a fully repaired system.	115
B.1 Efficiency of networks in different under different failure regimes	117
C.1 Tweet corpora summary. Description of events used along with the quantity of tweets and their acquisition methods and respective sources. . . .	119
C.2 Tweet Acquisition. Keywords, keyword phrases, and hashtags used to create the tweet datasets.	120
C.3 Categories of resilience and associated keywords. Keywords are manually coded based on conceptual definitions of resilience categories.	121
D.1 tiny bracketed caption	125
D.2 County and resilience values for each county in Florida during Hurricane Michael (2018)	132

LIST OF FIGURES

Figure	Page
1.1 (a) Shows the conceptual representation of the resilience triangle, as the area under the performance curve in tan, while (b) shows an empirically measured triangle in Bay County, FL during 2018 Hurricane Michael (Figure from Chapter 5)	13
2.1 The case study network situated in the Gulf Coast of the U.S. a The layout of the electric power grid placed over the county. b The density of customer-level power outages during Hurricane Katrina with the network overlain. c Census-tract level population density for the corresponding area. 25	25
2.2 Outage generation types. The results of three outage generation techniques, each inducing failures in 60% of the grid. Figure a is one instance of an outage generated randomly. Figure b is a an outage generated using a breadth-first algorithm, while c is a depth-first algorithm.	28
2.3 An overview of simulation methodology. The process here represents one simulation iteration.	31
2.4 Static disruption comparison. Relative density of network performance after 100 disruptions for each disruption generation method. a is the network efficiency for all three disruption generation methods while b is the size of the largest connected component.	32
2.5 Performance metrics measured after the disruption over time for each disruption method. In a-f , the bands of uncertainty represent 95% confidence intervals sampled from the empirical density at each point in time. The black line is the mean of the observations. The x-axis is the relative-completeness of the network repair scaled by the total restoration time for each replication.	36
3.1 Examples of system disruption types, orange vertices are operational, white have failed. Left is a network after <i>random</i> failures, center is a network after <i>BFS</i> failures, and right is a network after <i>DFS</i> failures. All three represent a failure of 60% of the vertices in the system.	43
3.2 Differences in initial disruptions. Left are density plots of both graphs subset by the corresponding distribution of failures in g_1 (rows) and g_2 (columns). Right shows the change in performance measure as a function of failure size (rows).	45

Figure	Page
3.3 Recovery of system over time subset by the distribution of failures in g_1 (rows) and g_2 (columns). Red points are the system efficiency for g_1 , and blue are the efficiency of g_2 . Black lines indicate the mean of all replications at a given percentage of the system repaired.	47
3.4 Differences in time to repair (TTR) by failure method subset by the size of failures (columns) and the distribution of failures in g_1 (rows). Each individual plot shows the difference in TTR for g_1 (red) and g_2 (blue) for changes in the distribution of failures in g_2 (sub-columns).	48
4.1 Visualizing social resilience fingerprints. Each heatmap represents the association between one category with another. The color red indicates the most association, and the color blue represents the least association. Diagonal values in the heatmap are indicative of how self-associative a category is. Bar-graphs show the relative occurrence of each category. Note the color scheme in this plot is based on log-normalization of A , as opposed to Sinkhorn-Knopp, to aid in visualization.	61
4.2 VAT assessment. Each element represents the distance between the event in the row and the event in the column, red indicating closer, blue farther away	62
4.3 PCA loading plot. The top 10 most contributing elements of each fingerprint are projected onto the first two principle components. Components and events along the negative x axis are shown in the bottom left inset for clarity.	65
4.4 Category-based difference between the average hurricane fingerprint and all average non-hurricane fingerprint. Each element represents the difference between the hurricane and non-hurricane association of categories. Blue indicates stronger association in non-hurricane events, while red indicates stronger association in hurricane events . Coloration based on log-normalization of A matrices rather than Sinkhorn-Knopp for visual clarity	67
5.1 Description of the baseline CCN for county clustering. (a) Values of four sample community risk factors across each CCN node. (b) Mapping of each county within the CCN. (c) Location of CCN clusters (colors) within Florida. Figures (a-c) created in R (v 3.2.1; https://www.r-project.org/) [179] using the <code>ggplot2</code> (v 3.3.0; https://ggplot2.tidyverse.org/) [180] and <code>kohonen</code> (v 3.0.10) [181]. Figure c additionally used <code>usmap</code> (v 0.5.0; https://github.com/pdil/usmap) [182]. Map shapefiles in b are from <code>usmap</code> and the US Census Bureau [183].	74

Figure	Page
5.2 Community change required for transformation. (a) The amount of change required to elicit transformation in Bay County, FL for each risk factor. Blue indicates a positive risk factor shift is required for transformation, while red indicates a negative shift, and green dots show community risk factors for which no transformation is possible. (b, c) As each community risk factor is perturbed (x-axis of b,c), we compute where Bay County is mapped in the CCN. The values of the risk factors occurring at each jump indicate the degree of risk factor-perturbation leading to a reconfiguring of the CCN.	83
5.3 Community risk factors contributing to baseline engineering resilience. (a) Shows the results of the machine-learning variable selection indicating which community aspects have the highest relative contribution to resilience and whether they make a positive (yellow) or negative (blue) contribution to resilience. (b) a map colored by the resilience of the power grid to Hurricane Michael along with the storm's track. Darker counties were <i>more</i> resilient to the storm. Figures (a-b) created in R (v 3.2.1; https://www.r-project.org/) [179] using the <code>ggplot2</code> package (v 3.3.0; https://ggplot2.tidyverse.org/) [180]. Plot b additionally used <code>usmap</code> (v 0.5.0; https://github.com/pdil/usmap) [182]. Map shapefiles in b are from <code>usmap</code> and the US Census Bureau [183].	87
5.4 Risk Factor Change Required to Elicit Transformation. Each boxplot shows the percentage change in each risk factor required for a county to be mapped to an alternative node in the CCN, across all counties in the case study of Florida. (a) through (d) have increasing degrees of change required for transformation. Risk factors with no possibility of transformation are listed in Table 5.2. Note the shift in y-axis values moving from left to right.	94
D.1 Power Outages During Hurricane Michael for Bay County, FL	122
D.2 Power Outages During Hurricane Michael for Holmes County, FL	123
D.3 Out of sample RMSE and R^2 for the prediction of engineering resilience as a function of the selected community risk factors. Note the null model's out of sample R^2 is 0 because the standard deviation of predictions is 0.	124
D.4 Example County Movement. (a) shows the location of nodes with no perturbation, (b) shows the temporal trajectory of Bay County associated with the shift to the lowest alternative- equilibrium seen in Figure 5.2 c. (c) is the temporal trajectory for the change to the highest alternative-equilibrium seen in Figure 5.2c. The length of the arrows in (b) and (c) are the magnitudes of transformation for each perturbation.	135

ABSTRACT

Rachunok, Benjamin A. Ph.D., Purdue University, December 2020. Analytical Methods for Computing the Resilience, Recovery, and Transformation of Communities and Their Constituent Systems in the Age of Big Data. Major Professor: Roshanak Nateghi.

Communities are increasingly vulnerable to climatic risks which are estimated to cost \$1.8 trillion and lead to 2 million deaths annually by the end of the century [1]. To minimize this vulnerability in the face of the increasing climatic risks, resilience is used as an organizing principal by all scale of governments, decision makers, and international organizations to address climatic risks. Resilience is conceptualized across many fields and is broadly meant to represent the ability of a system to maintain critical functionality, adapt, and ‘bounce back’ after a shock or disruption [2].

Moving from theoretical conceptualizations of resilience to operational decisions which aim to foster adaptive capacity in communities, requires consideration of the dynamics of engineered, social, ecological, economic, and political systems among others. This dissertation develops analytical techniques to leverage ‘big data’ to understand the multifaceted aspects of how communities and engineered systems are impacted by and recover from major disruptions in an effort to bridge the gap between resilience in theory and resilience in practice.

In the light of the disciplinary variations in conceptualization and operationalization of resilience, the introduction to this dissertation begins by unpacking the myriad of resilience definitions and how they relate to communities and engineered systems; describing analytical techniques which are used to model and quantify communities and engineered systems.

Chapters 2-5 summarize the articles included as a component of the dissertation. First (Chapter 2) I analyze the characteristics of large-scale disruptions in

network-based infrastructure systems. There is a large body of work which utilizes graph-theoretic representations of engineered systems to model resilience to shocks. However, the way by which shocks or disruptions are simulated in the system are either based on random failures –indicative of component aging– or targeted failures –based on an intentional threat like terrorism– and do not reflect the explicit spatial structure of natural hazards. To address this gap, I propose two methods for generating failures in network based infrastructure models which have a connected, spatial structure similar to that of a large-scale natural disaster such as a hurricane. When evaluating the performance of the system after a disruption using network-based performance metrics, the networks with spatially-distributed outages show statistically different measures of performance compared with similarly sized randomly-distributed outages. Additionally, when simulating the recovery of the system; the spatial characteristics of the outages drastically alter the way in which the network recovers. Of note, systems disrupted with random outages showed antifragile properties, while spatially-distributed outages do not. This work is extended to interdependent infrastructure systems in Chapter 3.

In Chapter 4, I contribute to the nascent literature on harnessing social media data for resilience analytics. Specifically, I develop algorithms for analyzing how community members perceive the dynamics of their community during a crisis event, using twitter data during 14 major crises events. Grounded in theories of community resilience and sociological risk appraisal, these algorithms —called the Social Resilience Fingerprint— capture the patterns of discourse in communities related to the attributes of communities which contribute to its resilience, such as infrastructure, economic, and ecological systems. Using this framework, I show how different types of major disruptions (hurricanes, earthquakes, political events etc) have signatures identifiable in social media data and discuss the trends driving these similarities.

Finally, in Chapter 5 I formulate machine learning methods for evaluating the potential of communities to transform after major disruptions. The current paradigm of community resilience modeling aims to rapidly return to normal-operation follow-

ing a disruption. By promoting the status quo, however, this modeling technique may be counteracting itself by reinforcing persistent, maladaptive states which inhibit the ability of communities to grow and transform. With this gap in mind, I have developed an alternative method for measuring community resilience, termed a Contrastive Community Network (CNN), which identifies key drivers of community transformation and quantifies how communities reorganize after major disruptions into alternative, stable equilibria. Using this improved methodology, I identify *resilience traps*: risk factors which, while critical for rapid recovery to the status quo, do not allow for any possibility of transformation and long-term adaptation. These traps clearly demonstrate some of the pitfalls present in current methodologies for quantifying community resilience.

The methodologies and algorithms developed in this dissertation can improve the ability of stakeholders and decision makers to understand and analyze how communities adapt and respond to major crisis events, allowing for data-driven decisions to be made to bolster the resilience of communities in response to climate change.

1. INTRODUCTION

Two societal inflection points in the early 21st century are likely to be the proliferation of connected technology and the data created by it, as well as the advent of anthropogenic climatic change. *Big Data* has served as the basis of new industries, enabled a greater degree of connectedness throughout the world, and fostered a new generation of analytical tools to extract insights from vast quantities of data. The data volume coupled with these techniques has fundamentally changed the relationship between technology and society more broadly, creating trillions of dollars in economic value [3].

Climate change on the other hand serves as a pressing crisis, with the potential to disrupt all aspects of our current society. Natural hazards are on the rise globally, a large fraction of the world's population stands to be displaced from their communities; and while climate change mitigation efforts are beginning to be implemented, they may not be enough to prevent global economic loss and disruption [4]. While climate change and big data are not opposing forces, this dissertation aims to utilize the latter to mitigate the impacts of the former. This section introduces how big data can be utilized to combat the impacts of climate change, and to describe how the technical work in the later sections fits among the larger issues facing the world.

1.1 Climate Change and Communities

The IPCC's 6th assessment report identifies many of the most severe impacts of climate change on societies. These range from melting snow and ice are disrupting hydrological systems, to the reduction in global crop yields, and the creation of significant burdens to worldwide public health [4]. Of these, one of the most significant global risks faced by climate change is the vulnerability to climate-related extremes.

Climate extremes are perturbations in climatic trends as a results of climate change and range from short-term changes in daily temperature and precipitation levels, to long-term changes to the frequency and intensity of major natural hazards like floods, droughts, and hurricanes [5]. During 2015-2020 in the US alone, these natural hazards cost an average of \$107 billion and claimed 700 lives per year [6]

A community is made up of a geographically linked groups of interacting individuals with shared norms and interests [7]. When natural hazards strike, they impact all aspects of communities. To understand these impacts, communities are often broken down into the systems and subsystems that comprise them. For example, Cutter (2008), breaks a community down into dimensions along ecological, social, economic, institutional, infrastructure, and community competence based on how communities are impacted by major disruptions [8]. With the numerous of threats facing communities from natural hazards, as well as the diverse methods for building adaptive capacity, analyzing the resilience of communities is frequently done at the system or subsystem level.

It is important to consider climate vulnerability when designing and managing the built environment and communities. Management and design decisions can be explicit such as considering a certain magnitude or frequency of natural hazard when designing the strength of components in physical infrastructure, or implicit in the formulation of governance and policy interventions to mitigate impacts from disruptions in institutional or economic systems.

With constraints imposed due to limited resources and the inherent uncertainty in predicting future climate impacts, the process of making engineering decisions to improve a community's ability to withstand climate impacts comes with fundamental questions about about where improvements should be made. As an example, consider designing a regional power grid to mitigate the impact of a major hurricane. Given a fixed budget, one avenue could be to strengthen the components of the grid to withstand high winds; making components less likely to fail, with a trade off that it may take more time to repair if they fail and could be more vulnerable to other modes

of disruption such as flooding. Alternatively, a design criteria could be minimizing the impact in a community due to the failure of any one component of the grid, with the trade off that redundancies are added into the system which create operational expense during non-disaster times.

This process of identifying potential threats and responses represents a risk-based or 'fail-safe' approach to designing for uncertainty [9], and highlight the multifaceted nature of selecting high-level objectives when aiming to design communities and systems to which can withstand climate impacts. A contrasting approach to incorporating uncertain future outcomes into design criteria is based around *resilience*. Broadly defined, resilience is the concept of a system 'bouncing back'; that is the system is 'safe-fail' rather than 'fail-safe' [9]. But as with the previous examples, the underlying objective of promoting resilient systems comes with trade-offs and should be carefully considered. The technical material in this dissertation focuses on resilience in physical and social systems and the subsequent section explain how the concept of resilience has come to be used in the study of systems as well as nuances which are relevant to applying resilience to communities.

1.2 Introduction to Resilience Theory

Resilience as an academic concept related to large-scale systems was introduced by ecologist C.S. Holling in reference to how ecosystems respond to disruption, calling resilience "An ecosystem's ability to maintain basic functional characteristics in the face of disturbance" [10]. The aim was largely to shift the prevailing views of ecologists from studying and quantifying ecological systems as they are *in equilibrium* toward viewing the behavior of ecological systems "in terms of the probability of extinction of their elements." [10] Related to the concept of bouncing back, this conceptualization of resilience "is a measure of the ability of a system to absorb changes of state variables, driving variables, and parameters and still persist." [11]

By focusing on the capacity of a system to adapt, Holling makes a key distinction which differentiates resilience from *stability*. A faster return to normalcy, in Holling's conceptualization, is a *more stable* system [11]. By separately defining resilience and stability, Holling has created two axes in which to evaluate a system. *i.e.*, a high-resilience, low-stability system may fluctuate greatly but ultimately persist in the face of disruptions. Likewise, a high-stability low-resilience system would have capacity to respond in the short-term but may lack the ability to adapt in the long term.

While one of the earliest definitions of resilience, Holling's work highlights three key properties of resilience which are important for considering how to design communities which are prepared for climate impacts. The first is that resilience is a system-wide phenomenon and should be evaluated accordingly. The second is how resilience changes temporally, and the importance of both short and long-term changes to systems. The third is the separation of resilience from stability: emphasizing that a rapid return to the status quo does not inherently mean a system is resilient. The following sections describe how these properties influence decision-making when resilience thinking is applied to communities and engineered systems.

1.3 Urban & Community Resilience

Geographers, urban theorists, and urban planners have applied the ecological concept of resilience to the study of urban systems and communities [12–14]. As previously described, communities are a geographically linked group of interacting individuals with shared norms and interests [7]. Urban systems are closely related, defined broadly as the systems and subsystems which make up a city [12, 15–17]. The study of urban and community resilience evaluates how *communities* adapt to adversity and disruption. Broadly, the concept of community resilience is a positive attribute of a community associated with increasing the local capacity, social support and resources associated with responding to crisis events combined with decreasing

risk, miscommunication, and trauma [8,18]. Urban and community resilience exist as interlinked concepts, with the focus of urban resilience being the planning of physical aspects of towns and cities while community resilience is focused on the inhabitants of these towns and cities [19].

Urban and community resilience is largely characterized as a *process based* attribute of a community [18,20,21]. In this way, community resilience is thought of as the linking of adaptive capacities from all community systems together to *continuously adapt* to disruptions and changes [9,22]. Accordingly, the resilience of a community is increased when aspects of a community work together and the community or urban system broadly shares a capacity to manage and adapt to major events [23]. Similarly, a non-resilient community can be thought of as one which is *disoriented* while a resilient community is one which *reorients* itself quickly [23], emphasizing the process of adaptation in communities.

The uses and conceptualizations of resilience in urban and community systems inherit Holling's separation of resilience and stability, however with the explicit focus on cities and communities, these terms get re-mapped such that *resilience* is referred to as *dynamic* or *radical* resilience where stability is called *equilibrium* or *conservative* resilience [15,24]. This distinction is particularly important in urban systems and communities as it reflects how resilience must be intentionally designed into communities and community systems [15]. The intentional process of incorporating resilience into urban and community systems also precipitates a focus on the causes of system disruptions [12,14]. This focus on the specifics of disruptions –and the threats and hazards that cause them– is a key aspect of community and urban resilience to climate impacts.

Relevant to this dissertation, the important takeaways from the application of resilience to communities and urban systems (1) that resilience as a theoretical concept is focused on continual (both long and short-term) adaptation, (2) that community and urban resilience are the result of the entirety of communities acting and interact-

ing together, and (3) that the specifics of individual threats and hazards need to be accounted for when improving the adaptive capacity of a community.

1.3.1 Measurement

When aiming to measure community resilience, many current methods utilize a *range of attributes*, or *indicator-based* approach [8, 18, 25]. In these methods, constructs which are understood to contribute to increased community resilience (such as socioeconomic status, social capital, and access to critical goods and services) are measured by proxy [26]. Community resilience is then estimated as a weighted sum of these variables. For example, if education is hypothesized to increase community resilience, then a proxy measurement –such as average educational attainment– will be used to identify communities which are *more resilient* based on higher aggregate level of household education [27]. There are growing number of proposed methods for analyzing community resilience based on the range of attributes approach [7], with the primary differences being the community attributes hypothesized to contribute to resilience, as well as the mathematical methods for combining them [7].

1.3.2 Transformation in Communities

Recalling Holling’s original conceptualization of resilience as a system’s ability to persist in the face of disruption, the ability of a system to *transform* is a vital part of the resilience process [28]. Implementations of community resilience ideas based on indicator-based approaches for improving the resilience of communities, however, have been criticized for neglecting to incorporate or quantify the aspects of communities which transform [16]. This criticism in operational models of resilience is due to a focus on returning systems to their original state rapidly and efficiently. Returning to the key points from Holling’s first work on resilience, these models promote *stability*.

From a governance perspective, the lack of consideration of transformation has been identified as a negative aspect of resilience-based policy and has been criticized as continuing to promote the status quo [15]. This entrenches policies and practices which may be untenable and unsustainable given the pressures of a changing climate [28].

1.3.3 Social Media and Community Resilience

A nascent approach to estimating community resilience which addresses the overemphasis on stability found in current indicator-based approaches utilizes social media as a tool to analyze how individuals respond to disruption [29, 30]. Social media serves as a key conduit of information flow between community members during crisis events [31, 32]. As content and interaction is user-driven, social media represents a ‘bottom-up’ look at dominant topics, themes, and events present in popular discourse [33]. When analyzed during crisis events, this allows for the characterization of emergent patterns of human self-organization inherent in community resilience, promoting for better understanding how communities adapt and respond to major disruptions. Techniques for analyzing social media data to understand community resilience are based primarily on three approaches: social media content, social network analysis, and metadata analysis.

Content-based analyses of social media utilize the user-broadcast information in social media posts during crisis events (*e.g.* a tweet, Facebook post, image posted etc.) to provide a broader situational understanding of a community disruption [34–36]. If the quantity of social media data is small and predominantly text-based (such as Twitter or Facebook), text can be manually coded to identify commonalities and topics of discussion [34, 35]. In larger text-based social media datasets, unsupervised statistical learning methods such as *Latent Dirichlet Allocation* (LDA) can be utilized to understand emergent trends and themes in large bodies of text data [37]. Finally, the linguistic sentiment –a method for characterizing the mood of the text such as

happy, angry, positive etc. – can be calculated about individual or aggregate social media posts [38]. All of these methods serve to quantify the key topics, themes, and events present in social media data during crisis events as a way to understand the collective opinion dynamics occurring during a disruption.

Another method of analyzing community dynamics during crisis events is by studying the interactions of individuals on social media, specifically the social network. Networks are “a collection of points [referred to as *vertices* or *nodes*] joined together by pairs of lines [referred to as *edges* or *links*]” [39]. On many popular social media platforms (*e.g.* Twitter, Facebook, Instagram), users choose the content they wish to receive by choosing the users they interact with by ‘following’ or subscribing to their content. Subsequently, users can interact with the content they are viewing (by ‘liking’ a Facebook or Instagram post or ‘retweeting’ an Tweet). By studying the flow of interaction and content as a network we can gain insight into the dynamics of communities during crisis events [40].

Networks constructed from social media data can be represented in different ways depending on the type of analysis performed. To study information flow, a graph can be constructed in which nodes in the graph are users, and edges represent a follower/followee relationship [41]. In this network, the propagation of content such as a Tweet or Facebook post can be studied to quantify how information spreads through the social network. Similarly, a graph can be constructed in which users are nodes, and edges represent one user interacting (through a retweet, like, or share) with each other’s content [42, 43]. In both of these network constructions, graphical properties of the networks, such as centrality and connectidness can be used to provide estimates of disaster outcomes [42, 44, 45].

Finally, analyses can be conducted based on social media metadata. Metadata is the non-content information which accompanies social media data, such as the time of posting, user location, and language the post is written in. The location of the user is particularly valuable to understanding community resilience as it allows for direct, place-based analyses of how communities are responding to crisis events [31].

By studying the content of social media data along with precise user location, accurate predictions about the vulnerability of communities can be made [46–48]. User location, however, is becoming increasingly unavailable to researchers because of concerns about user privacy [49]. In the absence of location data, other metadata has been extracted from social media data such as the time of day [50,51] in an effort to understand temporal changes in how communities respond to crisis events.

All three of these methods allow for particular aspects of communities to be analyzed as they are impacted by crisis events, providing insight into their resilience. The key benefits of social media analyses for understanding community resilience are (1) the flexibility of temporal and spatial scales of analysis, (2) the ability to quantify community transformation. These two points are paraphrased based on a paper by Rachunok et al; written during my graduate studies –currently under review– but not included in this dissertation.

The first key benefit of social media for understanding community resilience is the ability to align the spatial and temporal scales of analysis with the spatial and temporal scales of decisionmaking. A critique of resilience as a design goal is that decision and policy making can occur at scales and in ways incompatible with the scales of impact [16,52]. The flexibility of social media analysis can be tailored to provide specific insights which are useful for immediate and real-time decision making. The second benefit of social media data is the flexibility in analysis with respect to how a community is being impacted. Community resilience as measured by a range of attributes approach is inherently limited to the extremes of the attributes. For example if owning a vehicle is considered as a positive aspect of a community, then community changes focused on improving public transportation –which is generally beneficial to communities but may reduce vehicle ownership– may have counter intuitive or negative impacts on a 'range of attributes' approach. With the extensive volume of social media data available, analyses can be tailored to quantify specific

policy aims.

1.4 Engineering Systems Resilience

One area in which the effects of climate change and climate extremes are extremely costly –and is a domain in which utilizing resilience as a goal should be carefully considered– is in the design of engineering systems. This section provides background on how the resilience of engineered systems can be quantified by first discussing how engineering systems can be modeled, followed by an overview of how disruptions can be modeled in networked system.

1.4.1 System Representations

The breadth of systems which fall under the category of infrastructure and the importance of protecting the integrity of these systems leads to a large body of existing literature modeling the resilience of infrastructure. Approaches used to model the resilience of infrastructure to disruption from hazards can be generally categorized into (1) statistical machine learning techniques, (2) engineering fragility based models or (3) graph (or network) theoretic approaches [53–55]. This section outlines specific methods and approaches used to model infrastructure resilience.

Statistical machine learning techniques use spatio-temporal information about a given region combined with information about disruptions to train statistical predictive models which estimate the loss in infrastructure performance [56]. Information such as area population and demographics, infrastructure component types, land use, and hazard intensity are used as predictors for infrastructure performance models such as service loss in a region, or duration of service outage [56–58]. These models accurately predict a priori damages due to a disaster, however they are unable to characterize damage in non-stationary conditions such as an uncertain climate, changing infrastructure design, or fundamental shifts in demographics. Fragility models use the

physical properties of the components to study how they respond to perturbation [54]. This general concept has been applied to gas distribution systems [59, 60], cellular telecommunications network [61], and power transmission and distribution [62, 63].

Fragility-based models quantify the response of individual components in a larger system to outside stress. To do this, a mapping between particular parameters of threats—such as windspeed in hurricanes or ground acceleration in earthquakes—and the probability of a component failing is created based on the specifics of the component. In this way, the operational response of system components can be estimated under steady-state, or perturbed conditions [54, 61, 64, 65]. Fragility-based models provide insights into specific relationships between disruptions and an engineering system, however owing to the system-specific information used for analysis, they provide limited scalability and generalization to larger, systems-level analyses.

Graph-theoretic models of infrastructure represent infrastructure systems as a network, allowing for simple—and in most cases tractable—calculation of performance measures within the system. Measurements of the overall size, degree of connectivity, length of paths between vertices, and degree of clustering are easily computed from a network-model and provide insight as to the general performance of the system [66–68]. Network-based models have been used to model the electricity transmission and distribution systems [69], urban water drainage and water distribution systems [70, 71], as well as transportation systems, and social mobility [72]. The relative simplicity of network-based representations make them well suited to studying complexities in the perturbation of systems due to disruptions [62, 73]. Graphs representing a system in which the components interact can be used to model how the failure of one vertex may propagate through the network [74]. Theoretical analysis has been done in which failure likelihoods are drawn from certain probability distributions, showing the existence of a critical fractions of node failures for which the failure will cascade to the entire network. This holds in single, and interdependent infrastructure systems [75].

Resilience Measurement Generally, engineering systems resilience is conceptualized as a systems ability to prepare for, respond to, and recover from some manor of disruption, which encompasses a large time horizon [55, 76]. Accordingly, resilience as an attribute or process of an engineered system is based on the system's responses to disruption [77], the dynamics of the system throughout the disruption, and its post-disruption behavior [78]. Definitions of engineering-systems are most typically associated with a performance metric which allows for the quantification of the resilience of the system. These measures can be deterministic or stochastic, and vary in their time horizon, but they ultimately look to provide a quantitative assessment of a system as it is disrupted, broadly defined [55].

One highly-cited definition of engineering resilience is Bruneau et. al.'s four R's: *robustness, rapidity, resourcefulness, and redundancy* [79]. Robustness is the strength of the system and its ability to prevent disruption; rapidity is the speed with which systems return to original states (the rate of return to normal), resourcefulness is the system's ability to apply capital (tangible and intangible) in response to a disruptive event, and finally redundancy which is the extent with which the system can sustain damage and still function. All of these are quantifiable metrics of a system which can be used to measure its engineering-system resilience.

Many extensions of this exist in similar frameworks, which either, (1) change the performance metric or (2) change the calculation done with the performance metric. Performance metrics are generally defined as the ability of a specific system to perform a user-defined function. Performance metrics implicitly defines the boundaries of the system as everything which can affect the performance metric and excludes all else. Typically, these performance metrics are system-specific, having previously been defined for waterways, [80] transportation systems [81], electric power infrastructure [82], economic systems [83] and many others. A given performance metric can be monitored throughout the life cycle of a disruption, and calculations can be made based on performance measures to create a numerical representation of engineering systems resilience.

There are a significant number of methods which propose calculating system resilience from performance metrics. [29,84–90] Bruneau specifically creates a tractable formula for resilience based on a system performance metric, in which the resilience is the complement of the integral of the performance drop for a system throughout the disaster cycle. This creates the concept of the “Resilience Triangle” as the area under the performance-measure curve throughout a disruption (Fig. 1.1).

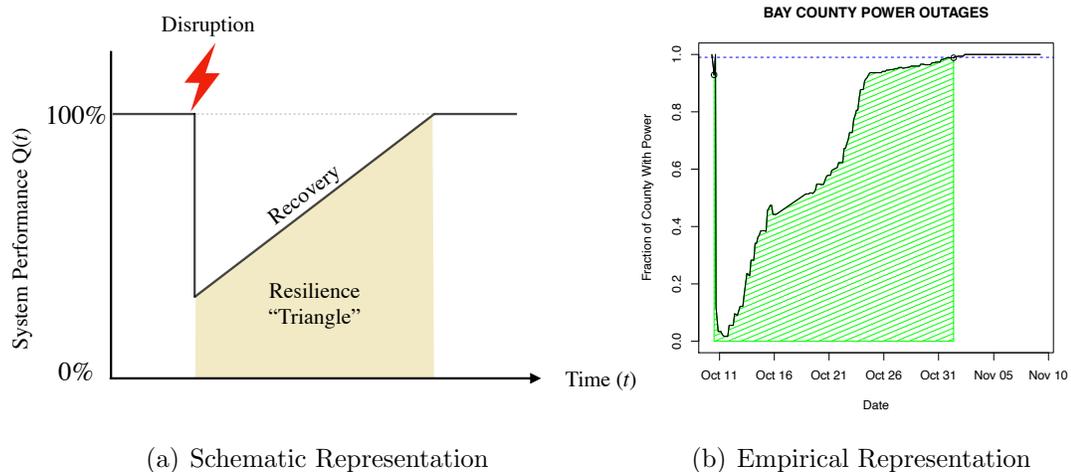


Figure 1.1. (a) Shows the conceptual representation of the resilience triangle, as the area under the performance curve in tan, while (b) shows an empirically measured triangle in Bay County, FL during 2018 Hurricane Michael (Figure from Chapter 5)

As an example of how this is used, the data shown in Figure 1.1(b) was originally recorded by the Florida Division of Emergency Management approximately every 3 hours during Hurricane Michael [91] and shows the fraction of Bay County Florida with access to power during the hurricane. In this case, the performance metric chosen to represent the status of the power system at time t , $Q(t)$ is the fraction of customers with power (the y-axis in Figure 1.1(b)).

Using calculations from Attoh-Okine et.al and Ayyub et. al. [89,92], the engineering resilience R , is calculated as is the area under $Q(t)$ from point at which the

system is impacted, t_0 , until the system is fully restored, t_f . This integral is scaled by the difference between t_f and t_0 . Formally, R is calculated as:

$$R = \frac{\int_{t_0}^{t_f} Q(t)}{|t_f - t_0|} \quad (1.1)$$

Using the fraction of a county with access to power as a performance metric and calculating resilience in this way, a county with $R = 0$ would be one for which $\int_{t_0}^{t_f} Q(t) = 0$, the numerator of R . In practical terms, this would mean the county lost all power immediately upon any sort of disruption, and the power remained inoperable until it was all restored at once. At the other extreme, $R = 1$ occurs when there is no deviation in the performance. Accordingly, improvements to this calculated measure of resilience could be made by increasing the ‘slope’ of the triangle (by shortening the recovery time), decreasing the magnitude of the disruption (shortening the initial drop in performance), and by changing the functional form of the recovery from a straight line to a concave down curve. Of these aspects of system performance which can contribute to overall resilience, characterizing the *impact* of a major disruption (such as a natural hazard or crisis event) is a component that this dissertation makes significant contributions to. The following sections describe how the impacts of major disasters on these systems can be represented mathematically.

1.4.2 Disruption Impacts in Engineering Systems

Critical to studying the resilience of infrastructure is an understanding of the method by which disruptions are represented in the system. In infrastructure resilience studies, a disruption is simulated, the disruption is mapped to its impact in the system and a partial or total failure is induced in the system as a result [55]. Characterizing the disruption is critical to contextualizing how resilient a system is to given failures. As one of the previously mentioned key takeaways from the measurement of community resilience, the specifics of individual threats and hazards must

accounted for when improving the adaptive capacity of a community. The methods for simulating disruptions and failures vary in their complexity and the degree of information included in them. Fragility-based models utilize representations of threats based on hazard parameters based on the specific fragility mapping. For example, if fragility curves are being utilized to relate hurricane wind speed to the probability of telephone poles failing, then the disruption representation will be a spatial distribution of windspeeds [64]. A disruption -or series of disruptions- is/are created and properties of such are used to probabilistically generate failures in the system.

Many infrastructure systems analyses use random component (a physical component in a fragility model, or nodes and edges in a graphical model) failures as the general form of the disruption [59,60,73]. Components in the network fail with a fixed and predetermined likelihood. This is representative of general system degradation - in which components are assumed to fail independently- or a naive attack on a system in which an actor only effects a fixed number of elements in the system. Targeting is another commonly used technique in which failures are induced in the network in accordance with a network or component property. Examples include targeting the component with the highest degree, graphical betweenness, maximum system flow, or a pre-determined ‘most important’ component in the system [62,64,65,69,74]. Targeting is representative of an omnipotent attack in which an agent wishes to cause maximum harm to a system, and additionally can create a worst-case scenario for system failures. Other studies have considered ‘localized’ disruptions in which failures are initialized in small connected components but they have been in an effort to replicate previous incidents [60,74].

1.4.3 Interdependence

An additional component of the study of resilient infrastructure is the inclusion of interdependence between infrastructure systems. Interdependence between systems

can increase vulnerability to natural or man-made disasters in multiple ways such as (1) propagation of failures across system couplings, (2) simultaneous failures affecting all systems, and (3) poor endogenous system performance causing excess stress in the coupled systems [93]. The interdependence does not need to be exclusively a physical dependence. Interdependence between infrastructure has been previously categorized by the degree and manor by which the systems interact [54]. Among the existing categories of interdependence are physical, geographical, cyber, and logical [94], functional and spatial [95], and budgetary and economic [93]. The breadth of the types of interdependence lead to numerous methods for studying their effect. Ouyang (2014) categorizes these modeling approaches into the following categories (1) empirical, (2) agent based, (3) system dynamics, (4) economic theory, and (5) network-based [54].

Network-based models are particularly common when studying interdependent engineering systems. Percolation theory and network diffusion are early examples of studying the spread of information (in this case disruptions) through interdependent networks [39]. Recent works have demonstrated how failures can propagate through interdependent networks in situations where the interconnected graphs both follow certain theoretical constructions, namely exponential degree distribution of equal size [96]. Other work has studied the effect of interdependence between the electric power grid and other utilities and concluded the topology of the inter dependencies can slow failure propagation during disasters but hinders daily operational performance [65]. Nan (2017) proposes an integrated framework for measuring the resilience of coupled infrastructure by measuring the *absorptive* and *restorative* capacity of interdependent networks, however the focus is on developing a single metric for resilience evaluation [97]. Interdependence has additionally been studied in network models which include resource buffers -such as a stored commodity dependence between systems [98]. Dueñas-Osorio (2007) performed an analysis of the degree of interdependence between networks, however the procedure focused on the static sen-

sitivity of network properties to the degree of interdependence between systems [99].

1.4.4 Transformation in Engineered Systems

In the engineering systems resilience domain, there is scant body of literature in understanding how a engineering systems may exhibit transformation. Generally this is captured in the concept of antifragility. An antifragile system is one demonstrates higher performance after a disruption than it did before; indicating a particular type of adaptation to the disruption [97,100]. These systems re-organize and adapt swiftly leading to the improved post-disruption performance [100]. However, as antifragility is still within the bounds of measuring systems via performance metrics; it is limited in the degree of transformation it can capture. Transformation still exists within the confines of the performance metric; that is to say a fundamental regime shift may occur but will not be identified unless it directly impacts the metrics used to evaluate system performance [16].

1.5 Summary of Included Papers

1.5.1 The Sensitivity of Electric Power Infrastructure Resilience to the Spatial Distribution of Disaster Impacts

As discussed in Section 1.4.1, network-based representations of infrastructure are utilized to model their resilience to outside disruption. Methods for simulating major disruptions and natural hazards are based on either random attacks –representative of general system degradation– or targeted attacks –representative of specific threats such as terrorism. The first paper included in this dissertation develops improved characterizations of network disruptions by simulating outages which follow a spatial patterns [101]. Specifically, the hypothesizes that changing the *spatial distribution* of

impacts through the network will have significant impact on measurements of system performance.

A network-based model of the electric power distribution grid in Mobile, Alabama is used as a test bed and three methods for generating disruptions in the system are tested. The first method simulates the impact of disruptions randomly in the system in line with existing studies; the other two use tree-based algorithms to create patterns of outages which are spatially *clustered*. Three performance measures are used to evaluate the network immediately after each type of disruption and after repairs are simulated in the system.

Controlling for the size of the disruption, the results indicate significant differences in the every performance metric evaluated between random and spatially-distributed outages throughout the failure and recovery process. Additionally, the system's failure and recovery after random outages showed antifragile properties, but that did not occur when utilizing spatially arranged outages.

Put in the context of the Mobile case study, a random disruption which impacts 60% of the components in the power distribution network leave 33-48% of the population with electricity after the initial impact compared to 26-53% using spatially configured outages. This work demonstrates that randomly generated outages may fundamentally mis-characterize the impacts of disasters in network-based infrastructure models and contributes fundamentally to improving the fidelity of network-based infrastructure models for representing the impacts of major disasters.

1.5.2 Interdependent Infrastructure System Risk Resilience to Natural Hazards

A follow-on work was completed [102] which considered the impact of the spatial distribution of outages in interdependent networks. The results are similar to the single-system case in which the effects of the spatial distribution of impacts are

detectable in both the network with outages, and any interdependently connected networks.

1.5.3 Twitter and Disasters: A Social Resilience Fingerprint

The third work included as a component of this dissertation provides a novel method for quantifying how communities respond to major disasters by analyzing social media [103]. As discussed in Section 1.3.3, social media analysis is a nascent area providing methods to quantify how communities respond to disruptions and crisis events. Existing methods for quantifying community resilience are based primarily on using indicators to measure attributes of communities hypothesized to contribute to resilience –called the ‘range of attributes’ or ‘indicator-based’ approaches in Section 1.3.1. These approaches, however, are limited in that they only capture community change if it occurs in the specified indicators, and are limited in the scale of analysis they can provide by the scales of available indicators. Utilizing social media-based analysis of communities respectively addresses these gaps by capturing emergent trends present in the discourse of communities, and allowing for analyses at multiple temporal and spatial scales.

The third work included in this dissertation develop which examine social media data and quantify public discourse during crisis events. The method, called the *Social Resilience Fingerprint*, measures public perception of a crisis event through social media through by leveraging a theoretical understanding of the attributes of a community which make it resilient. I compute the resilience fingerprint for 14 different crisis events and evaluate how different types of events (hurricanes, earthquakes, political events etc.) manifest in the social media discussion and identify themes present in the discourse. The results indicate that there are identifiable signatures present as a result of each type of crisis, as well as highlight the aspects of community discourse which contribute to this identifiability.

1.5.4 Improving Operational Measures of Community Resilience

The final work included in this dissertation address many of the issues outlined in Section 1.3.2 regarding the inability of current methods to capture how communities transform as a component of their resilience. Present methods for operationalizing resilience concepts in communities and engineered systems focus primarily on the efficient and rapid recovery of systems back to the status quo. This focus on stability, has been criticized for limiting the potential of communities to transform. To address this gap, this work develops a technique called the *Contrastive Community Network* (CCN), which is utilized to quantify how communities transform as a result of disruption. The CCN uses unsupervised statistical learning techniques to understand the relationship between communities and quantifies what aspects must change for those relationships to fundamentally change. In the paper, I compare the technique to existing methods for identifying factors contributing to the resilience of the community and find that 55% of risk factors found to contribute to resilience utilizing existing methodologies provide no avenue for communities to transform.

2. THE SENSITIVITY OF ELECTRIC POWER INFRASTRUCTURE RESILIENCE TO THE SPATIAL DISTRIBUTION OF DISASTER IMPACTS

Chapter 2 has been previously published in the journal *Reliability Engineering & System Safety*.

Rachunok, Benjamin, and Roshanak Nateghi. “The sensitivity of electric power infrastructure resilience to the spatial distribution of disaster impacts.” *Reliability Engineering System Safety* 193 (2020): 106658.

2.1 Introduction

Defined broadly, resilience is an emergent property of a system which manifests as the result of an iterative process of sensing, anticipation, learning, and adaptation to all types of disruptions [9]. Using this definition, resilience must be studied at a system-wide level, where the resilience of an entire system is studied in the context of hazards and disruptions. Characterization of the resilience of a complex system, therefore, is inherently a comprehensive analysis of that which acts against it. This system–disruption paradigm allows for the study of a wide range of interaction-based entities from ecological plant–pollinator relationships [104, 105] to the psychological resilience of families to trauma [106].

In the context of engineering urban systems, the resilience of a critical infrastructure (e.g., the electric power grid, telecommunication networks, natural gas, water network, etc.) *includes* study of the recovery from failures induced by hydro-climatic extremes and seismic events as well as acts of terrorism. Critical urban networked infrastructure is well-represented by a graph [39]. Subsequently, disrupting a graph requires removing or disabling fractions of the system consistent with an exogenous threat or hazard.

In this paper, we use a graph-theoretic approach to show that small changes in the spatial characteristics of a disruption to a system radically change the characteristics of system performance as a disruption is repaired over time. Whether the recovery is measured in-terms of network-based performance metrics or by the extent of impact on stakeholders, our results indicate that the measured resilience of a system is heavily dependant on the spatial characteristics of the initial disruption. We conduct this study in the case of an electric power distribution grid impacted by a major landfalling hurricane. We generate different *spatial* distributions of initial disruptions to a power grid and study their impact on graph-theoretic measures of network connectivity as well as the number of customers without power. The remainder of this paper is as follows: Section 2.2 introduces relevant other works, Section 2.3 outlines the data and methods used for this analysis, and finally Sections 2.4 and 2.5 detail the results and conclusion respectively.

2.2 Background

Network analysis deals with the study of graphs or networks. Networks are “a collection of points [referred to as *vertices* or *nodes*] joined together by pairs of lines [referred to as *edges* or *links*].” [39] The edge-vertex pairing lends itself to be an intuitive mathematical object for which to model phenomenon such as animal and plant interactions [107], academic authorship, urban infrastructure design [67] [108] and—most relevant to this work—electric power infrastructure [62,73,99,109]. Representing a system as a network allows for simple—and in most cases tractable—estimations of system performance. Measurements of the overall size, degree of connectivity, length of paths between vertices, and degree of clustering are all easily computed from a network model and can provide a myriad of insights about the system being represented [66]. Graphs representing a system in which the components interact can be used to model how the failure of one vertex may propagate through the network [74]. If failure likelihoods are drawn from certain probability distributions, there can ex-

ist critical fractions of node failures for which the failure will cascade to the entire network. This holds when multiple networks are coupled together [96].

Network-based approaches have been widely used to model the resilience of infrastructure [67, 110, 111]. This is in addition to conceptual frameworks [9, 79, 112], highly detailed hazard simulations [58, 61, 113, 114], and statistical and machine learning approaches [56, 57, 115–117]. All of this work contributes greatly toward improving the resilience of infrastructure by advancing theoretical understandings in networks science [111], addressing particular infrastructure inefficiencies [118], and improving policy decisions [53].

Generalized graph-theoretic resilience analyses commonly model disruptions by assigning a probability of failure to each vertex in the graph [74, 96, 111]. The random pattern of outages fits within a probabilistic formalism allowing for a theoretical understanding of network properties, but provides little realism in the spatial pattern of disruptions. Many of the infrastructure systems analyses continue to use random vertex failures as the general form of the disruption [59, 73, 119]. Degree targeting is another commonly used technique in which failures are initiated at vertices with the highest degree [62, 74, 99, 120, 121]. This method is representative of a targeted attack in which an agent wishes to remove nodes which connect to a large portion of the network, however, there is no restriction on the spatial distribution of the failures. Similarly, other vertex properties have been used to motivate targeting such as betweenness [62] or maximum flow [121]. Localized failures—in which failures are initialized in small connected components—have been previously studied, however with limited scope; focusing primarily on repair strategies [74], or to replicate previous incidents [119].

It should be noted that many previous studies consider disruptions to infrastructure which are -in some way- spatially organized either through explicit specification [88], fragility curves [122], or reliance on historical data [123]. However, to our knowledge the inclusion of spatially structured and non-spatially structured disruptions is secondary to the development of an optimization [124–126] or recovery model,

or resilience measurement algorithm [123, 127]. This work is the first to focus on the explicit impact of the spatial distribution of outages, which we perform by using general, network-based modeling paradigms.

In this work, we isolate the importance of accounting for the spatial distribution of a disruption and show that inducing changes in *only* the spatial distribution significantly impacts measurements of system performance. Specifically, the goal of the analysis is not so much to propose a particular spatial pattern of disruption over another, but to demonstrate the importance of considering the shape of disruptions in estimating infrastructure recovery. We present the results in a case study of an electric power distribution grid’s response to a hurricane. The electric power distribution system has been identified as a critical component of assessing the vulnerability of the electric power grid to severe-weather disruptions such as hurricanes, with approximately 90% of outages occurring at the distribution level [63].

2.3 Methods

As previously mentioned, to investigate the sensitivity of infrastructure system performance to the spatial distribution of disruptions, we present the case of an electric power distribution system’s recovery after a major landfalling hurricane. Specifically, we focus on the impact of the *spatial* distribution of hurricane-induced disruptions on the performance of an electric power grid located in the Gulf Coast of the U.S. (Figure 2.1)¹. We do this by simulating large-scale disruptions in the distribution grid, mapping the hurricane-induced disruptions to component failures (outages) in a distribution-level power grid and studying the sensitivity of the resilience of the system to the spatial distribution of the disruption. The simulated outages are subsequently repaired over time, replicating the actual recovery of the power grid from the hurricane disruption so as to study the dynamics of the system’s recovery.

¹The specific community on the Gulf-Coast is withheld for privacy reasons but represents a mid-sized metropolitan area

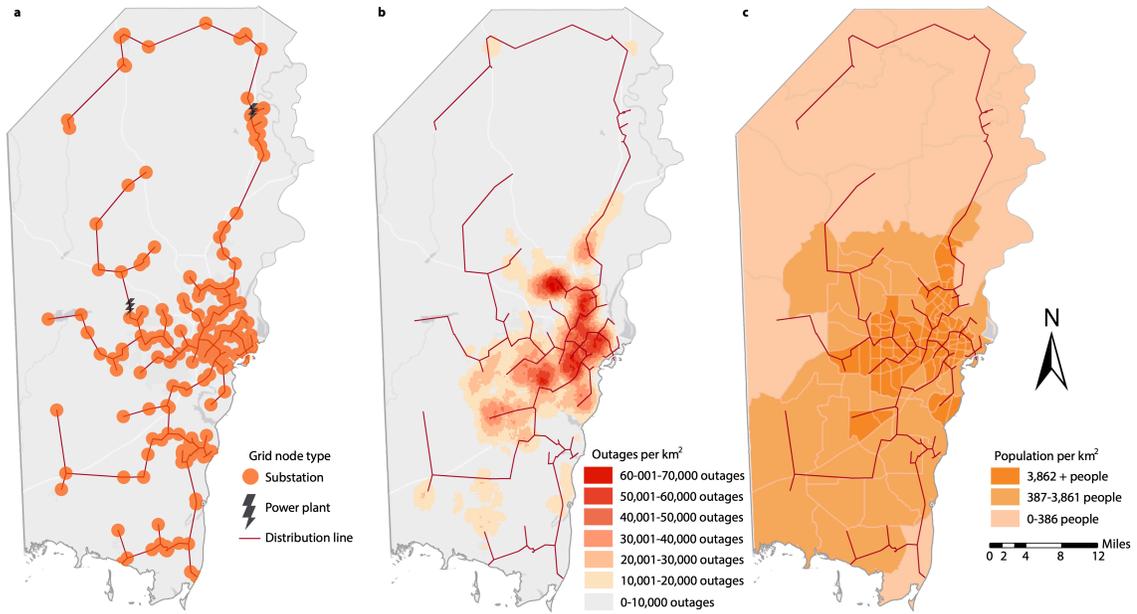


Figure 2.1. The case study network situated in the Gulf Coast of the U.S. **a** The layout of the electric power grid placed over the county. **b** The density of customer-level power outages during Hurricane Katrina with the network overlain. **c** Census-tract level population density for the corresponding area.

2.3.1 Electric Power Network

The city for which this analysis is being performed provided GIS files including the location of all of the county’s power substations. These are used to locate the position of the nodes in the test network. There are 221 substations and 2 power plants in this data. As we were unable to retrieve information on the connections between the substations, nodes are connected using a minimum spanning tree to establish the edges of the graph. A minimum spanning tree represents a radial network, common among electric power distribution systems [128] The resulting graph has 223 vertices and 222 edges.

Algorithm 1 Breadth-First Search

```

1: procedure BFS(graph =  $G$ , root =  $r$ , size =  $n$ )
2:    $Q \leftarrow$  empty list of vertices to search
3:    $T \leftarrow$  empty list of vertices in the tree
4:   append  $r$  to  $Q$ 
5:   while  $|T| < n$  do
6:     consider  $v$ , the first element of  $Q$ 
7:     remove  $v$  from  $Q$ 
8:     append  $v$  to  $T$ 
9:     for all  $w$  in  $neighbors(v)$  do
10:      if  $w$  is not in  $T$  then
11:        append  $w$  to  $Q$ 
return  $T$ 

```

2.3.2 Disruption generation algorithms

In this section, we describe the different disruption patterns evaluated in this study. All cases described cause failures in 60% of the vertices, and this failure proportion is kept constant through all trials. This is in accordance with the actual impact of Hurricane Katrina on the electric power distribution network under study. As previous work primarily focuses on analyzing randomized failures, we use random outages as a base for comparison with previous studies. In simulation replication, a different set of vertices is chosen at random such that 60% of the network is inoperable. The random disruptions form a *control sample* as there is explicitly no spatial association among the initial disruption.

To evaluate how the spatial characteristics of the disruption impact the network, additional simulation trials are performed using disruptions generated by search trees. Disruptions are generated using both a Breadth-First search (BFS) and a Depth-First search (DFS) tree [129] as both create spatially constrained patterns of outages while using no intrinsic information about the individual vertices. Details of the algorithms used to generate the disruptions are listed in Algorithms 1 and 2.

A BFS begins at a random vertex in the network and failures propagate to all neighbors of that vertex before extending to neighbors-of-neighbors. As the size of

Algorithm 2 Depth-First Search

```

1: procedure DFS(graph =  $G$ , root =  $r$ , size =  $n$ )
2:    $Q \leftarrow$  empty list of vertices to search
3:    $T \leftarrow$  empty list of vertices in the tree
4:   append  $r$  to  $Q$ 
5:   while  $|T| < n$  do
6:     consider  $v$ , the first element of  $Q$ 
7:     remove  $v$  from  $Q$ 
8:     append  $v$  to  $T$ 
9:     if  $w \in neighbors(v), w \notin T$  then
10:      append  $w$  to front of  $Q$ 
return  $T$ 

```

the failure is pre-specified, the failures continue until the BFS tree is the required size. This provides a method for generating localized clusters of failures. Similarly, a DFS outage pattern begins at a random vertex and progresses away from the root node as far as possible within the network before searching additional root-node neighbors. The spatial pattern of DFS trees are connected, but far less localized. These are referred to as the the BFS and DFS disruption methods for the remainder of the paper.

The search tree generation methods are computationally cheap, and are built entirely using the spatial structure of the network. The selection of these algorithms are motivated by existing research supporting the existence of tree-shaped outages in distribution systems owing to the hierarchical nature of electric power distribution [63, 130]. Here, we do not validate actual spatial distributions of outages against the BFS and DFS generation methods, but instead use these methods to isolate the significance of different spatial configurations of outages in the network on measurements of system performance. The initial distribution of outages for one simulation replication are seen in Figure 2.2.

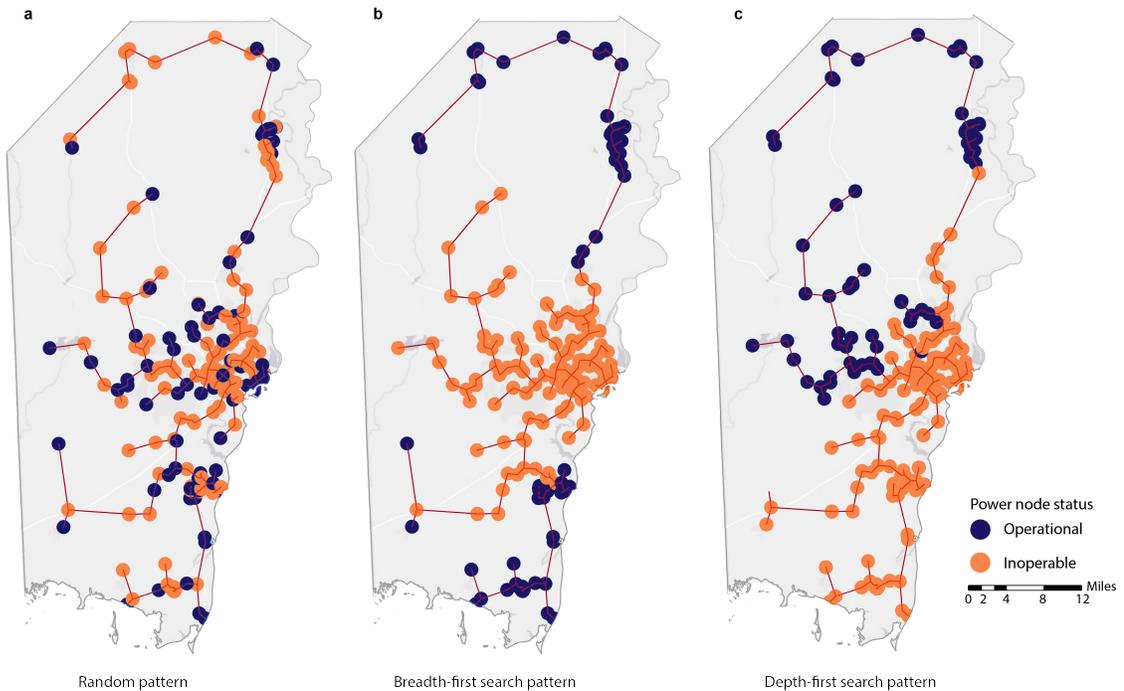


Figure 2.2. Outage generation types. The results of three outage generation techniques, each inducing failures in 60% of the grid. Figure **a** is one instance of an outage generated randomly. Figure **b** is a an outage generated using a breadth-first algorithm, while **c** is a depth-first algorithm.

2.3.3 Performance metric calculation

In order to characterize the networks as they fail and recover, we use two network-based measurements of system performance: *network efficiency* and *largest connected component*. We measure the *global efficiency* of the electric power network as it fails and recovers as one dimension of network performance. Global efficiency is defined as

$$\text{Eff}(G) = \frac{1}{n(n-1)} \sum_{i < j \in G} \frac{1}{d(i, j)} \quad (2.1)$$

where $d(i, j)$ is the distance between vertex pair i and j . Network efficiency as a concept was proposed as a measure of how efficiently a network exchanges information

[131] and has been previously used the context of power system resilience evaluation [73, 132] and used as a proxy for network performance [120, 133].

Additionally we measure the size of the largest connected component (LCC). This is defined as the number of vertices in the largest connected subgraph [39]. A connected subgraph is a subset of the vertices and edges for which a path exists between all pairs of vertices. LCC has previously been used to evaluate topological models [73] and provides a measure of the connectedness of the network (*ie* a fully connected network has a maximal LCC because every vertex is included in the largest cluster). LCC and efficiency have both been previously studied as performance measurements for network representations of power systems, and have been validated as system performance measurements when a broad range of vulnerability scenarios are evaluated [73].

2.3.4 Simulation Methodology

The recovery simulation generates initial disruptions via random, BFS and DFS methods then subsequently repairs vertices in the network. The rate of repair (*i.e.*, repaired vertices per time unit) is derived from the rate of outages seen in the gulf-coast power operator data. This rate is kept constant through all experiments. At every time step, the vertices to be repaired are chosen based on their contribution to the total network efficiency. The number of vertices to be repaired is first fixed based on the time dependent repair rate, then the set of vertices chosen for repair are selected from the subset of inoperable vertices which—if repaired—would maximally improve the network efficiency. Vertices are selected in a greedy fashion such that the selected subset maximally improves the efficiency of the network. The heuristic search is detailed in Algorithm 3.

Network statistics are recorded at each step and vertices are repaired until the network is fully operational. The simulation procedure is depicted in Figure 2.3. The process of creating disruptions and repairing is repeated 100 times for each disruption

Algorithm 3 Local-optimal search.

Here, GE is the global efficiency of a graph, and $F - R$ indicates the removal of vertices R from F .

```

1: procedure LOCALOPT(graph =  $G$ , failed vertices =  $F$ , repair =  $n$ )
2:    $R \leftarrow$  empty list of vertices to be repaired
3:   if  $|V(F)| = |V(G)|$  then
4:      $R =$  vertex with maximum degree
5:      $F = F - R$ 
6:     LocalOpt( $G, F, n-1$ )
7:   else  $|V(F)| < |V(G)|$ 
8:     if  $|V(F)| + n \geq |V(G)|$  then
9:        $R = F$ 
10:    else  $|V(F)| + n < |V(G)|$ 
11:       $R = f \in F$  s/t  $GE(G + f) \geq GE(G + f') \forall f' \in F$  and  $f' \neq f$ 
return  $T$ 

```

generation method to account for the inherent randomness in the generation of the initial distributions. The analyses were performed on a 16-core Intel Xeon W-2145 processor, each operating at 3.7GHz with 32GB of ram. Simulation, analysis, and resulting plots were all generated in R version 3.4.4 [134]. Network statistics were calculated using `igraph` [135].

2.4 Results

2.4.1 Static measures of impact

We first evaluate the sensitivity of the *static measure of performance*—i.e., the performance of the system at the moment the disruption occurs—to the spatial distribution of the disruption generated randomly as well as via BFS and DFS algorithms (Figure 2.4). To provide an equal comparison—and in accordance to real data from Hurricane Katrina—we present results which impact 60% of the network regardless of the method of outage generation. However, our extensive sensitivity analysis suggests that the results remained consistent when evaluating network failures ranging from 10% to 90%.

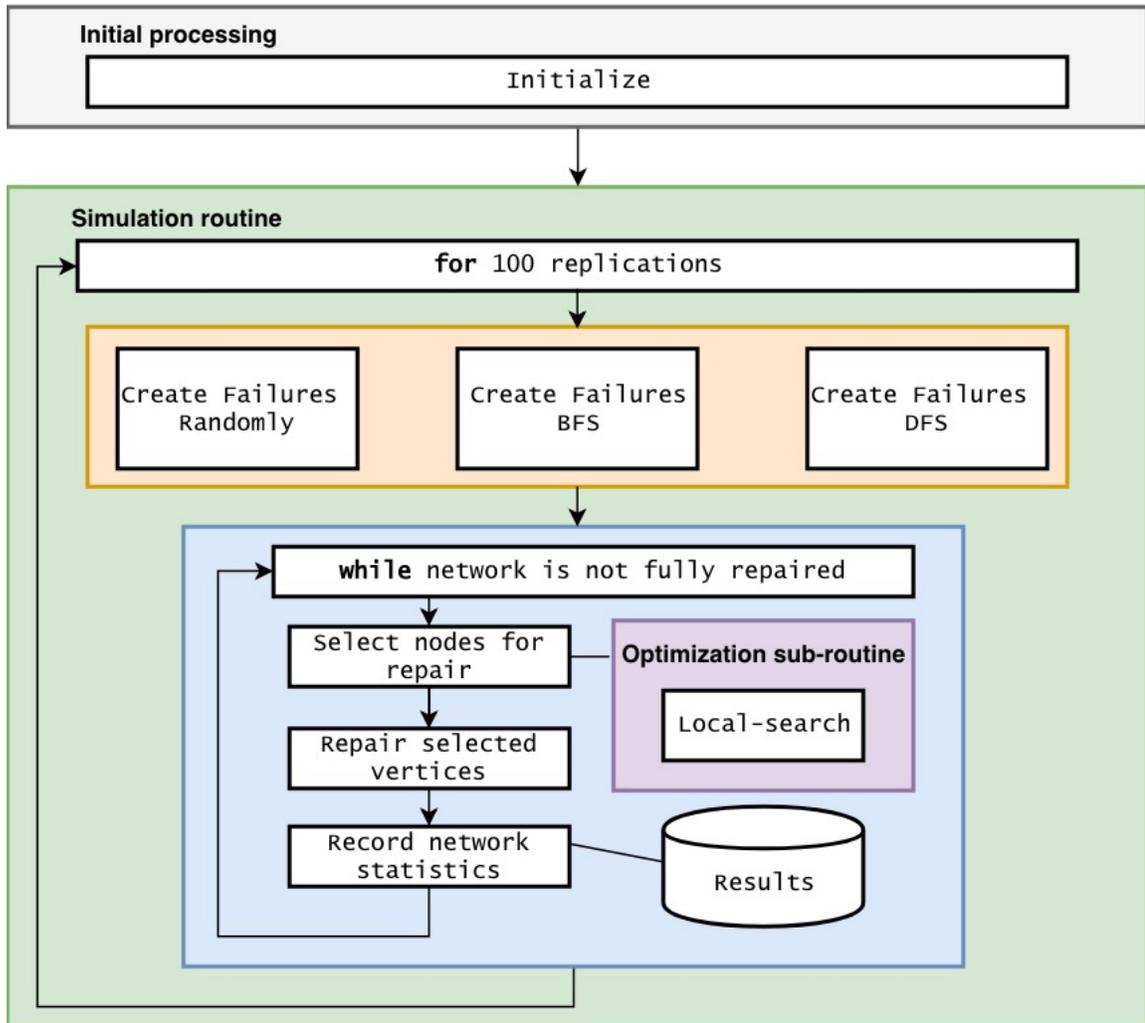


Figure 2.3. An overview of simulation methodology. The process here represents one simulation iteration.

Computed for 100 stochastic disruptions of each type, there is significant evidence that the disruption methods alter the resilience of the system. The mean efficiency of BFS- and DFS-constructed disruptions are 485% and 457% higher than randomly constructed disruptions respectively. Mean values vary significantly at each failure size as seen in Table 2.1. Mean LCC increases similarly with BFS disruptions—BFS increase of 595% over random, DFS increase of 494% over random (Table 2.3). Results additionally indicate sample variance increases for tree-constructed disrup-

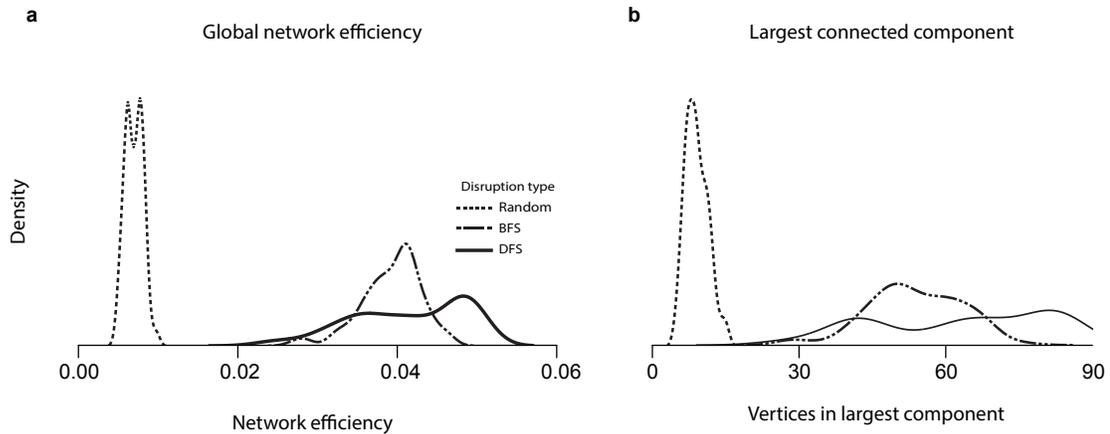


Figure 2.4. Static disruption comparison. Relative density of network performance after 100 disruptions for each disruption generation method. **a** is the network efficiency for all three disruption generation methods while **b** is the size of the largest connected component.

tions in both performance metrics as seen in Tables 2.1 and 2.3. In the case of the mean comparison, the distributions of efficiency and LCC values are compared using Kolmogorov-Smirnov (KS) two-sample tests and all comparisons are found to be statistically significant at a significance level of 0.01. Results of the KS tests are seen in Table 2.2.

The lower efficiency values and LCC of the random disruption method indicate greater disruption in the system. Lower network efficiency is representative of lower communicability among the network concomitant with greater static resilience to a disruption. Likewise lower LCC values indicate geographic sparsity among the network's operable vertices. While neither of these performance metrics directly map to the performance of a high-fidelity power-system simulation, they demonstrate the sensitivity of the spatial distribution of a disruption on generalizable measurements of system performance in a network model. Consequently any claim resulting from a measure of resilience is sensitive to the spatial characteristics of the initial disruption. Likewise,

Table 2.1.

Summary statistics for the distribution of efficiency for respective failure modes with the percentage of optimal network efficiency listed in parentheses. Failure fraction represents the fraction of the network which was induced as failed in each iteration. Results presented here are for failures in 60% of the network. Complete results are presented in Appendix Table A.1 and A.2.

Generation method	Mean	Standard deviation	Median	Min	Max
Random	0.0070 (20.68)	0.0011	0.0070 (20.68)	0.0047 (13.90)	0.0100 (29.33)
BFS	0.0414 (121.50)	0.0071	0.0420 (123.33)	0.0240 (70.62)	0.0494 (145.10)
DFS	0.0393 (115.33)	0.0038	0.0401 (117.80)	0.0270 (79.41)	0.0463 (136.13)

accounting for the spatial distribution of disruptions introduces greater uncertainty into our estimation of the resilience of a system.

The sensitivity of the resilience to disruption method additionally manifests when measuring the number of customers with restored power. Mapping the geographical location of each of the vertices in our network to their respective census tract allows us to allocate customers to each substation relative to their population density. Using this this approximation, an average of 40.60% of the customers retain power when disrupted randomly, versus 39.21% and 39.47% for BFS and DFS outages respectively. This similarity is expected as the disruptions are constructed to disconnect 60% of the substations in the network, leaving approximately 40% of the network operational. However similar to measurements of efficiency and LCC, the variance among population affected is higher for tree-based disruptions. Table 2.4 shows the distribution of the number of customers without power after the network is made inoperable. After random outages are induced in the system 33.57%–48.35% of the population’s distribution level power remains operational, while after BFS and DFS outages 26.54%–53.77% and 26.94%–48.95% of the population’s power remain

Table 2.2.

P-values for two-sample, two tailed, Kolmogorov-Smirnov tests between the efficiency and LCC of given initial failure methods and failure fraction. Results at the 0.6 failure fraction are presented in this article. Results use a significance level of $\alpha = 0.05$. Values of zero listed with one significant digit indicate $p < 1.11022e - 16$; this cutoff is the numerical precision of the machine used for computations.

Failure fraction	Efficiency			LCC		
	Random vs BFS	Random vs DFS	BFS vs DFS	Random vs BFS	Random vs DFS	BFS vs DFS
0.1	0	0	0.0039	0	0	0.0541
0.2	0	0	0.0004	0	0	0.0001
0.3	0	0	0.0014	0	0	0.0000
0.4	0	0	0.0014	0	0	0.0000
0.5	0	0	0.0001	0	0	0.0000
0.6	0	0	0.0000	0	0	0.0000
0.7	0	0	0.0000	0	0	0.0008
0.8	0	0	0.0000	0	0	0.0000
0.9	0	0	0.0000	0	0	0.0000

Table 2.3.

Summary statistics for the distribution of largest connected component (LCC) for respective failure modes with percentage of the optimal value listed in parentheses. Results presented here are for failures in 60% of the network. Complete results are presented in Appendix Table A.3 and A.4.

Generation method	Mean	Standard deviation	Median	Min	Max
Random	9.05 (4.058)	2.32	9.00 (4.036)	5.00 (2.242)	15.00 (6.726)
BFS	62.94 (28.22)	17.71	66.50 (29.83)	28.00 (12.56)	83.00 (37.22)
DFS	53.75 (24.10)	9.54	53.00 (23.77)	28.00 (12.56)	76.00 (34.08)

operational respectively. This represents an 88% increase in the uncertainty of the performance estimates. Providing estimates of uncertainty is critical to decision makers for the accurate characterization of the resilience of a system [136].

Table 2.4.

Summary statistics for the distribution of percent of county customers without power in a static analysis. All numbers represent the fraction of the total population of the county without power.

	Mean	Std Dev	Median	Min	Max
Random	0.5928	0.0351	0.5940	0.5165	0.6643
BFS	0.5909	0.0676	0.6079	0.4623	0.7346
DFS	0.6151	0.0651	0.6053	0.5105	0.7306

2.4.2 Dynamic measures of impact

We also evaluate the *dynamic performance*—i.e., time dependant performance metrics—under separate initial disruption methods as the power grid is repaired (Figure 2.5). The system performance—characterized by efficiency and LCC—is then measured over time as the system recovers. This is done to characterize the dynamic resilience of the grid under each disruption generation method, *ceteris paribus*.

Despite holding the recovery process constant, these results show the efficiency of the network differs greatly in overall functional form between random and spatially generated disruptions, indicating the recovery is significantly coupled to the spatial distribution of disruptions. Recovery from a random disruption pattern increases over time, reaching a maximum prior to all nodes being repaired (Figure 2.5e). This is an indication of the network exhibiting *antifragile* properties. Antifragility is a property by which a full reconstruction of the network is not optimal with respect to the chosen performance metric [100, 137]. In the context of network-performance measurements of an electric power distribution grid, antifragility indicates that a performance measurement rises above the optimal value prior to the system returning to its original

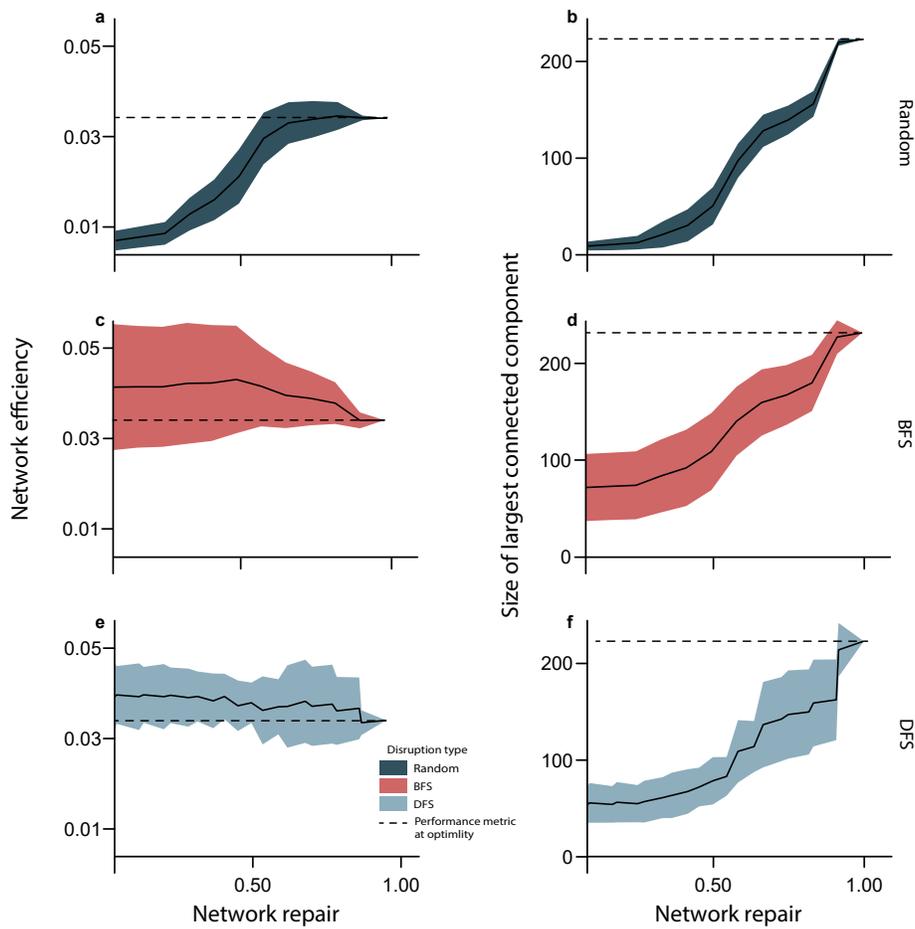


Figure 2.5. Performance metrics measured after the disruption over time for each disruption method. In **a-f**, the bands of uncertainty represent 95% confidence intervals sampled from the empirical density at each point in time. The black line is the mean of the observations. The x-axis is the relative-completeness of the network repair scaled by the total restoration time for each replication.

state, as evident by the concave response seen in Figure 2.5a [118]. As antifragility is considered an inherent property of a system [100], the lack of antifragility in spatially-constructed outage systems indicates that it is conditional on the choice of outage distribution. Spatially-constructed outages generally have a much higher efficiency

throughout but follow an entirely different functional form than the recovery from random disruptions. The deviation between mean efficiency is highest at the initial disruption and decreases over time. Similar to the static analysis, the variance is larger in the recovery from spatially characterized outages. Thus, failing to account for the spatial characteristics of the network disruption can drastically change implications drawn from the associated resilience analysis. A key difference is the lack of antifragility in the distribution electric power network with spatially characterized outages.

The difference between the disruption generation techniques is diminished when comparing the dynamics of the mean LCC rather than mean network efficiency (Figure 2.5 b,d,f). Beyond the initial value of the LCC at the time of failure, there is little difference in the functional form of the recovery of the network. The size of the LCC in the network generally increases at an increasing rate when vertices are repaired in the network, the primary difference being the initial size of the LCC after failures are generated in the network. These estimates of system recovery are therefore dependant on the spatial characteristics of the initial disruption; however, this result is sensitive to the performance metric used to measure recovery.

2.5 Conclusion

A key element of resilience is the ability of a system to respond to and recover from disruptions of unprecedented magnitude or unforeseen cause. By their nature, *all* disruptions will require recovery. This positions system recovery as a critical measurement in evaluating the multifaceted resilience of infrastructure systems. A holistic understanding of all types of community recovery is imperative for the continued adaptation to unforeseen challenges. However, these holistic understandings must be built upon a foundational knowledge of the interaction of disasters with the built environment. We contribute to the knowledge related to the interaction of the power distribution grid and hurricanes by providing a novel framework for

network resilience analysis which is agnostic to the specifics of the system, allowing for general insights about all facets of community recovery. Our framework for considering spatially-constrained disruptions can be applied to any hierarchical network within a community adversely effected by natural hazards. We plan to extend the work presented here by evaluating the impact of spatial distributions of outages on high-fidelity models of infrastructure systems.

We show that the post-disruption network-performance of the electrical power distribution grid is highly sensitive to the spatial characteristics of disruptions in the system. Consequently, any insights about general grid resilience which fail to account for the spatial characteristics of the hazard significantly misrepresent the impact of natural hazards on distribution-level electric power infrastructure. More specifically, through the repeated simulation of multiple methods of failure and recovery, we show that previous methods of evaluating disaster impact overestimate the certainty associated with the measurements of system recovery. We show via multiple avenues that improved characterizations of disaster impact significantly increase *both* the magnitude and uncertainty of the initial impact in the system. This difference holds through the duration of the recovery process; and when considering the dynamics of the system we find that emergent system properties such as antifragility are also dependant on the characteristics of the initial disruption. These differences are most striking when contextualized by their impact on the power distribution grid at a customer level. Our estimates indicate that the estimated range of customers with access to electricity varies from 33-48% of the county using previous methods, and up to 26-53% when using improved outage characterizations, highlighting the need for continued study of both the pattern of impacts due to natural disasters and the vulnerability of the electric power distribution grid. By demonstrating the sensitivity of the spatial distribution of outages on the electric power grid, we hope to encourage consideration of the spatial distribution of disruptions in conducting infrastructure resilience analytics.

3. INTERDEPENDENT INFRASTRUCTURE SYSTEM RISK & RESILIENCE TO NATURAL HAZARDS

Chapter 3 has been previously published in *Proceedings of the Institute of Industrial & Systems Engineering Annual Conference*. A post-review version is hosted on arXiv:1904.05763.

3.1 Introduction

Interdependence is inherent in many critical systems vital to the continuation of a nation's well-being [138]. Electricity, natural gas, transportation, and telecommunication are all provided by infrastructure systems which require bi-directional inter- and intra-system connection for optimal functionality. For example, telecommunication grids require continued power for operation, while the electric grid requires telecommunication networks to function [61, 120, 138]. The criticality of the goods and services provided by these systems necessitates the design of resilient interdependent systems. In this work, we study the disruption and recovery of interdependent systems after a major disturbance and quantify the influence of changes in the spatial distribution of hazards on overall system resilience.

Much attention has been given to the study of failures in interdependent networks from both a theoretical and applied perspective. Interdependence has been shown previously to improve overall system robustness to disruption [121] at the expense of reducing steady-state performance. Previous work has shown that interdependent infrastructure systems will respond differently to an identical hazard or disruption due to their individual components and their topology (*e.g.* telecommunications networks and water distributions will not be impacted similarly by a hurricane) [128]. Previous work has considered disruptions to the network which occur randomly [59, 121]

-indicative of general system aging and degradation- or via targeting [74] in which vertices are removed because of their importance. This work improves upon previous studies by considering the impact of changes to the *spatial distribution* of failures on system performance while controlling for the influence of the size of the disruption. We hypothesize that –contrary to previous analyses– the impact of hazards on interdependent systems does not follow a random pattern and may be clustered locally. To test our hypothesis, we change the spatial distribution of failures in each system and compare the resulting system performance immediately after failure and while the simulated systems are being repaired. This provides evidence to indicate that –when controlling for the size of the impact– system performance is significantly influenced by the spatial distribution of outages. We further show that a significant change in system performance can be measured in both systems if disruptions are assumed to impact each system with a separate spatial distribution.

3.2 Methods and Data

To evaluate the impact of different outages on measurements of system performance, we simulate the failure and recovery of two interdependent systems in response to different sizes and spatial distributions of disruptions. The performance of each system is measured as it fails and is repaired. What follows is an overview of the simulation methodology, the calculation of performance metrics, and the data used to construct the systems.

3.2.1 Methods

Our analysis of the interdependent systems uses two graphs -representative of two infrastructure systems- and couples them to create interdependencies among the systems. The two networks, g_1 and g_2 are generated such that

$$g_1 = G(V_1, E_1) \quad g_2 = G(V_2, E_2)$$

each are made up an edge set, E , and a vertex set, V . The size of the edge set and vertex set (*i.e.* the number of edges and vertices) of g_1 are $|E_1|$, and $|V_1|$ respectively [139]. The degree of each vertex is the number of edges to which it connects, and here is represented as.

$$\text{Degree of vertex } i = d(v_i), v_i \in V(g)$$

To generalize the connections between the vertices in opposing graphs, a dependence matrix D_{g_1, g_2} is used to relate elements of g_1 to elements of g_2 . D_{g_1, g_2} is defined as a matrix of size $|V_1| \times |V_2|$. Elements of the matrix represent individual component-level dependencies. Consequently, $D_{g_1, g_2}(i, j) = 1$ if v_j depends on v_i to function and $v_i \in g_1$ and $v_j \in g_2$. This allows for the representation of directional dependence in failures and recovery. if $D_{g_1, g_2}(i, j) = D_{g_1, g_2}(j, i)$, then we have an interdependence between components, and if $D_{g_1, g_2} = D_{g_1, g_2}^t$ then the systems are fully coupled insofar as $D_{g_1, g_2}(i, j) = D_{g_1, g_2}(j, i) \forall i \in [1, |V_1|]$ and $j \in [1, |V_2|]$. An example would be:

$$D_{g_1, g_2} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

in which v_1 through v_4 in g_2 depend on v_3 to function. This provides a general framework to relate one network to the other in the failure and recovery of the systems.

3.2.2 Failures

After g_1 and g_2 are coupled via D , outages are generated in each system. The methods of outage generation are discussed in subsequent sections. In the disruptions, the set of failures is comprised of two sets. First are the failures directly induced by the disruption's impact on the system. Second is the dependent impacts within or across systems. The initial set of failures -those induced by the hazard- in graph g are

denoted f_g^f and the subsequent dependent failures in graph g are denoted f_g^d . After the initial set of failures are generated in g_1 and added to $f_{g_1}^f$, dependencies in g_2 are identified via D . $f_{g_2}^d$ is updated to reflect the elements of g_2 which fail as a result of a failure in g_1 . $f_{g_1}^d$ is then updated based on $f_{g_2}^d$. The failures cascade across the two networks until no more dependencies are found. The process is repeated starting with $f_{g_2}^f$ and propagates until equilibrium. The results of the failure generation represent the total, initial impact of a disruption on the interdependent system.

3.2.3 Failure Generation Methods

We aim to evaluate how asymmetry in the impact across networks influences measurements of system performance, and to do so we evaluate three methods of disruption generation. The first are random disruptions in the system in which each node has an independent and identical probability of failure. Random failures are representative of system aging or general degradation. The second and third methods are derived from search trees and generate disruptions in the graphs which are spatially connected. The second method (BFS) uses a *breadth-first search* tree to create locally clustered distributions of failures around a randomly selected root node. The third method (DFS) uses a *depth-first search* tree to create a connected cluster of failures propagating away from a randomly selected root node and progressing away from the root to maximal length. Examples of each failure generation method are seen in Figure 3.1. The three disruption types are selected to isolate the impact of the spatial distribution of failures on the interdependent systems. In this way, we can evaluate how a disruption which induces failures asymmetric to the two systems, impacts overall system performance and measurements of system resilience. The three failure generation methods listed here are all used to generate the set of initial failures $f_{g_1}^f$ and $f_{g_2}^f$.

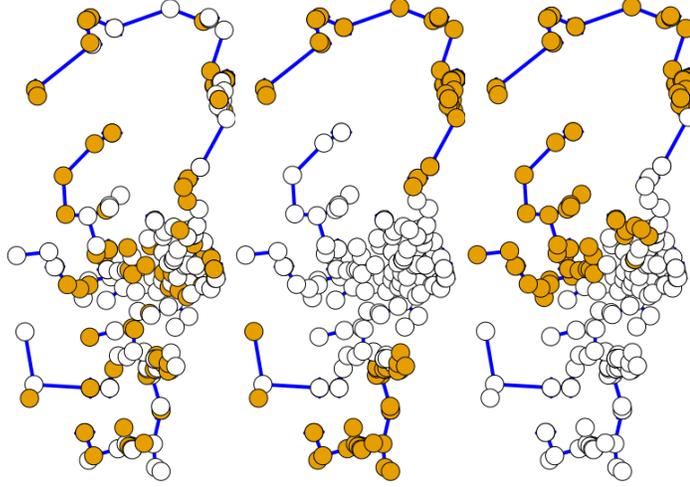


Figure 3.1. Examples of system disruption types, orange vertices are operational, white have failed. Left is a network after *random* failures, center is a network after *BFS* failures, and right is a network after *DFS* failures. All three represent a failure of 60% of the vertices in the system.

3.2.4 Recovery

After the initial failures in the system are generated, the system is repaired sequentially. $r_{g_1}^n$ is the n^{th} node repaired in the graph g_1 such that $r_{g_1}^n \in \{f_{g_1}^f \setminus \{r_{g_1}^1, \dots, r_{g_1}^{n-1}\}\}^1$. That is, the eligible nodes for repair in g_1 at step n are those which have failed directly ($f_{g_1}^f$) but have not yet been repaired ($r_{g_1}^1, \dots, r_{g_1}^{n-1}$). At step n , r_{g_1} and r_{g_2} are selected such to maximize the total system performance improvement. Total system performance is simply the sum of each individual network's performance. After r_{g_1} and r_{g_2} are repaired, any dependent failures (elements of $f_{g_1}^d \cup f_{g_2}^d$ connected to r_{g_1} and r_{g_2} via D) are also repaired. In this way, we can differentiate between repairs of directly failed elements of the systems and repair of elements which have only failed because of their dependency. The recovery and repair procedure is continued until both networks are fully operational.

¹In this notation, \setminus indicates the removal of the vertices $r_{g_1}^1, \dots, r_{g_1}^{n-1}$ from the set $f_{g_1}^f$

3.2.5 Data and parameters

The networks in our model are based on publicly available electric power distribution grid location and the natural gas pipeline layout of Mobile County, Alabama. The electric power distribution system contains 223 vertices and 222 edges, while the natural gas system contains approximately 25 vertices and 35 edges ². In this analysis, the failure and recovery of the system is simulated on randomly generated graphs of equivalent degree. The degree of interdependence is estimated based on the physical proximity of nodes in the system which results D having a matrix density of 0.01. Both networks are assigned a failure generation method (Random, BFS, DFS), and failures are generated such that 10, 20, 60, and 90% of the components fail - holding the size constant in each replication. Every parameter combination (failure size, generation method in g_1 , and generation method in g_2) is simulated 250 times, wherein each trial generates the failures randomly from one of the three methods.

Network performance is measured after the initial failures and is recorded throughout the recovery process. System performance is measured as the *global efficiency* of each network. Global efficiency is defined as

$$\text{Eff}(G) = \frac{1}{n(n-1)} \sum_{i < j \in G} \frac{1}{d(i, j)}$$

where $d(i, j)$ is the distance between vertex pair i and j . Network efficiency as a concept was introduced by Latora (2001) as a measure of how efficiently a network exchanges information [131]. It has been evaluated in the context of power system resilience evaluation [73] and used as a proxy for network performance [120, 133].

²Estimates of the gas pipeline network are taken from public-level aggregated pipeline locations available through the National Pipeline Management System

3.3 Results

3.3.1 Spatial Differences in Initial Disruption

Immediately after the failures in the system have completed propagating, we measure the efficiency of both systems in each replication. The distribution of the efficiency is listed for both systems in Appendix Table B.1 and density plots of the respective efficiency can be seen in Figure 3.2. At a fixed size of disruption, changing the spatial distribution of outages (or the *shape* of the outages) in either network impacts the overall system performance for both networks. Table 3.1 shows the results of two-sample Kolmogorov-Smirnov tests comparing the distributions of network efficiency for both systems after failures induced by different methods. The results of the KS tests show that there is a statistically significant difference in the performance of g_1 and g_2 when changing the spatial distribution of either network away from a random field of outages.

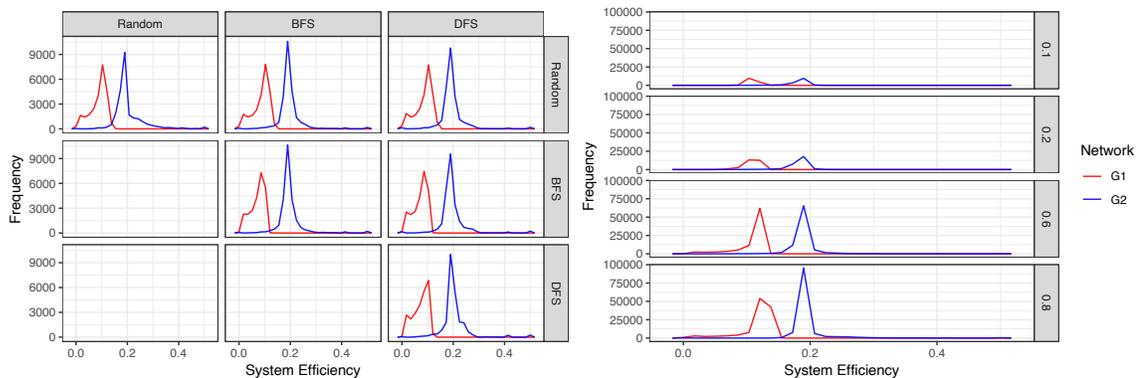


Figure 3.2. Differences in initial disruptions. Left are density plots of both graphs subset by the corresponding distribution of failures in g_1 (rows) and g_2 (columns). Right shows the change in performance measure as a function of failure size (rows).

Table 3.1.

Comparison of initial disruption types. P-values taken from 2-sample Kolmogorov-Smirnov tests. Values of 2.2×10^{-16} represent a p-value smaller than the numerical precision of R

Statistical test	P-value for difference in g_1	P-value for difference in g_2
Random-Random vs Random-BFS	0.00089	2.2×10^{-16}
Random-Random vs Random-DFS	1.576×10^{-9}	2.2×10^{-16}
Random-Random vs BFS-BFS	2.2×10^{-16}	2.2×10^{-16}
Random-Random vs BFS-DFS	2.2×10^{-16}	2.2×10^{-16}
Random-Random vs DFS-DFS	2.2×10^{-16}	2.2×10^{-16}

3.3.2 Changes in recovery of systems

As the network is repaired, we measure changes in the performance of both systems. Figure 3.4 shows the recovery of the systems as they are repaired subset by the distribution of outages in g_1 and g_2 . Similar to in the previous section, changes in the distribution of failures in either system induce changes in the overall recovery of the system. Additionally, changes in the distributions of outages - from random to a spatially-constrained outage- effect the variability of observations, with random outages exhibiting the highest variability among outage types.

In each simulation replication, the time is measured after failure until the system is exhibiting full performance. Because of redundancies in the network, it is frequently the case that the the system is fully operational *prior to* all elements being repaired. The Time to Repair (TTR) in this case is measured as the first time a system is performing optimally in a given replication. Figure 3.4 shows the TTR broken down by failure size and disruption type. As expected, larger failure sizes have higher TTR -corresponding to a longer time to repair. However changes in the time to repair can be observed when the distribution of failures is altered in either network.

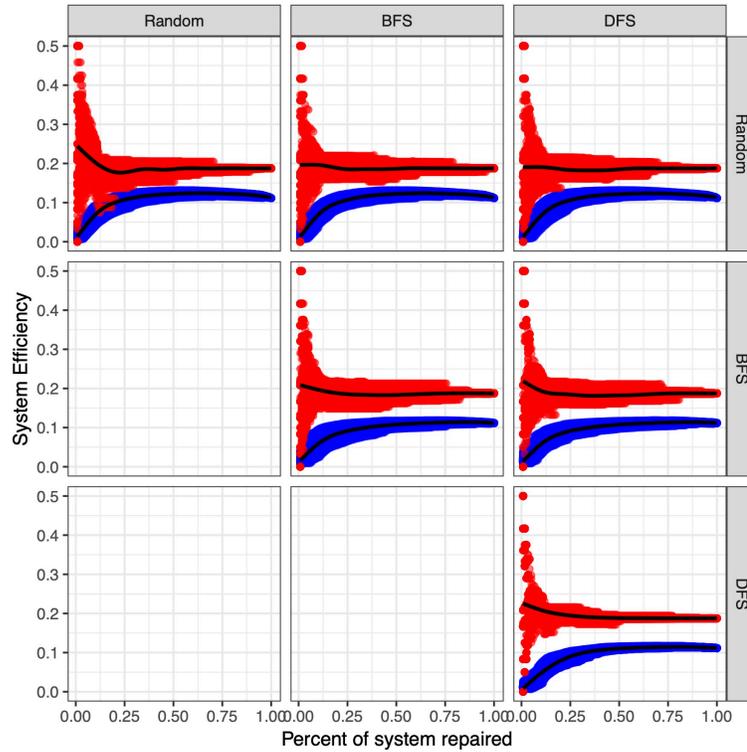


Figure 3.3. Recovery of system over time subset by the distribution of failures in g_1 (rows) and g_2 (columns). Red points are the system efficiency for g_1 , and blue are the efficiency of g_2 . Black lines indicate the mean of all replications at a given percentage of the system repaired.

3.4 Conclusion

In this work, we construct a simulation of interdependent networks, representing coupled infrastructure, which are subsequently disrupted and repaired. We hypothesize that a major hazard which disrupts interdependent systems will impact the constituent systems asymmetrically, inducing different magnitudes of failures and different spatial distributions of failures in each system. Via leveraging a rigorous simulation methodology to test our hypothesis, we provide evidence that the differences in the system performance can be observed when the spatial distribution of failures is changed; this is done while also controlling for the effect of the disruption

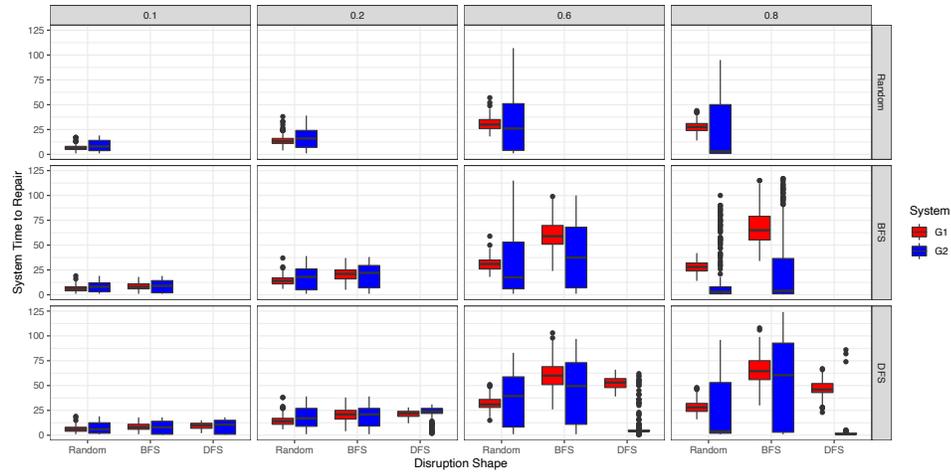


Figure 3.4. Differences in time to repair (TTR) by failure method subset by the size of failures (columns) and the distribution of failures in g_1 (rows). Each individual plot shows the difference in TTR for g_1 (red) and g_2 (blue) for changes in the distribution of failures in g_2 (sub-columns).

size. The spatial distribution of the failures additionally changes the recovery of the system- measured by the time to system repair and functional recovery form. Consideration the impact of a disturbance on each network within an interdependent system can provide better assessments of infrastructure and system risk and resilience.

4. TWITTER AND DISASTERS: A SOCIAL RESILIENCE FINGERPRINT

Chapter 4 has been previously published in *IEEE Access*.

B. A. Rachunok, J. B. Bennett and R. Nateghi, “Twitter and Disasters: A Social Resilience Fingerprint,” in *IEEE Access*, vol. 7, pp. 58495-58506, 2019, doi: 10.1109/ACCESS.2019.2914797.

4.1 Introduction

There is a temporal trend toward more frequent and more unexpectedly intense natural disasters [140]. To prepare for uncertain future disasters, it is fundamental to question what constitutes a *resilient community* so as to build a body of knowledge useful in enhancing communities’ adaptive capacity in the face of the next generation of unforeseen disasters. Resilience is a concept with multiple definitions, all of which stem from understanding how elements of a community protect against, respond to, and recover from a disruption [8, 77, 79, 83, 87, 103, 105, 141]. At their core, these definitions establish how an exogenous disruption bears on the dynamic interactions and responses inside a community whether through ecological, infrastructure, social, or economic mechanisms. However, previous analyses do not directly incorporate the experience of individuals during disasters when measuring the totality of a community’s resilience. Instead, (community) resilience analyses examine the impact of a disaster or disruption on individuals as manifested through an existing social, physical, economic, or ecological systems [83, 104, 105, 138]. Recent work has hypothesized that *online social networks* (OSNs) can fill the this gap in the study of resilience by incorporating the direct measurement of individuals in a community throughout the response to a major disruption [29, 142].

In this work, we formulate measurements of the resilience of a community by augmenting existing conceptualizations of community resilience with data from online social networks, namely the microblogging platform Twitter. In 2017, 80% of the US population is estimated to have a social media account; of those Twitter is among the most popular with 62 million monthly active users in the US in 2018 [143,144]. Twitter is a platform for disseminating and consuming content at an unprecedented scale, providing a direct conduit into the response of individuals to major events. Interactions on Twitter are based on short messages of 280 characters. These messages (called *tweets*) are broadcast to a user’s *followers*. Particularly during major events, the follower–followee relationships leads to emergent social properties at a macro-scale which are driven by a bottom-up self-organization of information [33], thus providing unique access to information deemed important by the community. Consequently when the resilience of a community is tested by a major event, the self-organization of Twitter discourse indicates that topics which are relevant to the resilience of a community are detectable. In this paper, we leverage this bottom-up information to develop a multi-dimensional *social resilience fingerprint* which analytically captures the interactions within pillars of community resilience during a disruption.

We introduce the *resilience fingerprint* as a multi-dimensional concept for understanding community resilience. A resilience fingerprint is the unique combination of components of community resilience in response to a major event or disruption. We use the analogy of a fingerprint to emphasize the identifiability of components critical to community resilience. In this way, we move away from evaluating resilience in one dimension and instead propose a relative-mapping of the interrelated aspects of resilience to one another. Rather than asking *how resilient is a community* we ask *what constitutes a resilient community*. We subsequently describe methods for measuring the resilience fingerprint of communities impacted by major events through analysis of the social media discourse surrounding the event thus establishing a *social resilience fingerprint*.

A social resilience fingerprint is an analytical method for understanding the interactions between components of community resilience as observed through social media. This is calculated first by defining community resilience as a set of resilience *components* suitable for measurement by social media, then categorizing the macro-scale Twitter response of a community before, during, and after a major event by its impact on the individual components. The relative measurements of each resilience component –along with the interaction between components– form the basis of the social resilience fingerprint.

The remainder of this paper is as follows: Section 4.2 provides background on community resilience and describes our categorization of community resilience in the context of online social network analysis; Section 4.3 describe the data used in this analysis, as well as the methods used to turn large corpora of tweets into a social resilience fingerprint. Finally, Section 4.4 applies the techniques presented to 14 events with a significant Twitter response, the results of which are presented in Section 4.5.

4.2 Background

Externally, communities are the “totality of social system interactions within a defined geographic space such as a neighborhood, census tract, city, or county” [8], and can be characterized by internal dynamics which comprise combinations of individuals and groups with multiple –potentially competing– interests and associations [2, 145]. The broad scope of communities leads to a vast number of approaches and methods for the study of their resilience. In this section we discuss how conceptualizing resilience as a multidimensional fingerprint fits within context of existing studies of resilience and online social networks.

4.2.1 Community resilience

In order to understand how multiple dimensions of disaster resilience can be studied through social media, we establish a definition of community resilience based on

previous constructions and in alignment with evaluation through online social networks. Community resilience has been formalized as a comparative assessment of the resilience of community *components* or *categories* [8,87,146]. Category-based definitions of community resilience share substantial overlap. One such definition is given by the Multidisciplinary Center for Earthquake Engineering Research, which categorize community resilience with the acronym PEOPLES: **P**opulations, **E**nvironment and ecosystem, **O**rganized government, **P**hysical infrastructure, **L**ifesyle and community, **E**conomic development, and **S**ocial-cultural capital [87]. A similar definition proposes a framework which distinguishes categories of resilience by how they are measured [8]. They include ecological resilience, social resilience, economic resilience, institutional resilience, infrastructure resilience, and community competence [8]. We leverage a multi-dimensional categorization of community resilience, defined as a set of components which are derived from previous definitions of community resilience so as to theoretically ground our analysis [8]. We define the categories of community resilience in an OSN context as the Ecological, Economic, Institutional, Social, Infrastructure, and Quality of life categories. These categorizations are not mutually exclusive, but are collectively exhaustive. Table 4.1 lists high level descriptions of the components of a social resilience fingerprint and the topics they encompass through Twitter.

4.2.2 Twitter

Since its inception in 2006, Twitter has been a common source of academic inquiry particularly relating to its use during disasters and major events. Since Twitter is a platform for sharing and consuming media, early work in the evaluation of tweet content established relationships between public Twitter posts and internal sentiment, situational awareness during disaster, and psychological trauma [33,147].

Twitter has also been studied as a form of sensing network which can augment more traditional analyses performed during a disaster such as the study of vulnerability or resilience [29,142]. Understanding how online social networks can be used

Table 4.1.
Resilience components, their description, and community elements from that category

Component	Description	Example elements
Ecology	Related to natural systems and features of the environment and ecosystem	Coasts, marshes, streams, beaches, wetland
Economy	Financial, economic, and business aspects within a community	Currency, business operation, labor
Institutions	Government and service-based institutions providing community function and care	Police, hospital, FEMA, government officials
Social	Non-institutional support systems within a community	Humanitarian aid, volunteerism, neighbors
Infrastructure	Physical infrastructure systems and their dependencies	Pipelines, power systems, cell communication
Quality of life	The health and wellbeing of the community	Health, hospital, mental well-being

to derive meaningful insight has been defined as *social media analytics* [29,40]. Work in this area is typically broken down into multiple dimensions based on how social media is used for analysis (e.g., tweet location, tweet content etc.) [40]. What follows is a review of literature relating to understanding disasters and communities through social media.

Social media analytics has been previously used in many disaster-related contexts to gather information about the spatial distribution of disasters in an attempt to correlate measurable elements of a disaster with measurable elements of social media. Tweets related to a topic of a disaster were shown to be more likely to occur near disaster-related areas during a flood of the Elbe river [31,46]. There is also significant evidence to suggest GIS and remote-sensing applications can be significantly improved by augmentation with social media data [32]. The primary benefit of this augmentation is that social media provides a ground-up network of sensors which can allow for hyper-local and rapid updating of geographic systems [47].

Temporal associations between tweets and disasters have also been investigated. A study of Hurricane Sandy found the time for an individual to learn about a disaster through social media was proportional to an individual's distance from the impact [148]. In a different context, the role of individuals in a disaster is found to be temporally-dependent [149]. During times of disasters, individuals are observed to transition toward an information-sharing role on Twitter, broadcasting and exchanging information [150].

Another thrust of social media analytics is an analysis of tweet content, in which a semantic understanding of a tweet is used to make assessments of the tweet author [40]. Related to disasters, the 'mood' of tweets was tracked through multiple disasters affecting North America as a proxy for how individuals recover psychologically from disasters [51]. Other analyses use the content of social media networks to understand the patterns of information diffusion in disaster [151].

4.3 Data and Methods

The accessibility of tweets issued prior to 7 days in the past as well as Twitter’s terms of service make acquiring corpora of tweets a non-trivial task. In this section we first briefly discuss the process of tweet acquisition, and follow with the methods used to analyze the Twitter corpora.

4.3.1 Tweet Acquisition

Our tweet datasets were retrieved from various archival sources described at a high-level in Table 4.2, with more details presented in Appendix Table C.1. Over 14 million tweets were analyzed spanning 14 major events. The major events include 5 hurricanes, 2 events of public violence, 2 political referendums, 2 earthquakes, 1 public health crisis, 1 death of a celebrity, and 1 solar eclipse. Events were chosen based on the scale of the social-media response to the event, but little other restriction was placed on inclusion in our study. This results in a corpus of tweets which spans multiple years, sizes, event types, and archival methods.

As of early 2019, Twitter limits access to the entire body of published tweets via a paid subscription service. Additionally, Twitter’s Terms of Service prohibit the reproduction or distribution of datasets of *whole* tweets and instead only allow for the distribution of lists of numerical serial numbers corresponding to each tweet, called tweet *IDs* [152]. Hence, the medium of tweet compilation and sharing is the tweet ID, which can be used to re-construct the original tweets. IDs are simply serial numbers corresponding to each tweet and provide no actionable information, therefore, the process of *hydrating* tweets must be carried out to convert tweet IDs into a full tweet as it would be seen on the platform. Hydrating repeatedly calls the Twitter API with a specified tweet ID and returns the associated tweet content as well as additional meta-data such as the author, the date of publication, whether it is a retweet of someone else, etc. As this is a process of retroactively accessing data, there may be a loss of data. Tweets may not be available due to deletion of

Table 4.2.

Size, date, and type of twitter events analyzed. Ordered by starting date of event.

Event	Dates	Total corpus IDs ^a	Final tweet count ^b	Retention ^c
Hurricane Sandy	10/22/2012 to 11/02/2012	6,554,744	3,252,011	49.61%
Ebola Outbreak	08/18/2014 to 01/19/2015	5,085,767	993,905	19.54%
California Earthquake	08/24/2014 to 08/30/2014	254,529	50,414	19.81%
Nepalese Earthquake	04/25/2015 to 05/19/2015	4,223,983	509,299	12.06%
Brexit	05/05/2016 to 08/24/2016	23,733,133	3,884,599	16.36%
Charlottesville Riots	08/14/2017 to 10/23/2017	3,015,437	207,098	6.87%
Eclipse	08/17/2017 to 08/23/2017	13,548,321	1,211,729	8.94%
Hurricane Harvey	08/25/2017 to 10/23/2017	18,352,142	1,062,127	5.78%
Hurricane Irma	09/01/2017 to 10/23/2017	17,244,139	976,294	5.66%
Hurricane Maria	09/20/2017 to 10/03/2017	1,096,335	87,160	7.95%
Las Vegas shooting	09/29/2017 to 10/07/2017	14,108,104	866,758	6.14%
Ireland 8th Amendment	04/13/2018 to 06/04/2018	2,279,396	195,050	8.56%
Aretha Franklin's death	08/08/2018 to 08/18/2018	2,832,128	252,433	8.91%
Hurricane Florence	09/05/2018 to 10/03/2018	4,971,575	488,106	9.82%
Total		117,299,733	14,036,983	11.97%

^a Original number of tweet ids published in corpus.

^b Total number of tweets used in this analysis after deleted tweets are accounted for and retweets removed.

^c Tweet retention is calculated as the final tweet count divided by the total corpus ids for each event.

the previous tweet, tweet-author’s user account, or change in privacy settings of a user account. Recent work has shown that despite this data loss, remaining samples of tweets are still representative of the data published in real time [152]. Based on previous findings which indicate that Twitter messages sent during consequential events are more focused on information-broadcasting and information sharing [149], we additionally remove retweets (*i.e.* a user re-broadcasting the tweet originally authored by someone else) from our data to focus on originally produced content.

4.3.2 Data processing

After hydrating and removal of retweets, the text data of each tweet are processed to remove abnormalities. First, URLs, and non-ASCII characters are removed using customized regular expressions [153]¹. English and Spanish stop words are then removed. Stop words are non-informative, frequently-used words which do not contribute to a semantic understanding of text [154]. In this case stop words are defined using the popular `stopwords` R package [155–157]. Each event’s tweets are then processed to remove words occurring less than 10 times through all tweets related to an event. Additionally -if the dataset was compiled based on keyword filtering- the words used for filtering were removed from the corpora, as they would otherwise be included in all tweets by construction. Finally, the remaining words are *stemmed* to remove word endings using the Porter stemming algorithm, implemented in R [158–160]. Stemming removes word endings to avoid differentiating between words of similar meaning used in different tenses, conjugations etc. For example **ecological** and **ecology** would both stem to the same root: **ecolog**. Word stemming has been previously shown to greatly improve text processing and analysis [154].

¹This is increasingly important in recent datasets as the use of emojis becomes more prevalent

4.3.3 Social Resilience Fingerprinting

At the core of the methodology proposed in this paper is understanding how individual *components* of community resilience can be measured and understood through the lens of social media. Formally we have a set of all events E comprised of n individual events E such that $E_1, E_2, \dots, E_n \in E$. For a given event E^* , we have a set of tweets, $t_1^{E^*}, t_2^{E^*}, \dots, t_m^{E^*} \in E^*$, where m is the total number of tweets compiled for each event after hydration and processing. Each tweet is subsequently comprised of a series of *features*, f , which are the individual words in each tweet such that for a given tweet t^* , $f_1^{t^*}, f_2^{t^*}, \dots, f_l^{t^*} \in t^*$. As each tweet can contain multiple copies of the same word, we also have a set of all features F^E for a given event E .

Additionally, we manually coded a set of words for each category in order to map the set of features to our pre-defined resilience categories. Thus, each category of resilience contains a set of words which indicate associated discourse. For example, $C_{\text{infrastructure}} = \{\text{power}, \text{water}, \text{cell}, \text{outage}, \text{road}, \dots\}$. The words were manually selected by two groups individually, then consensus was established between the two sets. The full listing of words coded for each category is listed in Appendix C.3.

To parse the features for a given event into categories we construct a *category co-association* matrix, A for each event. A is a symmetric 6 by 6 matrix with each row and column corresponding to a resilience category. A_{ij} is then the co-association of category i with category j . The co-association of a given category is based on the co-occurrence of words from categories. As such, for resilience categories i and j ,

$$A_{ij} = \sum_{\substack{r \in C_i \\ s \in C_j}} \sum_t \sum_{r \in t} \text{occ}(r, s) \quad (4.1)$$

Where

$$\text{occ}(r, s) = \begin{cases} 1 & \text{if word } r \text{ occurs with } s \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

After fixing a word r from one category, and word s from another, the `occ` function is an indicator function taking a value of 1 each time word r occurs with word s in a given tweet. This is summed over all occurrences of r in a given tweet (innermost summation of (4.1)), then subsequently summed over every tweet. This is done for all combinations of words in category i and category j . Thus A_{ij} is the total times a word from category i occurs in the same tweet as category j . The matrix of values A –one for each event– form the social resilience fingerprint. Off diagonal values of A represent the frequency of resilience categories appearing together in Twitter discourse. The diagonals of A are less intuitive, representing the relative frequency of words from the same category appearing in a tweet. This is a modified version of a *co-occurrence* matrix, used for term clustering in natural language processing [161–163]. This extension uses apriori categorizations –grounded in the theoretical definitions of community resilience– to find associations within topics to determine the relative association of categories of resilience. In the following section, we apply this fingerprinting methodology to multiple major events and discuss the feasibility of extracting category-based insights using this method. For the 14 events listed in Table 4.2, the categorical binning described in (4.1) and (4.2) are used to establish the social resilience fingerprint. As the total number of tweets gathered for each event vary substantially, the A matrices are scaled. This allows for a more balanced comparison between events as it removes information regarding the total number of tweets from the fingerprint so that any comparison made between events is based solely on the pattern of interactions among the components of the resilience fingerprint. Sinkhorn-Knopp matrix regularization is used on A matrices [164]; this preserves the structure of the fingerprint while allowing the relations between categories to be compared across events.

4.4 Results

Visual representations of the social resilience fingerprints are shown in Figure 4.1. Each heatmap and associated bargraph show the relative association of each category and the frequency of each category respectively. The respective heatmaps are visual examples of the matrices A , representing the co-association of discourse related to components of community resilience. Because of the self-organization of tweets in response to major events, we hypothesize that stronger textual association of categories indicate a stronger underlying relationship between the categories in the community and by extension in the resilience of the community. This is in line with previous findings which found that event-related keywords were indicative of a major event’s impact on an individual [148].

4.4.1 Event Similarity

From the wide range of the events studied, we hypothesize that the Twitter discourse in reaction to similar events will itself be similar, as measurable through the resilience fingerprint. To evaluate this, we measure the component-wise Spearman distance between scaled A matrices for all events. The result is a numerical measure of similarity among the structure of the resilience fingerprints in which a smaller distance represents a more-similar pattern of Twitter discourse between two events. In Figure 4.2, the resulting pair-wise distances are visualized in a heatmap after hierarchical clustering is performed on the rows and columns –a technique called VAT or a Visual assessment of Cluster Tendency [134, 165, 166]. Each element in Figure 4.2 represents the distance between the row and column event.

The VAT methodology is formulated to allow for visual identification of trends in data [165]. A VAT cluster appears visually as a square block along the lower-left to upper-right diagonal of the heatmap. In Figure 4.2, there are clear clusters corresponding to Hurricanes Florence, Irma, Sandy, Harvey, Maria as well as the 2018 Eclipse. Additionally, a case could be made for the clustering of the Nepalese earth-

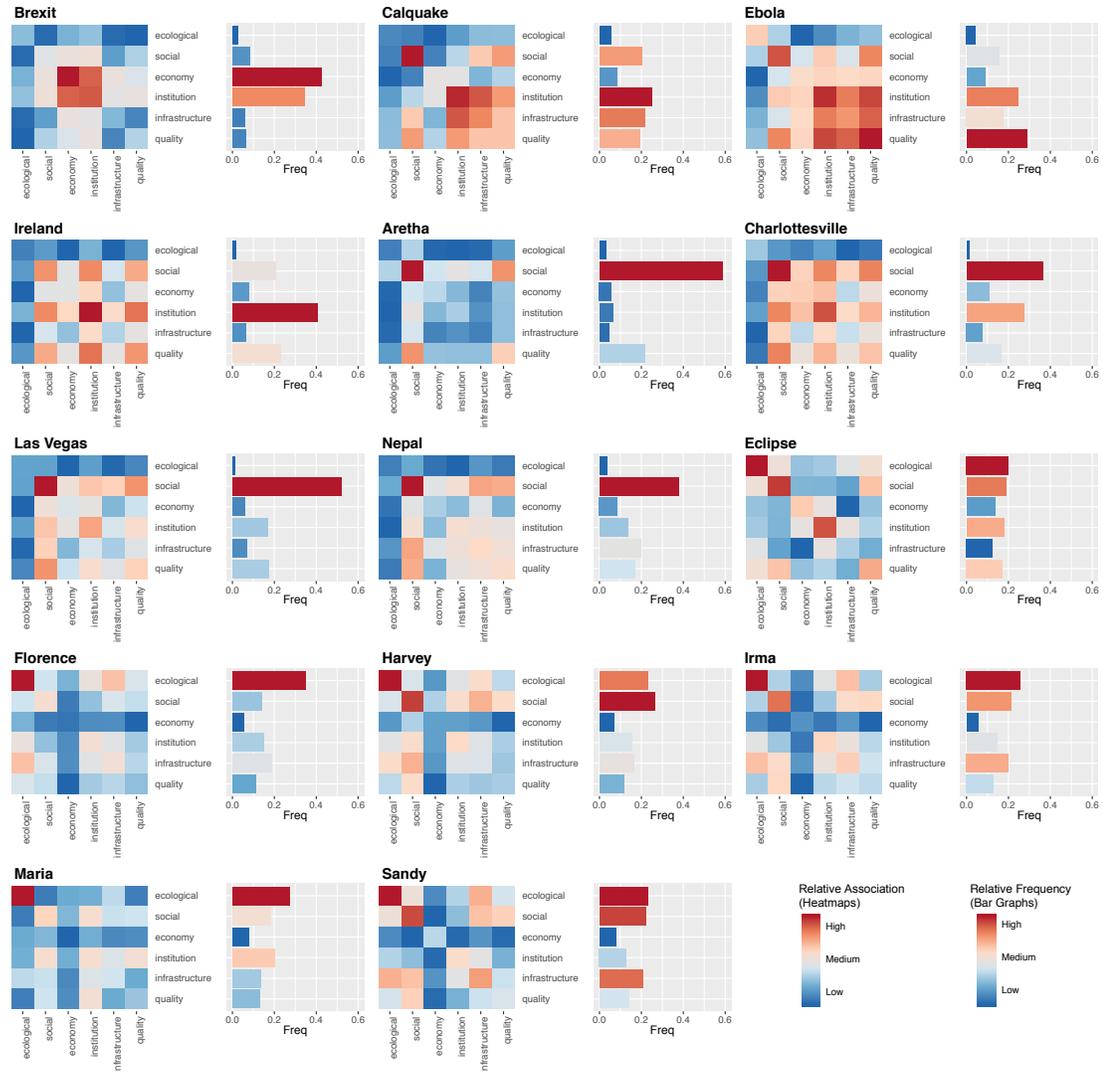


Figure 4.1. Visualizing social resilience fingerprints. Each heatmap represents the association between one category with another. The color red indicates the most association, and the color blue represents the least association. Diagonal values in the heatmap are indicative of how self-associative a category is. Bar-graphs show the relative occurrence of each category. Note the color scheme in this plot is based on log-normalization of A , as opposed to Sinkhorn-Knopp, to aid in visualization.

quake and the California earthquake. Finally, the upper right of Figure 4.2 provides evidence of clustering of the Las Vegas shootings, Charlottesville riots, Ireland's 8th

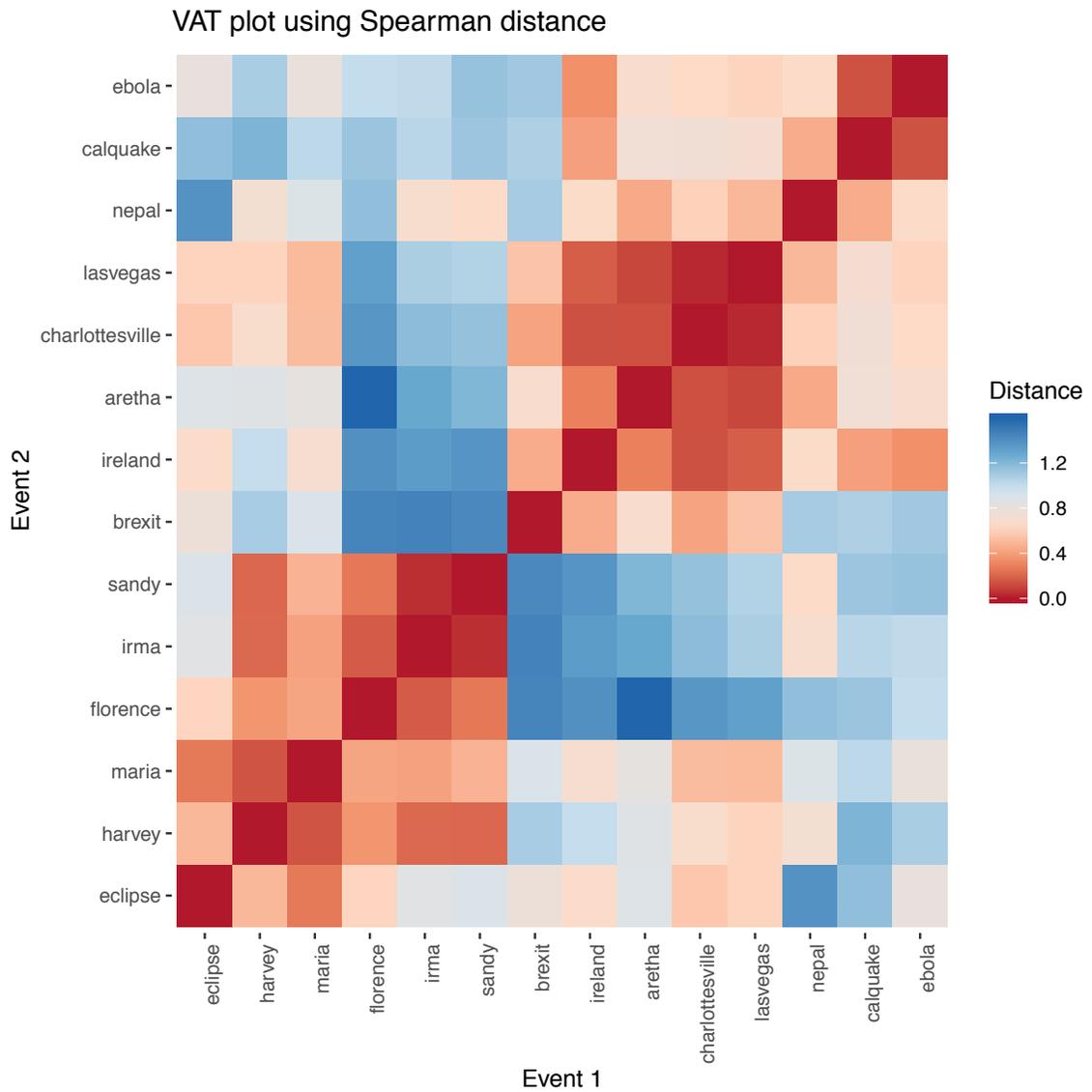


Figure 4.2. VAT assessment. Each element represents the distance between the event in the row and the event in the column, red indicating closer, blue farther away

constitutional amendment, and the death of Aretha Franklin. As we calculated the distance between the events by summing component-wise distances between two fingerprints –each scaled from their original counts– these clusters are representative of similarity in the pattern of associations between components of resilience. From this, we see a similarity in the social resilience fingerprints of alike events, providing evidence that our proposed methodology has discriminating power.

Based on this distance measure, we subsequently analyzed each event’s closest match using an alternative distance measure, namely, the Pearson’s correlation coefficient. This is also computed between each pair of fingerprints. The closest-correlated event to each event are listed in Table 4.3, along with the associated correlation. The results paint a similar picture to the VAT comparison. Natural disasters, such as hurricanes and earthquakes, pair closely with one another, as do acts of violence like the Las Vegas shooting and Charlottesville riots.

We perform another similarity measurement by comparing the clusters generated via the k -means clustering algorithm. We select 3 clusters as the marginal within-cluster error does not improve greatly with additional number of clusters. Figure 4.3 shows the three clusters plotted on the axes of the first two principle components of the data. The cluster containing the hurricanes and eclipse differs most greatly in the direction of the first principle component, while the remaining two clusters differ based on the second component. The first principle component is driven by differences in ecological categories of resilience while the second is a difference in social, economic, and institutional resilience.

From the results of these clustering methods, we hypothesize that the similarity between the fingerprints of similar events indicates that much of the emergent properties of the resilience of a community is driven by the specific disaster. Through this hypothesis, we propose that the elements of community resilience common to each type of event are distinct enough to affect the Twitter discourse of the individual communities to an extent that it is measurable at a macro scale.

Table 4.3.

Closest Events. The Pearson correlation is calculated between all pairs of events, with the closest match listed. The correlations above 80% have been highlighted in bold.

Event	Best Match	Correlation
aretha	lasvegas	0.77
brexit	charlottesville	0.50
calquake	nepal	0.85
charlottesville	lasvegas	0.86
ebola	irma	0.56
eclipse	charlottesville	0.51
florence	irma	0.94
harvey	irma	0.89
ireland	calquake	0.78
irma	florence	0.94
lasvegas	charlottesville	0.86
maria	florence	0.90
nepal	calquake	0.85
sandy	irma	0.88

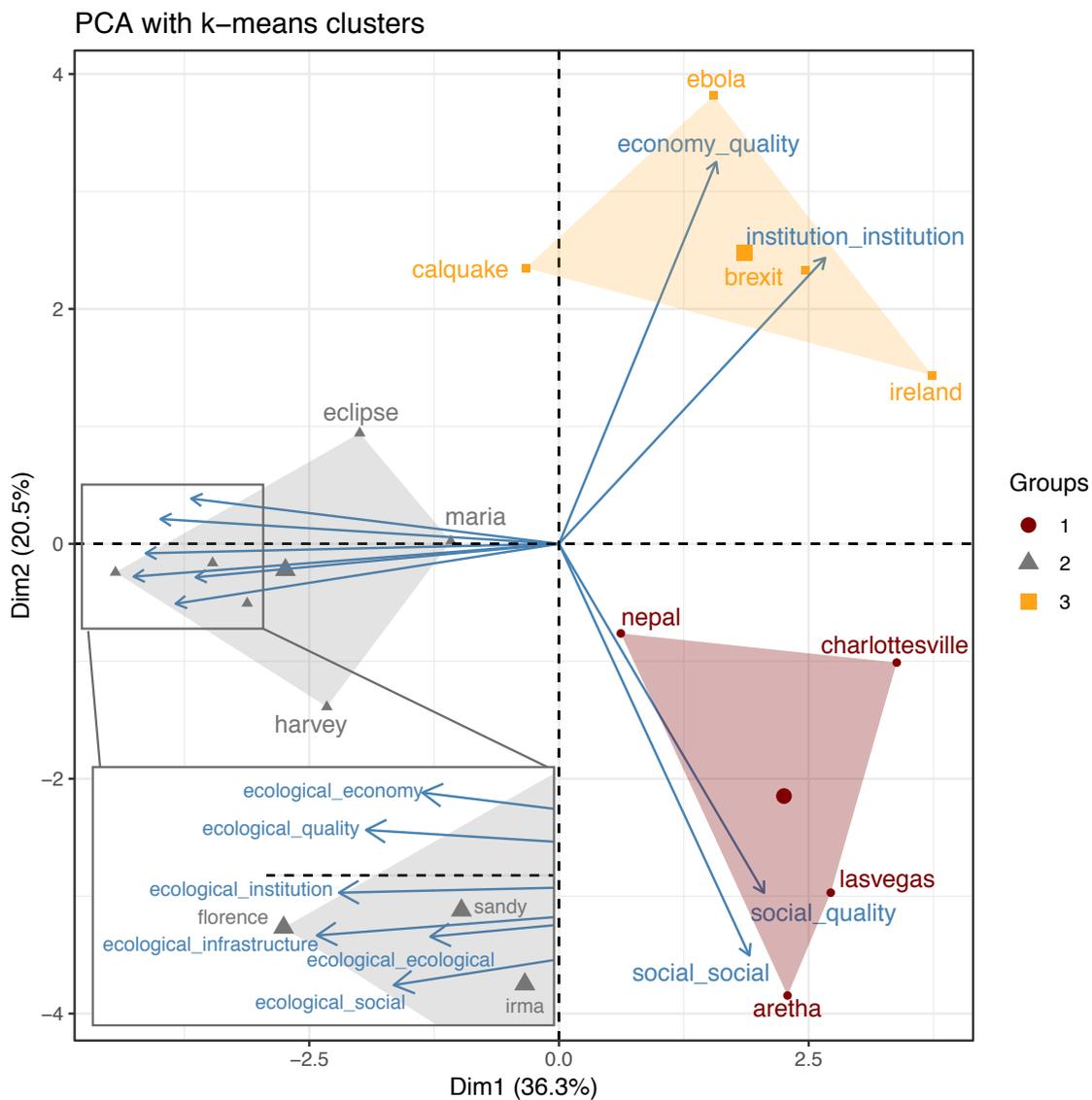


Figure 4.3. PCA loading plot. The top 10 most contributing elements of each fingerprint are projected onto the first two principle components. Components and events along the negative x axis are shown in the bottom left inset for clarity.

4.4.2 Critical Components of Community Resilience

To further understand the importance of the categories of community resilience, we now ask which elements of community resilience drive the similarity among events by looking at the loading of each variable –corresponding to an i, j element of the fingerprint across all events– as projected onto the first two principle components. The variable loading for the 10 most contributing variables are plotted in Figure 4.3 along with the events.

From the variable loadings, we can see that –as expected– the vectors with similar directionality have overlapping categories. Along the x-axis of Figure 4.3 are the associations of the ecological category with all others, indicating they are strong contributors to the similarity of the hurricane-events. Likewise, institutional and economic category dominate the first quadrant. Finally, social components tend in the direction of the cluster associated with the Charlottesville riots and Las Vegas shootings. In Figure 4.3, a small angle between vector loading indicate high correlation between variables. From this we can generally infer a positive correlation within the ecological and social categories as well as between the institution and economy categories.

From Figure 4.3 we can also interpret that the first principle component is driven by changes in the ecological categories indicating this may be primary drivers behind the clustering of the hurricanes, and consequently a significant component of community resilience.

4.4.3 Posterior Analysis

To further investigate the components most influential in the social resilience fingerprint, we look at the explicit difference between events of different types. The most apparent cluster of events are Hurricanes Florence, Irma, Sandy, Harvey, and Maria. As such, we compute the element-wise mean fingerprint of the hurricane-events and non-hurricane-events. The element-wise difference –calculated as the hurricane

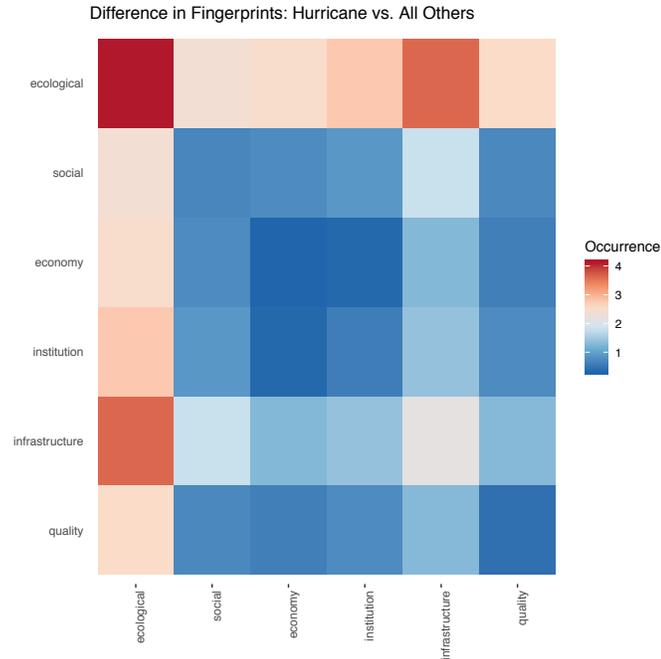


Figure 4.4. Category-based difference between the average hurricane fingerprint and all average non-hurricane fingerprint. Each element represents the difference between the hurricane and non-hurricane association of categories. Blue indicates stronger association in non-hurricane events, while red indicates stronger association in hurricane events. Coloration based on log-normalization of A matrices rather than Sinkhorn-Knopp for visual clarity

mean minus the non-hurricane mean— is visualized in a heatmap in Figure 4.4. For each pair of categories, the color of the cell value indicates whether those categories have a stronger association among the hurricane fingerprints (colored red), or the non-hurricane fingerprints (colored blue).

Figure 4.4 confirms the results of the PCA analysis and indicates the ecological and infrastructure categories of resilience are much stronger in the hurricane fingerprints than in the non-hurricane fingerprints. The interaction of infrastructure and ecological categories are the strongest for the hurricane category among the non-diagonal elements. At the same time, the economic-institutional relationship is most strong among the non-hurricane events.

4.5 Discussion

After clustering the social resilience fingerprints for all events and analyzing what drives their similarity, we identify two major trends: first is the strong distinction between hurricane and non-hurricane events with respect to fingerprint similarity, and second is the importance of ecological and infrastructure resilience in making that distinction.

We see a strong association, not just of one hurricane with another, but among *all* hurricanes for which we could collect data. The hurricane-related tweet corpora were collected in a variety of ways and span distinct spatial and temporal scales. Despite these differences, the similarity in the fingerprints indicate generalizable patterns in community resilience in the face of hurricane impacts. Moreover, it provides a strong evidence supporting the fingerprinting methodology. It also suggests that Twitter is a persistent source of data about individual responses to a disaster within a community, establishing Twitter as a valuable tool for measuring disaster resilience across communities.

Additionally, general similarity among specific non-hurricane events indicates emergent themes in the Twitter responses manifesting as similar social resilience fingerprints of related events, and thus similarities in the underlying resilience. The relative similarity of the California and Nepalese earthquakes, as well as the public violence in Charlottesville and Las Vegas, both indicate that other types of major events may also have fundamental, emergent themes decodable through Twitter discourse. We conjecture that similarity in the social resilience fingerprints of related events is indicative of fundamental similarity in the resilience of the communities facing such events. That is, there are emergent similarities between the way different communities respond to the same event across all types of events. However, we recognize the limitation of drawing conclusions from the similarity of only two events studied in this paper and intend to expand upon this analysis to test our conjecture.

The second major trend in the analysis of the social resilience fingerprints is the influence of individual components of resilience in the separation of one event from another. Ecology, infrastructure, and economic categories drive much of the separation between the emergent clusters in the data. Economic resilience is intuitively intertwined with all other categories in our definition [77, 83, 87], and is seen in the Principle Component Analysis to contribute greatly to the distinction between clusters of non-hurricane events.

The significance of infrastructure resilience in differentiating between hurricane and non-hurricane events –as seen in Figure 4.4– is likely due to the significance of infrastructure damage in communities affected by hurricanes. Ecological resilience and its close ties to sustainability, have been previously shown to be strong drivers of community resilience at all levels [8, 77]. We see the distinction in Figure 4.3, manifesting as the ecological loadings in the direction of the first principle component –indicating that ecology explains the largest degree of variance among the fingerprints. This reveals that the resilience fingerprint method is not limited by what has hampered the previous attempts in quantifying community resilience –namely the difficulty in acquiring data related to specific ecosystems. Due to the difficulties in finding relevant measurement indicators, ecological resilience has previously been excluded from resilience assessments [26].

The resilience fingerprints of three events were not revealed as expected: The Irish constitutional amendment, Brexit, and the Ebola outbreak. The authors hypothesized that the Irish constitutional amendment and Brexit would be similar events due to their close physical proximity and the general political nature of the event; a trend which did not emerge from our analysis. One explanation for the difference are in the specificity of search terms used for the generation of the Irish amendment tweet dataset. The Irish amendment tweet dataset used 52 terms to filter by, the most filter terms used by almost a factor of 2 (See Appendix Table C.2 for terms); the Brexit dataset was built on only one search term: `brexit`. The terms used to filter the Irish referendum dataset are also more specific than the others, leading to

a corpus of tweet text which may be overly specific to the Irish political system and the issues of the referendum, lacking substantive information about the community's response in favor of the individuals. Tweets related to the Ebola virus additionally showed little relation with other events. In this case, we hypothesize that the location of the event relative to major Twitter-adoptive societies may affect the ability of fingerprinting to detect a signal. International Twitter use is lower than that of the US [144]. As such we hypothesize that someone tweeting about Hurricane Florence was more likely witnessing community impacts due the storm than someone tweeting about the ebola outbreak.

4.6 Conclusion and Future Research

In this paper, we present the *resilience fingerprint* as a concept for understanding community resilience as the relationship of individual components. We then calculate a *social resilience fingerprint* by leveraging social media analytics guided by the community resilience theory. We find evidence that resilience fingerprinting can highlight the different community responses to a variety of major events and identify the components of community resilience which most contribute to the overall response. We leverage a category-based definition of community resilience to classify the macro-scale response on Twitter to a disaster into elements of community resilience.

In summary, the resilience fingerprint provides a concept for the multi-dimensional analysis of the emergent responses of communities to major events. The rapid spread of information via social media makes social resilience fingerprinting a vital complement to existing resilience analyses, capable of categorizing the community response to a disaster.

In this work, the categories were manually coded, as guided by the literature in community resilience. However, an ongoing extension of this work is to use automated topic detection to both determine what individual words best comprise a resilience category, and to determine the emergent resilience categories in an unsupervised way.

Additionally, we aim to extend the classification of tweets beyond word-association based on recent developments in the classification of tweets related to disasters [167]. Finally, we are expanding the fingerprinting methods to allow for the creation of a resilience fingerprint in real time. This will provide a dynamic look at the interactions among communities as they respond to major disasters and events.

5. IMPROVING OPERATIONAL MEASURES OF COMMUNITY RESILIENCE

Chapter 5 is currently under review

5.1 Introduction

Accelerated urbanization and climate change have amplified the vulnerability of communities to climate disasters. Resilience has long served as an organizing principle for marshalling resources to reduce vulnerability and stimulate recovery in response to major natural hazards and disruptions [55,168]. Theoretical and analytical studies of resilience exist in the social sciences, ecology, urban planning and engineering [17,55]. Despite disciplinary differences, resilience is broadly conceptualized as capacities to bounce back after shocks and systematically adapt and transform to preserve system functionality [2,28,169]

More recently, there is a push to move beyond ontological discussions of resilience towards an operational paradigm at the community level, with a community understood as geographically linked groups of interacting individuals with shared norms and interests [170]. Despite recent advancements in operational models of community resilience [170–173], fundamental knowledge gaps remain. These gaps can be traced to an overwhelming focus on bouncing back after disruptions, thus preserving the status quo. Specifically, in existing paradigms, a resilient system deviates minimally from its current state and returns to the status quo rapidly upon disruption [29,55,80–84,86,87,89,90,168]. Referred to as *stable-equilibrium* or *engineering* resilience [11,174], this paradigm has been frequently applied to economic systems [83] and civil infrastructure [54].

Engineering resilience has served as the foundation for many decision and policymaking frameworks aimed at building resilient and sustainable systems and communities [88, 175–177]. At their core, these frameworks quantify how communities are disrupted and recover –typically through measuring reliable access to critical infrastructure such as the electric power grid– and seek to identify risk factors within communities and/or systems which mitigate disruptions and promote a rapid return to a pre-disruption equilibrium. These frameworks are beneficial for prioritizing relief and mitigation efforts, but are incomplete in operationalizing the concept of resilience as they focus only on a system returning to the status quo rather than systematically transforming.

In this work, we shrink the gap between conceptual and operational models of resilience by developing methods to quantify the potential of communities to transform. Based on a definition of transformation as a ‘systemic change of the urban system’ [28] which includes nonlinear reorganizations of infrastructure, ecosystems, lifestyles, institutions, and governance [28, 178], we measure and track the reorganization and transformation of communities in conjunction with quantifying their engineering resilience. Rather than solely focusing on identifying factors associated with rapid recovery, we use statistical machine learning to also identify key risk factors which can catalyze or inhibit transformation. We quantify threshold effects and conduct tipping point analyses by estimating the degree of change needed in risk factors to cause transformation using the 2018 Hurricane Michael in Florida as a case study.

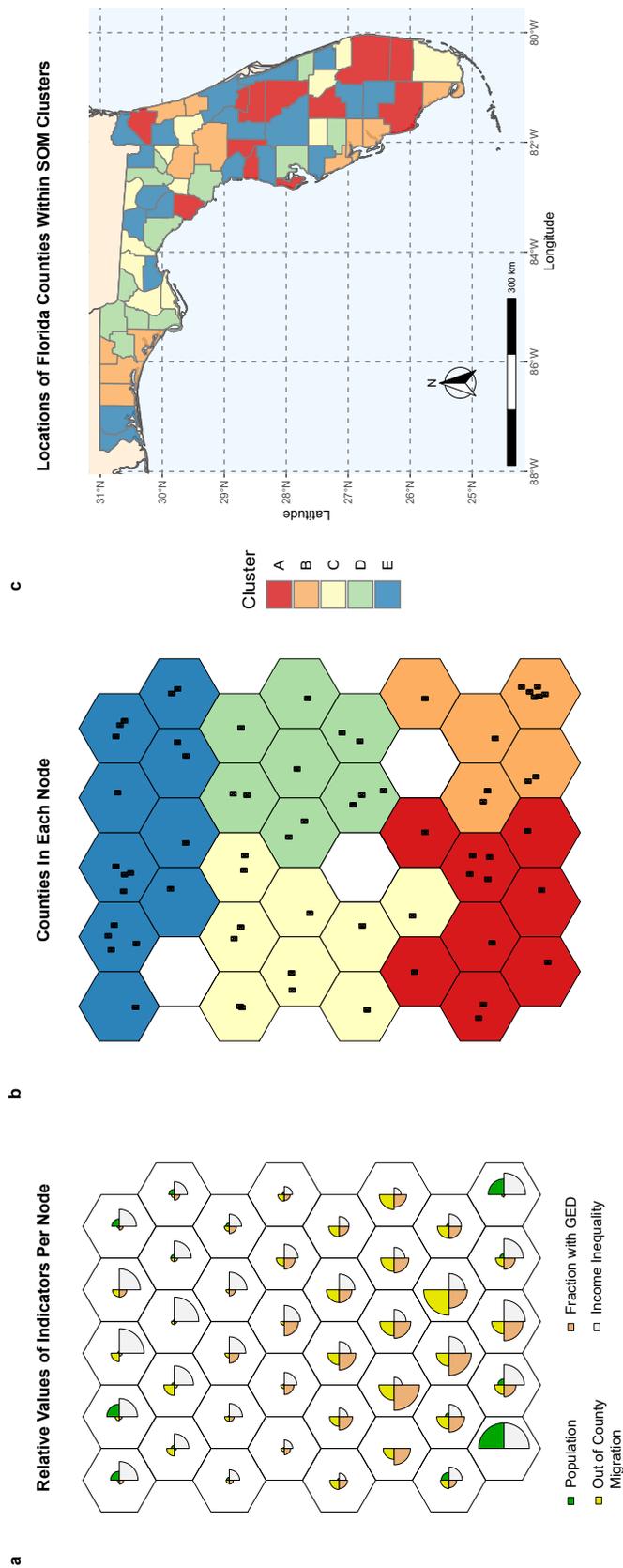


Figure 5.1. Description of the baseline CCN for county clustering. (a) Values of four sample community risk factors across each CCN node. (b) Mapping of each county within the CCN. (c) Location of CCN clusters (colors) within Florida. Figures (a-c) created in R (v 3.2.1; <https://www.r-project.org/>) [179] using the `ggplot2` (v 3.3.0; <https://ggplot2.tidyverse.org/>) [180] and `kohonen` (v 3.0.10) [181]. Figure c additionally used `usmap` (v 0.5.0; <https://github.com/pdil/usmap>) [182]. Map shapefiles in b are from `usmap` and the US Census Bureau [183]

5.2 Contrastive Community Networks

To quantify how communities transform, we develop a new approach termed *Contrastive Community Networks* (CCN). CCN is grounded in Self Organized Maps (SOMs) which are a class of unsupervised learning techniques for dimension reduction and projection [184]. The CCN utilizes a SOM to create a relational network of communities –here counties in Florida (Fig. 5.1 a)– in which proximity in the network corresponds with similarity in demographic and socioeconomic risk factors between counties. In contrast to previous methods which measure the temporal difference in risk factor values as a proxy for community change [173], the CCN algorithm measures transformation by monitoring for changes in risk factors substantial enough to be detected as a change in the relational network. Thus, rather than assessing transformation on the basis of change in individual risk factors, we quantify transformation by measuring the degree of contrast between a county and its peers. This section outlines the data used as inputs to the CCN method (Section 5.2.1), the methods used to calculate engineering resilience (Section 5.2.2) identify how risk factors contribute to engineering resilience (Section 5.2.3), a description of the SOM algorithm (Section 5.2.4), and finally the CCN algorithm (Section 5.2.5).

5.2.1 Community Risk Factors

We identify a pool of county level risk factors related to the environmental opinions, sociodemographic, economic, housing, and mobility characteristics for each of the 67 counties in Florida [185, 186] (See Appendix Table D.1 for a list of the risk factors, their sources, and descriptions). We select 96 county level variables to describe communities in the case study of Florida. The initial pool of variables are drawn from the American Community Survey [185, 187] and Yale Program on Climate Change Communication [186]. The variables describe the sociodemographic, economic, housing, mobility and environmental opinions for every county for the pe-

riod of time surrounding Hurricane Michael. A full list of included variable names and sources are listed in Appendix Table D.1.

Storm exposure data is taken from the US National Centers for Environmental Information’s Storm Events Database [188]. county level exposure is included as a binary variable, labeled as true if the county is included in the Storm Events Database for Hurricane Michael, false otherwise. We also include a measure of distance to the storm center as a continuous variable. Distance is measured as the minimum distance between Hurricane Michael’s center and the mean population center of each county, calculated with the R package `STORMWINDMODEL` [189], with a maximum distance of 1000 miles.

5.2.2 System Resilience: Restored Access to Electricity

As a measurement of system resilience, we measure the performance of the Florida electric power grid as impacted by 2018 Hurricane Michael. county level power outages are taken from outage reports for the Florida Division of Emergency Management for October 10th through November 9th, 2018 [91, 190]. For each of the 67 counties in Florida, the Division of Emergency Management publishes the number of customers without power approximately every 3 hours. At a time t , $Q(t)$ is the fraction of the county with access to power and represents the service level of the power system. We leverage a formal quantification of engineering resilience for a given county [89, 92]. Resilience for a county, R_{county} is the area under the service level curve, $Q(t)$ from the time of first disruption t_0 to the time when all outages are restored t_f scaled by the difference between t_f and t_0 . R_{county} is defined as

$$R_{\text{county}} = \frac{\int_{t_0}^{t_f} Q(t)}{|t_f - t_0|} \quad (5.1)$$

In this way, a county which lost all power immediately and remained so until it was recovery would have a resilience value of 0 and one with no disruption would have a resilience value of 1. Examples of the calculated resilience along with visual

descriptions of R , t_0 , t_f , and $Q(t)$ are shown in Appendix Fig. D.1 and D.2. Calculated resilience values for Hurricane Michael are given in Appendix Table D.2.

5.2.3 Engineering Resilience Model

To identify the community risk factors which contribute to system resilience, we utilize a predictive modeling paradigm. Predictive modeling aims to find a function, $y = \hat{F}(X)$ which maps inputs (X) to outputs (y) so as to minimize a measure of the distance between the predicted values and true values. Here, y is the county level resilience of the power grid and X are the community risk factors, and R^2 and RMSE (Root Mean Square Error, Eq. 5.2) are used as measures of distance. RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5.2)$$

Here, n is the total number of observations in the test dataset, y_i is the i^{th} actual value of the response variable, and \hat{y}_i is the response estimated by the model trained on the test data and evaluated on the test data. We train 5 model classes: linear models [134], generalized linear models [191], Random Forest models [192], and Bayesian Additive Regression Trees [193]—all implemented in R [134]. Selecting model classes based on minimizing prediction error, however, can lead to *overfit* models in which the prediction error is reduced at the expense of generalization to non-training observations. To counteract this, we perform a 5-fold cross validation procedure in which data is partitioned into 5 roughly equivalently sized *folds* [194]. Each fold—corresponding to approximately 20% of the data—is removed from the dataset, while remaining 4 fold are utilized to train the statistical models. The withheld fold (the *test data*) is then utilized to evaluate the out-of-sample predictive quality of the model. Out of sample RMSE and R^2 are shown in Appendix Figure D.3.

Based on out-of-sample performance measures, we select a random forest model to relate community risk factors to system resilience. Random forest is a tree-based,

non-parametric statistical model [195]. To predict response values, the random forest algorithm builds B decision trees [196] on random subsets of the data. The data used in the tree creation is called the *in the bag* data, and the data not used is the *out of the bag* or *OOB* data. The random forest algorithm averages the output over B trees to create a final estimate of the predicted variable, $\hat{f}(x)$ such that

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (5.3)$$

To extract the importance Using this trained model, we investigate the relative importance of community risk factors using the random-forest-based, three-step variable selection process VSURF [197, 198] to determine which community risk factors most greatly contribute to single-equilibrium system resilience. VSURF, or *Variable Selection Using Random Forest* is an algorithmic process for selecting the importance of variables from random forest models which aims to simultaneously find variables most related to the response for the purposes of interpretation, and to do this with the smallest set of variables possible [199].

The importance of a variable, j , in a random forest model –denoted $VI(X^j)$ – is computed by permuting variables to determine their sensitivity to the calculated error. Formally, $\text{error}_{\text{OOB}}$ is the RMSE of a single tree on the data which was not used to construct it. For the variable j , X^j is perturbed and the error calculated on the perturbed dataset, called $\text{error}_{\text{OOB}}^j$. The importance of the variable, then, is denoted as:

$$VI(X^j) = \frac{1}{B} \sum_{b=1}^B (\text{error}_{\text{OOB}}^j - \text{error}_{\text{OOB}_t}) \quad (5.4)$$

VI for each variable is shown in Fig. 5.3. The VSURF procedure begins by calculating VI for every variable included in the model, and sorting them in decreasing order of importance. Those below a threshold, chosen to be $2.95e-5$ our procedure, are removed. A series of random forest models are then created with the step-wise addition

of variables in descending order of importance until the mean errorOOB decreases by less than a pre-defined threshold.

The selected community risk factors are used as inputs to the CCN to create a baseline: i.e., establishing the network of similarities between the communities. In this step, 48 input nodes are selected to form the baseline CCN and each county is mapped to one of the CCN nodes based on the values of the selected risk factors (Fig. 5.1 b). Counties which occupy adjacent or nearby nodes in the CCN have greater similarity in the 20 selected risk factors (Fig. 5.1 a). As communities transform, their similarity with others will morph; resulting in a reconfiguration of the CCN and subsequently a county being mapped to alternative nodes in the CCN.

5.2.4 Self Organized Maps

To develop contrastive community networks, we utilize Self Organized Maps (SOM) [181,184,200]. SOMs are an unsupervised learning algorithm, based on artificial neural networks, for producing a low-dimensional, nonlinear representations of complex high-dimensional data [184]. SOM models are a graph of adjacent vertices in which each element in high dimensions is mapped to a node in the network. The process of assigning input data to nodes is done iteratively through a competitive learning process detailed. The result is a graph (Fig. 5.1 a) which preserves the vectorial topology of the input data where closer nodes (called map units) within the map have higher similarity in the original input variables.

SOM models have been previously utilized for understanding the similarity between items in high-dimensional space without imposing assumptions on the structure of the data [201], and when looking for trends in spatiotemporal data relating to community and urban change [202–204].

What follows is a description of the SOM training process developed by Kohonen [184], and implemented in R [134,200]. For a fixed number of nodes (or map dimension), the training process assigns weights to each risk factor of the input data at

each node in the map. In our experiments, 40 nodes were selected with 6 connections between neighbors based on SOM size heuristics [205]. This creates the initial mapping between input space (original data) and output space (the SOM). The weights between nodes are initially assigned at random, then a random input data point is selected. The *winning* map node –defined as the node with mean input data which is closest to the selected point– is selected. The weights between winning node and all others are updated by a value $\Delta w_{j,i}$, based on the number of iterations and the mean risk factor values of nodes within the selected node’s topological neighborhood T . Eq. 5.5 shows the updating procedure of $\Delta w_{j,i}$

$$\Delta w_{j,i} = \eta(t) * T_{j,I(x)}(t) * (x_i - w_{j,i}) \quad \text{for all } i, j \quad (5.5)$$

where i and j refers to different neurons, x_i is the value of the input data for node i , t refers to iteration number, $I(x)$ refers to the winning neuron, and $w_{i,j}$ is the weight between node i and j . The learning rate as a function of iteration is $\eta(t)$, where

$$\eta(t) = \eta_0 \exp(-t/\tau_n) \quad (5.6)$$

and η decreases with t and based on a pre-assigned hyperparameter τ_n , chosen in our experiments to be 0.05 based on previous empirical studies [200]. The topological neighborhood, T , defines how many neighboring nodes contribute to updating the learning rate of the selected node and is defined where

$$T_{j,I(x)}(t) = \exp(-S_{j,I(x)}^2/2\sigma(t)^2) \quad (5.7)$$

and $S_{j,i}$ is the distance between weights such that $S_{j,i} = ||w_j - w_i||$ and $\sigma(t) = \sigma_0 \exp(-t/\tau_0)$, which shrinks the neighborhood size over successive iterations as well. This process of updating node weights is repeated for every input data point over a fixed number of iterations, chosen to be 10000 in our experiments based on empirically observing convergence of the distances between nodes.

5.2.5 Contrastive Community Networks

We utilize the SOM algorithm as the basis for developing a contrastive community network (CCN). The details of the CCN procedure are shown in Algorithm 4, and described in summary here. Input variables for the CCN are the community risk factors, r , selected as important in the VSURF procedure for each county in Florida with storm exposure variables removed so as to compare communities on the basis of their structure rather than their hazard exposure. Risk factors are scaled to a standard deviation of 1 with mean 0 to facilitate integration of input data of different magnitudes into the training of the SOM in line with previous empirical studies [200].

For each county, c , and each risk factor r , the initial node the county is mapped to in the SOM, n_0 is recorded. The value of the risk factor for the given county, $x_{r,c}$ is perturbed in increments of 0.01 (δ in Algorithm 4) which is in units of standard deviation of each risk factor. Each risk factor is perturbed starting from its lower limit, R_i^{\min} to its upper limit, R_i^{\max} .

For risk factors which implicitly have lower and/or upper limits based on the way they are calculated –like county level fractions of the population or income inequality which are defined on the range $[0, 1]$ – we scale the limits in the same way as the input data and utilize the scaled values as limits to the perturbation of each risk factor. For risk factors without explicit limits –such as income deficit– we perturb values within a range of 1.5 times the minimum and maximum risk factor observed across the counties.

At each perturbation iteration, the perturbed risk factor $x'_{r,c}$ is included in the set of all risk factors across all counties, and an updated SOM, S' is calculated. The node the county is mapped to with the updated risk factor values, n_1 , is then compared against n_0 . If the new node, n_1 is different than the original node n_0 , the euclidean distance between them is denoted $t_{r,c}$ which represents the length of the transformation trajectory, and $x'_{r,c}$ at the value of the change is the transformation threshold.

Algorithm 4 Contrastive Community Network

1: r_i is a risk factor, where $r_i \in [R_i^{\min}, R_i^{\max}]$ and $X = \{r_i\} \forall i$ 2: c is a county where $c = \{1, \dots, C\}$ 3: $N(S, c)$ is the node of the self organized map, S which county c maps to 4: n_i is an arbitrary node i in the SOM 5: $t_{r,c}$ is the temporal trajectory length for factor r and county c , and $T = [t_{r,c}] \forall r, c$ 6: $ n_i, n_j _S$ as the euclidean distance in the SOM S between nodes i and j 7: Select predictive model F using cross- validation 8: Select important features, x using vari- able selection on F	9: Train SOM, $S = \text{SOM}(X')$ 10: Tune SOM hyperparameters to minimize mean distance from county to node 11: for all $r \in X$ do 12: for all $c \in C$ do 13: $n_0 = N(S, c)$ 14: $x_{r,c} = R^{\min}$ 15: while $x_{r,c} \in [R_i^{\min}, R_i^{\max}]$ do 16: $x'_{r,c} = x_{r,c} + \delta$ 17: $S' = \text{SOM}(X')$ 18: $n_1 = N(S', c)$ 19: if $n_0 \neq n_1$ then 20: $t_{r,c} = n_0, n_1 _{S'}$
---	--

CCN algorithm to return the set of temporal trajectories for each county and risk factor. In summary, the algorithm trains a self-organized map based on pre-selected community features, then systematically perturbs the values of the risk factors and remaps the counties to the SOM using the perturbed risk factors to determine if reconfiguration occurs.

Computing multiple SOMs with alternative input data and has previously been utilized to understand how high-dimensional data about the makeup of communities transform over time [202–204]. As neighboring nodes in the CCN are of higher similarity than those farther apart, a county being re-mapped to a node farther away indicates a greater degree of transformation; thus the length of the transformation trajectory represents the magnitude of reorganization as a result of the change in the community risk factor. This process is outlined in detail in Algorithm 4.

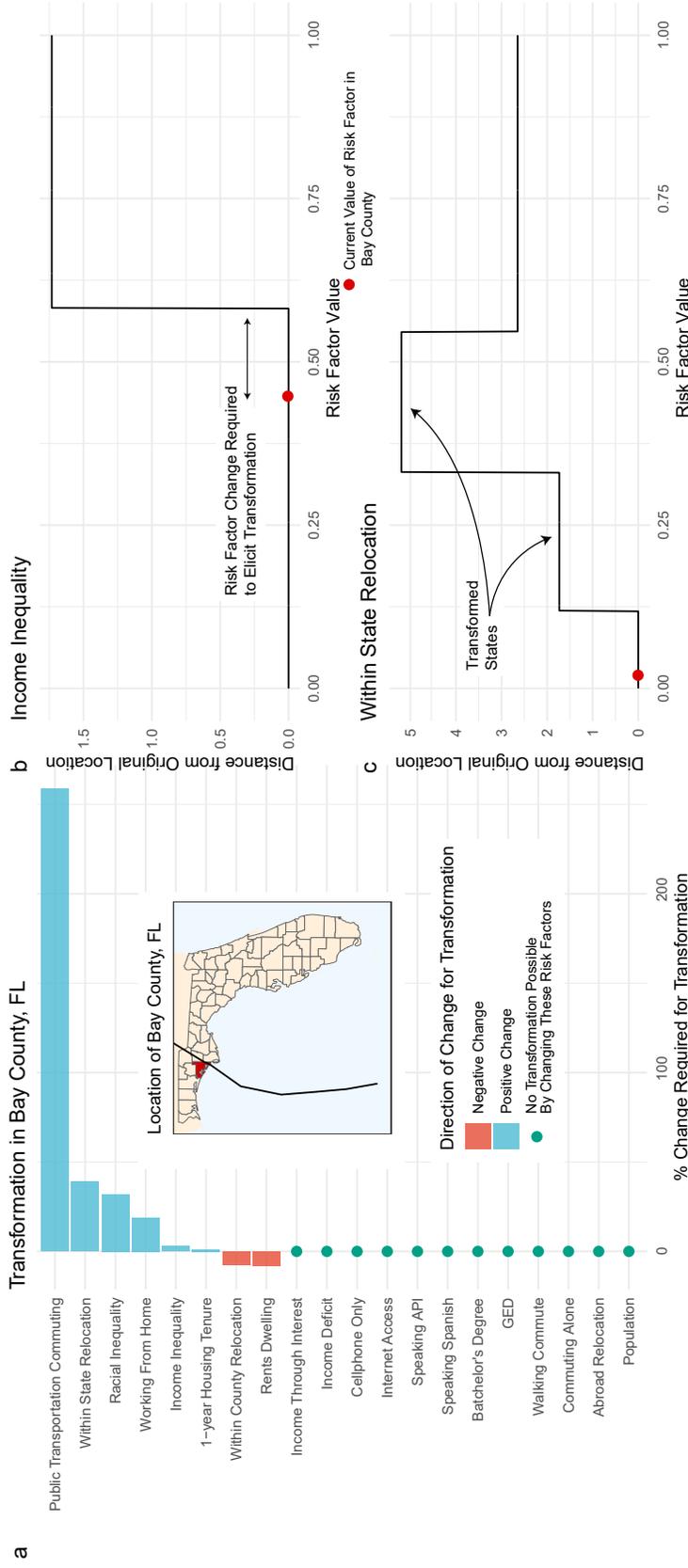


Figure 5.2. Community change required for transformation. (a) The amount of change required to elicit transformation in Bay County, FL for each risk factor. Blue indicates a positive risk factor shift is required for transformation, while red indicates a negative shift, and green dots show community risk factors for which no transformation is possible. (b, c) As each community risk factor is perturbed (x-axis of b,c), we compute where Bay County is mapped in the CCN. The values of the risk factors occurring at each jump indicate the degree of risk factor-perturbation leading to a reconfiguring of the CCN.

5.3 Results

5.3.1 Quantifying Community Transformation

After creating a baseline CCN, we perturb each community risk factor for each county, monitoring the configuration of the CCN at every updated value. We do this until the perturbation is great enough that the structure of the CCN re-organizes or ‘tips’ into an alternative configuration and county is mapped to a non-baseline node (Fig. 5.2b,c). Tracing the location of a county within the CCN as it re-organizes is called the county’s *transformation trajectory* [202, 203] (see Appendix Figure D.4), and the distance from the original to updated node within the CCN corresponds with the degree of transformation experienced. To illustrate the insights that can be drawn from this approach coupled with an engineering resilience model, we calculate the temporal trajectories for each risk factor in Bay County Florida (Fig. 5.2): a county which experienced extensive damage due to Hurricane Michael [206].

Results indicate that in Bay County, only 8 community risk factors (40% of those evaluated) have the possibility of triggering community transformation (Fig.5.2 a). For those which trigger transformation, we calculate the *transformation threshold*: the percentage increase or decrease in the risk factor associated with CCN reconfiguration. Transformation thresholds provide a relative comparison of the importance of risk factors as they contribute to transformation, such that counties or risk factors with lower thresholds are more sensitive to transforming.

In Bay County, two risk factors –the county level fraction of individuals who moved within the county and county level fraction of renter-occupied housing– have negative transformation thresholds while the other six (Fig.5.2 a) are positive. The six positive risk factors are the county level fraction of workers commuting primarily by public transportation, the county level fraction of the population who has moved to a given county from elsewhere in Florida in the past year, the county level measures of racial and income inequality, the county level fraction of the population who primarily works from home, and the county level fraction of the population that has lived in the same

residence for more than one year. These thresholds range from 11% (county level fraction of the population living in the same residence for more than one year) to 260% (county level fraction of workers commuting by public transportation). These transformation thresholds have two interpretations based on the normativity of the risk factor and the sign of the transformation.

5.3.2 Transformation vs. Degradation

In instances where the transformation threshold is positive and the risk factor is normatively good or neutral (*i.e.* risk factors for which an increase would be a community improvement and the threshold is positive), the transformation threshold represents a target for policy and decision makers. For example, in Bay County, a positive increase in the fraction of the population who commutes by public transportation –a normatively positive risk factor for improving the sustainability of a community [207]– will lead to community transformation (Fig. 5.2 a). This is in line with previous work which has found that access to public transportation provides sustainability and resilience benefits by improving individual health and equitable community connectivity [208,209]. This also applies to the inverse, in which normatively negative risk factors have negative transformation thresholds which serves as reduction targets.

Alternatively, community risk factors which are normatively negative but with positive transformation thresholds are indications of the potential for negative transformation or *degradation*. These degradation thresholds serve to highlight vulnerable community risk factors which could lead to a systematic transition toward negative outcomes. In Bay County, racial inequality and income inequality are both normatively negative risk factors which have positive transformation thresholds (Fig. 5.2a). A 29% increase in income inequality, for example, would lead to transformation, but represents a negative community outcome. As inequality in socioeconomic status is a

key contributor to vulnerability [25], this threshold outlines the relative degradation risk faced by Bay County as a result of changes in income inequality.

In instances where the transformation threshold is positive and the attribute is normatively good or neutral (*i.e.* attributes for which an increase would be a community improvement and the threshold is positive), the transformation threshold represents a target for policy and decision makers. For example in Bay County, a positive increase in the fraction of a population who commutes by public transportation –a normatively positive attribute for improving the sustainability of a community [207]– will lead to community transformation (Fig. 5.2 a), in line with previous work which has found that access to public transportation provides resilience benefits by improving individual health and equitable community connectivity [208, 209]. This also applies to the inverse, in which normatively negative attributes have negative transformation thresholds which serves as reduction targets.

5.3.3 Engineering Resilience and Transformation

To provide a more complete operational model of resilience which includes both engineering resilience and transformation, we calculate the relative contribution of risk factors toward transformation across the entire state of Florida, and compare those to risk factors associated with increased engineering resilience. Engineering resilience methods for assessing community resilience and access to critical services utilize predictive or explanatory modeling techniques to relate risk factors to disaster outcomes [82, 170–173, 210].

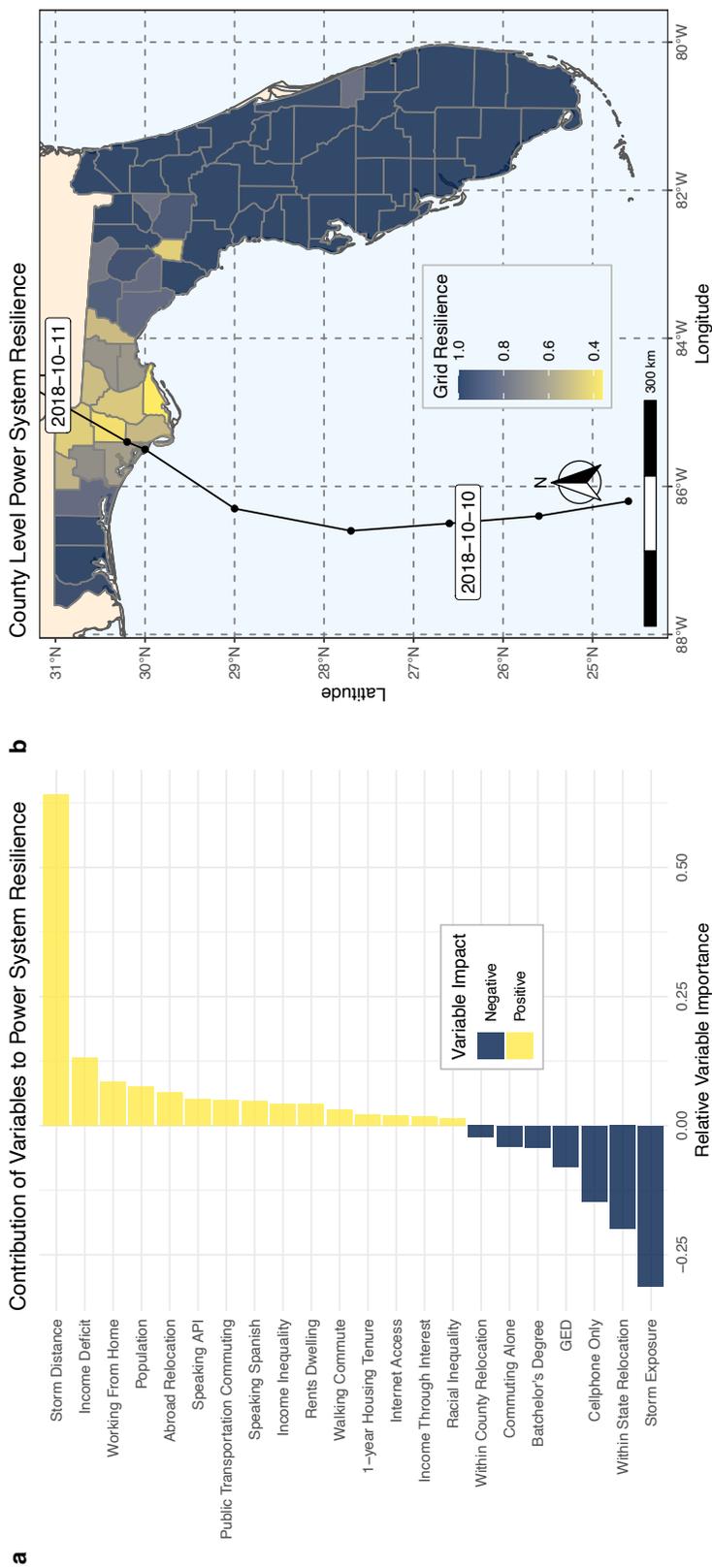


Figure 5.3. Community risk factors contributing to baseline engineering resilience. (a) Shows the results of the machine-learning variable selection indicating which community aspects have the highest relative contribution to resilience and whether they make a positive (yellow) or negative (blue) contribution to resilience. (b) a map colored by the resilience of the power grid to Hurricane Michael along with the storm's track. Darker counties were *more* resilient to the storm. Figures (a-b) created in R (v 3.2.1; <https://www.r-project.org/>) [179] using the `ggplot2` package (v 3.3.0; <https://ggplot2.tidyverse.org/>) [180]. Plot b additionally used `usmap` (v 0.5.0; <https://github.com/pdil/usmap>) [182]. Map shapefiles in b are from `usmap` and the US Census Bureau [183].

Table 5.1.
CCN importance and engineering resilience importance.

Risk Factor	CCN Importance Ranking (Fig. 5.2 a)	Engineering Resilience Importance Ranking (Fig. 5.3 a)
1-year Housing Tenure	1	17
Income Inequality	2	11
Internet Access	3	18
Within County Relocation	4	16
Within State Relocation	5	1
Working From Home	6	4
Rents Dwelling	7	12
Racial Inequality	8	20
Public Transportation Commuting	9	8

Given the localized and place-based nature of community resilience [211], and in line with previous studies [171, 173], we perform our analysis at a county level and use the same set of county level risk factors as the CCN. In the engineering resilience model, these indicators serve as independent variables in a model which predict the performance of critical infrastructure while controlling for population and hazard exposure (see Section 5.2.1); in our case study, this is the engineering resilience of the power grid at a county level in response to 2018 Hurricane Michael in Florida (Fig. 5.3b). By comparing the relative contribution of the risk factors toward improving engineering resilience with the contribution toward activating transformation, we demonstrate the importance of including both transformation and engineering resilience when operationalizing concepts of community resilience.

5.3.4 Triggering Transformation

Comparing the transformation thresholds and trajectories for all risk factors and counties in Florida (Fig. 5.4) against the risk factors identified as important in an engineering resilience model (Fig. 5.3), we find that the majority of risk factors identified as important for transformation are not identified as significant contributors to engineering resilience (Tab. 5.1). The risk factors with the lowest transformation thresholds are: the county level fraction of households who have lived in their current residence for over one year (10.3%), county level income inequality (15.8%), and county level internet access (43.9%). In the case of these three risk factors, they are the 17, 11, and 18th most important toward contributing to engineering resilience (Fig. 5.3 a).

Length of residence in a disaster-prone region is associated with decreased likelihoods of evacuation from major hurricanes, and greatly reduces perceptions of risk [212,213]. While the links between risk perception and community resilience are still being understood [214], we believe the importance of this risk factor in contributing to transformation comes from the place-based nature of community resilience and the social capital built with increased length of residence. Income inequality has also been identified as tightly linked to disaster outcomes; having been identified as both a consequence of major disasters [215], and a driver of more severe disaster outcomes [216], and individual behavior [217]. Access to communication technology has also been identified as a component of resilience in previous work [26,218].

5.3.5 Resilience Traps

The community risk factors which are identified as important through the engineering resilience model but do not allow for any possibility of transformation represent *resilience traps*. The term trap is used in many instances to describe feedback loops in which governance and interventions designed to rectify a larger societal problem contribute or exacerbate the problem, such as *poverty* traps in which individuals

are held in impoverished conditions by external forces [219], and *rigidity* traps when institutions and systems become self-reinforcing, and inflexible [220].

Contrary to poverty and rigidity traps —conceptualized as entirely social phenomena— resilience traps occur when mutually reinforcing socio-techno-ecological feedbacks drive systems towards persistent maladaptive states [221]; in essence emerging when short-term strategies are favored in the name of resilience over those which promote continual adaptation [222]. We argue that resilience traps occur because of an incomplete translation of resilience concepts into operational models, in which engineering resilience is the dominant paradigm. We identify risk factors which may be resilience traps in our case study by comparing the transformation thresholds for each risk factor and county in Florida with the relative importance of the risk factors contributing to engineering resilience. Risk factors which are candidates for being resilience traps are those which have a high relative importance in contributing to engineering resilience, but do not have any potential for transformation.

We find that of the risk factors included in the CCN, 11 of the 20 have no potential for transformation in *any* county evaluated in Florida while 9 allow for transformation in at least one (Fig. 5.4a-d). Of those which allow for transformation, county level income inequality has the smallest mean transformation threshold (6.25% across all counties), while the county level fraction of the population commuting by public transportation is the largest, with a mean transformation threshold of 12,042%. For the risk factors which do not allow for transformation, their importance as determined by an engineering resilience model is listed in Table 5.2.

The risk factors which do not allow for transformation range from the 4th to 20th most important variables as determined by the engineering resilience model (Tab. 5.2). The discrepancy in importance between engineering resilience and transformation highlight the possibility of resilience traps when aiming to operationalize the resilience of communities; and the potential barriers imposed by current resilience paradigms. Short-sighted policies, interventions, and investments motivated by solely prioritizing the risk factors which are associated with rapid recovery can entrench un-

tenable and non-sustainable aspects of the status quo [17] and inhibit transformation needed to promote a sustainable and resilient society.

5.3.6 Improved Operationalization of Community Resilience

In this section, we show the importance of including both engineering and transformation aspects of resilience when developing operational models by integrating the results of the CCN and engineering resilience models in our case study with previous work analyzing risk factors contributing to community resilience.

Income Deficit and *Income Through Interest* are two risk factors which do not contribute to transformation but are positively associated with engineering resilience. Income deficit quantifies the cumulative amount below the poverty line for all impoverished households the county [185]. This shows not just the number of households below the poverty line, but the degree of poverty experienced ¹ Income through interest is the fraction of households receiving income through interest, dividends or net rental income [185]. Previous work has found that higher income is associated with positive disaster outcomes [223] and is the most widely used resilience indicator [224]. There is also extensive evidence showing a greater impact of disasters on low income populations [26, 225]. Additionally, disasters create permanent increases rent in affected areas, while wealthy households expand their post-disaster real estate holdings [226]. Finally, federal disaster aid is primarily allotted to homeowners for disaster recovery [218]. Coupled these existing findings with CCN results which show that income-related risk factors do not contribute to transformation, we argue that the impact of income on disaster outcomes is based on the ability of wealthy communities to recover rapidly.

Additionally, the *Commuting Alone* risk factor –the fraction of a county who primarily commutes alone by car– is negatively associated with engineering resilience, while the fraction of a county who commutes by walking (*Walking Commute*) is

¹Note that deficit is a negative value so higher values of it correspond with less cumulative poverty.

positively associated with engineering resilience. This confirms previous results which show the prioritization of post-disaster recovery which favors higher-walkability urban areas as opposed to suburban or rural areas [115]. Finally, we note that two education-based risk factors –the fraction of a county with bachelor’s degrees and GEDs as their highest degrees earned (*Bachelor’s Degree*, and *GED*)– are negatively associated with engineering resilience. In this instance, we believe this is due to the proximity of the landfall location of Hurricane Michael to Leon County, FL. Leon County is home to Tallahassee and Florida State University both contributing to the county having the second highest proportion of Bachelor’s earners in the state.

Minority groups are frequently identified as vulnerable to disaster impacts [227, 228], however in our case study the fraction of communities speaking primarily Spanish (*Speaking Spanish*) or Asian and Pacific Island languages (*Speaking API*) are both positively associated with resilience outcomes in the engineering resilience model. We believe this indicates the presence of community organization and cohesion which is strongly associated with positive resilience outcomes [229].

Table 5.2.
Engineering resilience importance for non-transformation risk factors

Risk Factor	Engineering Resilience Importance Ranking (Fig. 5.3 a)
Cellphone Only	4
Income Deficit	5
GED	7
Abroad Relocation	9
Speaking API	10
Speaking Spanish	12
Bachelor’s Degree	13
Commuting Alone	16
Walking Commute	17
Income Through Interest	20

As is the case with all of the aforementioned risk factors, policy decisions and interventions based recovery-oriented risk factors will neglect the transformative aspects of resilience which are required for long-term sustainability. This echoes qualitative analyses of resilience-oriented policies which have found that the unilateral emphasis on restoring the status quo in engineering resilience models engenders norms and policies which inhibit the ability of communities to transform into alternative stable states [15, 24, 169, 230, 231]. Transformation as a process within communities is critical when a current system is untenable [232], and thus will be vital for attaining a high-sustainability future [28, 233].

By developing quantitative methods to assess the ability of communities to transform, we aim to shrink the gap between conceptual and operational models of resilience. We find that –when changed– only a subset of risk factors allow for transformation within a community and within those, certain risk factors will lead to positive transformation if they are improved while others will lead to negative transformation if they deteriorate. Furthermore, we identify resilience traps in which existing, engineering resilience models place importance on risk factors which ultimately do not allow for community transformation. The methods utilized in this work are scale and system agnostic. This allows for the systems analyzed and scale of analysis to be tailored to the scale and importance needed for effective decision-making. In this way, decision and policymakers can evaluate the level of transformation achievable through implementation policies and interventions to promote sustainable and resilient lifestyles, economies and societies.

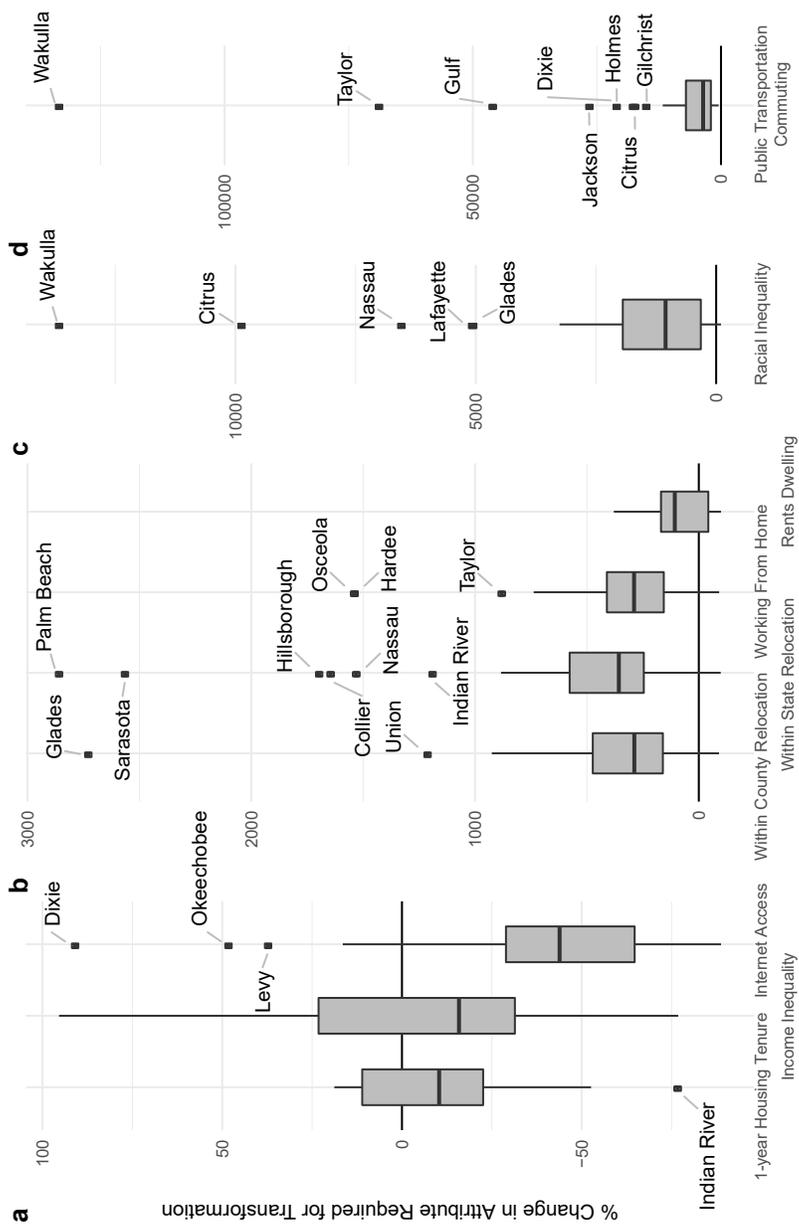


Figure 5.4. Risk Factor Change Required to Elicit Transformation. Each boxplot shows the percentage change in each risk factor required for a county to be mapped to an alternative node in the CCN, across all counties in the case study of Florida. (a) through (d) have increasing degrees of change required for transformation. Risk factors with no possibility of transformation are listed in Table 5.2. Note the shift in y-axis values moving from left to right.

REFERENCES

REFERENCES

- [1] Thomas Ackermann, Göran Andersson, and Lennart Söder. Distributed generation: a definition. *Electric power systems research*, 57(3):195–204, 2001.
- [2] Fikret Berkes and Helen Ross. Community Resilience: Toward an Integrated Approach. 26(1):5–20.
- [3] Angela Byers. Big Data, Big Economic Impact? 10:8.
- [4] Thomas F Stocker, Dahe Qin, Gian-Kasper Plattner, Melinda M B Tignor, Simon K Allen, Judith Boschung, Alexander Nauels, Yu Xia, Vincent Bex, and Pauline M Midgley. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. page 109.
- [5] David R. Easterling, Gerald A. Meehl, Camille Parmesan, Stanley A. Changnon, Thomas R. Karl, and Linda O. Mearns. Climate Extremes: Observations, Modeling, and Impacts. 289(5487):2068–2074.
- [6] Adam B. Smith and Richard W. Katz. US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. 67(2):387–410.
- [7] Ayyoob Sharifi. A critical review of selected tools for assessing community resilience. 69:629–647.
- [8] Susan L. Cutter, Lindsey Barnes, Melissa Berry, Christopher Burton, Elijah Evans, Eric Tate, and Jennifer Webb. A place-based model for understanding community resilience to natural disasters. 18(4):598–606.
- [9] J. Park, T. P. Seager, P. S. C. Rao, M. Convertino, and I. Linkov. Integrating Risk and Resilience Approaches to Catastrophe Management in Engineering Systems: Perspective. 33(3):356–367.
- [10] C S Holling. Resilience and Stability of Ecological Systems. 4(1):1–23.
- [11] C. S. Holling and Gary K. Meffe. Command and Control and the Pathology of Natural Resource Management. 10(2):328–337.
- [12] Philip R. Crowe, Karen Foley, and Marcus J. Collier. Operationalizing urban resilience through a framework for adaptive co-management and design: Five experiments in urban planning practice and policy. 62:112–119.
- [13] Robin Leichenko. Climate change and urban resilience. 3(3):164–168.
- [14] Barbara Pizzo. Problematizing resilience: Implications for planning theory and practice. 43:133–140.

- [15] Mike Raco and Emma Street. Resilience Planning, Economic Change and The Politics of Post-recession Development in London and Hong Kong:.
- [16] Sara Meerow, Joshua P. Newell, and Melissa Stults. Defining urban resilience: A review. 147:38–49.
- [17] Sara Meerow and Joshua P. Newell. Urban resilience for whom, what, when, where, and why? 40(3):309–329.
- [18] Sonny S. Patel, M. Brooke Rogers, Richard Amlôt, and G. James Rubin. What Do We Mean by 'Community Resilience'? A Systematic Literature Review of How It Is Defined in the Literature. 9.
- [19] Paul H. Gleye. City Planning versus Urban Planning: Resolving a Profession's Bifurcated Heritage. 30(1):3–17.
- [20] George A. Bonanno, Sandro Galea, Angela Bucciarelli, and David Vlahov. Psychological Resilience After Disaster: New York City in the Aftermath of the September 11th Terrorist Attack. 17(3):181–186.
- [21] Carl Gibson. An integrated approach to managing disruption-related risk: Life and death in a model community.
- [22] Fran H. Norris, Susan P. Stevens, Betty Pfefferbaum, Karen F. Wyche, and Rose L. Pfefferbaum. Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. 41(1-2):127–150.
- [23] Robin S. Cox and Karen-Marie Elah Perry. Like a Fish Out of Water: Reconsidering Disaster Recovery and the Role of Place and Social Capital in Community Disaster Resilience. 48(3-4):395–411.
- [24] Carl Folke. Resilience: The emergence of a perspective for social–ecological systems analyses. 16(3):253–267.
- [25] Susan L. Cutter, Bryan J. Boruff, and W. Lynn Shirley. Social Vulnerability to Environmental Hazards*. 84(2):242–261.
- [26] Susan L. Cutter, Christopher G. Burton, and Christopher T. Emrich. Disaster Resilience Indicators for Benchmarking Baseline Conditions. 7(1).
- [27] Paolo Gardoni and Colleen Murphy. Society-based design: Promoting societal well-being by designing sustainable and resilient infrastructure. 5(1-2):4–19.
- [28] Thomas Elmqvist, Erik Andersson, Niki Frantzeskaki, Timon McPhearson, Per Olsson, Owen Gaffney, Kazuhiko Takeuchi, and Carl Folke. Sustainability and resilience for transformation in the urban century. 2(4):267–273.
- [29] Kash Barker, James H. Lambert, Christopher W. Zobel, Andrea H. Tapia, Jose E. Ramirez-Marquez, Laura Albert, Charles D. Nicholson, and Cornelia Caragea. Defining resilience analytics for interdependent cyber-physical-social networks. 2(2):59–67.
- [30] Kate Starbird and Leysia Palen. "Voluntweeters": Self-organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1071–1080. ACM.

- [31] João Porto de Albuquerque, Benjamin Herfort, Alexander Brenning, and Alexander Zipf. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. 29(4):667–689.
- [32] Maged N. Kamel Boulos, Bernd Resch, David N. Crowley, John G. Breslin, Gunho Sohn, Russ Burtner, William A. Pike, Eduardo Jezierski, and Kuo-Yu Slayer Chuang. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. 10(1):67.
- [33] Kate Starbird, Leysia Palen, Amanda L. Hughes, and Sarah Vieweg. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*, page 241. ACM Press.
- [34] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, and Danah Boyd. The Arab Spring— The Revolutions Were Tweeted: Information Flows during the 2011 Tunisian and Egyptian Revolutions. 5(0):31.
- [35] Amanda L. Hughes, Lise A. A. St. Denis, Leysia Palen, and Kenneth M. Anderson. Online Public Communications by Police & Fire Services During the 2012 Hurricane Sandy. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*, pages 1505–1514. ACM.
- [36] Dhiraj Murthy and Alexander J. Gross. Social media processes in disasters: Implications of emergent technology use. 63:356–370.
- [37] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [38] Maite Taboada. Sentiment Analysis: An Overview from Linguistics. 2(1):325–347.
- [39] Mark Newman. *Networks*. Oxford University Press.
- [40] Zheye Wang and Xinyue Ye. Social media analytics for natural disaster management. 32(1):49–72.
- [41] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. 42(2):12.
- [42] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. 6(11):888–893.
- [43] C. Rossi, F. S. Acerbo, K. Ylinen, I. Juga, P. Nurmi, A. Bosca, F. Tarasconi, M. Cristoforetti, and A. Alikadic. Early detection and information extraction for weather-induced floods using social media streams. 30:145–157.
- [44] Yury Kryvasheyev, Haohui Chen, Esteban Moro, Pascal Van Hentenryck, and Manuel Cebrian. Performance of Social Network Sensors during Hurricane Sandy. 10(2):e0117288.

- [45] Yury Kryvasheyev, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. 2(3):e1500779.
- [46] Benjamin Herfort, João Porto de Albuquerque, Svend-Jonas Schelhorn, and Alexander Zipf. Exploring the Geographical Relations Between Social Media and Flood Phenomena to Improve Situational Awareness. In Joaquín Huerta, Sven Schade, and Carlos Granell, editors, *Connecting a Digital Europe Through Location and Place*, Lecture Notes in Geoinformation and Cartography, pages 55–71. Springer International Publishing.
- [47] Stephane Roche, Eliane Propeck-Zimmermann, and Boris Mericskay. GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. 78(1):21–40.
- [48] Sai Krishna Theja Bhavaraju, Cyril Beyney, and Charles Nicholson. Quantitative analysis of social media sensitivity to natural disasters. 39:101251.
- [49] Issie Lapowski. Your Old Tweets Give Away More Location Data Than You Think — WIRED.
- [50] Meredith T. Niles, Benjamin F. Emery, Andrew J. Reagan, Peter Sheridan Dodds, and Christopher M. Danforth. Social media usage patterns during natural hazards. 14(2):e0210484.
- [51] Armando López-Cuevas, José Ramírez-Márquez, Gildardo Sanchez-Ante, and Kash Barker. A Community Perspective on Resilience Analytics: A Visual Analysis of Community Mood. 37(8):1566–1579.
- [52] Nilufar Matin, John Forrester, and Jonathan Ensor. What is equitable resilience? 109:197–205.
- [53] Seth Guikema and Roshanak Nateghi. Modeling Power Outage Risk From Natural Hazards.
- [54] Min Ouyang. Review on modeling and simulation of interdependent critical infrastructure systems. 121:43–60.
- [55] Seyedmohsen Hosseini, Kash Barker, and Jose E. Ramirez-Marquez. A review of definitions and measures of system resilience. 145:47–61.
- [56] Roshanak Nateghi, Seth D. Guikema, and Steven M. Quiring. Comparison and Validation of Statistical Methods for Predicting Power Outage Durations in the Event of Hurricanes: Comparison and Validation of Statistical Methods. 31(12):1897–1906.
- [57] R. Nateghi. Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems. 6:13478–13489.
- [58] Andrea Staid, Seth D. Guikema, Roshanak Nateghi, Steven M. Quiring, and Michael Z. Gao. Simulation of tropical cyclone impacts to the U.S. power system under climate change scenarios. 127(3-4):535–546.
- [59] Burcin Cakir Erdener, Kwabena A. Pambour, Ricardo Bolado Lavin, and Berna Dengiz. An integrated simulation model for analysing electricity and gas systems. 61:410–420.

- [60] Pavel Praks, Vytis Kopustinskias, and Marcelo Masera. Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. 144:254–264.
- [61] Graham Booker, Jacob Torres, Seth Guikema, Alex Sprintson, and Kelly Brumbelow. Estimating cellular network performance during hurricanes. 95(4):337–344.
- [62] Paul Hines, Eduardo Cotilla-Sanchez, and Seth Blumsack. Do topological models provide good information about electricity infrastructure vulnerability? 20(3):033122.
- [63] Chuanyi Ji, Yun Wei, Henry Mei, Jorge Calzada, Matthew Carey, Steve Church, Timothy Hayes, Brian Nugent, Gregory Stella, Matthew Wallace, Joe White, and Robert Wilcox. Large-scale data analysis of power grid resilience across multiple US service regions. 1(5).
- [64] Leonardo Dueñas Osorio and Srivishnu Mohan Vemuru. Cascading failures in complex infrastructure systems. 31(2):157–167.
- [65] Winkler James, Dueñas-Osorio Leonardo, Stein Robert, and Subramanian Devika. Interface Network Models for Complex Urban Infrastructure Systems. 17(4):138–150.
- [66] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [67] Sybil Derrible. Complexity in future cities: The rise of networked infrastructure. 21:68–86.
- [68] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. 530(7590):307–312.
- [69] Leonardo Dueñas Osorio, James I. Craig, and Barry J. Goodno. Seismic response of critical interdependent networks. 36(2):285–306.
- [70] Jonatan Zischg, Christopher Klinkhamer, Xianyuan Zhan, P. Suresh C. Rao, and Robert Sitzenfrei. A Century of Topological Coevolution of Complex Infrastructure Networks in an Alpine City.
- [71] Soohyun Yang, Kyungrock Paik, Gavan S. McGrath, Christian Urich, Elisabeth Krueger, Praveen Kumar, and P. Suresh C. Rao. Functional Topology of Evolving Urban Drainage Networks. 53(11):8966–8979.
- [72] Takahiro Yabe, Satish V. Ukkusuri, and P. Suresh C. Rao. Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after Hurricane Maria. 4(1):98.
- [73] Sarah LaRocca, Jonas Johansson, Henrik Hassel, and Seth Guikema. Topological Performance Measures as Surrogates for Physical Flow Models for Risk and Vulnerability Analysis for Electric Power Systems. 35(4):608–623.
- [74] Fuyu Hu, Chi Ho Yeung, Saini Yang, Weiping Wang, and An Zeng. Recovery of infrastructure networks after localised attacks. *Scientific reports*, 6:24522, 2016.

- [75] Antonio Majdandzic, Lidia A. Braunstein, Chester Curme, Irena Vodenska, Sary Levy-Carciente, H. Eugene Stanley, and Shlomo Havlin. Multiple tipping points and optimal repairing in interacting networks. 7:10850.
- [76] United States Department of Homeland Security. *National Infrastructure Protection Plan*. U.S. Department of Homeland Security.
- [77] W. Neil Adger, Terry P. Hughes, Carl Folke, Stephen R. Carpenter, and Johan Rockström. Social-Ecological Resilience to Coastal Disasters. 309(5737):1036–1039.
- [78] Yossi Sheffi et al. The resilient enterprise: overcoming vulnerability for competitive advantage. *MIT Press Books*, 1, 2005.
- [79] Michel Bruneau, Stephanie E. Chang, Ronald T. Eguchi, George C. Lee, Thomas D. O'Rourke, Andrei M. Reinhorn, Masanobu Shinozuka, Kathleen Tierney, William A. Wallace, and Detlof von Winterfeldt. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. 19(4):733–752.
- [80] Hiba Baroud, Kash Barker, Jose E. Ramirez-Marquez, and Claudio M. Rocco S. Importance measures for inland waterway network resilience. 62:55–67.
- [81] Lichun Chen and Elise Miller-Hooks. Resilience: An Indicator of Recovery Capability in Intermodal Freight Transport. 46(1):109–123.
- [82] Seth D. Guikema and Terje Aven. Assessing risk from intelligent attacks: A perspective on approaches. 95(5):478–483.
- [83] Adam Rose and Shu-Yi Liao. Modeling Regional Economic Resilience to Disasters: A Computable General Equilibrium Analysis of Water Service Disruptions*. 45(1):75–112.
- [84] Devanandham Henry and Jose Emmanuel Ramirez-Marquez. Generic metrics and quantitative approaches for system resilience as a function of time. 99:114–122.
- [85] Scott Thacker, Raghav Pant, and Jim W. Hall. System-of-systems formulation and disruption analysis for multi-scale critical national infrastructures. 167:30–41.
- [86] Gian Paolo Cimellaro, Andrei M. Reinhorn, and Michel Bruneau. Framework for analytical quantification of disaster resilience. 32(11):3639–3649.
- [87] Cimellaro Gian Paolo, Renschler Chris, Reinhorn Andrei M., and Arendt Lucy. PEOPLES: A Framework for Evaluating Resilience. 142(10):04016063.
- [88] Min Ouyang. A mathematical framework to optimize resilience of interdependent critical infrastructure systems under spatially localized attacks. 262(3):1072–1084.
- [89] Bilal M. Ayyub. Systems Resilience for Multihazard Environments: Definition, Metrics, and Valuation for Decision Making: Systems Resilience for Multihazard Environments. 34(2):340–355.

- [90] Bilal M. Ayyub. Practical Resilience Metrics for Planning, Design, and Decision Making. 1(3):04015008.
- [91] Hurricane michael - power outage data.
- [92] N. O. Attoh-Okine, A. T. Cooper, and S. A. Mensah. Formulation of Resilience Index of Urban Infrastructure Using Belief Functions. 3(2):147–153.
- [93] Pengcheng Zhang and Srinivas Peeta. A generalized modeling framework to analyze interdependencies among infrastructure systems. 45(3):553–579.
- [94] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. 21(6):11–25.
- [95] Zimmerman Rae, Zhu Quanyan, de Leon Francisco, and Guo Zhan. Conceptual Modeling Framework to Integrate Resilient and Interdependent Infrastructure in Extreme Weather. 23(4):04017034.
- [96] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. 464(7291):1025–1028.
- [97] Cen Nan and Giovanni Sansavini. A quantitative method for assessing resilience of interdependent infrastructures. 157:35–53.
- [98] Nils Kalstad Svendsen and Stephen D. Wolthusen. Connectivity models of interdependency in mixed-type critical infrastructure networks. 12(1):44–55.
- [99] Dueñas-Osorio Leonardo, Craig James I., Goodno Barry J., and Bostrom Ann. Interdependent Response of Networked Systems. 13(3):185–194.
- [100] Nassim Nicholas Taleb. *Antifragile: Things That Gain from Disorder*. Random House.
- [101] Benjamin Rachunok and Roshanak Nateghi. The sensitivity of electric power infrastructure resilience to the spatial distribution of disaster impacts. 193:106658.
- [102] Benjamin Rachunok and Roshanak Nateghi. Interdependent Infrastructure System Risk and Resilience to Natural Hazards.
- [103] Benjamin Rachunok and Roshanak Nateghi. The sensitivity of electric power infrastructure resilience to the spatial distribution of disaster impacts. *arXiv preprint arXiv:1902.02879*, 2019.
- [104] Christopher N Kaiser-Bunbury, James Mougall, Andrew E Whittington, Terence Valentin, Ronny Gabriel, Jens M Olesen, and Nico Blüthgen. Ecosystem restoration strengthens pollination network resilience and function. *Nature*, 542(7640):223, 2017.
- [105] Crawford S Holling. Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4(1):1–23, 1973.
- [106] Shelley A Riggs and David S Riggs. Risk and resilience in military families experiencing deployment: The role of the family attachment network. *Journal of Family Psychology*, 25(5):675, 2011.

- [107] Michael C. Dietze, Andrew Fox, Lindsay M. Beck-Johnson, Julio L. Betancourt, Mevin B. Hooten, Catherine S. Jarnevich, Timothy H. Keitt, Melissa A. Kenney, Christine M. Laney, Laurel G. Larsen, Henry W. Loescher, Claire K. Lunch, Bryan C. Pijanowski, James T. Randerson, Emily K. Read, Andrew T. Tredennick, Rodrigo Vargas, Kathleen C. Weathers, and Ethan P. White. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. page 201710231.
- [108] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703.
- [109] Cen Nan and Giovanni Sansavini. A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering & System Safety*, 157:35–53.
- [110] Rae Zimmerman, Quanyan Zhu, and Carolyn Dimitri. Promoting resilience for food, energy, and water interdependencies. *Journal of Environmental Studies and Sciences*, 6(1):50–61, 2016.
- [111] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. *Nature*, 530(7590):307–312.
- [112] Igor Linkov, Todd Bridges, Felix Creutzig, Jennifer Decker, Cate Fox-Lent, Wolfgang Kröger, James H Lambert, Anders Levermann, Benoit Montreuil, Jatin Nathwani, et al. Changing the resilience paradigm. *Nature Climate Change*, 4(6):407, 2014.
- [113] Seung-Ryong Han, Seth D. Guikema, Steven M. Quiring, Kyung-Ho Lee, David Rosowsky, and Rachel A. Davidson. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. 94(2):199–210.
- [114] Min Ouyang, Leonardo Dueñas-Osorio, and Xing Min. A three-stage resilience analysis framework for urban infrastructure systems. *Structural safety*, 36:23–31, 2012.
- [115] Sayanti Mukherjee, Roshanak Nateghi, and Makarand Hastak. A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. 175:283–305.
- [116] A. Arab, A. Khodaei, Z. Han, and S. K. Khator. Proactive Recovery of Electric Power Assets for Resiliency Enhancement. 3:99–109.
- [117] S. Shashaani, S. D. Guikema, C. Zhai, J. V. Pino, and S. M. Quiring. Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages. 6:62432–62449.
- [118] Yiping Fang and Giovanni Sansavini. Emergence of antifragility by optimum postdisruption restoration planning of infrastructure networks. *Journal of Infrastructure Systems*, 23(4):04017024, 2017.
- [119] Pavel Praks, Vytis Kopustinskas, and Marcelo Masera. Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. *Reliability Engineering & System Safety*, 144:254–264.

- [120] Winkler James, Dueñas-Osorio Leonardo, Stein Robert, and Subramanian Devika. Interface Network Models for Complex Urban Infrastructure Systems. *Journal of Infrastructure Systems*, 17(4):138–150.
- [121] Leonardo Dueñas-Osorio and Srivishnu Mohan Vemuru. Cascading failures in complex infrastructure systems. *Structural Safety*, 31(2):157–167.
- [122] M. Panteli and P. Mancarella. The Grid: Stronger, Bigger, Smarter?: Presenting a Conceptual Framework of Power System Resilience. 13(3):58–66.
- [123] Sebastián Espinoza, Mathaios Panteli, Pierluigi Mancarella, and Hugh Rudnick. Multi-phase assessment and adaptation of power systems resilience to natural hazards. 136:352–361.
- [124] Yi-Ping Fang and Enrico Zio. An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards. 276(3):1119–1136.
- [125] Sarah G. Nurre, Burak Cavdaroglu, John E. Mitchell, Thomas C. Sharkey, and William A. Wallace. Restoring infrastructure systems: An integrated network design and scheduling (INDS) problem. *European Journal of Operational Research*, 223(3):794–806.
- [126] Andrés D. González, Leonardo Dueñas Osorio, Mauricio Sánchez-Silva, and Andrés L. Medaglia. The Interdependent Network Design Problem for Optimal Infrastructure System Restoration. 31(5):334–350.
- [127] Kash Barker, Jose Emmanuel Ramirez-Marquez, and Claudio M. Rocco. Resilience-based network component importance measures. 117:89–97.
- [128] Leonardo Dueñas-Osorio and Srivishnu Mohan Vemuru. Cascading failures in complex infrastructure systems. *Structural Safety*, 31(2):157–167.
- [129] Adrian Bondy and M. Ram Murty. *Graph Theory*. Graduate Texts in Mathematics. Springer-Verlag.
- [130] Ian Dobson. Electricity grid: When the lights go out. *Nature Energy*, 1(5):16059, 2016.
- [131] Vito Latora and Massimo Marchiori. Efficient Behavior of Small-World Networks. 87(19):198701.
- [132] Åke J. Holmgren. Using Graph Models to Analyze the Vulnerability of Electric Power Networks. 26(4):955–969.
- [133] Weiman Sun and An Zeng. Target recovery in complex networks. *The European Physical Journal B*, 90(1):10.
- [134] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [135] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*:1695, 2006.

- [136] Roger Flage, Terje Aven, Enrico Zio, and Piero Baraldi. Concerns, challenges, and directions of development for the issue of representing uncertainty in risk assessment. *Risk Analysis*, 34(7):1196–1207, 2014.
- [137] Terje Aven. The Concept of Antifragility and its Implications for the Practice of Risk Analysis. 35(3):476–483.
- [138] Min Ouyang. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering & System Safety*, 121:43–60.
- [139] Lenore Newman and Ann Dale. Network Structure, Diversity, and Proactive Resilience Building: A Response to Tompkins and Adger. 10(1).
- [140] Adam B. Smith and Richard W. Katz. US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Natural Hazards*, 67(2):387–410.
- [141] R. Nateghi. Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems. 6:13478–13489.
- [142] Patrick Meier. Human Computation for Disaster Response. In Pietro Michelucci, editor, *Handbook of Human Computation*, pages 95–104. Springer New York.
- [143] U.S. population with a social media profile 2018.
- [144] Top U.S. mobile social apps by users 2018 — Statistic.
- [145] Arun Agrawal and Clark C Gibson. Enchantment and Disenchantment: The Role of Community in Natural Resource Conservation. 27(4):629–649.
- [146] Stanley W Gilbert. *Disaster resilience: A guide to the literature*. CreateSpace Independent Publishing Platform, 2016.
- [147] Kimberly Glasgow, Clayton Fink, and Jordan Boyd-Graber. “Our Grief is Unspeakable”: Automatically Measuring the Community Impact of a Tragedy. page 9.
- [148] Yury Kryvasheyev, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science advances*, 2(3):e1500779, 2016.
- [149] Amanda Lee Hughes and Leysia Palen. Twitter adoption and use in mass convergence and emergency events. 6(3/4):248.
- [150] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. Microblogging during two natural hazards events: What twitter may contribute to situational awareness. page 10.
- [151] Jooho Kim and Makarand Hastak. Social network analysis: Characteristics of online social networks after a disaster. 38(1):86–96.
- [152] Arkaitz Zubiaga. A longitudinal assessment of the persistence of twitter datasets. 69(8):974–984.

- [153] Tyler W. Rinker. `qdapRegex`: Regular expression removal, extraction, and replacement tools. 2017. 0.7.3.
- [154] C. Silva and B. Ribeiro. The importance of stop word removal on recall values in text categorization. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 3, pages 1661–1666 vol.3.
- [155] Kenneth Benoit, David Muhr, and Kohei Watanabe. *stopwords: Multilingual Stopword Lists*, 2017. R package version 0.9.0.
- [156] Ingo Feinerer and Kurt Hornik. *tm: Text Mining Package*, 2018. R package version 0.7-5.
- [157] Ingo Feinerer, Kurt Hornik, and David Meyer. Text mining infrastructure in r. *Journal of Statistical Software*, 25(5):1–54, March 2008.
- [158] Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. `quanteda`: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774, 2018.
- [159] Martin F Porter. *Snowball: A language for stemming algorithms*, 2001.
- [160] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [161] Saeedeh Momtazi, Sanjeev Khudanpur, and Dietrich Klakow. A Comparative Study of Word Co-occurrence for Term Clustering in Language Model-based Sentence Retrieval. page 4.
- [162] Jurafsky, Daniel and Martin, James H. Vector Semantics. In *Speech and Language Processing*.
- [163] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics -*, pages 76–83. Association for Computational Linguistics.
- [164] Philip A Knight. The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275, 2008.
- [165] J. C. Bezdek and R. J. Hathaway. VAT: A tool for visual assessment of (cluster) tendency. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, volume 3, pages 2225–2230 vol.3.
- [166] Alboukadel Kassambara and Fabian Mundt. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2017. R package version 1.0.5.
- [167] Kevin Stowe, Jennings Anderson, Martha Palmer, Leysia Palen, and Ken Anderson. Improving Classification of Twitter Behavior During Hurricane Events. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 67–75. Association for Computational Linguistics.
- [168] Min Ouyang and Zhenghua Wang. Resilience assessment of interdependent infrastructure systems: With a focus on joint restoration modeling and analysis. 141:74–82.

- [169] Danny MacKinnon and Kate Driscoll Derickson. From resilience to resourcefulness: A critique of resilience policy and activism. 37(2):253–270.
- [170] Ayyoob Sharifi. A critical review of selected tools for assessing community resilience. *Ecological Indicators*, 69:629–647, 2016.
- [171] Susan L Cutter. The landscape of disaster resilience indicators in the usa. *Natural hazards*, 80(2):741–758, 2016.
- [172] Armin Tabandeh, Paolo Gardoni, Colleen Murphy, and Natalie Myers. Societal Risk and Resilience Analysis: Dynamic Bayesian Network Formulation of a Capability Approach. 5(1):04018046.
- [173] Susan L. Cutter and Sahar Derakhshan. Temporal and spatial change in disaster resilience in US counties, 2010–2015. 19(1):10–29.
- [174] Carl Folke, Thomas Hahn, Per Olsson, and Jon Norberg. Adaptive Governance of Social-Ecological Systems. 30(1):441–473.
- [175] Burak Cavdaroglu, Erik Hammel, John E. Mitchell, Thomas C. Sharkey, and William A. Wallace. Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems. 203(1):279–294.
- [176] Sarah G. Nurre, Burak Cavdaroglu, John E. Mitchell, Thomas C. Sharkey, and William A. Wallace. Restoring infrastructure systems: An integrated network design and scheduling (INDS) problem. 223(3):794–806.
- [177] Thomas C. Sharkey, Burak Cavdaroglu, Huy Nguyen, Jonathan Holman, John E. Mitchell, and William A. Wallace. Interdependent network restoration: On the value of information-sharing. 244(1):309–321.
- [178] Mark Pelling. *The Vulnerability of Cities: Natural Disasters and Social Resilience*. Earthscan Publications.
- [179] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [180] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [181] Ron Wehrens and Johannes Kruisselbrink. Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software*, 87(7):1–18, 2018.
- [182] Paolo Di Lorenzo. *usmap: US Maps Including Alaska and Hawaii*, 2019. R package version 0.5.0.
- [183] U.S. Census Bureau. *TIGER/Line Shapefiles (machinereadable data files)*, 2019. Date Accessed: 11/01/2019.
- [184] T. Kohonen. Exploration of very large databases by self-organizing maps. In *Proceedings of International Conference on Neural Networks (ICNN'97)*, volume 1, pages PL1–PL6 vol.1.
- [185] U.S. Census Bureau. *American Community Survey 5-year estimates*. Accessed June 23rd, 2020.

- [186] Peter D. Howe, Matto Mildenerger, Jennifer R. Marlon, and Anthony Leiserowitz. Geographic variation in opinions on climate change at state and local scales in the USA. *5(6):596–603*.
- [187] Kyle Walker. *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2020. R package version 0.9.9.2.
- [188] US National Centers for Environmental Information, National Oceanic and Atmospheric Administration. Storm events database.
- [189] Brooke Anderson, Andrea Schumacher, Seth Guikema, Steven Quiring, and Joshua Ferreri. *stormwindmodel: Model Tropical Cyclone Wind Speeds*, 2018. R package version 0.1.1.
- [190] Diana Mitsova, Monica Escaleras, Alka Sapat, Ann-Margaret Esnard, and Alberto J. Lamadrid. The Effects of Infrastructure Service Disruptions and Socio-Economic Vulnerability on Hurricane Recovery. *11(2):516*.
- [191] William N Venables and Brian D Ripley. *Modern applied statistics with S-PLUS*. Springer Science and Business Media, 2013.
- [192] Andy Liaw and Matthew Weiner. Classification and regression by randomforest. *R News*, *2(3):18–22*, 2002.
- [193] Adam Kapelner and Justin Bleich. bartMachine: Machine learning with Bayesian additive regression trees. *Journal of Statistical Software*, *70(4):1–40*, 2016.
- [194] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [195] Leo Breiman. Random forests. *Machine learning*, *45(1):5–32*, 2001.
- [196] S.R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *21(3):660–674*.
- [197] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. *VSURF: Variable Selection Using Random Forests*, 2019. R package version 1.1.0.
- [198] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, *7(2):19–33*, December 2015.
- [199] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. Variable selection using random forests. *31(14):2225–2236*.
- [200] Ron Wehrens and Lutgarde M. C. Buydens. Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, *21(5):1–19*, 2007.
- [201] Aiyshwariya Paulvannan Kanmani, Renee Obringer, Benjamin Rachunok, and Roshanak Nateghi. Assessing global environmental sustainability via an unsupervised clustering framework. *Sustainability*, *12(2):563*, 2020.
- [202] Andrew C. D. Lee and Claus Rinner. Visualizing urban social change with Self-Organizing Maps: Toronto neighbourhoods, 1996–2006. *45:92–98*.

- [203] Elizabeth Delmelle, Jean-Claude Thill, Owen Furuseth, and Thomas Ludden. Trajectories of multidimensional neighbourhood quality of life change. *Urban Studies*, 50(5):923–941, 2013.
- [204] Andr Skupin and Ron Hagelman. Visualizing Demographic Trajectories with Self-Organizing Maps. 9(2):159–179.
- [205] Jing Tian, Michael H. Azarian, and Michael Pecht. *Anomaly Detection Using Self-Organizing Maps-Based K-Nearest Neighbor Algorithm*.
- [206] Division of Water Resource Management. Hurricane Michael Post-Storm Beach Conditions and Coastal Impact Report, 04-2019.
- [207] L. Hockstad and L. Hanel. Inventory of u.s. greenhouse gas emissions and sinks. *OSTI Dataset*, 1 2018.
- [208] Susan L. Cutter and Sahar Derakhshan. Implementing Disaster Policy: Exploring Scale and Measurement Schemes for Disaster Resilience. 16(3).
- [209] {Mitigation Framework Leadership Group}. Draft Interagency Concept for Community Resilience Indicators and National-Level Measures.
- [210] Jin-Zhu Yu and Hiba Baroud. Quantifying Community Resilience Using Hierarchical Bayesian Kernel Methods: A Case Study on Recovery from Power Outages. 39(9):1930–1948.
- [211] Sarah Henly-Shepard, Cheryl Anderson, Kimberly Burnett, Linda J. Cox, John N. Kittinger, and Maka‘ala Ka‘aumoana. Quantifying household social resilience: A place-based approach in a rapidly transforming community. 75(1):343–363.
- [212] Vinayak V Dixit, Chester Wilmot, and Brian Wolshon. Modeling risk attitudes in evacuation departure choices. *Transportation research record*, 2312(1):159–163, 2012.
- [213] Jennifer A Horney, Pia DM MacDonald, Marieke Van Willigen, Philip R Berke, and Jay S Kaufman. Individual actual or perceived property flood risk: Did it predict evacuation from hurricane isabel in north carolina, 2003? *Risk Analysis: An International Journal*, 30(3):501–511, 2010.
- [214] Wanyun Shao, Maaz Gardezi, and Siyuan Xian. Examining the effects of objective hurricane risks and community resilience on risk perceptions of hurricanes at the county level in the us gulf coast: an innovative approach. *Annals of the American Association of Geographers*, 108(5):1389–1405, 2018.
- [215] Eiji Yamamura. The impact of natural disasters on income inequality: analysis using panel data during the period 1970 to 2004. *International Economic Journal*, 29(3):359–374, 2015.
- [216] Vassilis Tselios and Emma L Tompkins. What causes nations to recover from disasters? an inquiry into the role of wealth, income inequality, and social welfare provisioning. *International journal of disaster risk reduction*, 33:162–180, 2019.

- [217] Takahiro Yabe and Satish V Ukkusuri. Effects of income inequality on evacuation, reentry and segregation after disasters. *Transportation Research Part D: Transport and Environment*, page 102260, 2020.
- [218] Mitigation Framework Leadership Group. *Draft Interagency Concept for Community Resilience Indicators and National-Level Measures*. Jun 2016.
- [219] Stephen R Carpenter and William A Brock. Adaptive capacity and traps. *Ecology and society*, 13(2), 2008.
- [220] Crawford Stanley Holling and Lance H Gunderson. *Panarchy: understanding transformations in human and natural systems*. Washington, DC: Island Press, 2002.
- [221] Sarah Laborde, Alfonso Fernández, Sui Chian Phang, Ian M. Hamilton, Nathaniel Henry, Hahn Chul Jung, Aboukar Mahamat, Mouadjamou Ahmadou, Bruno K. Labara, Saïdou Kari, Michael Durand, Bryan Mark, Paul Scholte, Ningchuan Xiao, Roland Ziebe, and Mark Moritz. Social-ecological feedbacks lead to unsustainable lock-in in an inland fishery. 41:13–25.
- [222] Andrew P. Kythreotis and Gillian I. Bristow. The ‘resilience trap’: Exploring the practical utility of resilience for climate change adaptation in UK city-regions. 51(10):1530–1541.
- [223] Matthew E. Kahn. The Death Toll from Natural Disasters: The Role of Income, Geography, and Institutions. 87(2):271–284.
- [224] Heng Cai, Nina S. N. Lam, Yi Qiang, Lei Zou, Rachel M. Correll, and Volodymyr Mihunov. A synthesis of disaster resilience measurement methods and indices. 31:844–855.
- [225] Michel Masozera, Melissa Bailey, and Charles Kerchner. Distribution of impacts of natural disasters across income groups: A case study of New Orleans. 63(2-3):299–306.
- [226] Robin L Dillon-Merrill, Lei Ge, and Pedro Gete. Natural disasters and housing markets. the tenure choice channel. 2018. Working Paper.
- [227] Steven M Southwick, Brett T Litz, Dennis Charney, and Matthew J Friedman. *Resilience and mental health: Challenges across the lifespan*. Cambridge University Press, 2011.
- [228] Bob Bolin and Liza C. Kurtz. Race, class, ethnicity, and disaster vulnerability. In Havidán Rodríguez, William Donner, and Joseph E. Trainor, editors, *Handbook of Disaster Research*, pages 181–203. Springer International Publishing.
- [229] Kirk Chang. Community cohesion after a natural disaster: insights from a Carlisle flood. *Disasters*, 34(2):289–302, 2010.
- [230] Charles L. Redman. *Human Impact on Ancient Environments*. University of Arizona Press.

- [231] Thilo Lang. Urban Resilience and New Institutional Theory – A Happy Couple for Urban and Regional Studies? In Bernhard Müller, editor, *German Annual of Spatial Research and Policy 2010: Urban Regional Resilience: How Do Cities and Regions Deal with Change?*, German Annual of Spatial Research and Policy, pages 15–24. Springer.
- [232] Brian Walker, C. S. Holling, Stephen R. Carpenter, and Ann Kinzig. Resilience, Adaptability and Transformability in Social–ecological Systems. 9(2).
- [233] Paolo Bocchini, Dan M. Frangopol, Thomas Ummenhofer, and Tim Zinke. Resilience and Sustainability of Civil Infrastructure: Toward a Unified Approach. 20(2):04014004.
- [234] Ed Summers. aretha-tweets.
- [235] Dmitris Milajevs. brexit tweets collected from the 5th of May to the 24th August 2016, January 2017.
- [236] Muhammad Imran, Prasenjit Mitra, and Carlos Castillo. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may 2016. European Language Resources Association (ELRA).
- [237] Justin Littman. Charlottesville Tweet Ids, 2018.
- [238] Ed Summers. Eclipse tweet ids.
- [239] Mark Philips. Hurricane Florence twitter dataset, 2018. Accessed 12-01-2018.
- [240] Justin Littman. Hurricanes Harvey and Irma Tweet ids, 2017.
- [241] Justin Littman. Ireland 8th Tweet Ids, 2018.
- [242] University of Nevada. Web archive on the october 1, 2017 shooting in las vegas, nevada, 2017. Accessed 12-01-2018.
- [243] Firoj Alam, Ferda Ofli, and Muhammad Imran. A twitter tale of three hurricanes: Harvey, irma, and maria. *arXiv preprint arXiv:1805.05144*, 2018.
- [244] Haoyu Wang, Eduard Hovy, and Mark Dredze. The hurricane sandy twitter corpus. In *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, 2015.
- [245] Achim Zeileis. *ineq: Measuring Inequality, Concentration, and Poverty*, 2014. R package version 0.2-13.

APPENDICES

APPENDIX A
APPENDIX FOR CHAPTER 1

Table A.1.

Summary statistics for the distribution of efficiency for respective failure modes. Failure fraction represents the fraction of the network which was induced as failed in each iteration. Results presented in the body of the work represent a failure fraction of 0.6

Generation method	Failure fraction	Mean	Standard deviation	Median	Min	Max
Random	0.1	0.0178	0.0025	0.0174	0.0138	0.0245
BFS	0.1	0.0298	0.0046	0.0297	0.0215	0.0363
DFS	0.1	0.0308	0.0042	0.0320	0.0217	0.0363
Random	0.2	0.0124	0.0016	0.0121	0.0095	0.0191
BFS	0.2	0.0294	0.0050	0.0278	0.0213	0.0408
DFS	0.2	0.0312	0.0066	0.0293	0.0216	0.0410
Random	0.3	0.0099	0.0013	0.0097	0.0076	0.0134
BFS	0.3	0.0307	0.0056	0.0295	0.0236	0.0427
DFS	0.3	0.0324	0.0075	0.0286	0.0232	0.0443
Random	0.4	0.0084	0.0010	0.0083	0.0059	0.0119
BFS	0.4	0.0313	0.0043	0.0294	0.0257	0.0410
DFS	0.4	0.0310	0.0033	0.0304	0.0243	0.0390
Random	0.5	0.0076	0.0009	0.0075	0.0058	0.0099
BFS	0.5	0.0360	0.0055	0.0358	0.0255	0.0467
DFS	0.5	0.0340	0.0023	0.0343	0.0279	0.0384
Random	0.6	0.0070	0.0011	0.0070	0.0047	0.0100
BFS	0.6	0.0414	0.0071	0.0420	0.0240	0.0494
DFS	0.6	0.0393	0.0038	0.0401	0.0270	0.0463
Random	0.7	0.0068	0.0015	0.0069	0.0043	0.0119
BFS	0.7	0.0479	0.0127	0.0462	0.0264	0.0751
DFS	0.7	0.0506	0.0052	0.0508	0.0322	0.0659
Random	0.8	0.0071	0.0016	0.0069	0.0035	0.0126
BFS	0.8	0.0561	0.0174	0.0533	0.0294	0.0880
DFS	0.8	0.0699	0.0126	0.0724	0.0427	0.0867
Random	0.9	0.0097	0.0031	0.0097	0.0045	0.0184
BFS	0.9	0.0659	0.0208	0.0609	0.0392	0.1389
DFS	0.9	0.0971	0.0140	0.0961	0.0600	0.1372

Table A.2.

Summary statistics for the distribution of efficiency for respective failure modes scaled by the efficiency when the network is fully repaired. In this table, a value of 100 is the same performance metric seen at a fully repaired system.

Generation method	Failure fraction	Mean	Median	Min	Max
Random	0.1	52.29	51.12	40.54	71.97
BFS	0.1	87.54	87.25	63.16	106.64
DFS	0.1	90.48	94.01	63.75	106.64
Random	0.2	36.43	35.55	27.91	56.11
BFS	0.2	86.37	81.67	62.57	119.86
DFS	0.2	91.66	86.08	63.45	120.45
Random	0.3	29.08	28.50	22.33	39.37
BFS	0.3	90.19	86.66	69.33	125.44
DFS	0.3	95.18	84.02	68.16	130.14
Random	0.4	24.68	24.38	17.33	34.96
BFS	0.4	91.95	86.37	75.50	120.45
DFS	0.4	91.07	89.31	71.39	114.57
Random	0.5	22.33	22.03	17.04	29.08
BFS	0.5	105.76	105.17	74.91	137.19
DFS	0.5	99.88	100.76	81.96	112.81
Random	0.6	20.56	20.56	13.81	29.38
BFS	0.6	121.62	123.38	70.51	145.12
DFS	0.6	115.45	117.80	79.32	136.02
Random	0.7	19.98	20.27	12.63	34.96
BFS	0.7	140.72	135.72	77.56	220.62
DFS	0.7	148.65	149.24	94.59	193.60
Random	0.8	20.86	20.27	10.28	37.02
BFS	0.8	164.81	156.58	86.37	258.52
DFS	0.8	205.35	212.69	125.44	254.70
Random	0.9	28.50	28.50	13.22	54.05
BFS	0.9	193.60	178.91	115.16	408.05
DFS	0.9	285.25	282.31	176.26	403.06

Table A.3.

Summary statistics for the distribution of largest connected component (LCC) for respective failure modes.

Generation method	Failure fraction	Mean	Standard deviation	Median	Min	Max
Random	0.1	84.46	23.86	83.00	42.00	138.00
BFS	0.1	158.22	35.67	160.50	82.00	202.00
DFS	0.1	164.34	35.36	181.00	80.00	202.00
Random	0.2	41.89	12.49	40.00	20.00	87.00
BFS	0.2	122.21	32.35	125.00	82.00	179.00
DFS	0.2	125.23	43.74	111.50	66.00	180.00
Random	0.3	25.54	8.05	24.00	13.00	50.00
BFS	0.3	105.79	29.70	91.00	64.00	154.00
DFS	0.3	105.84	40.11	80.50	54.00	159.00
Random	0.4	16.90	4.33	16.00	9.00	29.00
BFS	0.4	85.47	19.73	82.00	44.00	129.00
DFS	0.4	73.91	13.19	70.00	44.00	124.00
Random	0.5	12.22	2.83	12.00	7.00	21.00
BFS	0.5	76.04	17.13	82.00	39.00	100.00
DFS	0.5	62.14	7.92	64.00	41.00	80.00
Random	0.6	9.05	2.32	9.00	5.00	15.00
BFS	0.6	62.94	17.71	66.50	28.00	83.00
DFS	0.6	53.75	9.54	53.00	28.00	76.00
Random	0.7	6.80	1.51	7.00	4.00	12.00
BFS	0.7	46.78	15.55	47.00	17.00	68.00
DFS	0.7	47.43	7.66	49.00	24.00	66.00
Random	0.8	5.03	1.01	5.00	3.00	8.00
BFS	0.8	28.57	9.98	28.00	10.00	46.00
DFS	0.8	35.42	9.02	38.00	16.00	46.00
Random	0.9	3.40	0.57	3.00	3.00	5.00
BFS	0.9	11.49	4.49	10.00	5.00	24.00
DFS	0.9	16.25	3.35	16.00	9.00	24.00

Table A.4.

Summary statistics for the distribution of largest connected component (LCC) for respective failure modes, scaled by the total LCC when the network is fully repaired. In this table, a value of 100 is the same performance metric seen at a fully repaired system.

Generation method	Failure fraction	Mean	Median	Min	Max
Random	0.1	37.87	37.22	18.83	61.88
BFS	0.1	70.95	71.97	36.77	90.58
DFS	0.1	73.70	81.17	35.87	90.58
Random	0.2	18.78	17.94	8.97	39.01
BFS	0.2	54.80	56.05	36.77	80.27
DFS	0.2	56.16	50.00	29.60	80.72
Random	0.3	11.45	10.76	5.83	22.42
BFS	0.3	47.44	40.81	28.70	69.06
DFS	0.3	47.46	36.10	24.22	71.30
Random	0.4	7.58	7.17	4.04	13.00
BFS	0.4	38.33	36.77	19.73	57.85
DFS	0.4	33.14	31.39	19.73	55.61
Random	0.5	5.48	5.38	3.14	9.42
BFS	0.5	34.10	36.77	17.49	44.84
DFS	0.5	27.87	28.70	18.39	35.87
Random	0.6	4.06	4.04	2.24	6.73
BFS	0.6	28.22	29.82	12.56	37.22
DFS	0.6	24.10	23.77	12.56	34.08
Random	0.7	3.05	3.14	1.79	5.38
BFS	0.7	20.98	21.08	7.62	30.49
DFS	0.7	21.27	21.97	10.76	29.60
Random	0.8	2.26	2.24	1.35	3.59
BFS	0.8	12.81	12.56	4.48	20.63
DFS	0.8	15.88	17.04	7.17	20.63
Random	0.9	1.52	1.35	1.35	2.24
BFS	0.9	5.15	4.48	2.24	10.76
DFS	0.9	7.29	7.17	4.04	10.76

APPENDIX B
APPENDIX FOR CHAPTER 2

Table B.1.
Efficiency of networks in different under different failure regimes

g_1	Distribution	g_2	Distribution	Min, g_1	Mean, g_1	Max, g_1	Min, g_2	Mean, g_2	Max, g_2
Random		Random		0.004	0.108	0.132	0	0.189	0.5
Random		BFS		0.005	0.108	0.132	0	0.188	0.5
Random		DFS		0.006	0.108	0.132	0	0.186	0.5
BFS		BFS		0.007	0.097	0.119	0	0.188	0.5
BFS		DFS		0.007	0.096	0.119	0	0.187	0.5
DFS		DFS		0.007	0.097	0.117	0	0.193	0.5

APPENDIX C
APPENDIX FOR CHAPTER 3

Table C.1.
 Tweet corpora summary. Description of events used along with the quantity of tweets and their acquisition methods and respective sources.

Event	Event Description	Event Dates	Total Tweet IDs	Resulting Tweets	Tweet acquisition method	Reference
Aretha Franklin's death	The death of singer Aretha Franklin	08/08/2018-2,832,128 08/18/2018	252,433	252,433	Keyword filtering	[234]
Brexit	The referendum to remove the UK from the European Union	05/05/2016-23,733,133 08/24/2016	3,884,599	3,884,599	Keyword filtering	[235]
California earthquake	Magnitude 6.0 earthquake striking south of Napa, CA	08/24/2014-254,529 08/30/2014	50,414	50,414	Keyword filtering	[236]
Charlottesville	White supremacist rally which resulted in significant counter-protesting and violence in Charlottesville, VA	08/14/2017-3,015,437 10/23/2017	207,098	207,098	Keyword filtering	[237]
Ebola outbreak	Ebola epidemic in Guinea, Liberia, Sierra Leone and other parts of West Africa	08/18/2014-5,085,767 01/19/2015	993,905	993,905	Keyword filtering	[236]
Eclipse	2017 Solar eclipse passing over much of the United States	08/17/2017-13,548,321 08/23/2017	1,211,729	1,211,729	Keyword filtering	[238]
Hurricane Florence	Major Atlantic Hurricane impacting the Eastern United States	09/05/2018-4,971,575 10/03/2018	488,106	488,106	Keyword filtering	[239]
Hurricane Harvey	Major Atlantic Hurricane impacting the Gulf Coast	08/25/2017-18,352,142 10/23/2017	1,062,127	1,062,127	Keyword filtering	[240]
Ireland 8th	Referendum to remove the Eight Amendment from the Irish Constitution, governing the legality of abortion	04/13/2018-2,279,396 06/04/2018	195,050	195,050	Keyword filtering	[241]
Hurricane Irma	Major Atlantic hurricane impacting the Caribbean and Florida Keys	09/01/2017-17,244,139 10/23/2017	976,294	976,294	Keyword filtering	[240]
Las Vegas shooting	Lone-gunner attack on a music festival in Las Vegas, NV	09/01/2017-14,108,104 10/23/2017	866,758	866,758	Keyword filtering	[242]
Hurricane Maria	Major Atlantic hurricane severely impacting Puerto Rico	09/20/2017-1,096,335 10/03/2017	87,160	87,160	Keyword filtering	[243]
Nepal	Magnitude 7.8 earthquake in the Gorkha District of Nepal	04/25/2015-4,223,983 05/19/2015	509,299	509,299	Keyword filtering	[236]
Hurricane Sandy	Major Atlantic storm impacting much of the Caribbean and East Coast of the US	10/22/2012-6,554,744 11/02/2012	3,252,011	3,252,011	Bounding box surrounding CT, DE, MA, MD, NJ, NY, NC, OH, PA, RI, SC, VA, WV.	[244]

Table C.2.

Tweet Acquisition. Keywords, keyword phrases, and hashtags used to create the tweet datasets.

Event	Keywords
Aretha Franklin's death	aretha_franklin, queen_of_soul
Brexit	brexit
California earthquake	napa_earthquake, sonoma_earthquake, bay_area_earthquake, california_earthquake, ca_earthquake, sfearthquake, san_francisco_earthquake, napaeearthquake, sfquake, napaquake, napa_quake, sonoma_quake, bay_area_quake, california_quake, ca_quake, san_francisco_quake
Charlottesville	charlottesville, standwithcharlottesville, defendCville, heatherheyer, unitycville
Ebola Outbreak	ebola, ebola_virus
Eclipse	solareclipse2017, solareclipse, eclipse2017, eclipseday, eclipse
Florence	florence, hurricaneflorence, florencehurricane, hurricane_florence", florencenc, hurricaneflorence2018, hurricaneflorence
Harvey	hurricane_harvey, hurricaneharvey, harvey, hurricane
Ireland 8th	8thref, hometovote, jointherebellion, trustwomen, repealthe8th, together4yes, together4yes, voteyes, time4choice, knowyourpealers, mybodymychoice, savethe8th, loveboth, loveboothvoteno, votenotoabortion, standupforlife, lifecanvass, protectthe8th, 8thamendment, whoneedsyouryes, men4yes, register4yes, roadtorepeal, repealfacts, healthcarenotairfare, repeal, trustwomen, itstime, whyinvotingyes, deaftogether4yes, doctors4yes, repeal4betterbirth, togetherformo, men4no, whoneedsyourno, rallyforlife, votenotoabortion, bemyyes, academics4yes, hometovoteno, hometocanvass, abortionreferendum, savita, repealshield, farmers4yes, lawyersforchoice, lawyers4yes, studentsforchoice, archivingthe8th, repealedthe8th, wemadehistory, nowformi, wetrustwomen,
Irma	irma, hurricane_irma, irmastrong
Las Vegas shooting	vegas
Maria	hurricane_maria, hurricanemaria, tropical_storm_maria, maria_storm
Nepal	basantapur, patan, anamnagar, bhaktapur, durbar_square, nuwakot, dhara_hara_tower, gorkha, lamjung, khudi, kathmandu, sankhu, sunsari, solu_district, okhaldhunga, nepal, nepal_earthquake, ktmeearthquake, indiawithnepal, nepalquake, nepalquakerelief, nepalearthquake, kathmanduquake, kathmanduquakerelief, kathmanduearthquake, quakenepal, earthquakekenepal, quakekathmandu, earthquakekathmandu, prayformnepal
Sandy	NA

Table C.3.

Categories of resilience and associated keywords. Keywords are manually coded based on conceptual definitions of resilience categories.

Ecological	Social	Economy	Institution	Infrastructure	Quality
ecological	social	economy	institution	power	community
ecology	love	nation	nation	nation	love
erosion	peace	market	hospital	emergency	life
wetlands	prayer	business	vote	flight	home
biology	family	bank	poll	airplane	hospice
coast	life	trade	country	safe	hospital
marsh	bless	stock	police	water	protest
dune	spirit	politic	mayor	power	health
fish	protest	money	president	relief	school
bird	rally	dollar	governor	city	doctor
river	monument	credit	senator	coal	nurse
climate_change	god	job	flag	evacuate	medic
rainfall	church	jobs	doctor	airport	safe
nature	donate	work	nurse	cell	found
floodwater	aid	money	govern	water	aid
beach	network	wealth	school	outage	humanatarian
sun	church	property	medic	road	life
stream	faith	pay	fema	bus	health
flood	friend	employer	shelter	car	depression
storm	friends	employee	school	infrastructure	
wind	family		potus		
rain	pray		red_cross		
water	neighborhood		church		
weather	town		evacuate		
beach			homeland		
tropic			responders		
climate			fema		
			ems		
			police		
			fire		
			government		
			alderman		
			county		
			officials		

APPENDIX D

APPENDIX FOR CHAPTER 4

Resilience Calculation Figures

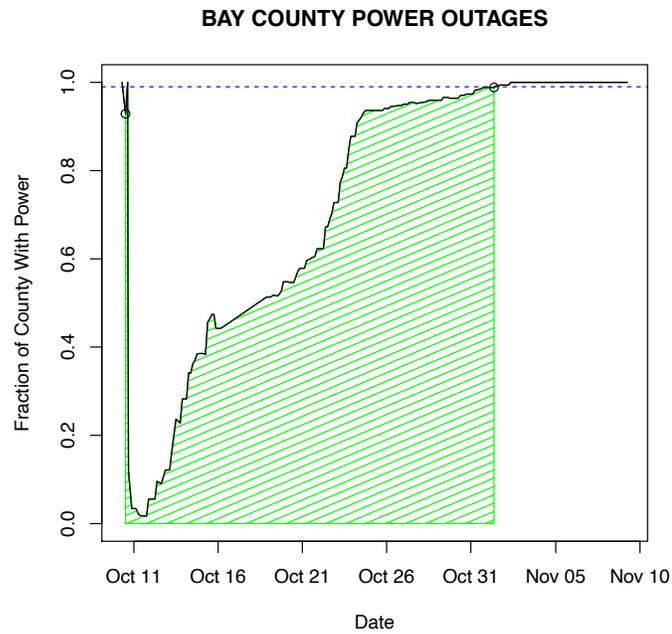


Figure D.1. Power Outages During Hurricane Michael for Bay County, FL

Examples of the resilience calculation performed for each county. For each county, the value $Q(t)$ is the fraction of the county with access to power at time t ; represented in these figures as the solid black lines. The time of initial disruption, t_0 is the first time in which $Q(t)$ drops below 0.99 (a pre-defined threshold; represented in Figs. D.1 and D.2 as the left-most black circle). Similarly, t_f is the first post-disruption time in which $Q(t)$ exceeds 0.99 and is shown in the above figures as the right-most

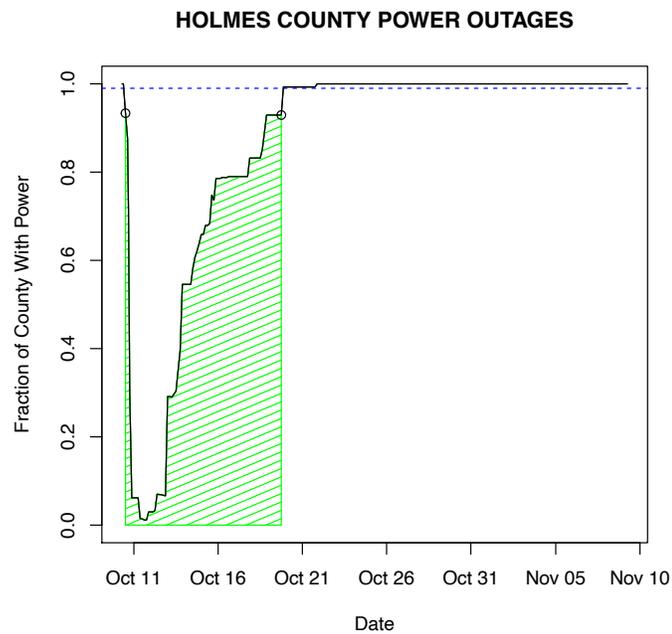


Figure D.2. Power Outages During Hurricane Michael for Holmes County, FL

black circles. The numerator of the Eq. 5.1 for calculating R_{county} , $\int_{t_0}^{t_f} Q(t)$, is shaded green in Figs.D.1 and D.2. $|t_f - t_0|$, the denominator of R_{county} , is the duration of time between the initial disruption and repair time.

Model selection plots and comparisons

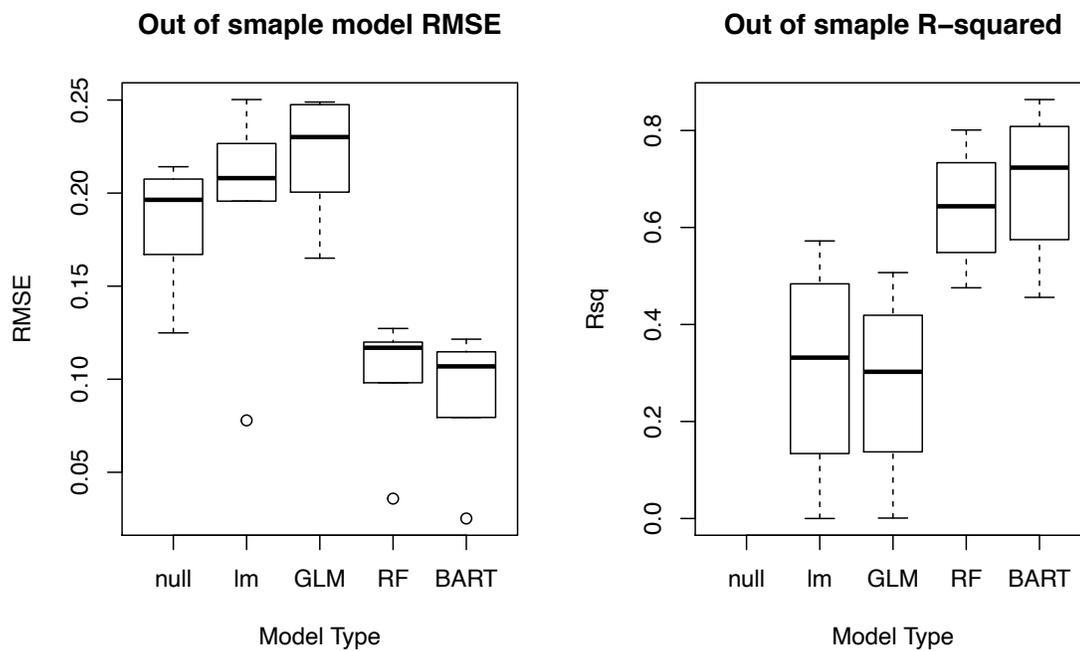


Figure D.3. Out of sample RMSE and R^2 for the prediction of engineering resilience as a function of the selected community risk factors. Note the null model's out of sample R^2 is 0 because the standard deviation of predictions is 0.

Community Risk Factors

Table D.1.: Original Input Variables

Category	Variable Name	Description	Year (s)	Source
Demographics	Population	County level population	2018	American Community Survey, Table: B01003 [185]
	Racial Inequity	Kolm's Inequality Measure of racial demographics of census tracts in each county	2018	American Community Survey, Table: B02001 [185]
	Citizenship	County level fraction of population who are citizens	2018	American Community Survey, Table: B05001 [185]
Housing	1-Year Housing Tenure	County level fraction of population living in the same home for more than 1 year	2018	American Community Survey, Table: B07001 [185]
	Within County Relocation	County level fraction of population who moved within the state in the past year	2018	American Community Survey, Table: B07001 [185]

Within State Relocation	County level fraction of population who moved from outside the county but within the state in the past year	2018	American Community Survey, Table: B07001 [185]
Out of State Relocation	County level fraction of population who moved from outside the state but within the US in the past year	2018	American Community Survey, Table: B07001 [185]
Abroad Relocation	County level fraction of population who moved from outside the US in the past year	2018	American Community Survey, Table: B07001 [185]
Household Size, Renters	County level average household size of renter-occupied housing units	2018	American Community Survey, Table: B25010 [185]
Household Size, Owners	County level average household size of owner-occupied housing units	2018	American Community Survey, Table: B25010 [185]

	Rents Dwelling	County level fraction of population in renter-occupied housing	2018	American Community Survey, Table: B25008 [185]
Mobility	Commuting Alone	County level fraction of population who primarily commutes in a car, truck or van alone	2018	American Community Survey, Table: B08101 [185]
	Carpool Commute	County level fraction of population who primarily commutes by carpooling in a car, truck, or van	2018	American Community Survey, Table: B08101 [185]
	Public Transportation Commuting	County level fraction of population who primarily commutes via public transportation (excluding taxis)	2018	American Community Survey, Table: B08101 [185]
	Walking Commute	County level fraction of population who primarily commutes by walking	2018	American Community Survey, Table: B08101 [185]

	Bike, cab, other commuting	County level fraction of population who primarily commutes by taxicab, motorcycle, bicycle, or other means	2018	American Community Survey, Table: B08101 [185]
	Working From Home	County level fraction of population who primarily works from home	2018	American Community Survey, Table: B08101 [185]
Educational Attainment	High School Degree	County level fraction of population over 25 with a regular high school diploma	2018	American Community Survey, Table: B15003 [185]
	GED	County level fraction of population over 25 with a GED or alternative credential	2018	American Community Survey, Table: B15003 [185]
	Associates Degree	County level fraction of population over 25 with an Associate's degree	2018	American Community Survey, Table: B15003 [185]

	Bachelor's Degree	County level fraction of population over 25 with a Bachelor's degree	2018	American Community Survey, Table: B15003 [185]
Language	Speaks English	County level fraction of population 5 years and older speaking only English at home	2018	American Community Survey, Table: B16007 [185]
	Speaks Spanish	County level fraction of population 5 years and older speaking Spanish at home	2018	American Community Survey, Table: B16007 [185]
	Speaks Indo-European	County level fraction of population 5 years and older speaking other Indo-European Languages at home	2018	American Community Survey, Table: B16007 [185]
	Speaks API	County level fraction of population 5 years and older speaking Asian and Pacific Island Languages at home	2018	American Community Survey, Table: B16007 [185]

Income	Income Inequality	County level Gini Index of income inequality	2018	American Community Survey, Table: B19083 [185]
	Aggregate Household Income	County level aggregate household income in the past 12 months in 2018 inflation-adjusted dollars	2018	American Community Survey, Table: B19025 [185]
	Income Through Earning	County level fraction of households with income from wage or salary income	2018	American Community Survey, Table: B19051 [185]
	Income Through Interest	County level fraction of households with income through interest, dividends or net rental income	2018	American Community Survey, Table: B19054 [185]
	Income Through SSI	County level fraction of households with Social Security income	2018	American Community Survey, Table: B19054 [185]
Communication	Internet Access	County level fraction of households with internet access	2018	American Community Survey, Table: B28002 [185]

	Cellphone Only	County level fraction of households with only a cellular data plan with no other type of internet subscription	2018	American Community Survey, Table: B28002 [185]
	No Internet Access	County level fraction of households without internet access	2018	American Community Survey, Table: B28002 [185]
Climate Change Opinions	Climate Opinions	County level PCA decomposition of positive responses to climate change-related polls	2019	Yale Program on Climate Change Communication [186]

In addition, the census-tract level data was collected for each ACS variable and Kolm's inequality measure [245] was computed for each county and variable. 1-year Data from the American Community Survey are collected from January 1st 2018 to December 31st 2018 for populations of 20,000 or more.

Table D.2.: County and resilience values for each county in Florida during Hurricane Michael (2018)

County	Resilience value
ALACHUA	0.9259779
BAKER	1.0000000
BAY	0.6339356
BRADFORD	0.8244461
BREVARD	0.9860843
BROWARD	1.0000000
CALHOUN	0.4019630
CHARLOTTE	1.0000000
CITRUS	1.0000000
CLAY	1.0000000
COLLIER	1.0000000
COLUMBIA	0.9881542
DESOTO	1.0000000
DIXIE	1.0000000
DUVAL	1.0000000
ESCAMBIA	0.9681062
FLAGLER	1.0000000
FRANKLIN	0.3629637
GADSDEN	0.5167954
GILCHRIST	0.4537804
GLADES	1.0000000
GULF	0.5326955

HAMILTON	0.7993632
HARDEE	1.0000000
HENDRY	1.0000000
HERNANDO	1.0000000
HIGHLANDS	1.0000000
HILLSBOROUGH	1.0000000
HOLMES	0.5542315
INDIAN	0.8273789
JACKSON	0.4582192
JEFFERSON	0.5209264
LAFAYETTE	0.8433915
LAKE	1.0000000
LEE	1.0000000
LEON	0.6739230
LEVY	1.0000000
LIBERTY	0.4777898
MADISON	0.8937640
MANATEE	1.0000000
MARION	1.0000000
MARTIN	1.0000000
MIAMI-DADE	1.0000000
MONROE	1.0000000
NASSAU	1.0000000
OKALOOSA	0.9813354
OKEECHOBEE	1.0000000
ORANGE	1.0000000

OSCEOLA	1.0000000
PALM	1.0000000
PASCO	1.0000000
PINELLAS	1.0000000
POLK	1.0000000
PUTNAM	1.0000000
SANTA	1.0000000
SARASOTA	1.0000000
SEMINOLE	1.0000000
ST JOHNS	1.0000000
ST LUCIE	1.0000000
SUMTER	1.0000000
SUWANNEE	0.9570933
TAYLOR	0.8497418
UNION	0.9877392
VOLUSIA	1.0000000
WAKULLA	0.6641285
WALTON	0.8435221
WASHINGTON	0.7034305

Example Temporal Trajectory

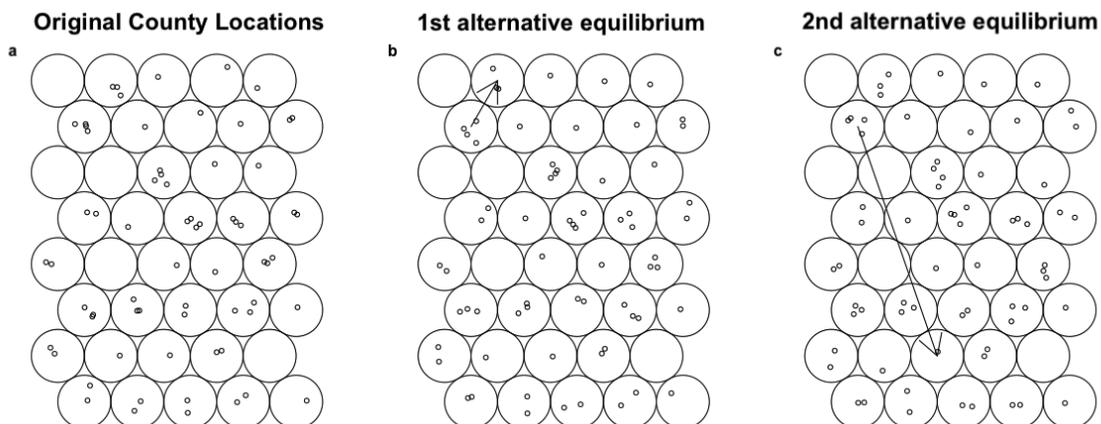


Figure D.4. Example County Movement. (a) shows the location of nodes with no perturbation, (b) shows the temporal trajectory of Bay County associated with the shift to the lowest alternative- equilibrium seen in Figure 5.2 c. (c) is the temporal trajectory for the change to the highest alternative-equilibrium seen in Figure 5.2c. The length of the arrows in (b) and (c) are the magnitudes of transformation for each perturbation.

VITA

BEN RACHUNOK

brachuno@purdue.edu

910.398.7206

www.brachunok.com

EDUCATION

- Purdue University**, West Lafayette, IN, USA 2020
Ph.D. in Industrial Engineering
- North Carolina State University**, Raleigh, NC, USA 2016
Bachelor of Science in Industrial & Systems Engineering

RESEARCH EXPERIENCE

- Stanford University**, Stanford, CA, USA 2020-
Postdoctoral Researcher (Supervisor: Dr. Sarah Fletcher)
Civil & Environmental Engineering
- Purdue University**, West Lafayette, IN, USA 2016-2020
Research Assistant (Advisor: Dr. Roshanak Nateghi)
Dissertation Title: *Analytical methods for Computing the Resilience, Recovery, And Transformation of Communities in the Age of Big Data*
- Sandia National Labs**, Albuquerque, NM, USA 2017-
R&D Intern (Supervisor: Dr. Andrea Staid)
Center for Computing Research, Discrete Math & Optimization

RESEARCH INTERESTS

Methodological

Statistical modeling, machine learning, risk and resilience assessment, sociotechnical systems analysis, decision-making under uncertainty

Application Areas

Climate change adaptation, civil systems engineering, natural hazard impacts, climate and environmental justice, maladaptive decision-making

JOURNAL PUBLICATIONS

Published

- Alemazkoor N, **Rachunok B**, Staid A, Chavas D, Tootkaboni M, 2020
Louhghalam A, Nateghi R. Hurricane-induced Power Outage Risk is Driven by Changes in Storm Frequency. *Scientific Reports*.
[DOI: 10.1038/s41598-020-72207-z]
- Kumar R*, **Rachunok B***, Silva D, Nateghi R*. Asymmetric 2020
Temperature Response of Electricity Demand Points to Severe Underestimation of Load Projections in California. *Scientific Reports*.
[DOI: 10.1038/s41598-020-67695-y] *(Equal contributions)
- Rachunok B**, Staid A, Watson D, Woodruff D. Evaluation of 2020
Parametric Wind Power Scenarios. *Applied Energy*.
[DOI: 10.1016/j.apenergy.2020.114986]
- Paulvannan Kanmani A, Obringer R, **Rachunok B**, Nateghi R. 2020
Assessing Global Environmental Sustainability Via an Unsupervised Clustering Framework. *Sustainability*.
[DOI: 10.3390/su12020563]

- Rachunok B**, Nateghi R. The Sensitivity of Electric Power Infrastructure Resilience to the Spatial Distribution of Disaster Impacts. *Reliability Engineering and System Safety*. [DOI: 10.1016/j.ress.2019.106658] 2019
- Rachunok B**, Bennett J, Nateghi R. Twitter and Disasters: A Social Resilience Fingerprint. *IEEE Access*. [DOI: 10.1109/ACCESS.2019.291479] 2019

Under Review

- Obringer R, **Rachunok B**, Silva D, Arbabzadeh M, Nateghi R, Madani K. The Overlooked Environmental Footprint of Internet Use. *Resources, Conservation and Recycling*.
- Rachunok B**, Nateghi R. Quantifying Community Transformation to Break Resilience Traps. *Nature Sustainability*.
- Rachunok B**, Bennett J, Flage R, Nateghi R. A Path Forward for Leveraging Social Media to Improve the Study of Community Resilience. *International Journal of Disaster Risk Reduction*.
- Choi M, **Rachunok B**, Nateghi R. Convolutional Neural Networks Predict Global Solar Irradiance. *Environmental Research Letters*.
- Bennett J, **Rachunok B**, Flage R, Nateghi R. Decoding Regional Climate Attitudes by Integrating Social Media and Survey Data. *PLoS ONE*.

CONFERENCE PUBLICATIONS

- Rachunok B**, Nateghi R. Interdependent Infrastructure System Risk and Resilience to Natural Hazards. *Proceedings of the 2019 IISE Annual Conference*. H.E. Romeijn, A. Schaefer, and R. Thomas (Eds.). [arXiv: 1904.05763] 2019
- Rachunok B**, Staid A, Watson JP, Woodruff D, Yang D. Stochastic Unit Commitment Performance Considering Monte Carlo Wind Power Scenarios. *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. [DOI: 10.1109/PMAPS.2018.8440563] 2018

INVITED PRESENTATIONS

- INFORMS Annual Meeting**. [Virtual]. *Quantifying Community Transformation*. 2020
- International Conference on Computational Social Science (IC²S²)**. Cambridge, MA [Virtual]. *Be Careful Where You Get Your Tweets: Bias in Analysis from Tweet Acquisition Source*. 2020
- Natural Hazards Center, Researcher's Meeting**. Boulder, CO [Virtual]. *Natural Hazard Twitter Analyses Depend Heavily On Tweet Sources*. 2020
- Oklahoma University INFORMS Student Chapter Summer Research Series**. Virtual. *Social Resilience Fingerprinting*. 2020
- Society for Risk Analysis Annual Meeting**. Crystal City, VA. *Social Resilience Fingerprinting*. 2019
- INFORMS Annual Meeting**. Seattle, WA. *Social Resilience Fingerprinting*. 2019
- Institute for Industrial and Systems Engineers Annual Conference**. Orlando, FL. *Interdependent Infrastructure System Risk & Resilience to Natural Hazards*. 2019
- Society for Risk Analysis Annual Meeting**. New Orleans, LA. *Electric Power Infrastructure Resilience*. 2018

INFORMS Annual Meeting. Phoenix, AZ. <i>The Sensitivity of Electric Power Infrastructure Resilience to the Spatial Distribution of Disaster Impacts.</i>	2018
IEEE International Conference on Probabilistic Methods Applied to Power Systems. Boise, ID. <i>Stochastic Unit Commitment Performance Considering Monte Carlo Wind Power Scenarios.</i>	2018
Society for Risk Analysis Annual Meeting. Crystal City, VA. <i>Transportation Network Recovery Analysis, an Equilibrated Perspective.</i>	2017
INFORMS Annual Meeting. Houston, TX. <i>Modeling Uncertainty, The Cost of Wrong Assumptions.</i>	2017
Purdue Environmental and Ecological Engineering Seminar Series. West Lafayette, IN. <i>Risk Analysis for Power Systems Planning.</i>	2017
Institute for Industrial and Systems Engineers Annual Conference. Anaheim, CA. <i>UAVs Provide Lifesaving Medical Care.</i>	2016
North Carolina State University Research Symposium. Raleigh, NC. <i>Simulating UAVs as a Healthcare Delivery Method.</i>	2016

MEDIA

“Twitter ‘fingerprint’ helps decode how individuals respond to crises” <i>Prevention Web: UN office of Disaster Risk Reduction Purdue University News ACM Technews Purdue Today Newsletter</i>	2019
“Parsing Tweets to Strengthen Community Disaster Resilience” <i>Purdue Engineering Review</i>	2019

TEACHING EXPERIENCE

Lecturer	
IE 590: Predictive Risk Analytics, Purdue University	2018
IE 330: Probability and Statistics II (Guest), Purdue University	2018-2019
Teaching Assistant	
IE 431: Industrial Engineering Design, Purdue University	2017-2018
IE 343: Engineering Economics, Purdue University	2016
Undergraduate Lecturer	
E115: Introduction to Computing Environments, NC State University	2015-2016
ISE110: Microsoft Excel & Visual Basic, NC State University	2015

WORK EXPERIENCE

Purdue University, Research Assistant, West Lafayette, IN Research related to NSF, and Purdue internal grants.	2017-
Sandia National Labs, R&D Intern, Albuquerque, NM Develop stochastic programming scenario generation methods Evaluate the effects of ramp events in stochastic unit commitment problem Maintain and test open source power grid evaluation software	2017-
Institute for Transportation Research and Education, Research Intern, Raleigh, NC Develop Java and PHP/HTML applications which implement research methods developed at the institute	2016
Robert Bosch GMBH, Process Analyst Co-Op, Charleston, SC Implemented SAP enterprise software in a large foreign trade zone Manage packaging logistics for a foreign trade zone Communicate between customers and suppliers to design safe, sustainable and economical packaging programs	2014

HONORS AND AWARDS

Outstanding Research Award, Honorable Mention, <i>Purdue University</i>	2020
Outstanding Service Award, <i>Purdue University</i>	2020
Student Spotlight, <i>SRA Engineering & Infrastructure Specialty Group</i>	2018
Travel Award, <i>Purdue University Graduate Student Government</i>	2018
Travel Award, <i>Society of Risk Analysis</i>	2018
Lee Chaden Fellowship, <i>Purdue University</i>	2018
Travel Grant, <i>Purdue Climate Change Research Center</i>	2018
Estus and Vashti Magoon Award, <i>Purdue University</i>	2018
Operations Research Track, Undergraduate Student Paper Competition, 1st place, <i>IISE Annual Conference, Anaheim, CA</i>	2016

SERVICE AND OUTREACH

Reviewer

<i>Journal of Risk Analysis</i>	2018-
<i>IEEE Access</i>	2018-
<i>Journal of Management in Engineering</i>	2019-
<i>Journal of Infrastructure Systems</i>	2019-
<i>Utility Policy</i>	2019-

University Service

<i>Graduate Student Government, Purdue University</i> Senate Chair, Executive Board Member	2017-
---	-------

OTHER CERTIFICATIONS AND TRAINING

<i>Open Science Grid Consortium</i> Open Science Grid User School	2018
--	------

AFFILIATIONS

<i>Society for Risk Analysis (SRA)</i> Engineering and Infrastructure Group Student & Young Professionals Committee, Founding Member	2017-
<i>Purdue Climate Change Research Center</i> Graduate Affiliate	2017-
<i>INFORMS</i> Decision Analysis Public Sector OR	2017- 2018-
<i>Institute of Industrial and Systems Engineers (IISE)</i> Member	2018-

SKILLS

Programming

Fluent: R, Java, Python, VBA, LaTeX, bash

Familiar: PHP, HTML, SQL, SAS

Tools/Applications

Adobe CS, Mac/Windows/UNIX systems, ArcGIS, MS Office

Updated 12/07/2020