Data-driven Planning For Food and Forest Supply Chains under Disruptions

By

Marie Pelagie Elimbi Moudio

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Professor Anil Aswani
Professor Phil Kaminsky
Professor Marta Gonzalez

Summer 2023

Data-driven Planning For Food and Forest Supply Chains under Disruptions

Abstract

Data-driven Planning For Food and Forest Supply Chains under Disruptions

By

Marie Pelagie Elimbi Moudio

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

Global disruptions such as the Covid-19 pandemic emphasize the fragile nature of connected supply chains and the impact of these interruptions on our daily lives. The alarming growth in the type and number of disruptions complicates planning efforts and will lead to increased losses in the absence of efficient systems that reduce complexity for policy makers. We investigate food and forestry supply chains, key systems for human survival. In the context of food and agricultural supply chains, disruptions have not only caused bottlenecks in different stages of food supply chains but also exposed the need for holistic solutions to food supply and allocation. To tackle this challenge, we model a two-stage stochastic resource allocation problem with non-linear connectivity costs to capture trade dynamics between countries. We compare model recommendations to historical trade flow data including coffee trade between countries, unveiling the value of centralized planning under potential disruption scenarios against the current practices. In another chapter, we design and incrementally update a data-driven network risk measure that focuses on including the downstream impact of nodes. starting with a hand crafted intuitive risk measure then the Downstream Protection Value (DPV) and finally proposing the Downstream Supply Risk Measure (DSRM). The consistency in results between DPV and DSRM measures is investigated and their limitations are discussed.

Additionally, the existence of a large variety of food products that are exchanged within and across countries renders the creation of separate policies for each product impractical. To aid this problem, we develop a framework to systematically group different food product supply chains by identifying their structure using world bank trade data between countries for 53 different products over a period of 25 years. Decision makers can thus create policies for food networks at the group level instead of individual policies for the large variety of food and agricultural products. In the context of Forestry systems, we extend an existing fuel treatment allocation model to include socio-environmental costs to the objective function. We investigate how to develop efficient action to mitigate future fires (fuel treatment plans)

so as to mitigate risk in Forestry supply chains. In the long run, developing an understanding of the impacts of risk and how they can be mitigated will be essential in resolving their lasting and usually devastating impacts to the world ecosystem.

To my friends and family, to God that directs my life , to my support circle: every is a blessing day because of you

Contents

List of Figures

List of Tables

Acknowledgments

First I would like to express my deepest gratitude to my advisor, Professor Zuo-Jun Max Shen. From the very beginning he has been my strongest supporter and has guided my projects and academic journey. He has advocated for me during all the various stages of my journey. I would like to thank my dissertation committee members Professor Marta C. Gonzalez, Professor Anil Aswani, and Professor Phil Kaminsky for their great support starting from my qualifying exam and being very flexible and generous with their time and suggestions. In addition, many thanks to all the Professors of the IEOR department for the great learning experience I had through the program and to the IEOR staff, for their daily support and continuous help. My deepest thanks to all my co-authors and collaborators: Cristobal Pais, Marta C. Gonzalez, Arman Shehabi, Prakash Rao . Working with such amazing researchers, and most importantly, great people has been an incredibly gratifying experience.

To my beloved writing group: Cara H. and Emily G., I would never have made it through the writing stage without your keen insights and unwavering support. Finally, to my family and support circle. I want to especially thank my sisters Agnes Elimbi Moudio and Claudia Ndenge for literally holding my hand though this journey from day 1. I am also very grateful to my parents (Samuel and Anne Elimbi Moudio) and siblings (Ida and Francis) for their support even when we were separated by thousand of miles and even an ocean.

Chapter 1

Challenges Facing Food and Forest Supply Chains

Global disruptions such as the Covid-19 pandemic have emphasized the fragile nature of connected supply chains and the impact of these interruptions on our daily lives. While the pandemic's effects are devastating, supply chain vulnerability to disruptions have been on a rise for a long time. According to the National Oceanic and Atmospheric Administration, over 332 natural disasters since 1980 resulting in approximately $2.275 trillion in overall damages to the United States [84]. In addition, the Centre for Research on the Epidemiology of Disasters has reported less than 200 global disasters per year in the 1980s and over 300 in the 2010s. Natural disasters alone in the United states have quadrupled in the last 40 years, moving from 2.7/yr to 10.5/yr [100]. These disruptions have wide range negative impacts including financial, social, and environmental impacts, among others. The Federal Emergency Management Agency (FEMA) report that roughly 40% of businesses impacted by disasters do not reopen and 90% fail within two years of the event [67, 2]. The alarming growth in the type and number of disruptions complicates planning efforts and will lead to increased losses in the absence of efficient systems that reduce complexity for policy makers.

These problems are only worsening as the world's supply chains are increasingly complex, extended and fragile to all kinds of interruptions such as weather events (e.g., hurricanes, flooding), climate change (fires, droughts), interruptions to shipping, cyberattacks, and industrial accidents [114, 77]. In this work, we focus on food and forest supply chains because we believe these are critical to long term survival and sustainability of our global ecosystem. Humanity constantly faces challenges around feeding its ever growing population despite the earth's limited resources and with increasingly complex Food supply chains that are more susceptible to disruption. [48] The Covid-19 pandemic for instance has highlighted the vulnerability of food supply chains to disruptions and exposed an increasing number of communities globally to food insecurity. [125] These disruptions have not only caused bottlenecks in different stages of food supply chains but also exposed the need for a holistic solutions to food supply and allocation. [23] Planning around food and agricultural supply chains is particularly challenging because of a number of factors including (1) the large product space

on the country scale. (2) Different storage and transport requirements for food and agricultural products. The existence of a large variety of food products that are exchanged within and across countries renders the creation of separate policies for each product impractical.

Additionally, the origin of most food and agricultural products is very constricted by factors such as soil, climate, topology etc. As such, many food products can only be grown in specific regions and climates. Unexpected shocks to these production heavy regions could entirely crippled the Agricultural supply networks and reduce food access globally. [123] Certain economies, especially regions with low agricultural yield, rely on imports for food and are susceptible to food insecurity and negative disruptions to the global food network. On the African continent, roughly half of the population faces food insecurity presently with more than 250 million people considered to be severely food insecure. [80, 93] According to reports from the world bank, a rising number of households are experiencing food insecurity due to the supply chain disruptions and reduced salaries resulting from the COVID-19 pandemic. [125] This is just one of the many instances where disruptions to food supply chains have directly impacted food insecurity. Understanding the global movement of different agricultural products and their vulnerability to disruptions is crucial in ensuring global accessibility to these food products.

When looking at Forest resources supply chains, the implementation of strategic and efficient fuel treatment plans can modify fire behavior and greatly help with fire suppression efforts [86, 57, 43, 42]. Fuel treatments include actions and procedures such as cutting and clearing wood, prescribed burns, commercial harvesting, and thinning, that can promote fire hazard reduction [5]. In [105], the authors discuss strategies to define treatments using burning probability maps, the area burned, or the flame length. Other research in this area focuses on finding the optimal spatial allocation of prescribed burning activities [6, 79], and designing fire breaks to control fire spreading [104]. Decision-makers that carry out these fuel treatments face questions about how to make such decisions. However, the problem of optimal fuel allocation is challenging due to various sources of uncertainties. In [34], the authors discuss challenges surrounding fuel treatment methods, timing, and the high uncertainty levels in climate and ignition areas over multiple time periods. As a result, deterministic fire simulators such as FARSITE, Prometheus, and Wildfire Analyst [42, 115, 99], which can reproduce fires with and without treatment activities are popular in practice. The main critique to using fuel treatment methods is that due to the difficulty in predicting wildfire occurrence and propagation, the allocation of these treatments generally do not match areas in which future fires occur, leading to wasted investments coupled with economic, human, and environmental losses [11].

In order to address the mismatch between fuel treatment allocation and fire occurrence, the authors in [90, 24] develop a framework that integrates fire spread, optimization, and simulation models. The study highlights an adaptable metric known as the Downstream Protection Value (DPV), that ranks the impact of treating a unit of the landscape, by modeling a forest as a network and the fire propagation as a tree graph. The framework requires weather and topography inputs of the forest to run and can be modified by users to incorporate region-specific forest data so as to provide more effective and targeted treatments.

Results from the first version of the model value the equivalent volumes of subsets of the forest equally. However, such an assumption is limited in real settings. Different parcels of a landscape may have different values based on multiple factors such as the existence of animal migration corridors, the amount of biodiversity hosted in that region, the presence of human settlements and infrastructure, or the amount of carbon sequestered.

In the long term, understanding the impacts of risk and how they can be mitigated will be essential in tackling their lasting and sometimes devastating impacts to the world ecosystem. This dissertation investigates quantifying and mitigating risk on food and forestry supply chains by answering the following questions:

1. How can we effectively distribute food resources at the global scale (at a country level) under supply disruption?

2. Can we develop a data-driven measure that captures downstream risk propagation effects global food networks?

3. Can we facilitate the development of scalable food supply chain policies and obtain new insights using their structural representations?

4. Can we quantify the socio-environmental costs of supply chain systems facing risk?

The rest of the dissertation is organised to answer these questions. In chapter 2, we propose a stochastic global food allocation model for planning under supply side disruptions. First we present a single stage allocation model, which aims to minimize expected transportation, lost demand and connections costs. The we follow up with a two-stage extension model that includes inventory and depreciation, factors that are prolific within agricultural and food supply chains.

In the previous chapter, we used a simplistic definition of disruption by only considering a disruption as the loss in supply of a single node. our second research question attempt relax this assumption by attempting to define a risk measure that captures the propagation of risk downstream of the supply chain, We present different and improved iterations of a data-driven risk measure for food supply chains in chapter 3. We propose three risk metrics starting with a hand crafted intuitive risk measure then the Downstream Protection Value (DPV) and finally proposing the Downstream Supply Risk Measure (DSRM) focusing in including downstream impact of nodes. Developing data-driven food regulation and policy at the global scale, while essential to ensure food access and limiting waste is still very challenging. In chapter 4, we propose a framework to support the development of scalable policies by aggregating the wide number and variety of food and agricultural supply chains based on the supply chain structure and characteristics.

More so, we find in the literature that risk when quantifies usually considers only economic costs, as such we were interested in quantifying the socio-environmental costs of supply chains facing disruptions. In Chapter 5, expands the integrated framework proposed by [90] to aid decision making under wildfire uncertainty by evaluating the sensitivity of the

objective function to key environmental and economic factors. We compare unweighted treatment plans against versions including environmental and demographic factors such as carbon sequestration, canopy height and density, population density, and accessibility of the area, as well as expected future fire behavior and discuss impact on proposed treatment policies. Finally chapter 6 covers our conclusions and future work .

Chapter 2

Global Food Allocation Model

2.1 Introduction

In this chapter, we present a data-driven allocation model for planning global food allocation under supply side disruptions. As noted in previous chapters, disruptions to supply chains have been on the rise in recent years and can have devastating effects on strategic supply chains such as those involving food and agricultural products. Global food networks have been growing in complexity with food dependencies between countries rendering the overall network extremely vulnerable. Modelling and understanding the impact of different disruptions at the country level on agricultural supply chains is thus essential. We propose a two-stage stochastic resource allocation problem with non-linear connectivity costs to capture trade dynamics between countries. We compare model recommendations to historical trade flow data including coffee trade between countries, unveiling the value of centralized planning under potential disruption scenarios against the current practices.

Supply chains' vulnerability to disruption has increased while demand has become increasingly volatile in most sectors [31, 32]. These changes are especially relevant for agricultural supply chains (ASCs) that already face high levels of stochasticity due to some their inherent characteristics such as long lead-times, seasonality, perishability, and even sanitary emergencies (e.g., diseases), among several others [17]. Many countries, especially regions with low agricultural yield, rely on imports for food and are susceptible to food insecurity and negative disruptions to the global food network. As such, unanticipated shocks to supply-heavy regions could entirely paralyze the ASC networks and diminish food access globally [123]. The COVID-19 pandemic, a disruption that impacted food systems and its surrounding infrastructure on a global scale is a recent example of such disruptions. A number of studies suggest that the impacts of this pandemic, highlight the importance of creating centralised policies that allow world supply chains to be flexible enough to respond to network interruptions, [3, 73].

The goal of this chapter is to investigate the effects of supply side disruptions at the country level to global food access and distribution and how to create systems to mitigate

these disruptions. In this work, we extend previous work by [81] detailed in Appendix B2, that developed a single stage stochastic global food allocation model to a two-stage version can be modeled by including inventory management of the commodities between periods. The two-stage stochastic resource allocation problem has non-linear connectivity costs in the objective accounting for the trade connections between countries. We use historical coffee trade data from a World bank database to illustrate the use of the proposed framework. The model provides recommendations to help decision-makers efficiently allocate their supply of agricultural products and inventory levels to promote food networks that are more resilient to disruption.

Agricultural Supply Chains and Risk

Robustness and resilience are prolific metrics in the supply chain community, especially among researchers analyzing risk management strategies. Resilience is described as the Supply Chain Network's (SCN's) ability to rapidly and effectively recover from disruption. Robustness can be defined as the SCN's ability to withstand disturbances, maintain its original structure, and remain functional in uncertainties[15]. Several researchers have explored these concepts in the context of industrial and manufacturing supply chains especially using network science.[95]. A number of metrics and methods have been proposed to increase the robustness of logistics networks with many researchers adopting a topological view when characterizing the robustness of supply chain networks [96, 14, 110, 66, 4].

Previously, work has been done to create a methodology that supports decision making by assessing the robustness of supply chain networks with different topologies when exposed to disruptive events [128]. one of this method's main limitations is its reliance on generic network properties that exclude various properties specific to agricultural supply chains. In [70], researchers look into methods for designing robust and sustainable fresh-food supply chains by surveying many scientific methods such that the waste caused by shrinkage is reduced. Overall, studies exploring agricultural supply chain risk have places emphasis on separately obtaining measures of risk in agricultural supply chains. A limited number of works have attempted to create measures that combine multiple and relevant uncertainty metrics. Additionally, most studies on risk analysis does not include any measures distinctive to the ASC's when analyzing their risk management strategies. A few seminal papers have combined the study of robustness and resilience in using an ASC as a case study[17]. In this paper, the authors explore effectiveness of a mixed set of robust and resilient strategies for managing rare high-impact harvest time and yield disruptions. They develop a two-stage stochastic programming model, that incorporates an exponential perishability function, a feature specific to ASC's. The study suggests that a mixed combination of robust and resilient strategies are most effective for mitigating supply-side disruption risks.

Inventory Models and Perishability

Items that through time become decayed, damaged, evaporated, expired, invalid, devaluated and so are known as perishable or deteriorating items [121]. Some of the early comprehensive literature reviews on deteriorating inventory items were made by Raafat [98] and Goyal and Giri [52] in 1991 and 2001 respectively. In [75], the authors propose key factors that should be considered in studies involving deteriorating inventory. They suggest that the current literature on deteriorating inventories are distinguished into two categories: the studies based on an enterprise and those based on supply chain. Several papers have explored inventory models with perishable items under different conditions including deterministic versus stochastic demand, supply uncertainty, lead times and batch orders among many others [12, 38, 89]. The authors in [13], focus on a continuous review (s, S) model of perishable items with lost sales analyzing the average cost criterion and finding the optimal re-order level, s, and order up-to level, S. They show that the effectiveness of a heuristic that does not include perishability decreases with the demand variability and that costs may either increase or decrease with this variability. In [109], the authors propose and solve a inventory model for perishable items with constant demand where holding cost is time dependent. A recent in-depth review of inventory models for perishable products was done in [28], highlighting the existence of many perishable inventory studies with divergent objectives and scope including factors such as demand, perishable rate, price discount, allowing shortage or not, inflation, time value of money and so.

While inventory perishability is an inherent characteristic of most ASCs, it has generally been studied independently due to the added complexity of this consideration in ASCs. The authors in [87] present a multi-period inventory routing problem with perishable products that consists of a single fresh food supplier, who owns a central warehouse that serves several retail centers. Another study focuses modeling the location of collection centers and companies processing perishable foods in mountainous regions, based on a multi-product and multi-echelon transport system [88]. Perishability in Agricultural Supply Chains has also been studied by a number of researchers using simulation tools and including various tracking technologies [18, 29, 116, 8, 74].

Modeling Disruption in Networks

Understanding network disruption and its propagation down the supply chain is an essential complementary analysis of the performance of supply chains. Several researchers have analyzed supply chain network disruptions via simulations [50, 27]. A fraction of these studies relies on topological network metrics for analyzing the impact of removing nodes and/or edges from the network [83, 4]. Optimization models have also been employed by researchers to study disruption and to design supply chains that are robust and/or resilient to these disruptions [82, 58, 30, 20].In [128], the authors use a flow optimization model coupled with network metrics in a framework that assesses the robustness of supply chain

networks with different topologies when exposed to disruptive events informs decision making. A recent study examines the structural relationships among entities in four fundamental supply network structures to help understand supply network disruption and resilience. The analysis shows that node and arc-level disruptions do not necessarily lead to network-level disruptions, and indicate that network structure significantly determines the likelihood of disruption [68]. In addition, in [85] the ripple effect of node disruption is evaluated using metrics like fragility, service level, inventory cost and lost sales by researchers using bayesian network theory to analyze the multi-echelon network faced with simultaneous disruptions.

2.2 Global Agricultural Supply Chains Definitions

The global Agricultural Supply Chains network consists of interactions and exchanges between aggregated farmers and wholesalers per country for single agricultural products. This network is essential when analyzing the global impact of disruptions to agricultural supply chains. The supply chain is abstracted into a graph, with vertices representing countries and edges denoting the quantity of material (i.e., tons of products) moving between the nodes.

We generate a series of scenarios capturing potential disruptions in the global supply chain at a node level in order to add uncertainty in our optimization framework. We define a disruption as a fractional loss in the production/supply capacity of a node during a single period. This is modeled by multiplying the production/supply capacity of each node by a parameter $\gamma \in [0, 1]$. When $\gamma = 1$, no disruption is included while $\gamma = 0$ represents a full disruption where all the node production capacity is lost. Perturbing the capacity will have a linear effect on our model because the disruption is defined using a linear function (see final section of Results and Discussion). As a result, we focus on the extreme cases in our experiments. We define P_j as the total capacity of node j without disruption. In the full disruption state, the node loses all production capacity such that for node j we have $P_j^{full} = P_j$. Similarly, the no disruption state is defined by $P_j^{none} = 0$ for every node j in the network. Therefore, a scenario consists of a vector of all the individual nodes with a selected capacity state. For our analysis, we assume that at most one node can experience a capacity disruption in each scenario. This leaves us with a number of scenarios upper bounded by the total number of nodes in the network + 1.

Not all nodes (countries) in our global supply chain have the capacity to produce and supply food products. It is essential to only include nodes with significant supply quantities such that their disruption can influence the performance of the ASC. To ensure this, we run our model using scenarios where a single country on the top 25th production percentile is fully disrupted. This analysis results in a set of 46 different scenarios, allowing us to quantify the impact on the costs compared to the baseline (deterministic, no disruptions) case. In this initial study, all scenarios are treated as equally probable.

Data: Global Agricultural Network

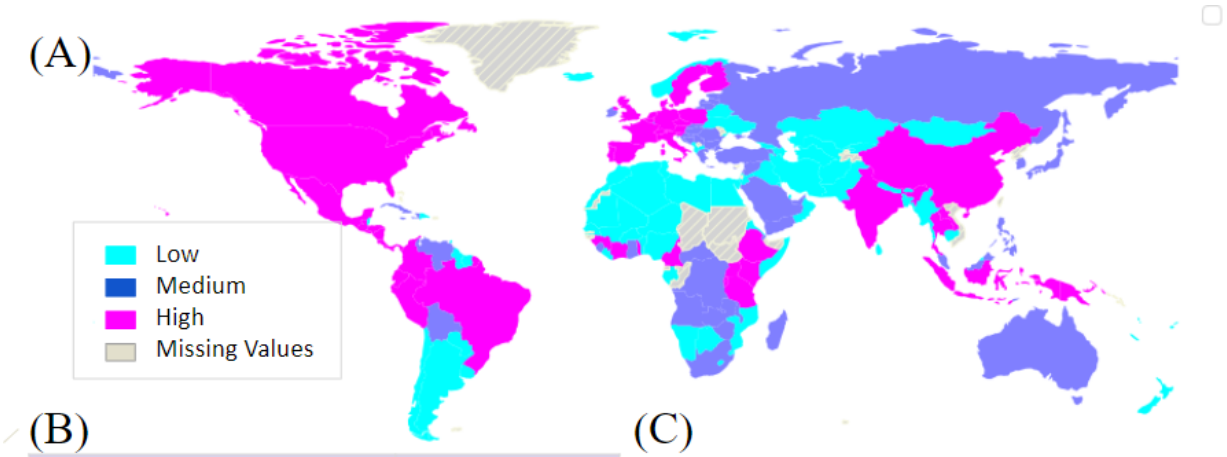
The data used for this analysis consists of global coffee trade data between countries from $2009 - 2019$ obtained from the the World Integrated Trade Solution database [10, 78] . We focus on the coffee trade data, because there is a long history of reliable and available open source data for coffee trade. In addition, at the global scale, Coffee is an interesting product because its network has a clear subset of of relevant suppliers with multiple customers. We use data from recent years (2015 –> single stage model and 2019 –> two stage model) before the COVID-19 pandemic so as to not capture irregular patterns from the COVID-19 pandemic starting in 2020. A summary the statistics for coffee demand and supply data in Kilotons[kt] in 2019 can be found in Table 2.1. In Figure 2.1-(A), we highlight four supplier classes on a world map. Using statistics from the network (Table 2.1), we divide the suppliers into high, medium, and low capacity based on their relative supply capacities. We use the following definitions: high capacity is considered greater than the 75th percentile values, medium-capacity is close to the 50th percentile (median) values, and finally, low capacity is less than the 25th percentile values. This, excluding the zero supply countries which are grouped in a separate category (no supply). In addition, we create a special category for countries with missing supply information. A summary table characterizing the 2019 coffee network is shown in panel (B). In Figure 2.1-(C), we present the top five coffee-supplying countries and their corresponding market share for reference.

Table 2.1: Summary statistics of coffee demand and supply data in Kilotons[kt] in 2019

Characteristic	Demand [kt]	Supply [kt]
Min	0	0
25th percentile	0.378	0.008
Median	9.076	7.419
75th percentile	303.9	142.6
Max	31,070	27,700
Mean	754.6	754.6
Standard dev.	3,041	3,187

2.3 Single-stage Stochastic Allocation Model

We introduce the formulation of the global agricultural allocation model for one commodity under disruption uncertainty. Some key structural assumptions for our model include (1) Use a single product with deterministic demand at each node. (2) We do not allow self loops in the network. i.e we ignore demand of nodes satisfied locally and (3) Unmet demands are lost and assigned to dummy node,$\overline{k}$ We define a directed graph, $G = (\mathcal{N}, \mathcal{A})$ with nodes $i \in \mathcal{N}$



(B)

Number of Nodes	182
Number of Edges	5176
Network Density	0.157
Clustering Coefficient	0.582
Time Resolution	1 year
Spatial Resolution	Country level

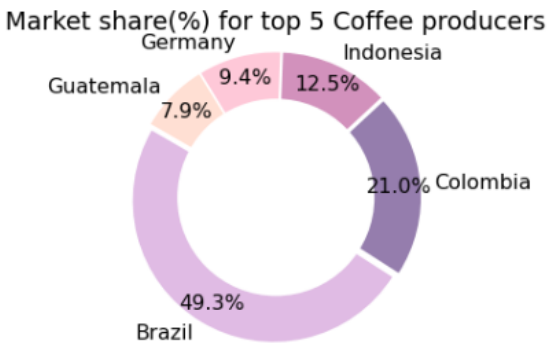


Figure 2.1: Panel (A) illustrates the world map with countries coded by supplier category. They are classified into four categories: no supply, high, medium, and low volume suppliers. Countries with missing data are represented by the missing values category. Panels (B) and (C), show a table summarizing the 2019 global coffee supply chain network characteristics and a pie chart highlights the top five coffee supplying countries and their corresponding market share, respectively.

representing countries and the arcs $(i, j) \in \mathcal{A}$ denoting the flow of agricultural products from node i to node j.
Sets

- I: set of nodes (countries)
- S: set of scenarios (disruptions)

Decision Variables

- X_{ij}^{s} = Fraction of demand node j allocated to node i under scenario s

Parameters

- P_j^s: Production capacity of node j under scenario s
- D_i: Demand at node i
- c_{ij}: Transportation costs from node i to node j
- π: Penalty cost for lost demand
- $N = |I|$: Total number of nodes
- p_s: Probability of scenario s ($p_s = 1/|S|$).
- $f_{i,j}(X_{ij}^s)$: non-linear function (connectivity costs)

Then, we define the stochastic allocation problem $SAP(X)$ including node disruption (scenarios $s \in S$) as follows:

$$\min_x \sum_s p_s \left(\sum_{i,j} (f_{ij}(X_{ij}^s) + c_{ij} X_{ij}^s D_i) + \sum_i \pi D_i X_{i\bar{k}}^s \right) (**)$$

$$\text{s.t} \quad \sum_i X_{ij}^s D_i \leq P_{j,s} \quad \forall j \in I, s \in S \tag{2.1}$$

$$\sum_j X_{ij}^s + X_{i\bar{k}}^s = 1 \quad \forall i \in I, s \in S \tag{2.2}$$

$$X_{ij}^s \in \{0,1\} \qquad \forall i,j \in I, s \in S \tag{2.3}$$

In Eq. (**), the objective is to minimize i) connectivity costs associated with each node which is a function of allocation variables, X_{ij}^s (see Section D) and ii) penalty costs resulting from lost demand allocated to the dummy node $\bar{k}$. Eq. (1) ensures that we do not allocate beyond the production capacity of each node. In Eq. (2) we ensure that for each node, either demand is allocated to another node or lost (sent to the dummy node $\bar{k}$). Binary constraints are imposed in Eq. (3).

Results and Discussion - Single Stage Model

First, we run a baseline case in which no country experiences a capacity disruption, representing the most basic planning situation. Thus, we obtain a solution (lower bound) that will allow us to measure the impact and value of incorporating uncertainty in the planning process. In order to set the different components of our objective function on an equal scale, we normalize the demand and supply quantities, in the $[0, 1]$ interval. As such, we obtain relative expected cost values for comparative purposes.

Solving the baseline case, we find an optimal objective equals 154.74 using $\pi = 10$ (deduced empirically from data; can be modified by decision-makers) and assuming there are no adjusted edge maintenance and transport costs, i.e. setting $\alpha_{ij} = 0$. Since we assume no transport or maintenance costs (subject of future studies), the model focuses on the trade-off between minimizing loss demand while ensuring that allocations are made with as few connections as possible. From the results, we observe how Brazil, Colombia, and Indonesia act as the main suppliers in the network, covering 35.3% of the total global demand. On average, each source supplies coffee to 23 countries, with negligible deviation from this number between suppliers. On the other hand, coffee-demanding countries tend to absorb an almost equal percentage of the total coffee demand (0.8%) while being connected to 8.6 suppliers on average with a standard deviation of 16.7 countries. From the results, we note a complex supplier assignment structure where some countries are connected to as little as 1 supplier whereas others are connected to multiple suppliers. This pattern differs from the more uniform distribution we observe for the number of customers each coffee supplier can access.

We compare our model suggestions to real demand and supply allocations. The top three suppliers (Brazil, Colombia, and Indonesia) in our model correspond with the data. However, it underestimates the total contributions of the top coffee suppliers, with real contributions accounting for 63% of the total supply versus the 35.3% suggested by our model. We find greater diversity in allocations (both on the demand and supply side) within the real data with the top supplier (Brazil) having as many as 43 customers and Canada, the top importer, having being connected to 63 suppliers. The lack of diversity in supply allocations in our results suggest that the current connection function does not fully capture the intricacies of trade relationships between countries. Overall, we observe that our model is demand-driven. This can be a limitation as it prioritizes satisfying high-demand countries and would rather not allocate supply to countries with small demands.

Disrupted network: the value of information

When including uncertainty, we randomly select subsets of scenarios, incrementally increasing their number from 2 to 30. This procedure allows us to observe (Figure 2.6) the evolution and distribution of the objective costs as the number of scenarios increase. From the results (without including the two major outliers, Brazil and Colombia, to facilitate the visualization), we find that these high-impact countries coincide with the high-capacity suppliers (Figure 2.1), as expected. We note from this analysis that there exists a small set of very high-impact countries, such as Brazil and Colombia, that are responsible for providing a huge volume of coffee supply globally. As such, any disruption to the capacity of these nodes is translated into major cost increments. This unbalanced network structure makes the coffee supply chain network significantly vulnerable to disruptions, as any attacks on these critical nodes can severely reduce the overall coffee availability worldwide, significantly altering the market (the supply, demand, and price of the product, its substitutes, and complementary goods).

In Table 2.2, we present the expected objective values, their mean deviations from the baseline (deterministic) case, and the standard deviation of the solutions as a function of the number of scenarios. We note that there is a spike in the expected objective (due to the disruptions of Brazil and Colombia) before the expected solutions begin to converge towards 163 (a 5% deviation above the baseline) as we increase the number of scenarios. We also observe similar patterns in the standard deviation values. From the data, we observe that the peak in the results is the consequence of the successive addition of the two extreme cases (Brazil and Colombia), translated in significantly higher values.

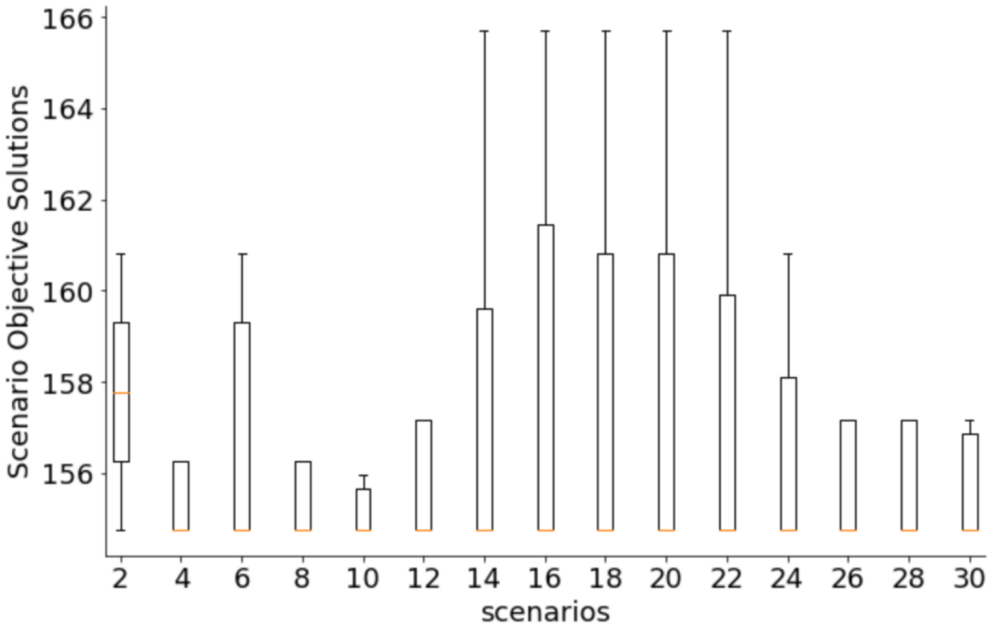


Figure 2.2: Boxplot denoting Changes in objective costs distribution with increasing number of scenarios. The orange line represents the objective value for each instance.

At this point, we also note major differences between the deterministic and stochastic solutions. The limitations of the deterministic solution to adjust the allocation plan when dealing with uncertainty are translated into complex decision-making scenarios. For example, costly distribution alternatives would need to be explored to satisfy the demand and supply contracts of specific sections of the ASC. Moreover, countries could require to increase the consumption of substitute products to cover the existing demand. These situations are translated into a higher vulnerability and poor efficiency of the whole network when facing unexpected disruptions, significantly impacting multiple actors of the ASC. The inclusion of scenarios allows the planner to explore potential effects of disruption and make more robust and risk-sensitive allocation decisions.

2.4 Two Stage Stochastic Allocation Model

We introduce the formulation of the global agricultural allocation model for one commodity under disruption uncertainty with the following key structural assumptions (1) use a single product with deterministic demand at each node; (2) demand can be satisfied by external and/or local production, with no special incentive of covering it locally; and (3) unmet demands are lost and assigned to a dummy sink node, $\overline{k}$. We define a directed graph, $G = (\mathcal{N}, \mathcal{A})$ with nodes $i \in \mathcal{N}$ representing countries and the arcs $(i, j) \in \mathcal{A}$ denoting the flow of agricultural products from node i to node j.

Decision Variables

- X_{ij} = Fraction of demand node i satisfied by node j in first stage
- I_i = Inventory held at node i in first stage
- W_{ij}^s = Fraction of demand node i allocated to node j under scenario s in second stage

Sets

- I: set of nodes (countries: i, j)
- S: set of scenarios (disruptions: s)

Parameters

- $\overline{P_j}$: Historical production capacity of node j
- P_j: Production capacity of node j
- $P_j^s = \gamma * P_j$: Production capacity of node j under scenario s
- D_i: Demand at node i in period t
- c_{ij}: Transportation costs from node i to node j
- π: Penalty cost for lost demand
- $N = |I|$: Total number of nodes
- p_s: Probability of scenario s ($p_s = 1/|S|$).
- C_i^I: per unit inventory holding cost at node I
- $f(X_{ij}), f(W_{ij}^s)$: non-linear function (connectivity costs)
- $\beta \in [0, 1]$: depreciation parameter

- $\zeta \in [0, 1]$: Inventory Level Multiplier
- $In_i^s = \zeta P_j$ Inventory held at node i under scenario s in second stage

Then, we define the extensive form of our two-stage stochastic allocation problem $SAP(X)$ including node disruption (scenarios $s \in S$) as follows:

$$\min_{X,W,I,In} F(X_{ij}) + \sum_j C_j^I(I_j) + E[F(W_{ij}^s) + \sum_j C_j^I(In_j^s)] \tag{2.4}$$

$$\text{s.t} \sum_i W_{ij}^s D_i + In_j^s = P_j^s + \beta I_j \quad \forall j \in I, s \in S \tag{2.5}$$

$$\sum_i X_{ij} D_i + I_j = \overline{P_j} \quad \forall j \in I, s \in S \tag{2.6}$$

$$\sum_j W_{ij}^s + W_{i\bar{k}}^s = 1 \quad \forall i \in I, s \in S \tag{2.7}$$

$$\sum_j X_{ij} + X_{i\bar{k}} = 1 \quad \forall i \in I \tag{2.8}$$

$$In_j^s = \zeta P_j \quad \forall j \in I, s \in S \tag{2.9}$$

$$I_j \geq 0 \quad \forall j \in I, s \in S \tag{2.10}$$

$$0 \leq X_{ij}, W_{ij}^s \leq 1 \quad \forall i, j \in I, s \in S \tag{2.11}$$

In Eq. (1), the objective is to minimize the inventory and non-linear connectivity costs across the ASC network. In the first stage, we solve for inventory levels, I_i and quantity flows X_{ij} for a system with no disruptions. For the second stage variables, given previous stage inventory levels, I_i we minimize the expected costs for different disruption scenarios. Eqs. (2) and (3) represent inventory balancing equations that ensure that the total allocated quantities and inventory from previous stage equals to total production values and current inventory for each node in each scenario, for both stages. We note that Eq. (2) includes the presence of inventory at the end of the second stage for each scenario (In_j^s). In Eq. (4) and (5), we ensure that for each node, either demand is allocated to another node or lost (sent to the dummy node $\bar{k}$) for the first and second stage variables. Eq. (5) defines the inventory required at the end of the second stage by the decision maker for each node and scenario. By default, this is set as a proportion – controlled by parameter *zeta* – of the expected production level of the node. Non-negativity and upper bound constraints are imposed in Eq. (7) and (8).

Connectivity Function

To estimate the connectivity costs we use a function $f(.)$, inspired by the Connections Model from social and economic theory [61]. In our model, making the value of each node derived

from the other proportional to shortest paths makes no sense because all nodes are assumed to be directly connected. Instead, we want to capture the added value of a new connection. As such, we model $\delta^{t_{ij}}$ parameter of the original model in an alternative way. More so, we are interested in costs instead of utility. Since we can think of costs as negative utilities, we negate the utility form. Thus, our connectivity costs function between node i and node j, under scenario s becomes:

$$f_{ij}(W_{ij}^s) = \bar{c}_{ij} - \delta_{ij}^s w_{ij} - w_{ii} \tag{2.12}$$

where,

- δ_{ij}^s: Relative value that node i derives from being connected to node j under scenario s
- w_{ij}: quality (intrinsic value) product node i expects to receive from node j
- $\bar{c}_{ij}$: cost of maintaining edge (i, j)

The δ_{ij}^s parameter captures the variation in value that each connection has, associated with scenario s. To estimate its value, we use the concept of degree centrality, C_i, that measures the number of links incident to a node to assess the importance each node has within a network:

$$C_i = \frac{1}{N-1} \sum_{j=1, i \neq j}^{N} a(i, j) \tag{2.13}$$

with $a(i, j) = 1$ if i is connected to j, 0 o/w.

In our setting, the value of node i will likely decrease as it becomes saturated with connections. Thus, we estimate that node i will experience a lower value from being connected with j as its degree centrality increases. Since we are solving for allocation, and thus, do not know the values of $a(i, j)$, we use the allocation variable X_{ij}^s, obtaining:

$$C_i^s = C_i^s(X_{ij}^s) = \frac{1}{N} \sum_{j=1, i \neq j}^{N} X_{ij}^s \;\; \forall s \in S \tag{2.14}$$

$$\delta_{ij}^s = (1 - C_i^s) \;\; \forall s \in S \tag{2.15}$$

We define w_{ij} as the quality of product node i expects to receive from node j. We assume that all nodes have the same constant intrinsic value: $A = 1$ to other nodes but 0 to themselves (i.e, $w_{ij} = A$ and $w_{ii} = 0$). We set $w_{ii} = 0 \; \forall i \in I$ because our data do not have information about demand fulfilled locally.

Replacing all the expressions in Eq. (5) and multiplying by X_{ij}^s to include only allocated edges, we get:

$$f_{i,j}(X_{ij}^s) \;\; = \;\; \bar{c}_{ij} X_{ij}^s - A(1 - C_i^s X_{ij}^s) \tag{2.16}$$

Thus our (non-linear) objective function becomes:

$$\min_x \sum_s p_s \left(\sum_{i,j} (AC_i^s X_{ij}^s + \alpha_{ij} X_{ij}^s) + \sum_i \pi D_i X_{i\bar{k}}^s \right) \tag{2.17}$$

where $\alpha_{ij} := \bar{c}_{ij} + c_{ij} D_i - A$ (interpreted as the adjusted edge maintenance and transport cost).

2.5 Results and Discussion - Two Stage Model

First, we run a baseline case in which no country experiences a capacity disruption, representing the most basic planning situation. Thus, we obtain a solution (lower bound) that will allow us to measure the impact and value of incorporating uncertainty in the planning process. In order to set the different components of our objective function on an equal scale, we normalize the demand and supply quantities in the $[0, 1]$ interval. As such, we obtain relative expected cost values for comparative purposes.

For the two stage analysis, we use different production values for each stage. In the first stage, the goal is to get a representative baseline inventory level for the start of the year. For this, we use historical average production values from the previous 10 years ($2009 - 2018$). In the second stage we decide to use the production levels for the year of interest, 2019. In Figure 2.3, we present the average and current (2019) supply for historically top and bottom suppliers. For the 50 largest suppliers, the average historical supply trends and 2019 supply values are consistent with an average deviation of less than 4%. For the bottom 30 suppliers, we observe more variations between the average supply and the historical supply performance (up to 105%). Therefore, we note that coffee production tends to follow the same historical patterns i.e., the same countries dominate the production and supply of coffee with most of the discrepancies appearing only in countries with relatively insignificant supply quantities (e.g., Uzbekistan and Solomon Islands). This consistency in supplier size is important to validate the assumption that historical production as a the first stage production level is a representative approximation for production levels in the coming year.

Solving the baseline case, we find an optimal objective equals to 162.07 using $\pi = 10$ (deduced empirically from data; can be modified by decision-makers). We also set out depreciation factor to $\beta = 0.75$ to capture the average shelf life of coffee, about $6 - 9$ months. In the final stage, as a simple inventory management rule, we aim to maintain a base inventory level proportional to the amount of expected supply available by each supplier. This inventory level is regulated by ζ parameter, which we set to 0.1 by default (i.e., 10% of the expected production level). Later on, we explore the sensitivity of the solutions to the perishability of the product, captured by the β parameter. We assume there are no adjusted edge maintenance and transportation costs, i.e., we set $\alpha_{ij} = 0$. The suppression of transport or maintenance costs (subject of future studies) means that our model focuses on the trade-off between minimizing loss demand while ensuring that allocations are made with

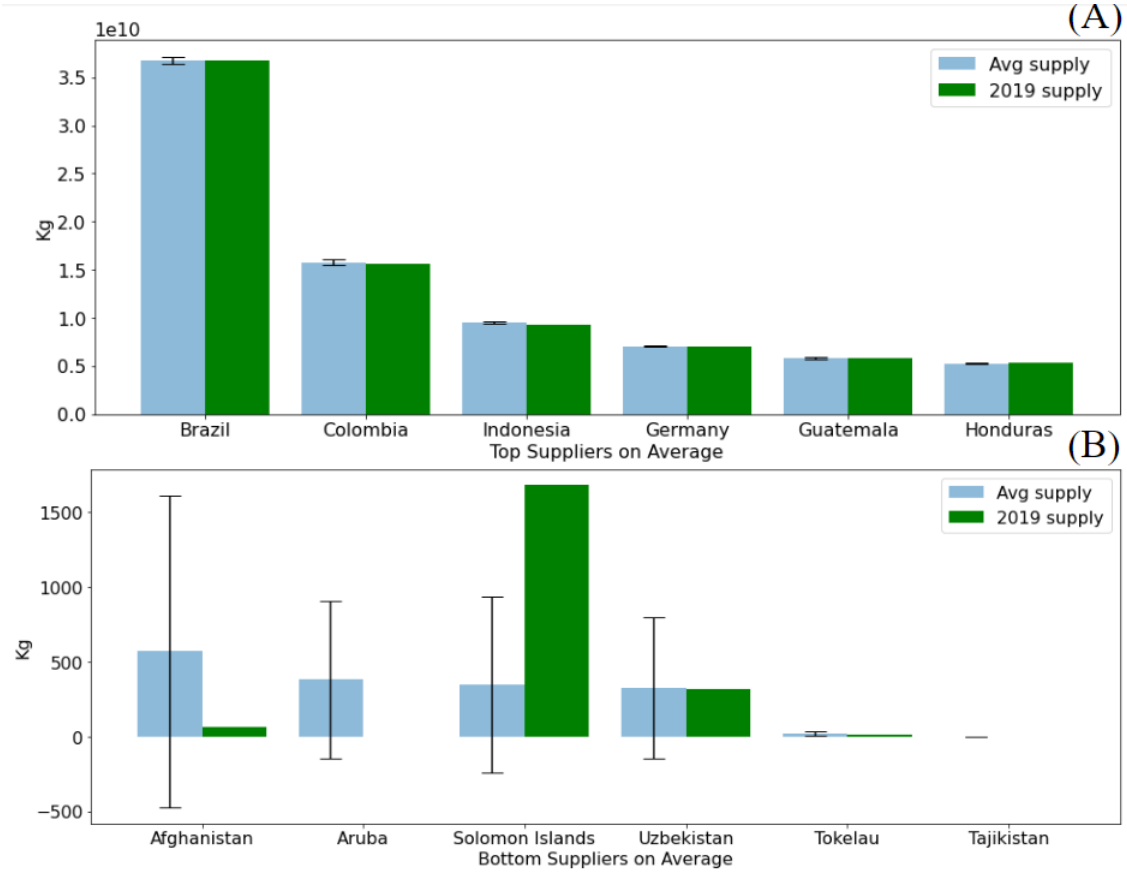


Figure 2.3: Average historical supply quantities from 2009 - 2018 versus supply values in 2019 for the six largest and smallest suppliers historically in panels (A) and (B), respectively.

as few connections as possible. In addition, we observe how Brazil, Colombia, and Indonesia act as the main suppliers in the network, covering 54.2% of the total global demand. On average, each source supplies coffee to about 80 countries, with negligible deviation from this number between suppliers. On the other hand, coffee-demanding countries tend to absorb an almost equal percentage of the total coffee demand ($0.55\% \pm 2.2\%$) while being connected to 37 suppliers on average with a standard deviation of 27 countries. From the results, we note a complex supplier assignment structure where some countries are connected to as less than 3 supplier (e.g.,Lesotho and Sao Tome and Principe) whereas others like Canada are connected to multiple suppliers. This pattern differs from the more uniform distribution we observe for the number of customers each coffee supplier can access.

We compare our model suggestions to real demand and supply allocations. The top three suppliers (Brazil, Colombia, and Indonesia) in our model correspond with the data. While there exists some degree of alignment between supplier allocations from our model and reality, our results tend to have less diversity in supplier allocations. This is explained

because the model underestimates contributions of the top suppliers, with real contributions accounting for 54.2% of the total supply versus the 44.8% of allocated supply suggested by our model. There is greater diversity in allocations (both on the demand and supply side) within the real data. Italy and Canada have the most export/import diversity being connected with as many as 126 customers and 112 suppliers, respectively, while our model assigns, on average, a uniform number of 1 supplier and 40 customers. The lack of diversity in supply allocations in our results suggests that the current connection function does not fully capture the intricacies of trade relationships between countries. In practice, trade between countries are usually influenced by several factors including custom free trade agreements between countries, specified quotas and even environmental and regional agreements [118, 49]. As such, the simplistic view represented by our connections model is one limitation in our model which we plan to improve in the future.

Inventory Analysis

In this section, we discuss the distribution on inventory and how the inventory allocation changes from the deterministic baseline model under different disruption scenarios. For this analysis, we use only single scenario cases so as to compare the impact of disruption on inventory at the country level. In Figure 2.4, we plot the evolution of the average relative inventory (normalized relative to demand and supply values) as the number of scenarios rise. In addition, we also include the maximum and minimum inventory levels. The top two disruption scenarios (Brazil and Colombia) are added first to observe their significant impact on the optimal network allocation. The rest of the scenarios are randomly added in increments of two. There is a massive drop (over 57%) in inventory values between the first two and after including the next set of scenarios. This drop shows that a disruption of the two largest suppliers has a significant impact on the coffee network's ability to supply coffee to other countries, i.e., the whole network is impacted by the absence of Brazil or Colombia's production. This results in large quantities of inventory during the first stage of the planning required to make up for these missing production quantities. As the number of scenarios increases, the range of inventory values converge to about the same levels with minimal or negligible impact on the objective function and network allocations. Additionally, inventory values tend towards smaller values as new scenarios (more information) with negligible impact across the network are included into the formulation, diluting the probability of occurrence equally across all scenarios in this study. This decrease in inventory values with increasing scenarios is due to the dilution of the objective with a larger number of scenarios with countries have lower impact when disrupted.

To measure the impact of the disruption on inventory at an individual country level, we compare single scenario disruptions to the deterministic baseline. In Figure 2.5, we observe the proportion of excess relative inventory from three single scenario disruption cases (Brazil, Colombia, and Cameroon) compared to the inventory needs during the baseline case with no disruptions.

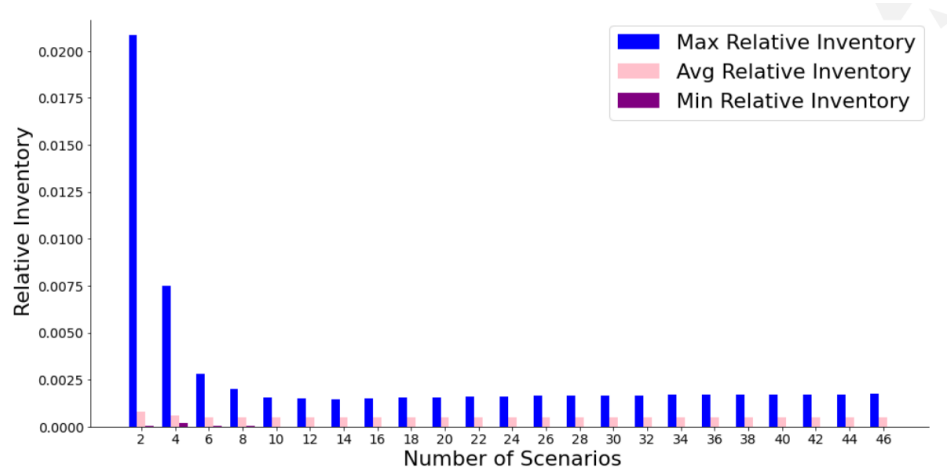


Figure 2.4: Evolution of relative inventory with increasing number of scenarios. In each case we, include minimum, average, and maximum inventory levels. Increasing the number of scenarios results in lower inventory levels as the objective function (expected cost) is diluted by less impactful scenarios.

We select large size suppliers because they have the strongest impact on the model results. We note that the inventory level of the disrupted country is significantly increased to balance the effect of the interruptions. For a small subset of top suppliers like Brazil and Colombia, disruptions to their supply quantities results in these countries having to store as much as 18 and 25 times their required inventory levels respectively for the baseline case. We also note in the case of a disruption in Brazil that the top $10 - 20$ suppliers almost double their inventory needs from the baseline level. This is because on average, the relative proportion of excess inventory needed is a factor of about one from the baseline value on average. There is a similar pattern observed in the case when Colombia's supply is disrupted. For disruptions to a medium-to-large supplier like Cameroon, ranked 27 by supply quantity, the trend is to observe a more localized effect on inventory levels. While Cameroon has its inventory increased by a factor of 12 times from its baseline value, there are little to no change in inventories of other countries.

Disrupted network: the value of information

When including uncertainty, we randomly select subsets of scenarios, incrementally increasing their number from 2 to 46. This procedure allows us to observe (Figure 2.6) the evolution and distribution of the ASC management expected costs as the number of scenarios increases, as well as the impact of disrupting different sections of the network. We zoom into Figure 2.6 such that the two major outliers (i.e., Brazil and Colombia) are not shown in the image to facilitate the visualization. From the results, we find that these high-impact countries

Table 2.2: Summary statistics of the stochastic solutions including mean expected costs (objective), mean deviations of objective from baseline deterministic case, and standard deviations as we increase the number of scenarios considered in the optimization model. We start with the top two most relevant disruptions with highest expected objective value. This decreases as the probability of experiencing disruptions in top suppliers is diluted when including scenarios affecting less relevant countries for the coffee network.

Number of Scenarios	Expected Costs	Standard Deviation	Mean Relative Deviation from Baseline
2	1325.2	514.4	7.18
4	772.3	663.0	3.77
6	571.1	615.7	2.52
8	469.6	561.4	1.90
10	422.6	511.9	1.61
12	379.1	477.3	1.34
14	362.0	445.4	1.23
16	337.0	421.8	1.08
18	317.6	401.5	0.96
20	302.0	383.7	0.86
22	289.3	368.1	0.78
24	279.0	354.0	0.72
26	274.0	340.9	0.69
28	270.5	329.1	0.67
30	263.6	319.0	0.63
32	258.8	309.5	0.60
34	257.1	300.5	0.59
36	251.9	292.8	0.55
38	247.1	285.7	0.52
40	252.9	281.7	0.56
42	248.6	275.6	0.53
44	244.7	269.8	0.51
46	244.8	264.5	0.51

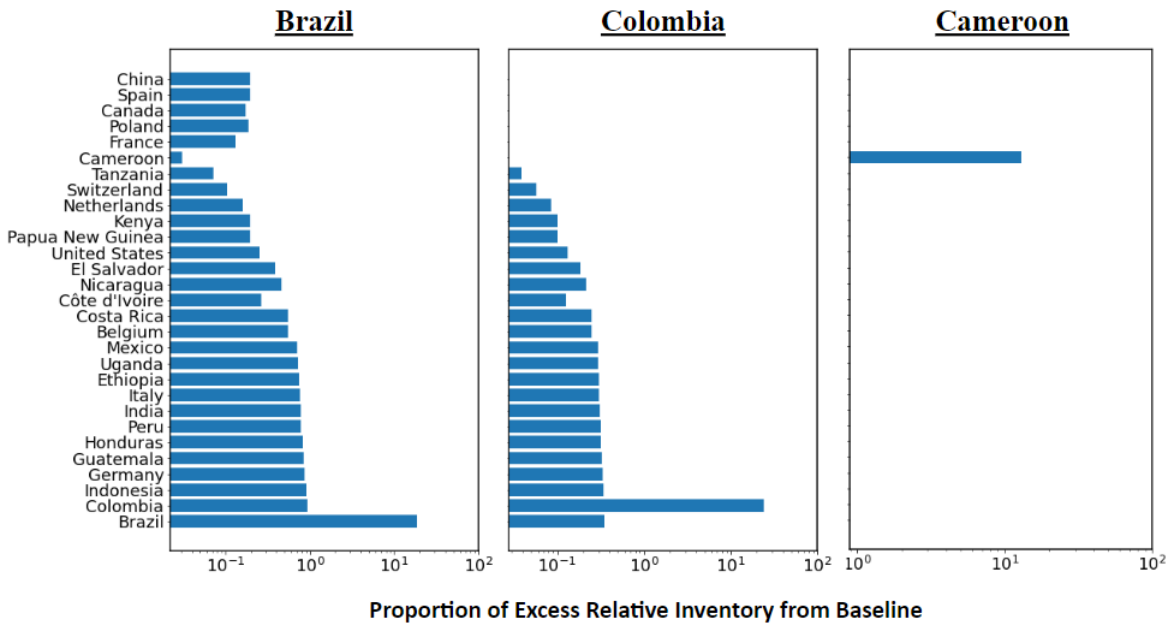


Figure 2.5: Proportion of excess relative inventory from baseline (no disruption scenario) for single three country disruption scenarios: Brazil, Colombia, and Cameroon (left to right).

coincide with the high-capacity suppliers identified in Figure 2.1, as expected. We note from this analysis that there exists a small set of high-impact countries, such as Brazil, Colombia and Indonesia, that are responsible for providing over (50%) of coffee supply globally. As such, any disruption to the capacity of these nodes is translated into major cost increments, triggering a chain of multiple demand re-allocations across the ASC network. This unbalanced network structure makes the coffee supply chain network significantly vulnerable to disruptions, as any attacks (or issue impacting production) on these critical nodes can severely reduce the overall coffee availability worldwide triggering uncertainty and reducing the reliability of the network, significantly altering the market at all levels: its supply, demand, price of the product, its substitutes, and even complementary goods.

In Figure 2.6, we note trends in the objective value which is the expected cost of the stochastic allocation model as the number of scenarios increase. As previously mentioned, the first two scenarios are intentionally selected to represent to top two most impactful scenarios. The rest of the scenarios are randomly selected and incrementally added in pairs. As expected, there is an attenuation in the objective values as the number of scenarios increase (i.e., capturing more uncertainty) because the impact of the disruptions of the large supply countries is diluted by the inclusion of low impact scenarios that have little or no significant effect on global coffee supply allocations (e.g., a disruption in Cameroon as seen in Figure 2.5).

In Table 2.2, we present the expected objective values, their mean deviations from the baseline (deterministic) case, and the standard deviation of the solutions as a function of the number of scenarios. Starting with the two most costly disruptions (Brazil and Colombia) having the highest expected objective value. There is a sharp drop in expected objective before the expected solutions begin to converge after more than 30 scenarios towards 252.1 on average (a 55% deviation above the baseline solution) as we increase the number of scenarios. From the data, the peak in the initial scenario results is the consequence of the successive addition of the two extreme cases (Brazil and Colombia), translated in significantly higher objective values. We also find similar patterns in the standard deviation for the same reasons previously mentioned.

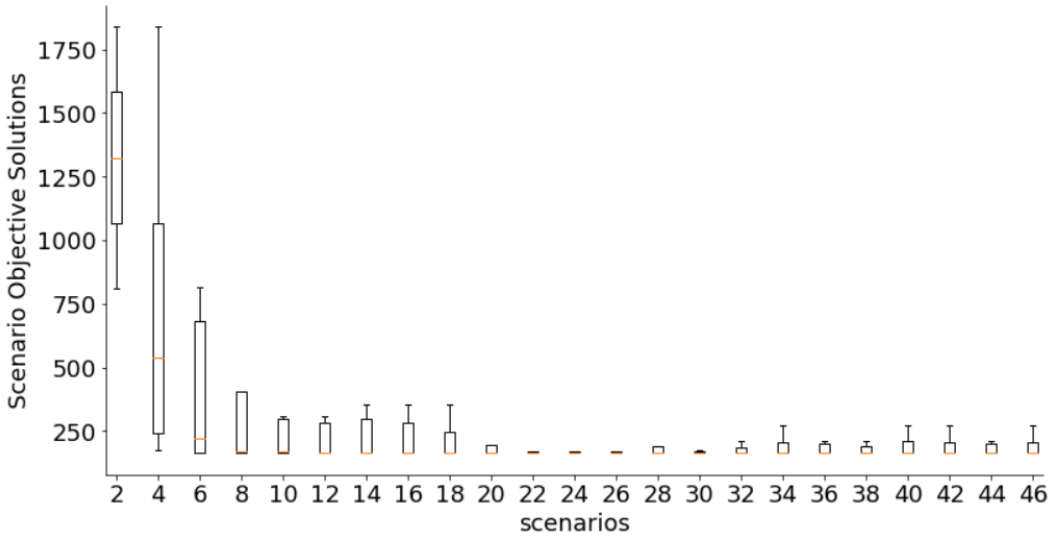


Figure 2.6: Boxplot denoting changes in objective costs distribution with increasing number of scenarios. We select the two most disruptive scenarios initially and the rest of the scenarios are added randomly in pairs. The orange line within the boxes represents the objective value for each instance. As the number of scenarios rise, there is a decline in objective values due to the dilution of probability of high impact scenarios.

We also observe important differences between the deterministic and stochastic solutions. The limitations of the deterministic solution to adjust the allocation plan when dealing with uncertainty are translated into complex decision-making scenarios. For example, costly distribution alternatives would need to be explored to satisfy the demand and supply contracts of specific sections of the ASC. Moreover, countries could require to increase the consumption of substitute products to cover the existing demand, triggering modifications in the whole system. These situations are translated into a higher vulnerability and poor efficiency of the whole network when facing unexpected disruptions, significantly impacting multiple

actors of the ASC. The inclusion of scenarios allows the planner to explore potential effects of disruption and make more robust and risk-sensitive allocation decisions.

Model Sensitivity to Depreciation Factor

In this section, we discuss the effects of varying the depreciation parameter, β on the model objective. We have noted from the single disruption analysis in previous sections that the impact of disruption in proportional to the size of the supplier. As such, for this analysis, we use a five scenario model, including disruptions to the five largest coffee suppliers: Brazil, Colombia, Germany, Guatemala, and Indonesia. These high impact scenarios are selected to enable us to better observe objective changes, if any, as we vary the β parameter. We

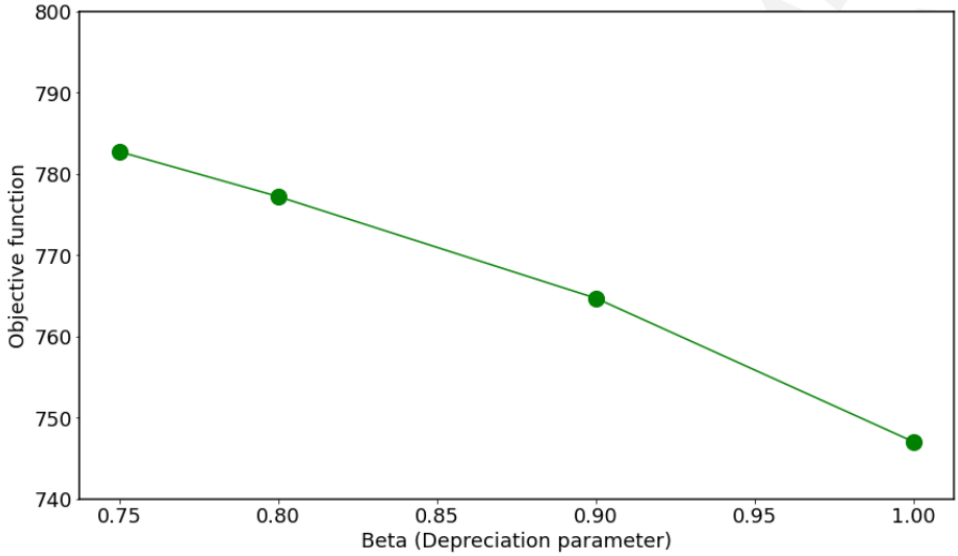


Figure 2.7: Evolution of expected objective cost for different β values using top 5 scenarios. There is a linear decrease in objective value with increase in depreciation factor due to the rise in available increase in inventory.

explore the changes in the depreciation parameter, β over a range of values between 0.75 and 1. These values are selected to mimic/represent the fact that coffee has a relatively long shelf life and can sometimes last longer than a year, which is our time resolution. In Figure 2.7, we plot the evolution of the expected objective costs for β values in range between 0.75 and 1, for the model incorporating the selected scenarios. We note that, as expected, a somewhat linear decrease in objective value as the depreciation factor, β, increases, as less inventory is lost between the two stages. This parameter can be further perturbed to provide insights to decision-makers about different optimal allocation policies according to their storage capacity/resources and even incorporate an extra source of uncertainty in future extensions of the model (e.g., it can be modeled as a random variable for specific products).

2.6 Summary of Contributions

In this work, we developed a general product agnostic framework to analyze the performance of global agricultural supply chains that can be easily applied to multi-commodity networks. Our framework attempts to expand the scale of previous studies where, traditionally, production disruptions (fractional or complete) are not included in the planning process for the allocation of agricultural goods. Planners are now able to explicitly incorporate uncertainty reflecting potential events that could affect the performance of the network by modifying the capacity of the nodes. This analysis is crucial as certain agricultural networks, such as the one analyzed in this study, present a significant dependency on a small subset of countries/nodes, being extremely vulnerable to perturbations depending on their location. Therefore, the analysis of potential solutions and policies to, e.g., evaluate alternative transportation systems, identify backup suppliers to cover their demand, or plan for substitute products, among many other potential scenarios, could be done by planners and decision-makers while including uncertainty.

Chapter 3

A global risk and network health measure

3.1 Introduction

As highlighted in previous chapters, the ever increasing frequency and severity of disruptions can have lasting negative impacts on global supply chains. As such, it is important to understand and assess how risk and its effects impact the supply chain and evolve over time. In this chapter, we aim to develop a global network risk measure and ultimately estimate a data-driven index to measure for health (risk trends) of global food networks.the network health is defined as a measure of trends in the distribution of a network's risk levels over time. We incrementally update our network risk measure starting with a hand crafted intuitive risk measure then the Downstream Protection Value (DPV) and finally proposing the Downstream Supply Risk Measure (DSRM) focusing in including downstream impact of nodes. The consistency in results between DPV and DSRM measures is investigated and their limitations are discussed.

Several studies have been conducted to develop conceptual and mathematical frameworks to identify and assess risk of food and agricultural supply chains. [16, 127, 112, 126, 62] A literature review of various methodologies for identifying and assessing supply chain risk specific to agri-food supply chain can be found in [108]. The authors in [117] suggest that in Food Supply Chains uncertainty, which can be considered from aspects such as time, location, quantity, quality, cost creates vulnerability in supply chains. In [94] the authors investigate and mitigate supply chain risks associated with organic rice in Thailand, by developing and implementing theBest-Worst method (BWM), a risk metric for ranking the critically of different factors, used to establish a hierarchy of importance of the 26 identified risk factors in the Thai rice supply chain. In [108],the authors provide a literature review of various methodologies for identifying and assessing supply chain risk specific to agri-food supply chain. While most of the work intersecting risk and food supply chains have been focused on single products, a few studies have explored the evolution of multiple products.

The Authors in [97] assess the evolution in fragility of 2 staple products (rice and wheat) over a period of 18 years.

3.2 Risk Measures and Numerical Analysis

In this section, the following three different risk measures were proposed (1) Intuitive risk measure (2) Downstream Protection Value (DPV), adapted from previous work (see chapter 6) and (3) Downstream Supply Risk Measure (DSRM). Each subsequent risk metric, is created to improve upon its predecessor. The data for this analysis is obtained from the World Integrated Trade Solution database, created by the World Bank in collaboration with the United Nations Conference on Trade and Development (UNCTAD), International Trade Center, United Nations Statistical Division (UNSD), and the World Trade Organization (WTO) [10, 78]. This data set includes historical ($2009 - 2019$) trade data for various agricultural and food products. A total of 53 different products were used for the analysis and also focus on a 2 groupts of products. The first subset of products (coffee, cocoa, oats, rice, beans) and the second subset consisting mainly of grains (corn, barley, oats, rice, rye and millet) were selected because their high consumption and utilization worldwide and long history or recorded global trade.

Intuitive Risk Measure

First an intuitive risk measure obtained by hand crafting features from the supply chain data that we believe capture the risk levels of each node and eventually the whole network was explored. It was hypothesized that the effects disruptions to a country's supply are more important for countries with (1) large supply capacity and (2) having multiple trade connections relying on their supply. To quantify these factors the following features defined below were extracted:

Definitions

- Supply_fraction (SF): Percentage of total supply lost from network when node removed
- num_supply_dependents (Df): Percentage of countries depending on the node for supply (out-degree)
- Network health threshold (beta): Cutoff value (%) for node risk measure
- Significance level: The probability that the event (risk level of network) could have occurred by chance. Our analysis, focused on the value = 0.05.
- Single product network: This assumes that each supply chain (movement of products between countries) consists of only 1 agricultural product e.g. coffee, cocoa.

For each node, the node risk measure is defined by finding the weighted average of the supply_fraction) and num_supply_dependents.

$$\text{node risk measure} = 0.5(\text{supply\_fraction}) + 0.5(\text{num\_supply\ \_dependents})$$

In order to estimate a measure that captures the networks health, we want to move from assessing risk at the individual node level to the whole single product network. We consider the distribution of the node risk within a and set different node thresholds. We expect higher risk networks to have a larger fraction of its nodes having high risk values. We define the network risk of product i for each time period, t by the probability that the node risk is greater than $1 - \beta$, where β is the cut off value decided prior to the analysis. This network risk is denoted by the formula below:

$$\text{Network risk(i,t)} = \text{Prob}\{\text{node risk measure (x) >= (1-beta)}\}$$

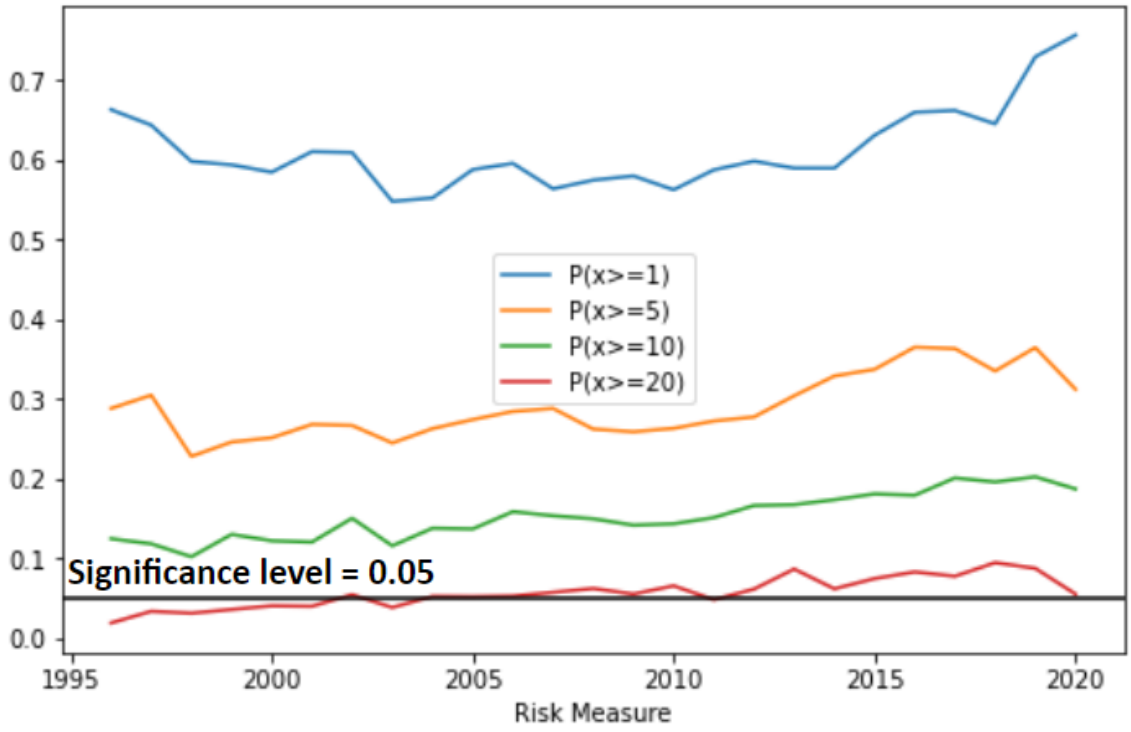
In Figure 3.1, we examine the evolution of node risk for beans at different threshold values. We set various threshold values: 1%, 5%, 10%, 20% but observe from the data that the red curve provides most differentiation between products at 0.05 significance level. The large proportion of curve above significance level line indicates higher risk. Based on the trends in the red line, we consider beans to be moderate to high risk network with an overall increasing supplier size and number of customers dependent on beans with time. We also look at the trends in general products including coffee, cocoa, rice and oats and grains such as corn, rye, millet and barley.

Figure 3.2 and figure 3.3 detail the trends in the intuitive risk measure for various products (cocoa,coffee, oats, rice, corn, rye, millet and barley), we observe the following behaviors

- Cocoa and Oats (lower risk networks): mostly flat risk rates across thresholds with red line sitting below significance lines. Increasing level of risk after 2015.
- Coffee (high risk network): Overall increasing supplier size and number of customers over time.
- Rice (moderately risky network): Mostly flat risk rates for larger thresholds (green, red) lines with red line sitting on significance lines.
- Rye, Barley and Millet (lower risk networks): mostly flat risk rates across thresholds with red line sitting below significance lines.
- Corn (high risk network): Overall increasing supplier size and number of customers over time.

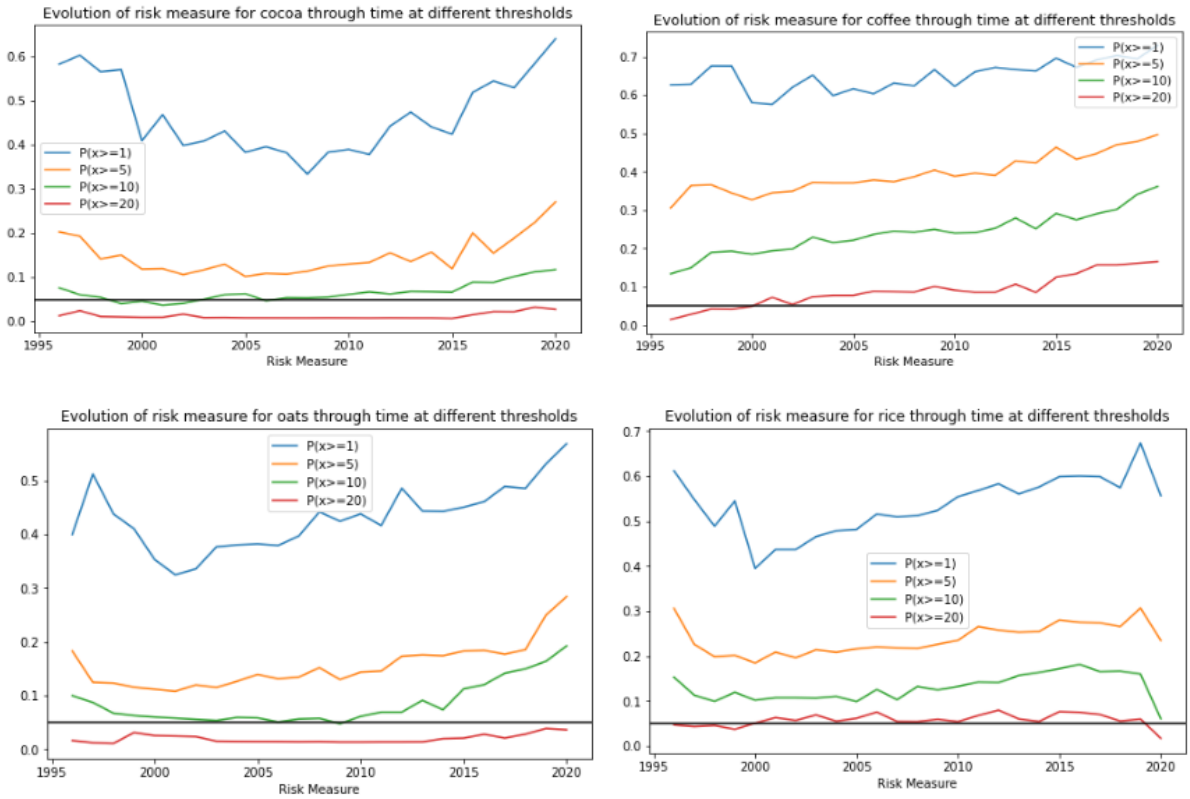
From our definitions, trends in risk can serve also serve as proxy for network health. We note that coffee and corn show curves above the line with steady increasing trends which may lead us to believe that these supply chain is getting more risky with time and may be

Figure 3.1: Evolution of the probability that node risk measure greater than threshold values (1%, 5%, 10%, 20%) for beans. Significance level, the cutoff for which probability values are significant is set to 0.05



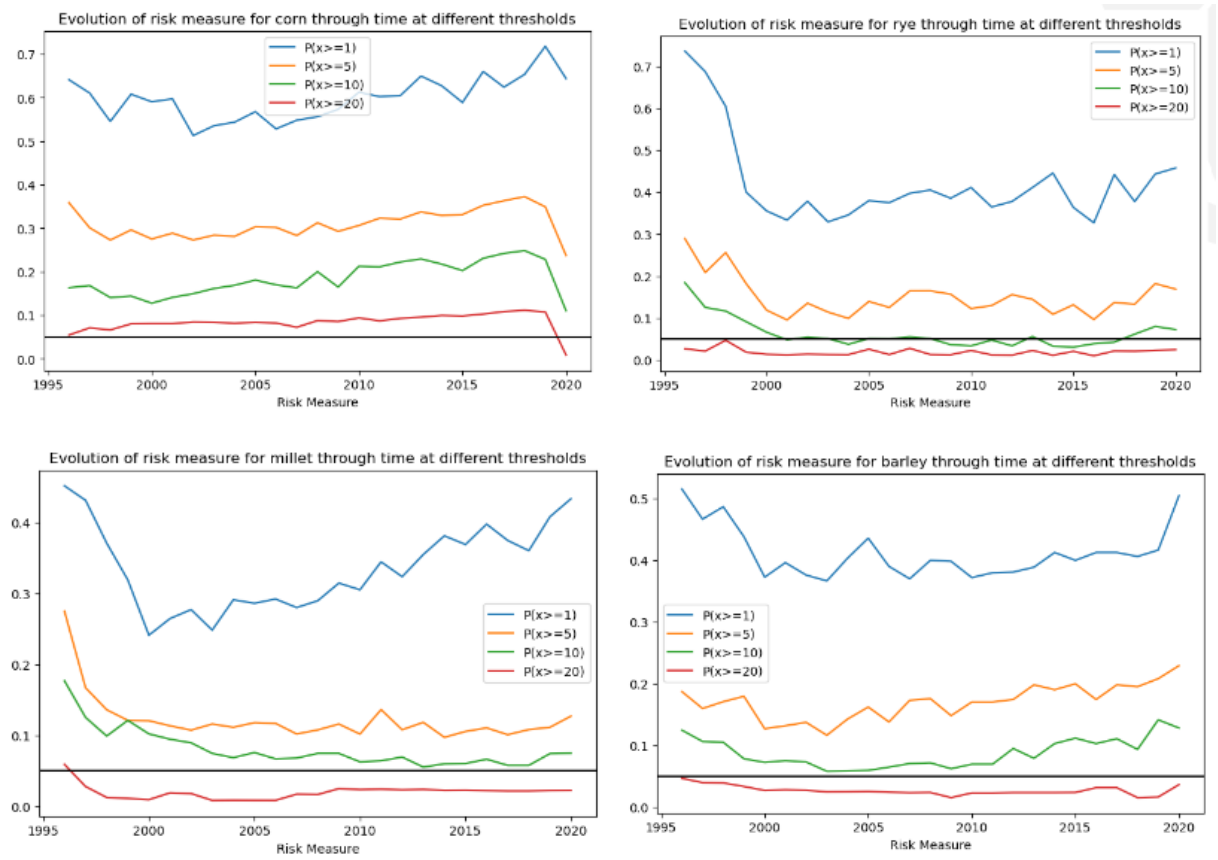
deteriorating in health. Other products such as barley, rye, cocoa and oats have relatively flat lines below the threshold indicating more neutral or stable network health. Finally a product like rice, has a relatively flat curve above the line indicating a high risk product with stable network health (neither deteriorating nor improving
In order to numerically qualify these trends,the slopes of the different curves over the whole time period investigated are obtained. Table 3.1 below details the slopes of the red line with threshold = 1% for the each product. The slope values do not provide enough differentiation and thus we are unable to use this number to qualify the risk. As such we also consider whether the curves lie above or below the significance line. We observe that cocoa and oats are lower risk networks with mostly flat risk rates across thresholds with red line sitting below significance lines. Increasing level of risk after 2015. Coffee on the other hand is a high risk network with an overall increasing supplier size and number of customers over time. Finally, rice is a moderately risky network similar to beans with with the red line sitting on significance lines. bookstab

Figure 3.2: Evolution of the probability that risk measure greater than threshold values: 1%, 5%, 10%, 20%, for various products (cocoa,coffee, oats and rice) with Significance level at 0.05. The products are listed from left to right and top to bottom.



Overall, we present a rudimentary method for assessing risk trends and ultimately network health of single product supply chains. This measure tries to capture the probability that node risk is larger than predetermined cutoff values. Our node risk tries to capture how much supply each node commands and how many nodes are dependent on this supply downstream. Increasing trends usually signal a deterioration in the health of the network whereas constant or decreasing trends signal healthier networks. From the five products we investigated, we found that coffee and were considered higher risk (least healthy) , rice was moderately risky. Finally products like cocoa and oats were lower risk thus more healthy. While this method does not summarize the metric into a single number, the observed trends in our risk evolution curves can provide us some insights about how well the network is doing.

Figure 3.3: Evolution of the probability that risk measure greater than threshold values: 1%, 5%, 10%, 20%, for grains (corn, rye, millet and barley) with Significance level at 0.05. The products are listed from left to right and top to bottom.



Downstream Protection Value (DPV)

One limitation of the previous method is our inability to easily summarize our intuitive risk measure into one numerical value. We attempt to use the slope but we find that this measure is insufficient to provide enough differentiation between the different products. The downstream protection value (DPV) is a flexible metric that estimates the impact of disruption to a country's supply. [92, 41] In this analysis, countries are represented by nodes and the edges indicate flow of material from country i to j (flow(i,j)). To calculate DPV, we use the concept of the shortest path tree of each node, i, defined as the set of all the nodes accessible from node i.

Definition 1 *Let $i \in \mathcal{N}$, and $\mathcal{T}_i = (\mathcal{N}_i, \mathcal{E}_i)$ its propagation tree with root node i. Then, the*

Table 3.1: Summary table for products (Beans, Cocoa, Coffeee, Oats and Rice) including the slope and corresponding standard error of the risk time series plots. In addition, the table also details the position of the curve relative to the significance line and a final column that provides an inferred risk label.

Products	Slope (standard error)	Curve Position Relative to significance line	Risk Label
Beans	0.0031 (0.0012)	Above	High
Cocoa	0.0008 (0.0025)	Below	Low
Coffee	0.0037 (0.0008)	Above	High
Oats	0.0052 (0.0013)	Below	Low to Moderate
Rice	0.0053 (0.0015)	Neutral	Moderate
Corn	0.0036 (0.0012)	Above	High
Rye	-0.0053 (0.0027)	Below	Low
Barley	-0.001 (0.0011)	Below	Low
Millet	0.0026 (0.0016)	Below	Low

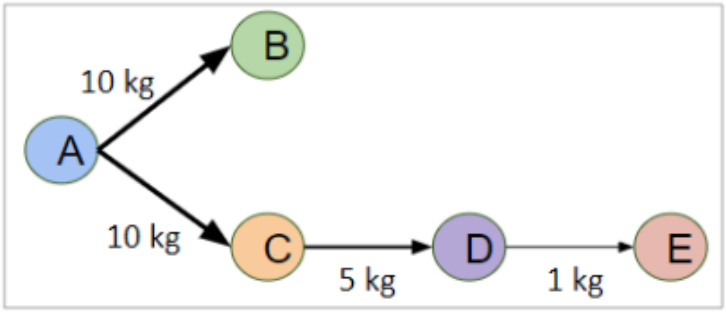
Downstream Protection Value, $DPV\,(i)$ is defined as

$$DPV\,(i) = \sum_{j \in \mathcal{N}_i} VaR_j$$

where VaR_j is the allocated supply from parent node to node j and $DPV(i)$: Accounts for dependence of nodes (countries) downstream of i represented by the shortest path tree of i.

In Figure 3.4 we illustrate an example of how the risk measure DPV is calculated for a node A by summing the downstream flow from its shortest path tree. The shortest path tree of A consists of four nodes downstream (B,C,D,E) plus itself. Each of these nodes receives a flow from its parent node in the tree. We sum all these flows in the tree and that is considered the dpv of A. In this example this leads us to a value of 26 kg.

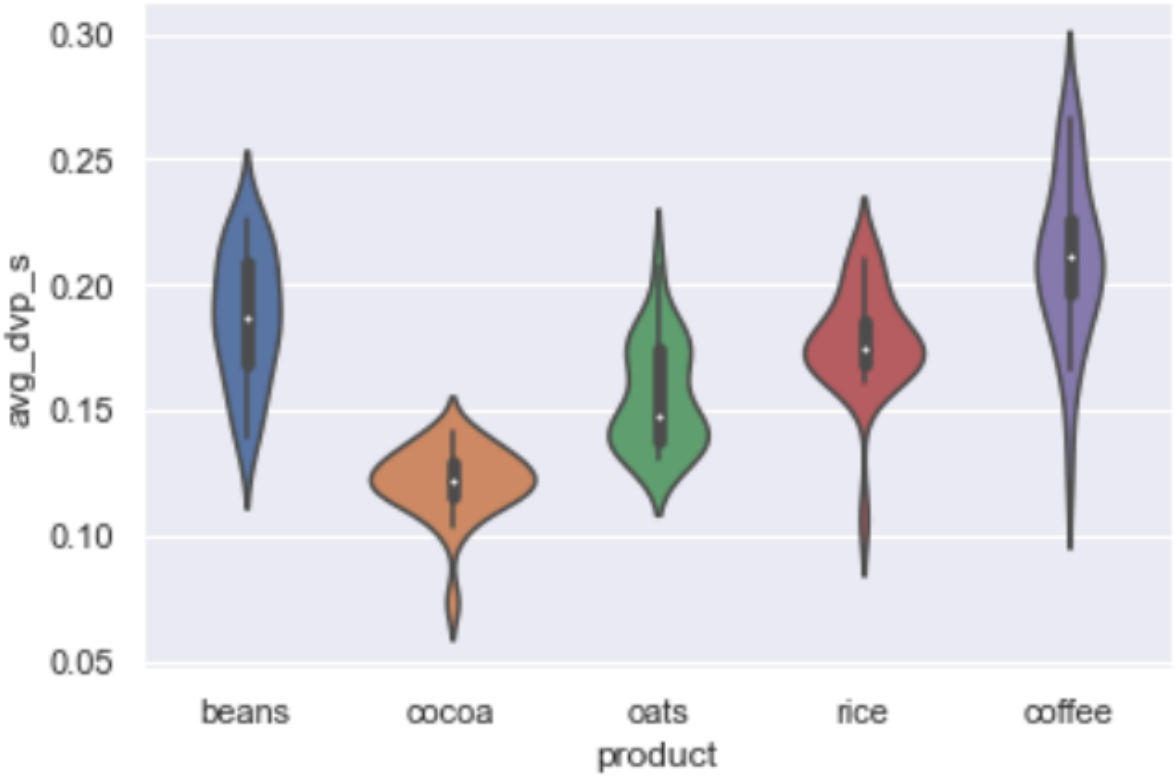
Figure 3.4: DPV calculation on a sample shortest path tree of node A shown on the left side of the figure by summing the downstream flow. The table to the right reflects the DPV value for each node in the tree



Node	DPV / Kg
A	26
B	0
C	5
D	1
E	0

For each yearly product network, we compute the DPV values for each node. We also generate the average DPV across all nodes within the network to summarize these values into a singular metric. Figure 3.5 shows the distribution of DPV values across five products of interest (coffee, cocoa, oats, rice, beans). Overall there are similar patterns in risk behaviours between DPV and intuitive risk measure with coffee and beans having higher risk values which cocoa and oats tend to have lower risk values. Figure B.1 in Appendix **B.2** shows increasing trends in risk Some limitations of the DPV risk measure include it is biased towards large flows and prone to overestimating impact of downstream supply since these are cycles within the network.

Figure 3.5: Violin plots showing the distribution of the Downstream Protection Value (DPV) risk metric across selected products (coffee, cocoa, oats, rice, beans). Corn and beans have relatively high risk values in general whereas cocoa networks tend to have pretty low risk values.



Downstream Supply Risk Measure (DSRM)

While the DPV metric provides a more systematic and dynamic way to measure the risk levels of our network. It still has a few limitations that need to be improved upon. Some of DPV's shortcomings include (1) its bias towards large flows, where countries with large supplies have dis-proportionally large risk values that influence risk distribution and (2) the method prone to overestimating impact of downstream supply as the downstream networks can include cycles that lead to double counting. To tackle this, we propose an updated metric Downstream Supply Risk Measure (DSRM) that solves challenges around including propagation effects of supply disruption and normalizing between nodes and across networks. We denote a directed graph, $G = (\mathcal{N}, \mathcal{A})$ with nodes $i \in \mathcal{N}$ representing countries and the arcs $(i, j) \in \mathcal{A}$

Table 3.2: Summary table including for each product average DPV values with the corresponding standard deviation and an inferred relative risk label based on the average DPV values.

Products	Average Normalized DPV (Standard Dev.)	Relative Risk Label
Beans	0.187 (0.026)	High
Cocoa	0.121 (0.014)	Low
Coffee	0.211 (0.032)	High
Oats	0.156 (0.021)	Low to Moderate
Rice	0.178 (0.021)	Moderate
Corn	0.205 (0.025)	High
Rye	0.123 (0.033)	Low
Barley	0.155 (0.017)	Low to Moderate
Millet	0.152 (0.015)	Moderate

Definitions

Definition 2 *We define the flow of products from node i to node j, $Flow(i,j)$ and define a measure of edge significance,*

$$\delta_{(j,k)} = Flow(j,k)/\max\{Supply(k) - Demand(k), Demand(k)\}$$

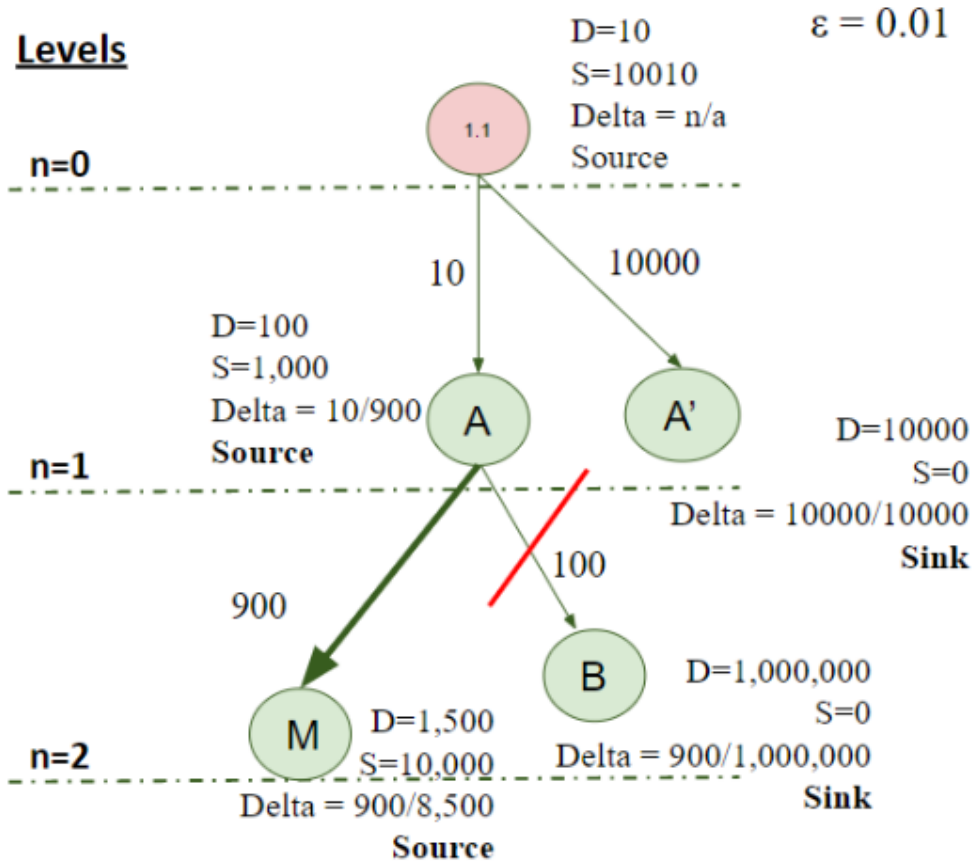
We say that j is accessible from i, denoted by $(i \to j)$ if there exists a path, $i_1, i_2, ..., i_p \in \mathcal{N}$ such that:

$$i \to_{t(i_1)} i_1 \to_{t(i_2)} i_2 \to \cdots \to i_p \to_{t(j)} j.$$

*We say that the graph $\mathcal{T}_i = (\mathcal{N}_i, \mathcal{E}_i)$ is the **"Significance Tree"** of the cell i if:*

- *$i \in \mathcal{N}_i$ is the root of the tree $\mathcal{T}_i$, where $\mathcal{N}_i = \{j : i \to j\} \cup \{i\}$.*
- *$\mathcal{E}_i = \{(j,k) : j,k \in N_i, j \leftrightarrow k\}$ where $\leftrightarrow$ represents the existence of an edge representing supply from j to k , given that k has not appeared previously in $\mathcal{E}_i$ and $\delta_{(j,k)} > \epsilon$, where ϵ is a threshold parameter.*
- *Each edge $e = (j,k) \in \mathcal{E}_i$ is associated with a weight, $w_{(j,k)} = \delta_{(j,k)}/\sum_i(\delta_{(j,i)})$*

Figure 3.6: Sample significance tree for Node 1.1 obtained from Definition 2 with detailed annotations around how each stage of computation for the DSRM measure of Node 1.1. Below the significance tree is a summary table detailing DSRM computation steps for node 1.1 using all its "significant" downstream nodes



Node	Parent	Delta	Weight	DSRM
M	A	0.11	1	0
B	A	0.0009	N/A	N/A
A	1.1	0.01	0.01	0.11(1) = 0.11
A'	1.1	1	0.99	0
1.1	none	n/a	n/a	1(0.99)+0.01(0.01)' +0.11(1) = 1.1001

In figure 3.6, a sample significance tree for node 1.1 is shown. The significance tree is created by following the steps detailed in **definition 2** and eliminating nodes that are not significant i.e. do not allow the and calculate the normalized measure for node 1.1. DSRM computed

from bottom to top using significance tree. Levels are defined as the number of edges between the downstream node and the root node. The table below the significance tree shows the analysis starting with the leaf nodes. In each level, we combine all the weighted delta values of each node and then add it to the next level. In our example, the furthest level is two and consists of leaf nodes M and B. Node B is eliminated after failing the significance test. Then we move up to level 1. Since node M is downstream of node A. The pseudocode capturing the detailed implementation of DSRM is presented in Appendix **B.3**).

3.3 Results and Discussion

In this section the performance of the updated DSRM metric is compared to our previous DPV measure. The goal of our new measure is to reduce double counting of downstream effects in the network and also improve the process of normalizing between nodes and across networks.

Figure 3.7: Comparing distribution of DSRM (left) vs DPV (right) across a subset of products. In the top panel, we compare difference between a general set of products (Coffee, Cocoa, Oats, Rice, Beans) while in the bottom panel we focus specifically on grains (Corn, Barley, Oats, Rice , Rye and Millet)

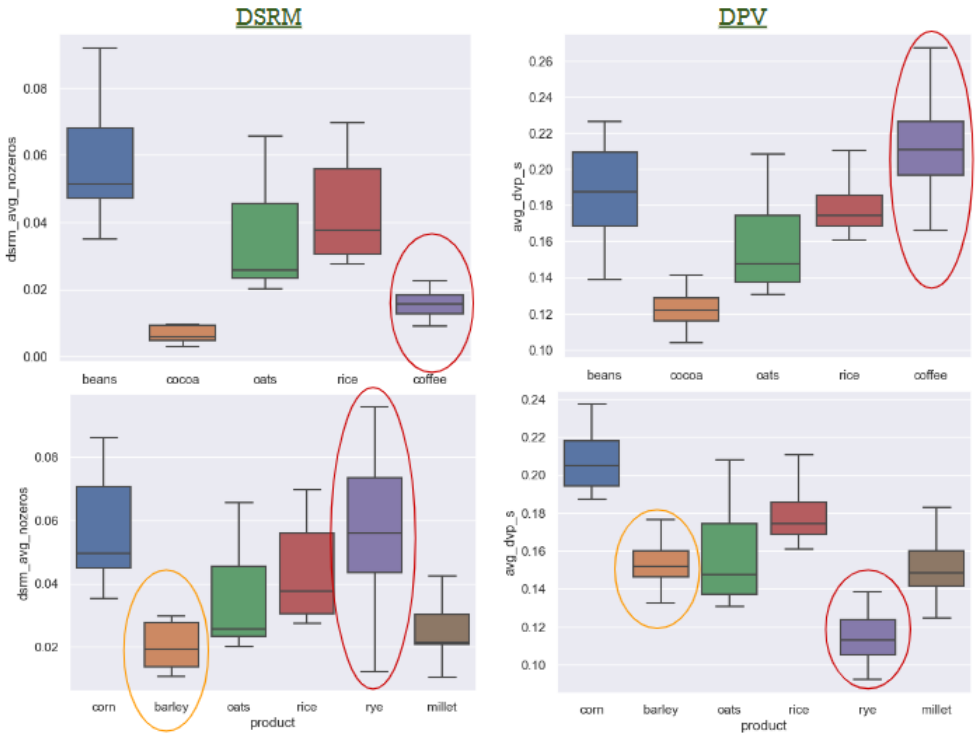


Figure 3.7 shows comparisons between the distribution of DSRM and DPV for two sets of products, a more general of commonly consumed foods and another set consisting mainly of grains. Grains in particular are selected because of the high dependence of most of the world's diet on these products. As expected the total Average DSRM measures were much lower than the normalised DPV because insignificant edges were removed. This result is due to the removal of redundant and/or insignificant downstream connections as well as the normalization at each level in DSRM compared to DPV. Overall there seems to be similar patterns in relative behaviours between DSRM and DPV across products but there are some products (e.g coffee and barley) that behave very differently between DPV and DSRM.

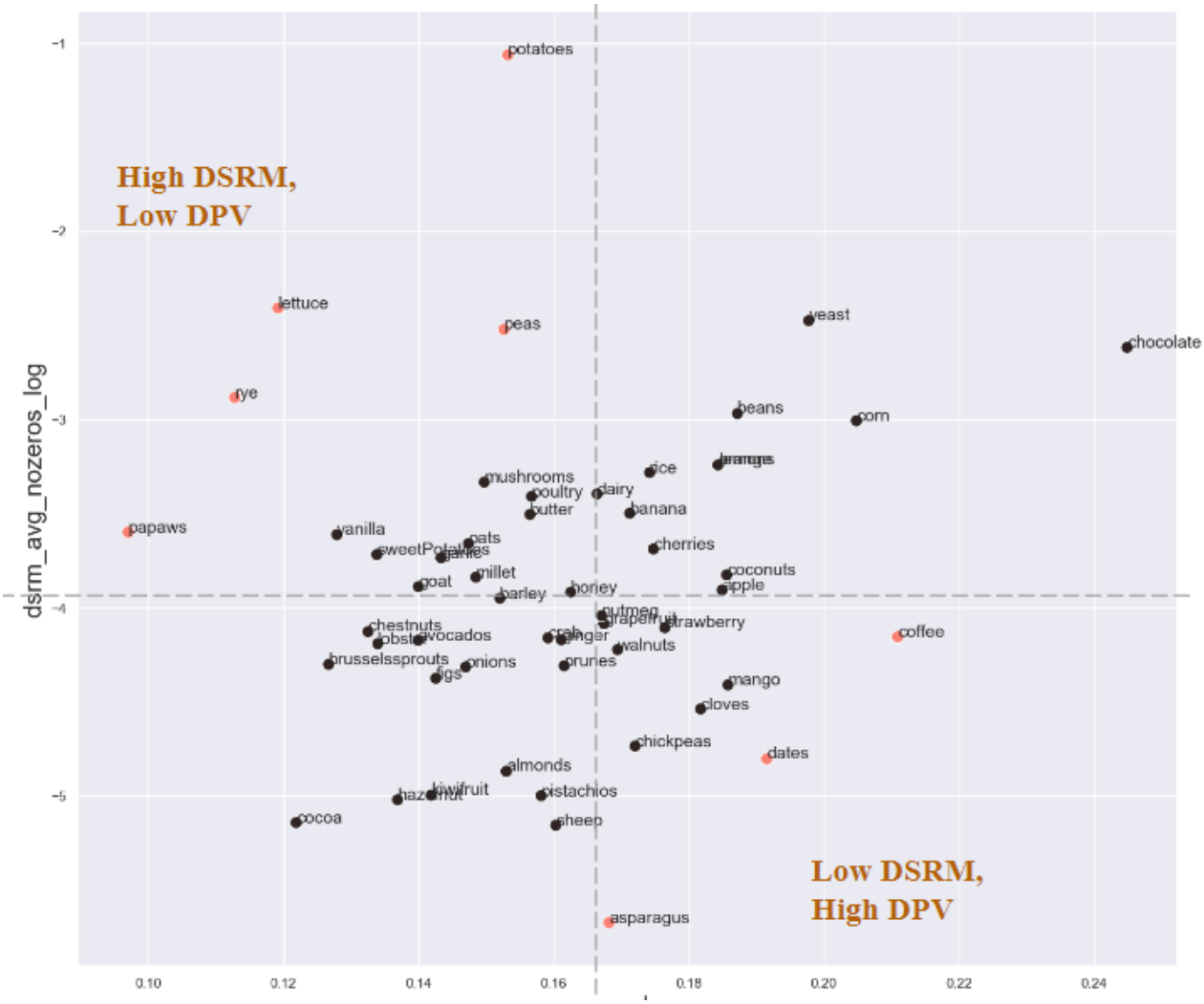
To further investigate the prevalence of outlier behaviors between the two metrics, we explore correlations between DSRM and DPV. In figure 3.8, we present the scatter plot of DSRM vs DPV. We re-scale the DSRM measure using the log scale for better visualization since the relative values are very small. For the whole data set including median values across all product we obtained a correlation of 0.13. The original data set contains multiple products data points for every year, each with slightly different values. The noise in the signal for correlation between 2 metrics across products is reduced by obtaining and using the median values for each product instead of the whole distribution. From figure 3.8, two sets of outliers groups are observed. The first group characterized by the high DSRM and low DPV group consists of potatoes, lettuce, rye, pawpaws, and peas. The second group which includes asparagus, dates and coffee is the low DSRM and high DPV group. Removing these outliers increases the correlation between the two measures to 0.54.

A closer look at trends in risk behaviors in DSRM reveals surprising differences compared to the DPV trends. The figure in Appendix **B.2**, shows increasing trends in the yearly distribution of Normalized DPV Across all Products. However, the overestimation of downstream effects when using the DPV metric render these trends less reliable. In Figure 3.9, we revisit risk trends, taking a closer look at boxplots showing the distribution of DSRM across all products through time. This time we get more complex risk behavior, with decreasing trends indicated ameliorating network health between 1996 and 2001. After 2015, reveals mildly increasing risk trends signally overall increasing vulnerability of networks globally in the more recent years. The period between 2002 and 2015 for the most part is relatively flat indicating stable network health.

3.4 Summary of Contributions

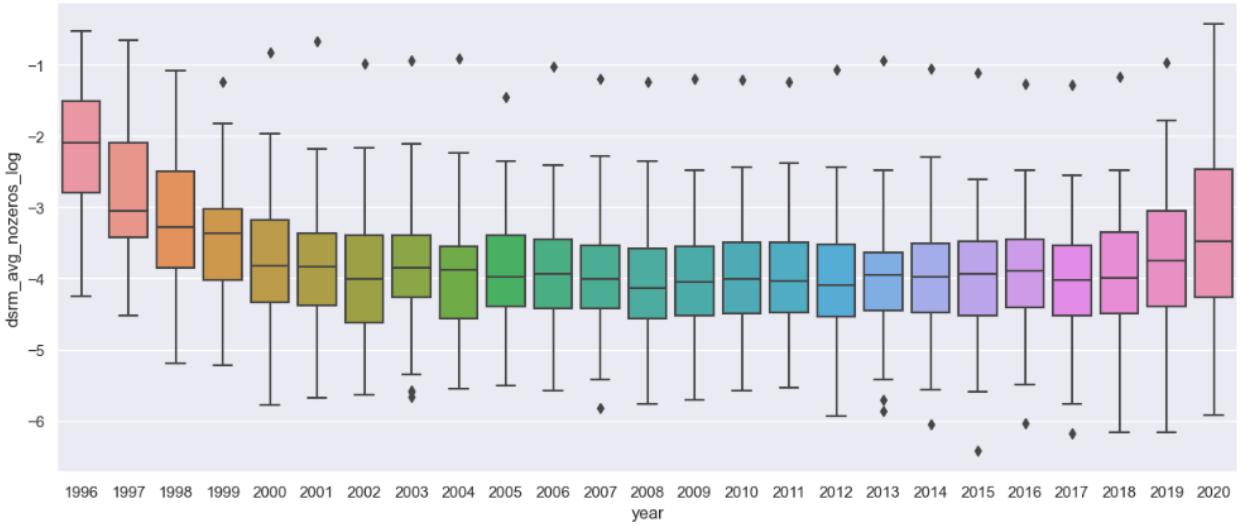
We designed and incrementally improved a data-driven measure that quantifies the network risk levels. We implemented DSRM, our most current measure on multiple products and compare trends in risk values to our previous measures (DPV and Intuitive risk measure). In general, the three metrics seem to capture similar relative behaviors. However, only DPV and DSRM can use a unique number to capture the entire network's performance. In addition, DSRM solves two challenges faced by its predecessor, DPV. The first challenge of overestimating impact of downstream supply due to double counting is solved by defining

Figure 3.8: Scatter plot of the median values across products for of the log(DSRM) versus Downstream Protection Value (DPV) metric for the values. The median values are used to reduce noise from the product networks through time. There is a trend to the top right corner indicating positive correlation between the two metrics. The orange dots indicate 2 groups of outliers, the high DSRM and low DPV group (top left quadrant) and the low DSRM and high DPV group (bottom right quadrant).



the significance tree to better track impacted nodes downstream. The second issue with DPV is the bias towards large flows which we combat by normalizing between the nodes in the significance tree at each level. While we observed similar relative risk patterns between DPV and DSRM for most products, there were some products (e.g coffee, rye) that had

Figure 3.9: Box-plot showing the distribution of log of Normalized DSRM values across all Products between 1996 and 2020.Between 1996 and 2001, the time series has a negative slope. For the most part there are no trends between 2002 and 2015. After 2015, there is a slight positive trend towards risk with time potentially signally overall increasing vulnerability of networks globally in the recent years.



completely different relative behaviors between the two metrics. More so, results from our measure align with our previous optimization work (see chapter 2) where coffee was shown to be a relatively high risk network due to its supply chain structure that was heavily reliant on a small subset of suppliers.

One of the goals of this work is to attempt to quantify the network health, which we define as trends in the risk levels through time. The DPV metric shows consistently rising trends in risk with time potentially supporting that supply chain risk has been steadily rising since 1996. However, these findings may not entirely correct due the double counting error that occurs when using DPV on graphs instead of trees. On the other hand, the trends in DSRM present more interesting results. Before 2001, the time series has a negative slope. For the most part there are no trends between 2002 and 2015. After 2015, there is a slight positive trend towards risk with time potentially signally overall increasing vulnerability of networks globally in the recent years.

Chapter 4

Framework for categorizing supply chains(food) using structure.

4.1 Introduction

This chapter further explores the concept of supply chain risk and investigates methods for grouping supply chains by using features that capture network structure to allow the development of more scalabable policies. We develop a framework to systematically group different food product supply chains by identifying their structural similarities and differences and vulnerability using Graph Neural Networks. These vector embeddings obtained from the graph neural network enable us to summarize the overall structure of the individual networks. This method uses trade data between countries from a World bank database for 53 different products over a period of 25 years. Decision makers can thus create policies for food networks at the group level instead of individual policies for the large variety of food and agricultural products.

Providing global and equitable access to food and nutrients is a longstanding problem that is important because food is a basic human right. This challenge requires the design of effective systems that regulate the movement of food between countries and access to nutrients by different populations. Studies suggest that global trade between countries is necessary to enable nutrient access to all populations. This is especially essential for nutrient poorer communities and countries thereby allowing them to nourish large proportions of their population.[124] The increased exchange of food between countries have rendered food supply chains increasingly complex. In addition, the need for many outsourced procedures such as food product storage and transportation services and lack of communication between stakeholders augment the already existing complexity in these supply chains.[60] In Agricultural and food supply chains, highly perishable grocery foods constitute up to 50% of all sales in retail food. [33] These products with widely varying life cycles make it difficult for decision makers to manage perishable food products because of deterioration, wastage,spoilage, and a short shelf life. [7, 120, 76] More so, the rising frequency and intensity of extreme climate

events with projected continued temperature increases, and shifting rainfall patterns have accentuated volatility of crop and livestock yields. All these factors contribute to rendering global agricultural and food supply chains more vulnerable to external shocks such as natural disasters, pandemic outbreaks etc.

In the long run, a country's income is determined by its product variety and sophistication. [55, 56] As such, Food policy makers face the challenge of creating solutions for a large number and type of food and agricultural products and varying life cycles. Usually, it is not feasible to create solutions for individual products. As such products tend to be grouped based on their origins and destinations, with little thought about the overall structure of the products supply network and how it is affected by disruptions. Our study aims to present a streamlined process of identifying grouping of products that are similar in network structure. This allows decision makers to create policies at the group level, where the groups are developed without prior assumptions about the product type and its supplier locations.

4.2 Methods

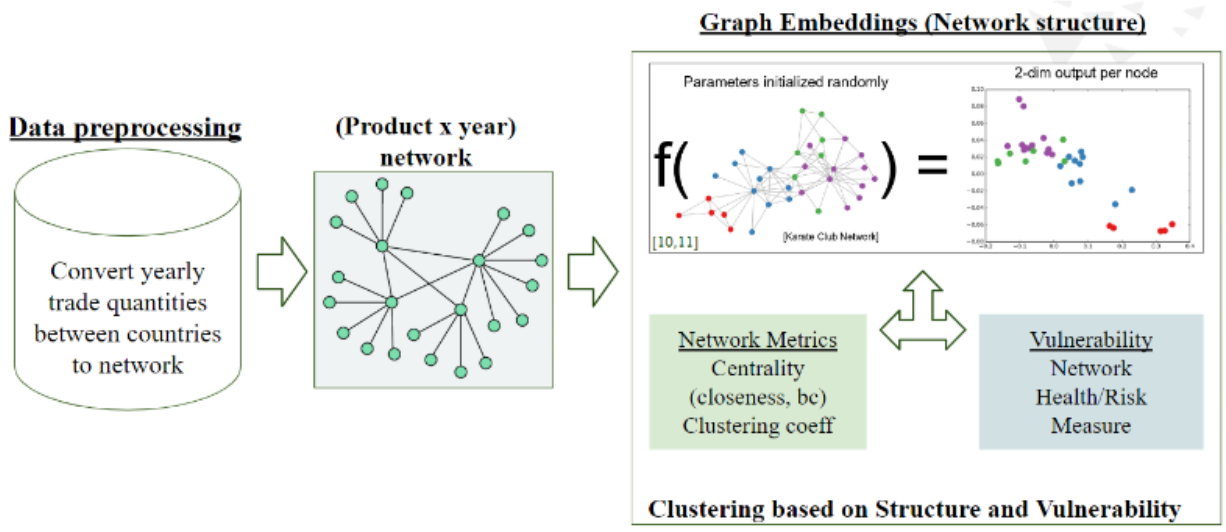
Clustering food networks based on structure

Our method aims to group the product supply chains based on their structural similarity and risk levels. The data consists of historical (1996−2020) trade data for various agricultural and food products. The data is found in the World Integrated Trade Solution database, created by the World Bank in collaboration with the United Nations Conference on Trade and Development (UNCTAD), International Trade Center, United Nations Statistical Division (UNSD), and the World Trade Organization (WTO) [10, 78]. Each data point consists of a product supply chain for a specific year. We model each product supply chain as a network with nodes consisting of countries and edges denoting the quantity of product flowing between countries. The network measures such as number of edges, centralities (degree, closeness, betweenness), average clustering coefficients and network density enable us to approximate the structure of each product supply chain. We use the concept of downstream protection value (DPV) introduced in [92, 41], to estimate the vulnerability of our networks. DPV is a flexible metric that ranks the impact of disruption to a country's supply. This measure of risk Accounts for dependence of nodes (countries) downstream of each node i, represented by the shortest path tree of i.

Extracting Network Embeddings

To better capture the network structure, we use graph neural networks (GNN) to create graph embeddings, a vector representation based on the structure of a network. These deep network models exploit information from the adjacency matrix (A), the vector of attributes associated with the nodes (X) and potentially, the vector of attributes attached to the edges (Y). This allows us to preserve the spatial structure and connectivity properties of the graph,

Figure 4.1: Framework for clustering Agricultural supply chain networks based on structure and vulnerability levels. For each product and year, trade data between countries is converted to a network. These networks are then converted into embeddings and clustered using convolution graph neural networks.



in contrast to traditional convolution neural networks (CNN) that could only process grid-like planar graphs due to their homogeneous sliding window scan approach. We use two methods to generate embeddings from our product networks.

First, the Graph2Vec algorithm is implemented to move from networks to embeddings. This method consists of first sampling and relabeling all sub-graphs from the graph. The skip-gram neural network consisting of an input layer, a hidden layer, and the output layer is then trained to maximize the probability of predicting sub-graph that exist in the graph. The network accepts the one-hot encoded words. The hidden layer has no activation function, its output presents an embedding of the word. The output layer is a softmax classifier that predicts neighborhood words. Finally, using the aforementioned inputs the graph neural network generates multidimensional embeddings of the original graph, summarizing its information in P-dimensional vectors, with P an hyper-parameter to be determined empirically during the research.

In the second method, we implement and train a three layer Graph Convolution Network (GCN) with Pytorch Geometric. GCNs allow us to generalizing well-established neural models like convolutional neural nets (CNNs) to work on arbitrarily structured graphs. The Pytorch Geometric implementations of GCNs are based on work from authors in [69]. Given a network, the goal of the GCN is to learn an underlying function that returns node-level

output features, that can also be converted to graph level features using pooling operations [39]. Our 3-layer GCN performs three propagation steps during the forward pass of the algorithm. This allows the GCN to convolve the 3rd-order neighborhood of every node (all nodes up to 3 "hops" away) generating embeddings that capture the structure individual networks.

We use a number of clustering methods and alogorithms to group our product supply chains and we perform a number of tests with different combinations of these algorithms before finally designing a community detection algorithm. We use agglomerative hierarchical clustering, an unsupervised clustering method that focuses on creating clusters that have predominant ordering from top to bottom. The algorithm works by initializing every product supply network in its own cluster, and then pairs of clusters are merged based on similarity as one moves up the hierarchy. Using hierarchical clustering allows the decision maker the flexibility to select the desired number of clusters which could be based on external constraints. The main limitation of this clustering approach is that the results can change based on how the clusters are initialized and changes to the data set (e.g. adding new data). We also validate our clusters by comparing with results from k-means clustering using the same number of clusters. While we only present results from clusters generated using one random state we also try different random start points and manually check approximately the same clusters generated. Spectral clustering is a method that considers clustering a graph partitioning problem, searching for nodes in a graph with a small distance between them. Spectral clustering relies on the kernel trick to introduce additional dimensions to the data thus overcoming the linear cluster boundary problem faced by the more common k-means clustering method.

Finally, we propose the community detection cluster algorithm to bypass the problem of selecting number of clusters using TSNE embeddings from the GCN embeddings to identify different groups of networks. The community detection algorithm groups networks based on their cosine similarity function. After generating graph embeddings for each network using graph neural nets we apply the community detection algorithm to group these networks. The method consists of first compute the cosine similarity between pairs of embeddings in an iterative fashion . At each stage, there are thresholds for the minimum size and mininmum level of similarity to be considered for group membership. The algorithm proceeds until convergence, where no new samples gets added to clusters and the remaining samples are clustered as noise.

4.3 Results and Discussion: Globalization Trends

We analyze of 53 food products across 25 years resulting in 1325 individual networks. Table 4.1 details the different product broken down by their food groups. In general, we observe an increase in the number of nodes (countries) and trade connections between these countries for every product with time. This highlights the increasing globalization as the number of countries in which products are traded and thus consumed have grown to include almost

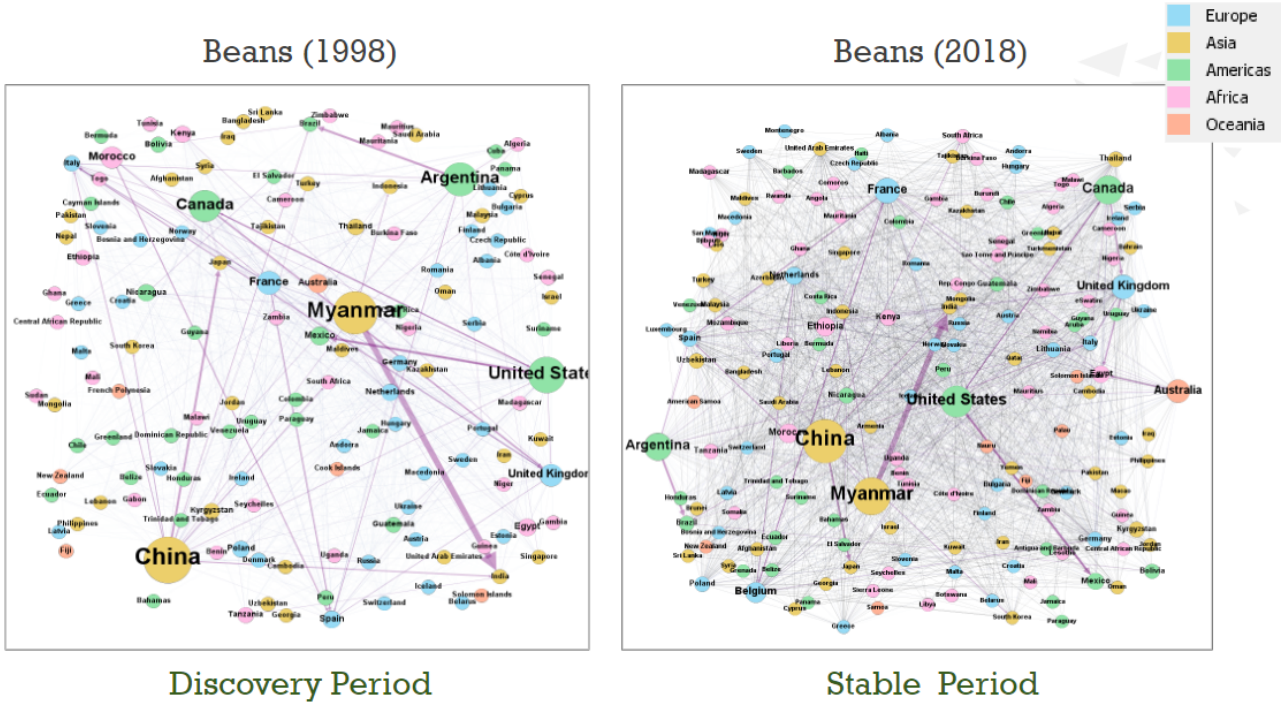
Table 4.1: Break down of 53 agricultural and food products representative of all groups in the food pyramid and across 25 years (1996 - 2020). This resulted in 1325 Networks.

Food Groups	**Products**
Animal Products	Crab, Lobster, Sheep, Goat, Butter, Honey,
Bread, Cereals and Pasta	Rice* , Barley, Oats, Corn*, barley, Millet, Rye
Fruits	Apples, Strawberries, Bananas, Mangos, Oranges, Coffee, Cocoa, Prunes, Pawpaws, Lemons, Grapefruits, Avocados, Figs, Dates, Kiwi, Cherries, Coconuts
Spices and Nuts	Ginger, Nutmeg, Cloves, Vanilla, Pistachios, Hazelnuts, Walnuts, Chestnuts, Almonds,
Legumes and Vegetables	Mushroom, Beans, Onions, Asparagus, Brussel sprouts, Chickpeas, Lettuce, Sweet Potatoes, Potatoes, Peas, Garlic,
Tertiary	Yeast, Chocolate
Milk , Yogurt and Cheese	Dairy

the entire word. We loosely divide the time line into 3 main periods (Discovery, Growing, Stable)based on the rate of change of the number of countries and their trade interactions. The Discovery period which happens between 1996 and 2000 is characterized by the fastest rise in number of nodes and connections. This period has a mean rate of change of 111.7 ± 77.5 connections and about 41.5 ± 26.1 countries. The Discovery period set between 2001 and 2015 has a average rate of change of 12 ± 2.5 connections and about 1.5 ± 0.4 countries. The Stable period after 2015, has average rate of change of -31.8 ± 98 connections and about -2.9 ± 2.9 countries has almost no change as the errors are wider than the changes in slope in both number of nodes or connections in the networks.

While the size of the networks grows, these changes happen differently for different classes of products. Connections in chocolate networks overtake and eventually dominate Coffee networks. Rice and Beans pretty much have the same growth in connections. Products such as oats, barley, cocoa start off at similar levels but diverge during the growing period with oats growing much faster. Figure 4.2, further highlights the effects of globalization on network structures. Over a 20 year period, there are changes in the structure of the beans network from 2008 and 2018. In 2018, a year in the stable period, beans has more distributed supply controlled by large numbers of small and medium sized suppliers and denser connections. While in 2008 , a year in the discovery period, there are much fewer connections indicating less countries trading on the global scale.

Figure 4.2: Evolution of beans networks in a twenty year time-span between 2008 (discovery period) and 2018 (stable period). The left image represents beans network in 2008 has fewer connection and less nodes. To the right we see the effects of globalization with beans network in 2018 having denser connections network with a large number of small and medium sized suppliers.



Preliminary Analysis using Graph2Vec

First we explored a subset of our data set to explore the feasibility of our approach. We used 16 products over a period of 25 years and using the Graph2Vec algorithm, which was the simplest method we used to generate graph embeddings. From our hierarchical clustering analysis we identify 7 separate groups. The breakdown of the products and the group statistics (mean vulnerability, number of networks) are shown in table 4.2.

The embeddings generated from the Graph neural networks together with our vulnerability measure (DPV) and other network topology metrics (e.g density, etc) are then passed into clustering algorithms for grouping. We prioritize hierarchical clustering to allow flexibility in selecting group size, but we also compared our results using clusters of the same size generated from Kmeans clustering. We set the number of clusters to between 5 and 8 clusters inferring from our dendrogram. Table 4.2 presents a breakdown of the products and the group statistics (mean vulnerability, number of networks) into 7 separate groups. We

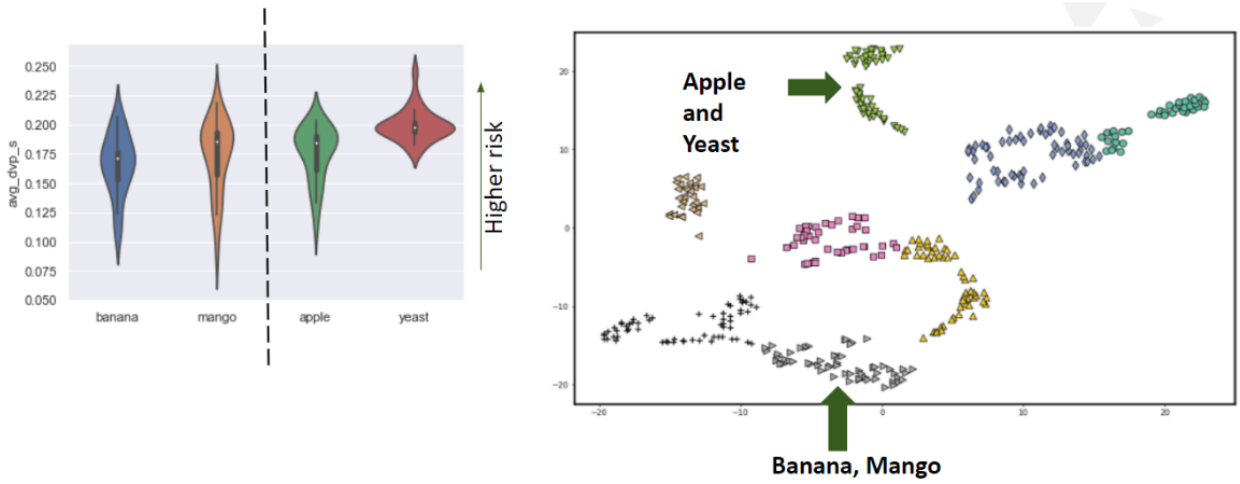
Table 4.2: Summary statistics including predominant products, size, average number of nodes and connections with data grouped into 7 clusters using hierarchical clustering. Groups vary slightly between different clustering algorithms and start points but overall group composition is relatively consistent.

Groups	**Predominant Products**	**Relative Size (%)**	**Average # of Countries**	**Average Connections**
0	Corn, Oats, Coffee	15.25	164	2,607
1	Orange, Rice	14.75	149	1,434
2	Dairy, strawberry, beans	14.75	157	1,801
3	Banana, mango, coffee	16.0	160	1,757
4	Barley, poultry	13.5	141	1,215
5	Apple, yeast	10.25	159	2,118
6	Chocolate, beans	8.0	173	3,985
7	Cocoa	7.5	135	827

note that while the groups generated vary slightly between different clustering algorithms and start points, the overall composition of the groups remained relatively consistent.

From preliminary analysis, the network structure of supply chain dominates during the assignment of product supply chains to clusters. When investigating the clusters deduced for our method, we find a few patterns in the groupings, some expected (eg. cocoa and banana having similar supplies and being of predominantly tropical origins) while others are counter intuitive(e.g. apple and yeast) which supports the need for such a framework in the first place. In addition, we observe that the differences in risk within groups, levels become important. When risk (DPV) values and distribution deviate significantly from each other, the products are large within the same cluster. In Figure 4.3, the left figure denotes two sets of products: banana and mango which fall in the same group have overlapping distributions for risk (DPV) values while apple and yeast have very different DPV distributions. In the right figure, which denotes the 2D representation of the product-year network embeddings color coordinated by their group. The networks in the former group (banana and mango) with similar DPV distributions have no clear separation in space. However, in the latter group (apple and yeast) there is a distinct separation between networks from both products. In the next steps, we want to expand the product network quantities and diversity.

Figure 4.3: Left figure shows the distribution of risk levels for products in 2 separate groups. The first containing mango and banana have overlapping risk distributions that are quite similar. Whereas, the products in the second group (apple and yeast) have very different risk distribution. The right figure shows the vector representations of the different networks in 2D space. networks in the former group with similar DPV distributions have no clear separation in space, whereas those in the second group show clear separation between networks from both products.



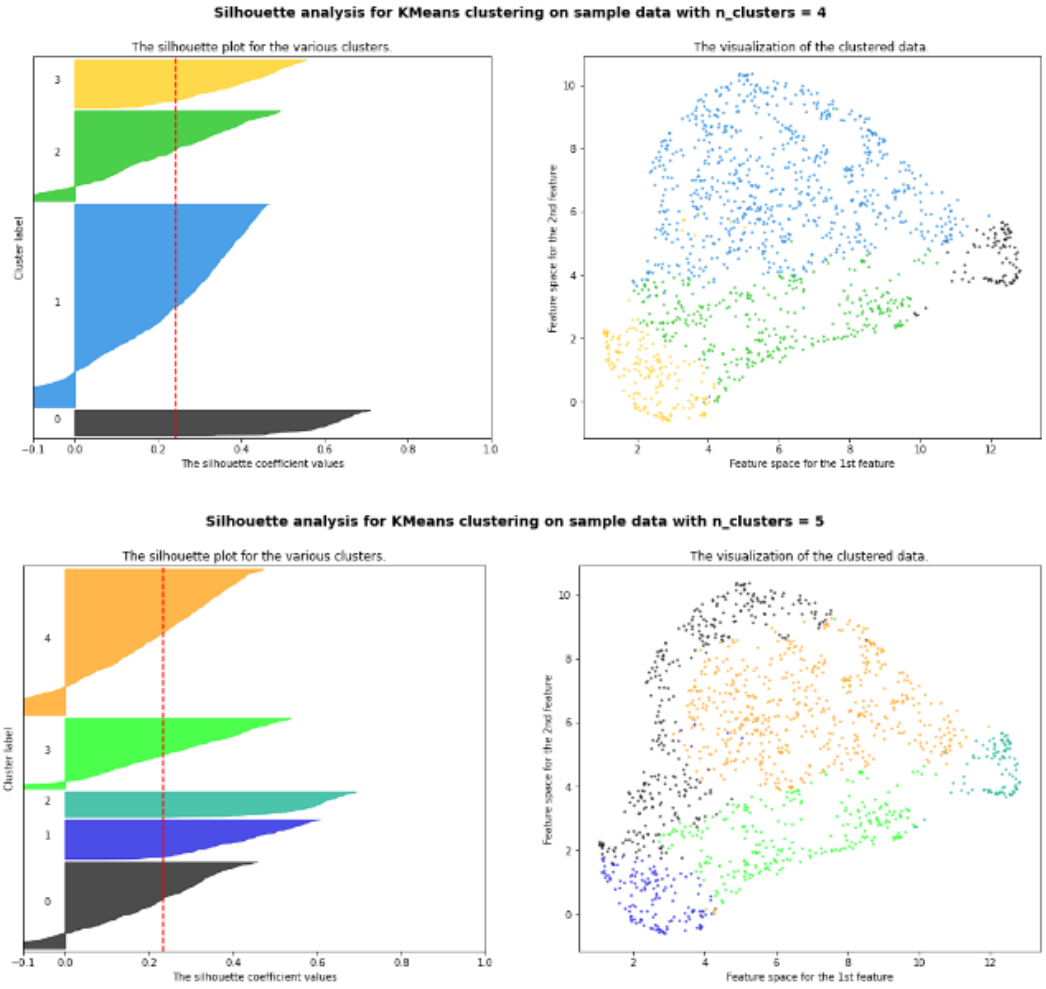
Selecting Number of clusters

In this section, we want to investigate how to select number of clusters using different clustering algorithms (spectral and hierarchical clustering) for our clustering framework. In this analysis we includes all 53 food products across 25 years using both. During our analysis, we generated 4 dimensional vector embedding of each individual graph to capture its structure using Graph2Vec. The embeddings generated from the Graph neural networks together with our vulnerability measure (DPV) are then passed into clustering algorithms for grouping. Hierarchical clustering allows the flexibility in selecting group size and we compare our results using clusters of the same size generated from spectral clustering. Using a combination of Silhouette plots generated from the using clusters from spectral clustering and the dendrogram from hierarchical clustering, we investigate the best cluster size with the range of 2 and 7.The range of values for number of clusters was selected using the dendrogram empirically to maximize difference between generated clusters.

The height of the dendrogram represents the distance between clusters. Appendix **C.5** shows the dendrogram used for this analysis. Observing the dendrogram from the top to bottom, we note that the big difference between clusters is between the components of the orange cluster versus the red and green clusters the vertical height (blue line) is longer for

the former.

Figure 4.4: silhouette plots for number of clusters ranging between 4 and 5 and their corresponding 2D visualization on the top and bottom respectively. The clusters are generated obtained from spectral clustering.



Our analysis using a combination of silhouette plots and a dendrogram suggests that using 4 or 5 clusters seem to be the better selection. In Figure 4.4, we examine the different silhouette plots for number of clusters equal to 4 and 5. From their silhouette profiles their clusters have uniform behavior across the cutoff and each cluster have sizes in the same order of magnitude (no very skinny or extremely large clusters). The rest of silhouette plots for number of clusters ranging between 2 and 7 can be found in Appendix **C**.

We note that increasing the number of products studies from 16 to 53 makes the embeddings generated using Graph2Vec less separable in 2D space. The two dimensional projec-

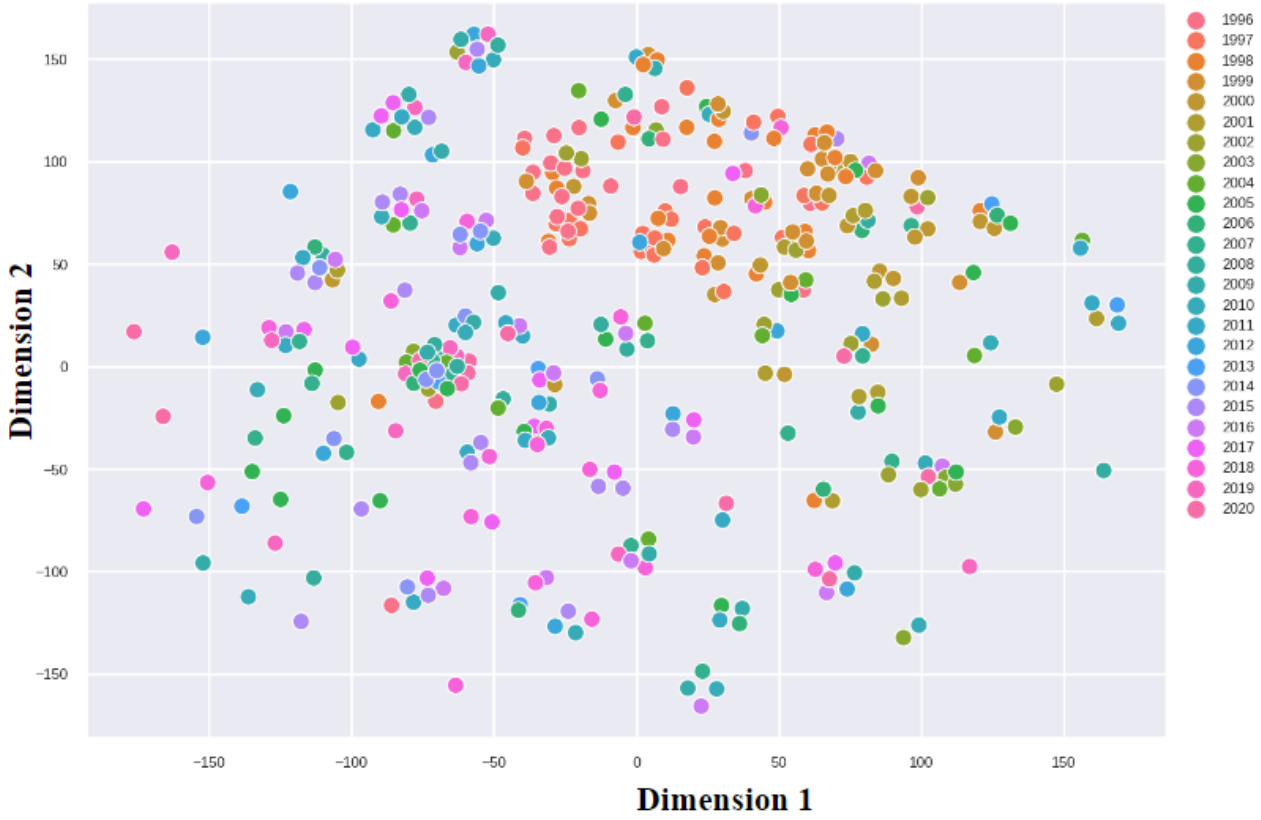
tions on the right side of figure 4.4 present points that are relatively evenly distributed on the space. These results led us to improve our graph representations (embeddings) by (1) using higher dimension embeddings (64-256) to better capture individual network structure (2) implement and train our own graph convolutional network (GCN). More so, to get around pre-selecting the number of clusters, we implement a community detection algorithm which allows us to run the algorithm without initially specifying the number of clusters.

4.4 Clustering Global Food Networks by Structure - Full Analysis

Following the preliminary results, we the extend the analysis to include all 53 products and graph embedding methods. During our analysis, we generate embeddings of sizes in the range of 64 to 256 of each individual graph to capture its structure using Graph2Vec and our trained GCN (more details on combination of parameters in Appendix D). With the increase in number of products, we note immediately that the graph2Vec model becomes much less expressive than the 3 layer GCN model. In Appendix D, we show figures of the 2D projection with TSNE of the Graph2Vec Emmbeddings (with and without attributes). The Graph2Vec embeddings (non-attributed) do not reveal any clear separations and the addition of attributes do not help. Figure 4.5 shows two dimensional projection with TSNE of 128 dimensional GCN embeddings (attributed). Each data point represents an individual network with colors representing the year in which that network was created. From the figure, we can already identify separations within the data indicating that some clustering may be feasible. From the distribution of colors on the image, we see some indications that yearly dependence is relevant for cluster formation. This suggests that the trained GCN models are more capable of separating networks by structure.

Both models are trained to generate embeddings that capture structure but also include certain node and edge attributes. The node attributes consist of our risk measure(DPV), degree, demand, supply while the edge attributes include amount of flow, the presence of a connection. We use t-distributed stochastic neighbor embedding (t-SNE), a statistical method to project the high-dimensional embeddings to a two dimensional map. We focus on the GCN model as it is more expressive than the Graph2Vec model. We expand on results generated from a 128 dimensional embeddings generated from the trained GCN. The community detection cluster algorithm is then applied to the t-SNE embeddings and the number of clusters are generated to ensure 99% coverage of the data. From our community detection clustering analysis of 53 food products across 25 years, we identify 11 groups consisting of 10 clusters and one group of non-clustered points. The breakdown of the products and the group statistics (top representative products by cluster time period and relative cluster size (%)) are shown in table 4.3 In figure 4.6, we present a two dimensional projection with TSNE projection of the 128 dimensional embeddings from the graph convolutional networks after the community detection algorithm has been implemented. The data points are color-labeled

Figure 4.5: Two dimensional projection with TSNE of 128 dimensional GCN embeddings (attributed).



representing individual clusters with an additional group denoting the non-clustered points obtained after the algorithm converges.

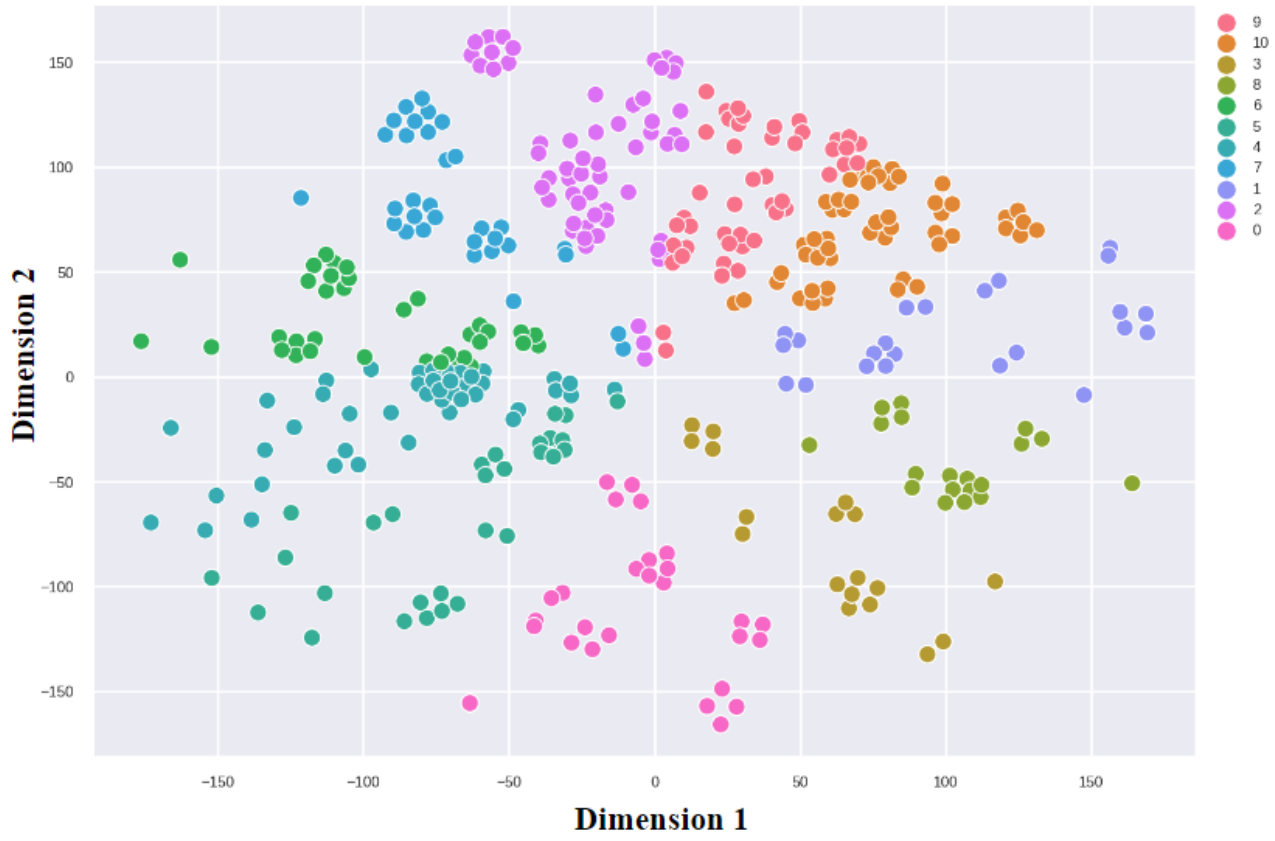
We identify several factors that contribute to cluster formation including time period, product group , climate (eg tropical vs temperate) with high regional concentration of suppliers products with shared function or use (complementary or substitute products) and network density. There is a very strong effect of time on cluster formation as products in same groups tended to fall within the same time periods (discovery, growing, stable). In addition, similar products grouped together, even when not the dominant groups with the cluster. For instance, staple products like rice, beans and corn were found in the same groups. Regional concentration of suppliers is another important factor that influences the cluster formation. We find that products such as with even demand across countries and characterized by having a small subset of countries responsible for over 50% of global supply.

Different food products such as Grains(rice, corn, barley) and nuts(pistachios, almonds,

Table 4.3: Summary statistics including top representative products by cluster time period, median network density and relative cluster size (%) with data separated into 11 groups(10 clusters + 1 group of non-clustered data points) using community detection clustering algorithm.

Cluster	**Top Percentile of Representative Networks by Time Period Buckets**				**Relative Cluster Size (%)**
	Discovery (1996-2000)	**Early Growing (2001-2008)**	**Late Growing (2009-2014)**	**Stable (2015-2020)**	
0	banana , chocolate, mango, orange		cocoa		12.7
1	beans, banana, orange	barley	barley	Cocoa	10.3
2	dairy, barley, oats, strawberry	mango	dairy, oats, poultry		9.8
3	corn	cocoa, oats	barley, oats		9.7
4		corn, beans, chocolate	corn, beans, chocolate		9.6
5		corn	corn	corn	8.8
6		corn, beans, banana	corn, beans, banana, chocolate		8.6
7			corn, beans, , banana	corn, beans, banana	7.9
8	corn, beans, chocolate		cocoa , barley	cocoa, barley	7.8
9	corn, chocolate, cocoa, oats , barley, rice, orange (1996)				7.7
10 (not grouped)					7.2

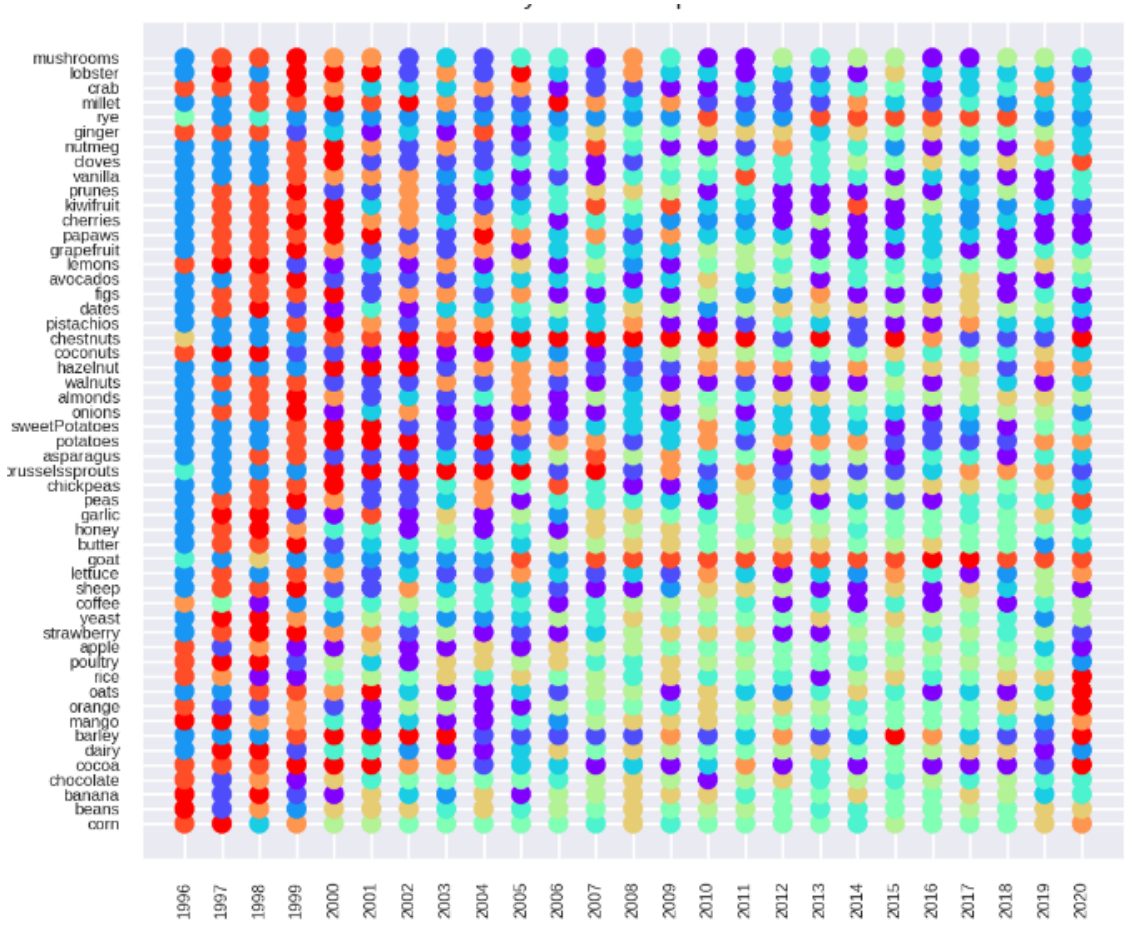
Figure 4.6: Two dimensional projection with TSNE of 128 dimensional GCV embeddings (attributed) with each color representing a different cluster from the community detection algorithm.



walnuts etc) with longer shelf lives compared to fruits (orange, mango and strawberry) and livestock (poultry, sheep) each require separate technologies for storage and transportation. As such, we assume that these separate food categories would probably not be intuitively grouped together. However, this assumption does not always reflect the reality of the groupings obtained from the community detection algorithms. We find unexpected groupings such as strawberry, apple and yeast that tend to have the same functionalities and be used together for baking and desserts. These findings highlight the importance of the methodology to group agricultural product networks using their structure to disturbances and without making assumptions about the product type or suppliers.

In addition, we see evidence of network density on group formation in figure 4.7 which shows the cluster dependencies across products and time periods that include Discovery

Figure 4.7: Scatter plot of product versus year with clusters denoted by different colors across the three periods: Discovery $(1996-2000)$, Growing $(2001-2014)$ and Stable$(2015-2020)$.



$(1996-2000)$, Growing $(2001-2014)$ and Stable$(2015-2020)$ periods. We note from the graph that several products are together in 1996 just because of topology, i.e., the differences in flow are not that relevant as the graphs were sparse.

4.5 Summary of Contributions

We present a flexible problem-agnostic framework for clustering networks based on risk and structure. The framework that allows stakeholders to group different food supply chains based on their structural similarity. We identify some distinctive network structures including those consisting of a small subset of high volume suppliers (e.g. cocoa, coffee) and also networks (e.g. chocolate) that have more distributed supply with a large number of medium volume suppliers. Insights from this method also allows users to learn from product clusters

and anticipate potential shifts which can facilitate the development of more time and cost effective policies. We find that seemingly unrelated products like apples and yeast are grouped similarly whereas tropical fruits like mango and oranges overlap in production origins and suppliers are structurally different. This reflects that our intuition about product groupings are not always the most effective.

Chapter 5

Quantifying the socio-environmental Impact of WildFire disruption to forest supply chains

5.1 Introduction

In the previous chapters, we focused on food supply chains and the impacts of disruption on global planning. In this chapter we delve into risk management when planning for forest supply chains. We specifically focus on wild fire risk as their frequency and severity have risen in recent years due to various environmental factors. Studies show that the strategic application of fuel treatments are effective at altering fire behavior and its spread patterns. Efficient planning for mitigating future expected losses under wildfire risk is a complex challenge that requires the integration of fire spread, simulation, and optimization models as well as the inclusion of multiple objectives into a unified framework.This analysis has been simplified in the past where researchers usually value the landscape regions using a unique objective (e.g., minimize the average expected area burned) . However, in reality various pieces of land have different values based on factors such as the presence of human settlements and infrastructure, availability of environmental services, and forest health, among others. We extend past research by developing an integrated framework that naturally includes and weights multiple objectives and analyzes the trade-off between present objectives and future protection against wildfire risk. We obtain treatment plans, focusing on three key regions based on their recent fire history, landscape diversity, and demographic variety and using various combinations of these layers reflecting how different priorities of the decision-makers could affect treatment policies.

It is important to highlight the changes in intensity and recurrence of wildfires and their relationship with climate change. Several studies provide clues suggesting that the ever rising global surge in the number and intensity of fires in the last decade are caused by the dehydrating effects of climate change [44, 65, 122, 1, 103]. Some of the largest and devastating

fires leading to heavy human, financial, and infrastructural losses have occurred in the last decade. A closer exploration of California reveals that despite a decrease in the total number of fires within the state, the total area burned by these fires has increased. California has experienced its most disastrous wildfires in the last ten years, [111]. The 2018 Camp Fire is California's most destructive fire recorded, where a single fire destroyed more structures than any other in modern history. This fire is also the most expensive natural disaster in the world in 2018 in terms of insured losses resulting in the loss of 13,696 and the death of over 88 people [40, 9, 47]. Following in level of destructiveness is the 2017 Tubbs Fire in the Napa and Sonoma counties, which destroyed thousands of structures leading to 46 human fatalities [36] at the Wild-land urban interface (WUI) and in large areas of forest [22] to protect human lives and maintain an adequate coexistence with nature. In [57], researchers propose a paradigm that considers opportunities in three main dimensions: i) decrease of the potential fire behavior of the landscape, ii) reduction of the potential impact of fire ignition, decreasing the expected losses and number of escape wildfires in fire-prone areas, and iii) increase the capability of fire suppression. The term Fire-Smart Forest Management (FSFM) has emerged and includes the above concepts.

In this chapter, we present a study that expands the integrated framework proposed by [90] to aid decision making under wildfire uncertainty by evaluating the sensitivity of the objective function to key environmental and economic factors. Three key regions in California (Napa Valley, Paradise, and Getty center) are chosen for analysis based on the existence of documented catastrophic events in the last five years and significant variance in vegetation types and demographic variables. We compare unweighted treatment plans against versions including environmental and demographic factors such as carbon sequestration, canopy height and density, population density, and accessibility of the area, as well as expected future fire behavior. Using multiple combinations of these layers, we generate different treatment plans that reflect how different priorities of the decision-makers could affect the treatment policies. We then analyze the trade-off between maximizing the decision-maker utility function and protecting the land against future expected losses due to wildfire with the aim of finding robust treatment plans.

5.2 Material and Methods

Data Extraction and Processing

The California region with its established seasonal fires is the focus region in our study because of the existence of documented and destructive fire events. The default framework relies on weather and topographic data obtained from various sources. In addition, we discuss how to generate the main layers that serve as weights for estimating the risk associated with each unit of landscape in our optimization module. We group the layers extracted for the objective function into Environmental factors (e.g., canopy density/height) and Demographic characteristics (e.g., population density, accessibility). Additionally, we split the data into

a training and test set. During training, we fine-tune key parameters in our model using the training data set. The test set allows us to measure the performance of our model by assessing how well our model results compare to real data.

The data consists of a combination of shapefiles and GeoTIFF files. For the most part, these layers are already aligned and have the same resolution (30 x 30 m). We transform the original data into a series of rasters to be ingested into our framework. Data enters into the framework using two main approaches: (1) local data provided by the user and (2) semi-automatic collection and processing of online assets (e.g., population density, climatic conditions that can play a fundamental role in the propagation of the fire) available in Google Earth Engine (GEE) [51], to easily generate a consolidated dataset (Fig. 5.1). This latter method relies on Python scripts and can be used as an independent module for performing any kind of query in GEE. Finally, the data is automatically processed and formatted to be ingested into the different modules of the proposed framework.

Environmental factors

The building blocks for the data layers in this category are obtained from the Landscape Fire and Resource Management Planning Tools (LANDFIRE) [102] open repository. This data hub consists of a shared program involving the participation of the U.S. Department of Agriculture Forest Service and the U.S. Department of the Interior, providing support for fire and vegetation simulators created and used by the US Forest Service. The Canopy bulk density (CBD) and the Canopy height (CH), each a grid with resolution (30x30 m) are obtained directly from the database. CBD is a measure of the density of the landscape canopy which is the portion of vegetation above ground. Canopy height as implied by its name measures the height in meters of the landscape canopy.

Forest. We combine the CBD and CH layers in various manners to obtain the different layers constituting the environmental factors category including estimating the area of green vegetation in the forest (forest health) and carbon sequestration volume. The CBD and CH layers are averaged together to form a new layer that serves as a proxy for forest health. The CBD layer indicating vegetation density and the CH measuring vegetation height provide insights about the health of the landscape vegetation.

Carbon Sequestration The carbon sequestration (CS) layer is estimated following the simple method proposed in [46] for each unit in our landscape grid as follows:

1. Estimate the total (green) weight of the trees. Using the canopy density, we obtain the canopy area by multiplying the density by the area of a unit cell. Then the canopy area is multiplied by the canopy height to obtain a volume. Once we have the volume of green in each cell, we use Huber's formula [119] to estimate the weight using the equation:

$$\text{Weight} = \text{volume} \times \text{green density}$$

where green density represents the total density of wood and bark combined (CBD).

2. According to previous research [37], the average dry weight for different temperate tree species in the United States is about 72.5% of regular weight. We use these results to approximate the average dry weight of green for each cell.

3. Next, we estimate the weight of carbon using findings from [19], which state that the carbon weight is about 50% of regular tree weight.

4. Finally, we calculate the average weight of carbon dioxide sequestered in the tree per cell using:

   a) The atomic weight of CO_2 is $C + 2 \times O = 43.999915$.

   b) The ratio of CO_2 to C is $43.999915/12.001115 = 3.6663$

Therefore, to determine the weight of carbon dioxide sequestered in the tree, we multiply the weight of carbon in the tree by a 3.6663 factor.

Fuel vegetation type. The fuel layer With a 30 by 30 m resolution that we require for our simulations is obtained from the LANDFIRE [102] publicly available repository[1]. It provides a categorical grid at a national level representing the forty Scott and Burgan fire behavior fuel models lastly updated in 2014. For a comprehensive analysis and description of all fuel types, their characteristics, and experimental parameters, see [107]. State-level data (California) is locally extracted and uploaded into GEE to consolidate it with the additional layers of the study. This is the main input for calculating the fire rate of spread in the simulation model.

Weather and moisture scenarios. The rate of Spread (ROS) mainly depends on the type of forest fuel that goes into combustion and the wind speed. However, another influencing factor is the moisture content of dead and live vegetation present in the forest. This variable is crucial for estimating the surface ROS as certain land-covers (e.g., grass types) tend to significantly modify their propagation patterns depending on their humidity levels. In order to capture and simulate interesting situations for the practitioners and researchers, multiple value thresholds are empirically studied. According to Scott & Burgan [107], these interesting humidity levels are represented by four scenarios: D1L1, D2L2, D3L3, and D4L4, from the driest (D1L1), where the fire tends to propagate faster, to the wettest (D4L4), with the opposite effect. Weather scenario files that describe the evolution of the temperature, wind speed, and wind direction are obtained from the historical time series of the closest weather station (with respect to the centroid of each instance) available for the simulated fire duration. Each fire is simulated for 12 hours under the D1L1 scenario, to be able to capture relevant propagation patterns.

[1] https://www.landfire.gov/bulk/downloadfile.php?FNAME=US_140_mosaic-US_140FBFM40_20180618.zip&TYPE=landfire

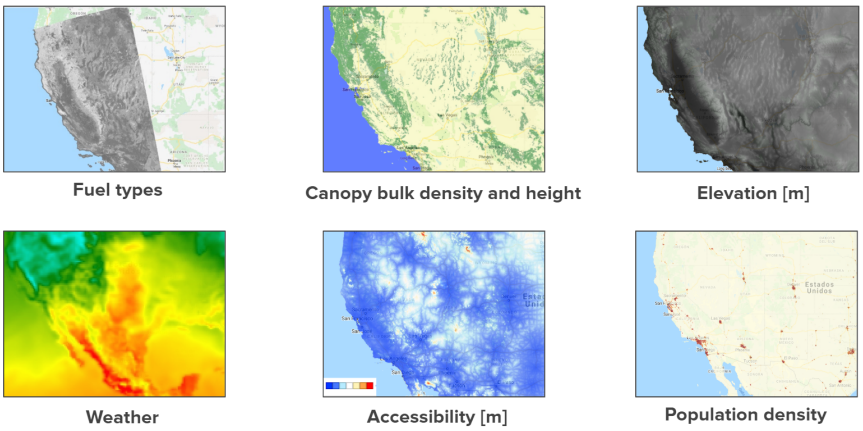


Figure 5.1: Example of layers included visualized in GEE. All layers are obtained for California and consolidated into a 30 by 30 m. multi-band raster.

Demographics. The estimated population densities (number of persons per square kilometer) for the years 2000, 2005, 2010, 2015, and 2020 are extracted from the Gridded Population of World Version 4 (GPWv4), Revision 11 dataset (https://sedac.ciesin.columbia.edu/data/collection/gpw-v4) at a resolution of 30 arc-second grid cell and averaged. Accessibility to cities, which measures the land-based travel time (minutes) to the nearest densely-populated areas with 1,500 or more inhabitants per square kilometer is obtained from the Malaria Atlas Project (https://malariaatlas.org/research-project/accessibility_to_cities/) at a 30 second-arc resolution for 2015.

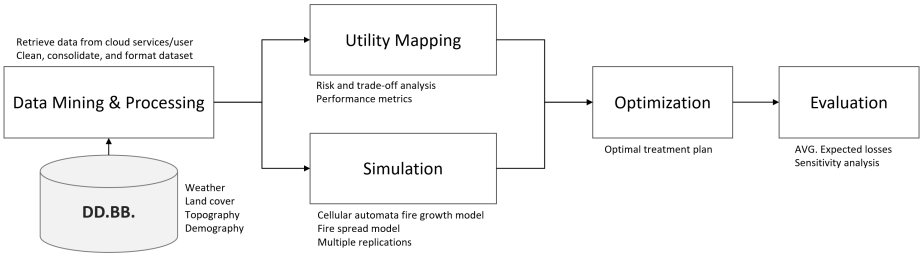


Figure 5.2: Framework schematic. Data is retrieved from cloud services and local user inputs. Decision-makers define relevant objectives by analyzing the trade-off between multiple variables. Once the data is processed, multiple simulations are performed to estimate the impact of future wildfires in the landscape. An optimization model is fed with the outputs from the utility mapping and simulation models. Finally, results are evaluated by estimating the average expected losses due to wildfire, as well as analyzing the sensitivity of the optimal treatment plan.

Wildfire simulation

A region of interest is modeled as a two-dimensional lattice with an underlying network structure to represent the connectivity between cells/nodes in the grid. Cells represent a homogeneous area with similar characteristics, focusing on points of interest such as potential propagation sources, densely populated areas, natural reservoirs, among others. Each cell from the input data corresponds to a node in an undirected grid graph $G = (V, E)$. Nodes $i \in V$ is associated with relevant attributes (e.g., population density, fuel type) included as part of the risk evaluation and further optimization models. Nodes are connected by edges $e \in E$ with weights (the distance between cells) generated by the structure of the region. In the current version of the study, each cell has at most 8 neighbors to simplify the analysis of the framework. However, the model can be easily extended to a general case (graph) where nodes can be connected with any other in the lattice as long as they share information or are related by relevant variables for the model.

Adapting the simulation framework discussed in [90, 91], the fire spreads following a messaging process between the cells of G. The intensity of these signals is represented by the rate of spread (ROS) obtained from an independent fire spread model (BEHAVE) integrated into the simulator [21] that models the fire behavior for static conditions in the U.S. based on empirical studies. It is used to update the fire progress between neighboring cells at every time-step t. From here, a directed tree graph – denoted as *Propagation Tree* – is obtained, with a root at the ignition node i, $T_i = (V_i, E_i)$ where $V_i \subseteq V$ contains the burned cells and E_i the directed edges representing the propagation trajectories of the fire within the region.

Simulator parameters may need calibration in order to reproduce realistic propagation patterns or capture the impact of previously unseen conditions. Multiple approximations during the implementation of the spread models and inherent noise within the data may inaccurately represent the expansion of the fire. To account for this situation, we propose an automatic adjustment following the work in [24] of the main parameters of the model via a hybrid AI-Optimization procedure that aims to minimize the differences between the simulated and historical fire scars of the region of interest. Using this approach, we can automatically adjust the ROS estimations to accurately represent observed fire perimeters. Therefore, decision-makers are able to automatically adjust the fire spread model to account for variations in the fire behavior and conditions of the area, obtaining more accurate simulation results.

Risk analysis

Relevant features are mapped onto a common scale and weighted to account for their relative importance for the decision-maker. In order to condense their information into a single utility function, each feature is mapped unto a common scale $[0, 1]$ using an adequate function (e.g., linear) according to the expected impact of the feature in the landscape. Thus creating a single matrix representing the original $n \times m$ grid containing the value for each node in

landscape ($NV \in \mathbb{R}^{n \times m}$), different convex combinations of the relevant values-at-risk (raster layers) are utilized. Let $\mu_k \in [0,1]\ \forall k \in K, \sum_{k \in K} \mu_k = 1$ with K being the total number of layers included in the study, we thus combine the layers as follows:

$$NV = \sum_{k \in K} \mu_k L_k \tag{5.1}$$

where $L_k \in \mathbb{R}^{n \times m}$ is the matrix of dimensions $n \times m$ containing the grid values of layer k. Following this framework, we obtain a consistent risk function across the entire graph. The detailed procedure is as follows:

1. Given a set of K features representing characteristics of the nodes, we map them into a common scale between $[0,1]$. The mapping function (e.g., linear) from raw feature values to the $[0,1]$ interval is selected by the researcher according to the impact of each feature in the construction of a global utility/cost function. For example, if the protection of nodes with higher population density is prioritized, an increasing non-linear function can be applied, where densely-populated nodes are associated with values near to one, while sparsely populated areas are mapped to near-zero values (see Fig. 5.3).

2. This procedure is repeated for all K features, obtaining a set of normalized variables.

3. Correlated and complementary variables are combined into meaningful categories by weighting individual features with weights. As an example, canopy bulk density and canopy height could be summarized into a *Forest* category that equally weights both variables. Similarly, accessibility and population density layers could be condensed into a *Demography* category.

4. Once all categories are generated and normalized, a global utility function is calculated repeating the weighting procedure.

5. Gaussian kernels are applied to smooth the distribution over the landscape/grid. This avoids abrupt changes in the utility function as well as accounts for the intrinsic correlation of the cells in the landscape dynamics (Fig. 5.6).

Using this framework, decision-makers are able to condense any number of features into unique values associated with each node of the graph (NV), as well as generate a series of scenarios to evaluate the impact of certain features and their weights. In this way, each node of the grid is characterized by a unique value incorporating all variables. Using different weight combinations allow us to control which factor we want to weight more in each instance (see Fig. 5.5). Next, we utilize the downstream-protection-value (DPV), a risk model that aims to measure the value of what is affected downstream in the network given the risk of a certain node [91]. This model captures the role of a node in propagating the risk through

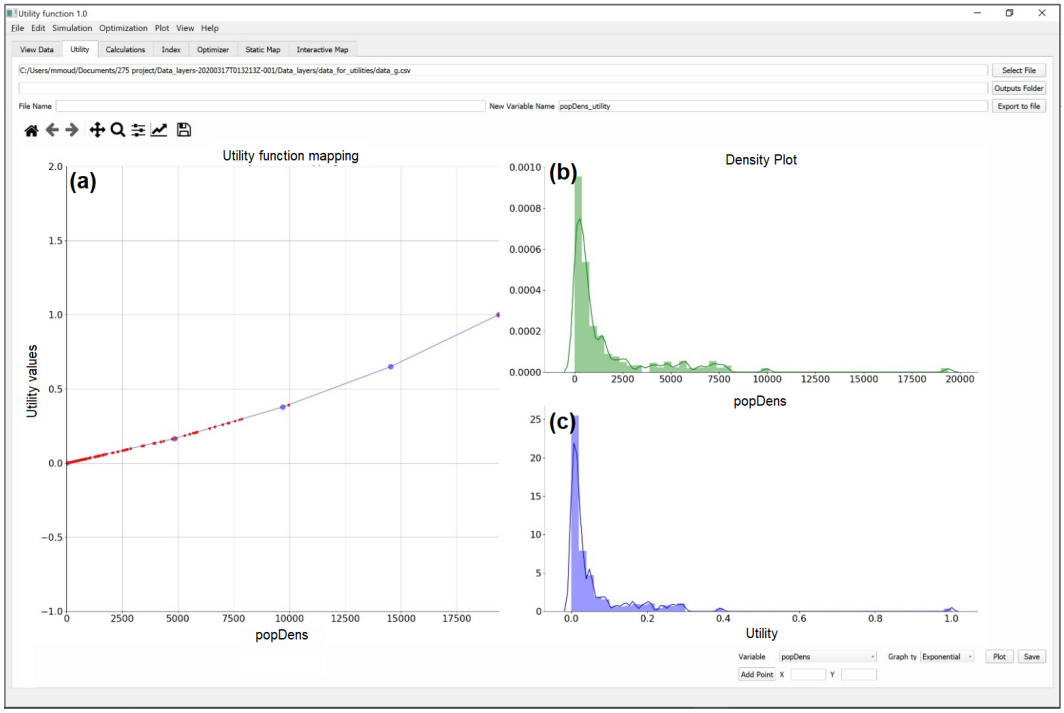


Figure 5.3: Utility mapper application. (a) The original values of the population density layer (x-axis, popDens) are mapped to the [0,1] interval following an exponential function (y-axis). (b) Density plot for the original variable values. (c) Distribution of the mapped [0,1] feature (called utility) following the applied transformation.

the landscape based on the connectivity and relative influence of different nodes in the entire system. We define the DPV of node i inside the landscape network as:

$$DPV\,(i) = \alpha_i \sum_{j \in S_i} NV_j \tag{5.2}$$

where α_i is a weight factor for node i, e.g., the number of connections inside the region, allowing the prioritization and management of different zones; S_i is the set of nodes of the network that are affected by fire propagation from node i, and NV_j the value of node j incorporating all the relevant variables included to capture the potential losses caused by wildfire. Researchers can represent and evaluate multiple scenarios by providing different weights to the components of the risk function or to the final node value, thus obtaining variations of DPV. For example, NV values of nodes playing a fundamental role in the connectivity of the network could be weighted, among several options, by the degree of the node to highlight its importance in the propagation of fire. In addition, thanks to the nodes' additive property, there exists a natural extension from nodes to larger units (e.g., stands). This goes in hand with a practical implementation since authorities take decisions over certain

areas with common characteristics instead of individual units, satisfying operational/logistic constraints.

An optimized resource allocation plan

Formulating an explicit optimization model (See Appendix, Section 4.1), we solve the problem of finding the connected cells that maximize the total utility/protection value considered under wildfire risk, subject to specific constraints provided by the decision-maker (e.g., budget). The connectivity constraints are imposed to mimic realistic scenarios where the protection of the landscape must be performed within connected patches to account for operational constraints. If needed, these constraints can easily be relaxed by the decision-maker, running a simplified version of the optimization model, or even implement his/her own algorithms. In this way, we obtain an optimized treatment plan that identifies the set of units that, once protected, will significantly disrupt the fire propagation while taking into account the impact on relevant features of each node via the provided NV values.

We introduce $\lambda \in [0, 1]$ as the trade-off factor between the utility function defined by the NV matrix and the DPV heatmap obtained after simulating R wildfires in the area. We define the objective function as the convex combination $U(\lambda) := \lambda DPV + (1-\lambda)NV$, where larger values of λ indicate that wildfire risk becomes more relevant to the decision-maker when allocating resources through the network ($\lambda \to 1$) while smaller values represent the case when the NV values play a more relevant role for prioritizing the treatments ($\lambda \to 0$). Therefore, different optimized resource allocation plans are obtained depending on the objective and risk assessment of the decision-maker represented by the λ parameter, balancing the maximization of the utility function and the protection of the region against the future propagation events (See Appendix, Fig. D.1).

Solving scheme

Although the problem is NP-hard, being a variant of the maximum-weight connected subgraph problem (MWCSP), a known NP-hard combinatorial problem [63], we are able to obtain optimal solutions in all the experiments using a two-stage solving approach. First, we provide a warm-start to the exact MIP model obtained from a greedy algorithm that, given an initial node, selects the adjacent nodes to the current solution with maximum values after sorting them by their $U_i(\lambda)$ until the maximum number of nodes (Q) is satisfied. Once the initial feasible solution is provided, we solve the optimization problem with CPLEX v12.9 solver using its default configuration. The two-step solution method approach enables the problem to converge faster because we start the second step with a high-quality feasible solution. This reduces the searching space of the method significantly (upper and lower bounds), improving the convergence and memory usage of the optimization algorithm. Thus, we obtain optimal solutions in reasonable (less than 1 minute on average) solving times in all our experiments. Our solution is guaranteed to be globally optimal because we obtain

0% optimality gap solutions from the exact MIP formulation. Although multiple solutions can exist, these are extremely rare in practice.

Case study areas

The areas in California selected for the case study are chosen based on factors including significant variation across fuel types, the existence of documented catastrophic fires in the last five years, and the presence of strongly influencing demographic factors. The three areas selected for this study are the Getty center, Napa Valley, and Paradise. Figure 5.4 denotes the terrain of the three case studies areas. In addition, a summary of the key features of our three study instances including area, mean elevation, dominating fuel types, and elevation range are presented in Table 5.1.

Napa valley.
Napa County, also known as Napa valley is recognized worldwide as a premium wine region. In addition, the region is also responsible for the production of many agricultural crops. According to the U.S. Census Bureau, the county has a total land area of about 748 square miles and a population of 137,744 as of 2019 [26]. However, this region has also suffered a number of destructive fire events. A recent article mapping all fires in the region from 1950 to 2019 shows that the fires have been getting larger and more destructive [71]. Most notably the 2017 "Tubbs" fire in the Napa and Sonoma counties is the second most destructive fire recorded in California's history [59]. The long fire history experienced by this region coupled with its landscape being suitable for the farming of different agricultural products particularly grapes from vineyards makes it a very relevant and interesting area to study in California. The total number of nodes and edges conforming this instance are 9,309 and 72,272, respectively.

Paradise.
Paradise is a small town located in Butte County in California that has experienced the most destructive fire in California's history at the end of 2018. In 2018, the population of Paradise was about 26,800 with a land area of about 18.31 square miles [26, 59]. However, the population numbers after the fire event are unknown as over 9300 were displaced and relocated during and after the fire [72]. We use population numbers before the fire as an estimate for our analysis. This town is a unique region to study because of its high fire risk, limited accessibility to, and relative isolation from neighboring towns. With a total number of 11,477 nodes and 81,581 edges, it is the largest instance of the study.

Getty center. The Getty center, a campus of the Getty Museum and its surrounding regions located in Los Angeles is selected due to its 2019 fire which forced the museum to temporarily close its doors [25]. This region is different from the other regions of interest because it is located in an area with a more built-up environment and less vegetation. We

speculate that the demographic factors will have a stronger influence on our analysis of this region. Modeling the instance as a network, it consists of 5,454 nodes connected by 41,166 edges.

Table 5.1: Summary of instances' main characteristics. For each instance, we provide the total area in hectares, the average elevation and its range in meters, the dominant flammable fuel of the terrain following the fuel type layer characterization, the total number of different fuels available in the region, and the number of edges conforming the network used for the optimization model connecting the flammable cells (in any direction).

Instance	**Area [ha.]**	**Mean elev. [m]**	**Elev. range [m]**	**Dominant flammable fuel**	**# Fuel types**	**# Edges**
Napa Valley	9,540	376.74	[131, 724]	SH2 is woody shrubs and shrub litter	16	72,272
Getty center	11,102	221.39	[39, 596]	GS2 is grass and shrubs combined	11	41,166
Paradise	13,433	449.71	[0, 740]	TL6 is moderate load broadleaf litter	17	81,581

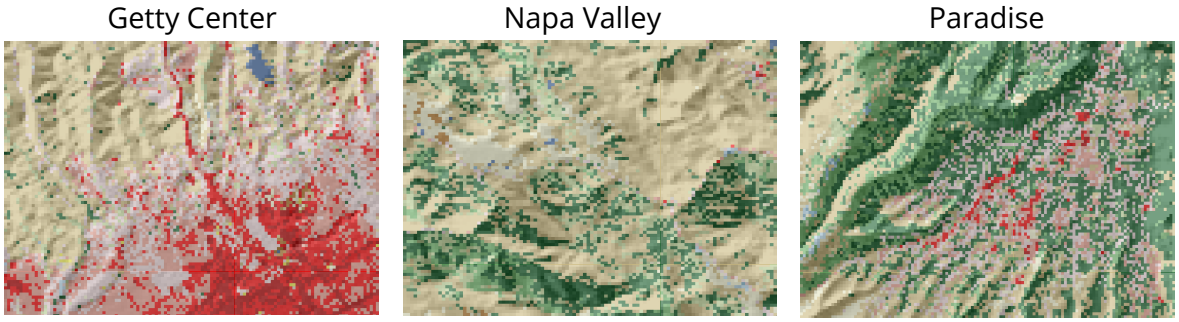


Figure 5.4: Land cover representations. The three case studies areas are depicted with a hill shade effect where different colors represent the fuel types characterizing the instances following the Scott & Burgan [107] classification system.

Utility Mapping. For each of the three instances we study, we utilize four layers in our utility function including forest, volume of carbon sequestered, accessibility, and population density layers. We select mapping functions to highlight the priorities of the decision-maker for every layer. These functions serve to map the values of our current layers to a range between 0 and 1 using a pre-specified distribution. We Choose an exponential mapping function for the population density layer because we want an increasingly high utility as the number of people in a region increases. In addition, we use the inverse function to map our accessibility values to the range between 0 and 1. This function is selected to reflect increased utility for regions that are closest to densely populated areas. Finally, a linear function is used to map the volume of carbon sequestered and forest layers to the

appropriate ranges. Our framework allows provides the option to change the distribution of the mapping functions and customize these functions based on the decision maker's most important considerations. In order to combine the different features into a single utility, we select five convex combinations of our different layers for our experiments. First, we use an average combination where all four layers are weighted equally. Then, we create four feature dominant layers, where the selected dominant feature contributes 75% to the final utility whereas the remaining three layers equally split the remaining. These five utility combinations are created for each of our three instances (Napa valley, Paradise, and Getty center).

Experiments

We model five different NV values for each instance by modifying the μ weights vector associated with the different four categories following the procedure described in Section 5.2. In this way, we obtain a balanced weight function $NV_{equal} = \sum_{k \in K} 0.25 L_k$ where all layers are weighted identically and four variations where a dominant layer is weighted by $\mu_{dom} = 0.7$ and $\mu_j = 0.1$ for $j \neq dom \in K$ obtaining $NV_{forest}, NV_{access}, NV_{carbon}$, and $NV_{population}$.

A total of $R = 100$ replications with random ignitions and defined D1L1 weather scenarios are performed in the simulation module to obtain the final DPV heatmaps using the previously generated NV values as the node weights, multiplied by $\alpha_i =$ number of neighbors connected to node i. For each instance and NV value combinations, we generate utilities $U(\lambda)$ with $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$. Each combination is then solved in the optimization module for all treatment levels $tf \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.5\}$, solving a total of 450 optimization problems. Optimal solutions $X^*(\lambda, tf)$ indicating the selected cells are recorded for each combination.

Finally, we evaluate the average expected losses, $\mathbb{E}\left[Losses(X^*(\lambda, tf))\right]$, due to future wildfire events in each instance, given the output of the resource allocation plan. For this, we estimate the expected damage provoked by future fires – discounted by a $\gamma \in (0, 1]$ factor set to $\gamma = 0.9$ – using $R = 100$ simulations in a modified landscape where the selected cells from the optimal solution of the $ORAP_\lambda(tf)$ model are transformed into non-flammable ones. We use this as a simplified version of a more realistic setting where certain fuel treatment actions (or resource allocation) will decrease the fire susceptibility (or effective ROS) of a certain cell but it will be still flammable. The pseudo-code summarizing all the experimental steps can be found in the Appendix (see Fig. 3).

Computational implementation

The Cell2Fire fire-growth simulator is implemented in C++ using the boost and omp libraries [106, 113] to allow shared memory parallel execution. The decision support system which wraps the simulation module and processes all relevant outputs is programmed in Python. Statistics and visualizations are processed using the known Pandas, Numpy, and Seaborn libraries. Network structures are managed with the networkx package [53], generat-

ing outputs such as propagation trees and providing the users a variety of complex network metrics out-of-the-box such as betweenness centrality for developing their own fire risk indexes. we implement derivative-free optimization algorithms [35] in our framework using the NLOPT package [64]. Meta-Heuristics (e.g., genetic algorithms) follow the DEAP [45] library framework. In addition, we rely on the PYOMO modeling language [54], a flexible package to generate linear/non-linear models and solve them via an open-source or commercial solver depending on the user needs to embed mathematical programming models in the framework's optimization module. The utility mapper standalone application and scripts are programmed in Python using the PyQt5 package and compiled using the Pyinstaller package.

Experiments are performed in a daily use laptop with I7-4200 2.1 GHz processor, RAM Memory 8 GB DDR3, and Ubuntu 14.0 OS. All codes are available for public use at http://www.github.com/cpaismz89/DPV_Utility.

5.3 Results and Discussion

The utility functions in our experiments are obtained by combining raster layers as described in Section 5.2 and the different utility combinations are explained in 5.2. We conceive these different combinations to mimic decision-makers having multiple objectives, where one of the goals is more important than the remaining ones in the decision-making process. In Figure 5.5, we present the utility maps for all proposed combinations of the primary layers for the three instances we explored. We observe that the combinations produce very different heatmaps which we hypothesize will be translated into different optimal treatment decisions.

Utility and wildfire risk trade-off

DPV values are obtained for all the generated utility heatmaps following Eq. (5.2) with NV_j representing the value of cell $j \in V$ from the calculated utility layer. Looking at the DPV matrices (Fig. 5.6), we clearly observe the most likely wildfire propagation patterns after performing $R = 100$ replications for each instance (first column) where lighter colors highlight those sections of the landscape with higher DPV, i.e., the nodes that play a fundamental role in propagating the fire between different areas of the land. As expected, the DPV matrix obtained for the Napa Valley instance covers a significant portion of the landscape (88.63%) since it is the one with the largest proportion of flammable fuel types, representing 97.5% of its total composition. On the other hand, Paradise and Getty center instances include a significant proportion of non-flammable nodes representing urban settlements/rocky areas (14.56% and 50.87%, respectively) leading to more focused DPV heatmaps surrounding those areas, covering 40.3% and 37.27% of their total size, respectively.

The expected area burned and expected utility losses under the current conditions, i.e., when no resource allocation plan is implemented, can be seen in Table 5.2. From the experi-

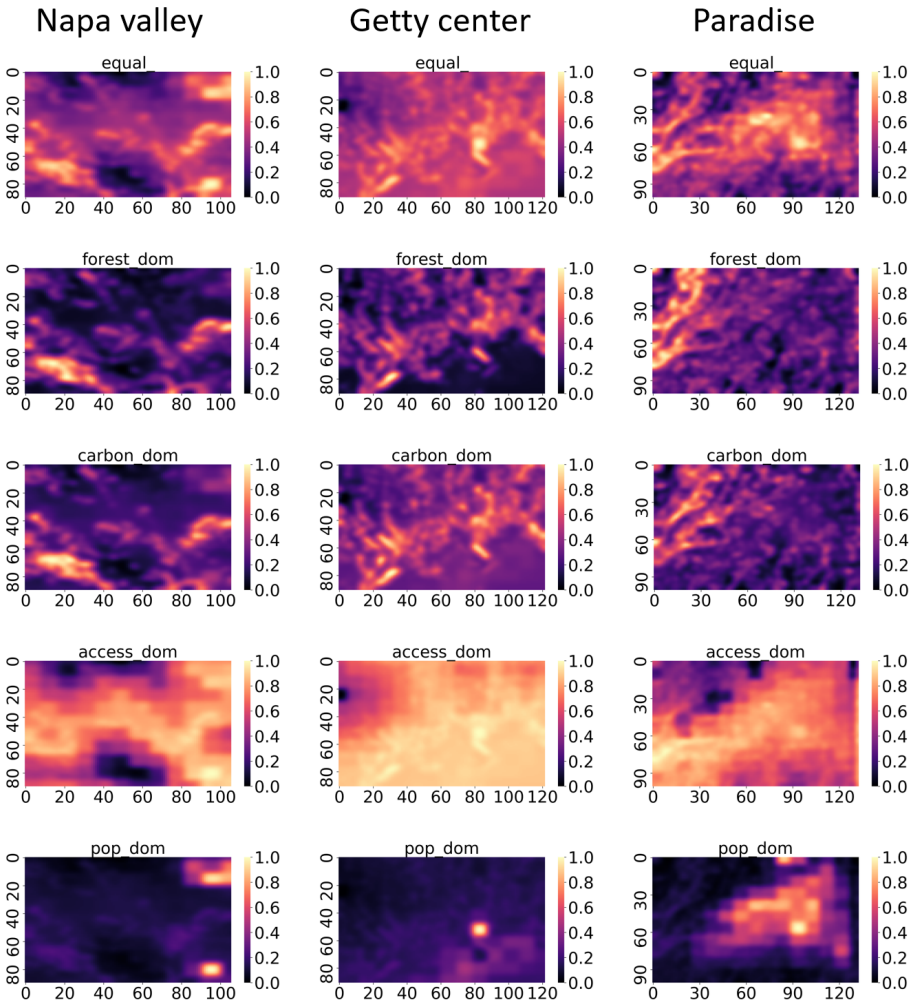


Figure 5.5: Utility heatmaps for all proposed convex combinations of the main four categories for each instance (columns). The first row represents a balanced combination of all four categories ($\mu_i = 0.25\ \forall i$). The suffix *dom* indicates that the dominant category was weighted by $\mu_{dom} = 0.7$ and the remaining three categories with $\mu_j = 0.1$, $\forall j \neq dom$.

ments, we observe a significant impact on the Napa instance with an expected area burned of 2,055 ha. representing 22% of the landscape. This is translated into expected losses close to 20% among all utility functions with respect to the total value available. In the case of Getty center, we expect a 19% of the area burned due to future wildfire events, with an impact on the utility functions varying from 14.32% (Forest dominant utility) to 16.18% (Access dominant utility). Finally, due to the characteristics of the Paradise instance – the distribution of the non-flammable fuel types – a 7% of the total area of the landscape is expected to be affected by future wildfire events under the tested conditions, leading to expected losses

between 3.91% (Population layer) to 7.58% (Forest layer).

Table 5.2: The expected area burned and expected losses for all utility functions as a percentage of the total instance area and the total utility available (heatmaps) per instance, respectively. Expected values are calculated from $R = 100$ independent wildfire replications, weighting all simulations equally, and without any intervention of the landscape.

Instance	$\mathbb{E}[Burned]$ %	$\mathbb{E}[L(U_{Equal})]$ %	$\mathbb{E}[L(U_{Forest})]$ %	$\mathbb{E}[L(U_{Carbon})]$ %	$\mathbb{E}[L(U_{Pop})]$ %	$\mathbb{E}[L(U_{Access})]$ %
Napa Valley	22.08%	20.73%	19.66%	20.06%	18.95%	21.75%
Getty center	18.97%	15.27%	14.32%	14.66%	14.43%	16.18%
Paradise	6.94%	6.33%	7.58%	6.83%	3.91%	6.94%

Analyzing the most common fuel types involved in the propagation patterns identified when calculating the DPV, we observe grass and shrubs combined (GS2, 36.84%), grass though small amounts of fine dead fuel (GR2, 27.18%), and woody shrubs and shrub litter (SH7, 20.40%) for Napa Valley; woody shrubs and shrub litter (SH7, 50.73%) and grass and shrubs combined (GS2, 46.14%) in the area near Getty center; and woody shrubs and shrub litter (SH7-SH5, 62.49%) and grass and shrubs combined (GS2, 23.14%) for the Paradise instance. This information allows the decision-maker to gather relevant insights about the most dangerous sections of the landscape in terms of wildfire risk and identify the set of potential actions to mitigate their impact when implementing the solution obtained from the optimal resource allocation plan.

Resource allocation plans

As described in Section 5.2, we analyze the trade-off between the expected losses due to future wildfires in the area after applying the optimal resource allocation plan – assuming full protection of the selected nodes for simplicity – and the protected value due to the implementation of this plan by combining the utility layer generated by the decision-maker and the DPV matrix obtained from the simulations via the λ parameter. Depending on the expectations of the decision-maker and his/her level of risk aversion, different λ values should be tested and selected for a particular region and context. In the rest of this section, we will focus our attention on three interesting results where the trade-off between present value and future protection plays a crucial role in the decision-making process. This analysis can be performed for all combinations of utilities, treatment fractions, and instances, providing the decision-maker with a comprehensive set of results (see Tables D.1, D.2, and D.3 in Appendix) and quantitative support for establishing the optimal point to balance the trade-off between present utility and future protection of the landscape.

To illustrate and analyze the inherent trade-off between present utility and the future protection of the landscape, we observe the results for the Napa valley instance when identical weights are provided for all categories conforming the utility layer (Fig. 5.7-(a)) and a 25%

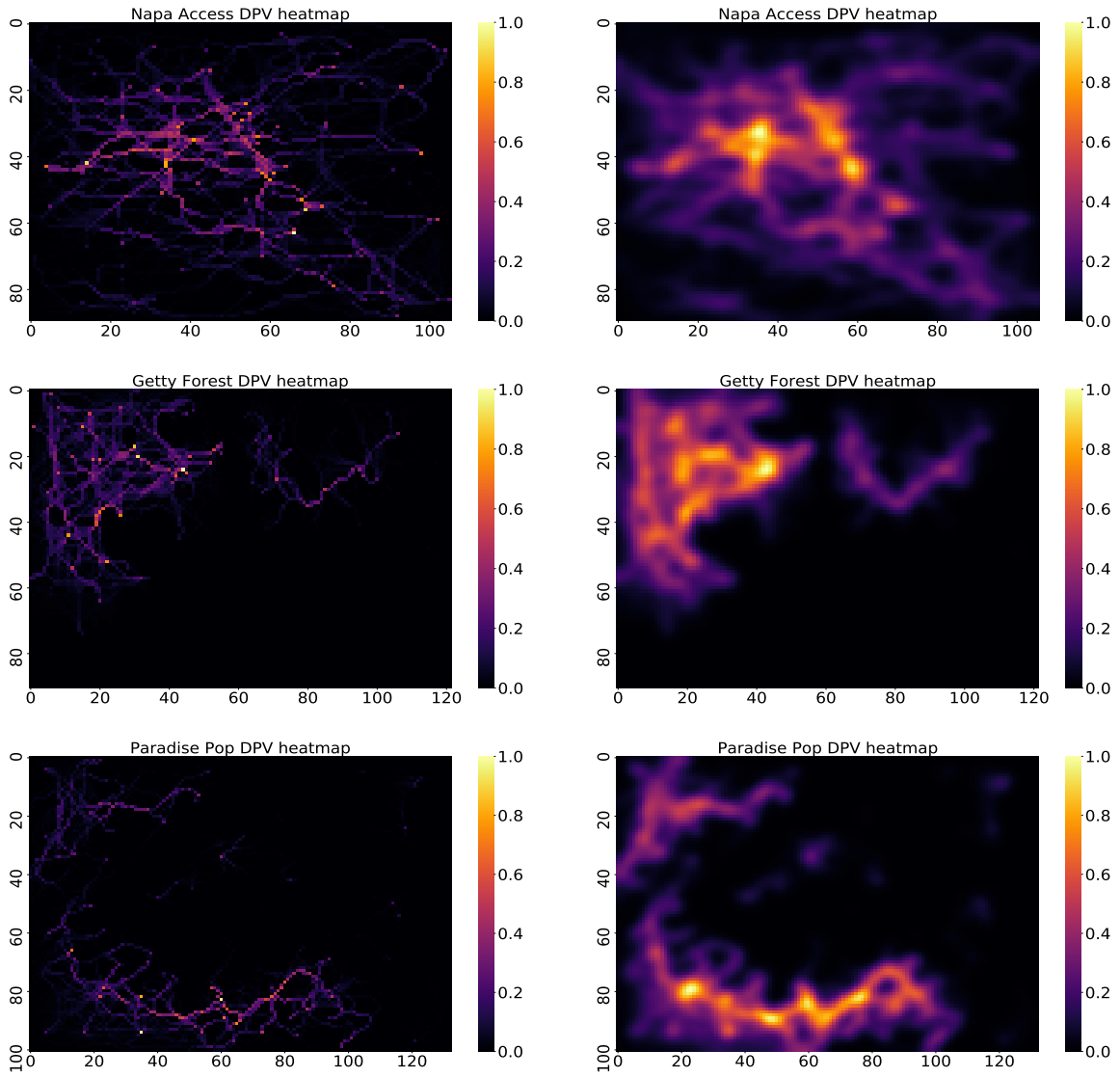


Figure 5.6: Raw (left) and smoothed (using a Gaussian kernel - right) DPV heatmaps calculated for Napa valley, Getty center, and Paradise instances using the Access, Forest, and Population density layers as the dominant layers for the NV function, respectively. Lighter cells increasingly highlight the nodes playing a fundamental role in propagating the fire to the rest of the landscape.

of the landscape is protected. As expected, the present/raw utility tends to decrease as $\lambda \to 1$ since the decision-maker is sacrificing present utility by focusing more resources in those locations where the fire will likely propagate, prioritizing the disruption of future wildfire events instead of the original utility function. However, present utility values are

overestimated when $\lambda \to 0$ as the decision-maker oversees the impact of future wildfires, not accounting for this risk. From the graph, we can observe that the best performance in terms of discounted utility is attained when $\lambda = 0.5$ (1017.39 ± 186.06), obtaining the best balance between raw utility and expected wildfire losses using a discount factor of $\gamma = 0.9$.

Analyzing the gaps between the present utility and the discounted function, we observe a clear decreasing pattern as $\lambda \to 1$ in terms of total utility value variations[2], with differences of 38.23%, 30.6%, 23.65%, 15.75%, and 7.41 %, respectively. This is aligned with our expectations: as λ is increased, the lands are better prepared for future wildfires, minimizing the gap between today's total utility and the discounted function by sacrificing present value. Looking at the distribution of the discounted utility as a function of λ (Fig. 5.7-(b)), we observe that increasing the weight of the expected wildfire risk ($\lambda \to 1$) results in a more compact distribution of the discounted utility $\Delta U(\lambda)$, as seen in the graph. This is consistent with the fact that higher λ values lead to better protection plans by sacrificing the value of the utility function, a trade-off that the decision-maker will analyze to decide which resource allocation plan is aligned with her expectations and goals.

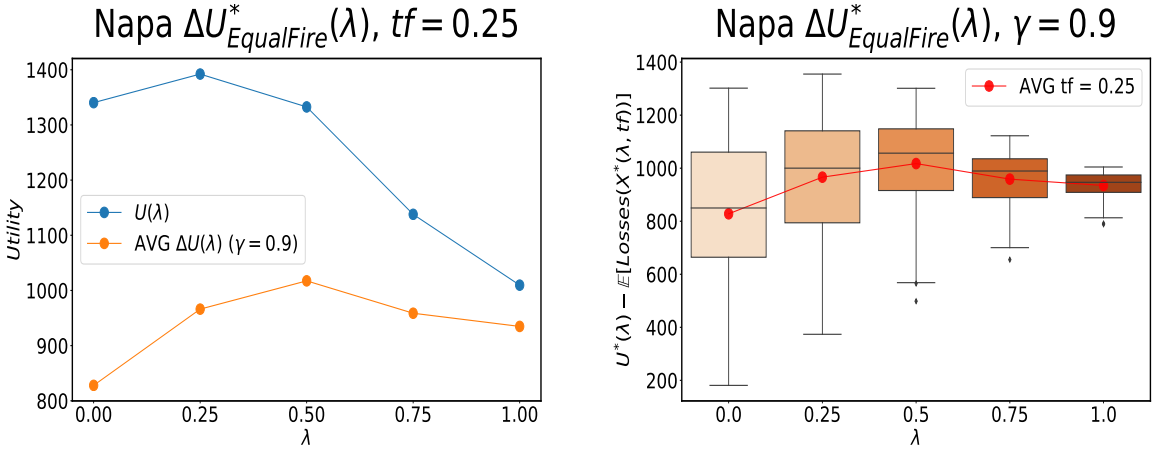


Figure 5.7: (a) Napa valley instance raw utility (blue) $U(\lambda)$ weighting all categories by identical weights and average discounted utility (orange) $\Delta_{tf}U(\lambda) = U^*(\lambda) - \mathbb{E}\left[Losses(X^*(\lambda, tf))\right]$ including future expected losses due to wildfire events as a function of λ. Treatment fraction is set to 25%. (b) Distribution of the optimal utility discounted by future expected wildfire losses ($\gamma = 0.9$) for different λ levels when protecting 25% of the landscape. Average values are highlighted with red dots.

We then focus our attention on the results obtained for Paradise with a carbon dominated utility when treating 50% of the total landscape. Visualizing the optimal plans for multiple λ values (Appendix, Fig. D.1), we observe that the resource allocation plan is not as sensitive as in the case of Napa Valley, but we can still observe differences across the different levels of

[2]Calculated as $\mathbb{E}\left[Losses(X^*(\lambda, tf))\right] / U^*(\lambda, tf)$

λ to analyze the trade-off between present value and the protection of the landscape. This is mainly associated with three factors: (1) we are treating 50% of the land so there exists a larger overlap between the optimal plan and the propagation patterns identified in the DPV matrix, (2) the distribution of the carbon dominated utility matches the most relevant DPV spread lines, and (3) the instance, similar to the results from Getty center, presents a significant amount of non-flammable nodes, thus limiting the potential fire spread paths.

This is translated into significantly smaller gaps between the present/raw utility value and the discounted one for all λ levels (3.15%, 2.45%, 0.86%, 0.4%, and 0.04%, respectively). As seen in Fig. 5.8-(a), both functions converge to an almost identical value for $\lambda \geq 0.5$. This situation indicates that optimal plans giving at least 50% of weight to the DPV layer are able to significantly control and mitigate the future expected losses due to wildfire events. Observing the distribution of the discounted utilities in Fig. 5.8-(b), it can be seen that the value and variation in the discounted utility (y-axis) are significantly affected by λ, obtaining different levels of risk. From the results, the plan obtained when $\lambda = 0.5$ arises as a good solution, balancing the raw utility and expected losses as well as being characterized by a compact distribution (2250.25 ± 17.58). Values of $\lambda < 0.5$ lead to greater average discounted utility (red dots) but incur in more risk and potential negative outcomes (e.g., low discounted utility value outliers in $\lambda = 0.25$). On the other hand, values with $\lambda > 0.5$ sacrifice a portion of present utility to decrease the impact of expected future losses (5.1% and 14.5% w.r.t. $\lambda = 0.5$, respectively).

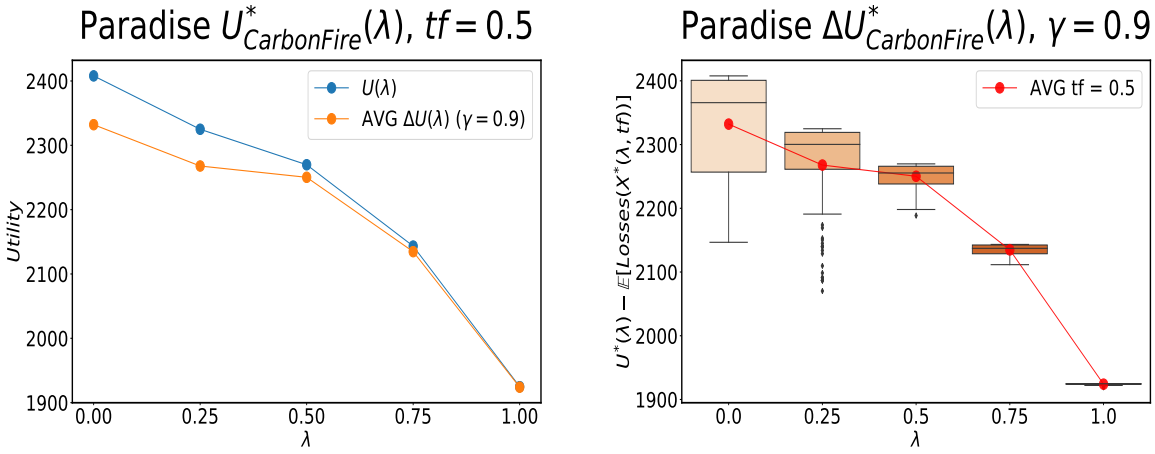


Figure 5.8: (a) Paradise instance raw utility (blue) $U(\lambda)$ with carbon as the dominant category and average discounted utility (orange) $\Delta_{tf}U(\lambda) = U^*(\lambda) - \mathbb{E}\left[Losses(X^*(\lambda, tf))\right]$ including future expected losses due to wildfire events as a function of λ. Treatment fraction is set to 50%. (b) Distribution of the optimal utility discounted by future expected wildfire losses ($\gamma = 0.9$) for different λ levels when protecting 50% of the landscape.

Finally, we analyze the results obtained for the Getty center instance, when the utility

function is dominated by the accessibility layer and only 15% of the landscape receives treatment. Contrary to the previous results, we observe an increasing pattern in the discounted utility (Fig. 5.9-(a)) as more weight is provided to the DPV matrix, this is, the optimal plan is mainly aligned with the mitigation of future wildfire losses instead of the current benefit ($\lambda \to 1$). These results indicate that, for this particular experiment, focusing the attention only on the present objective function value for selecting the nodes to be treated/protected is not the most efficient solution as there is no significant intersection between the treatment plan and the most likely propagation patterns experienced in the landscape, leading to larger expected losses as $\lambda \to 0$. Analyzing the gap between both curves, we observe differences of 148.48%, 119.77%, 74.86%, 49.17%, and 31.64% as we increase the value of λ, respectively. Therefore, $\lambda = 1$ arises as a robust option under the current experimental parameters.

Following the discussion, we observe how the distributions of the discounted utilities (Fig. 5.9-(b)) with $\lambda < 0.5$ are particularly wide with a significant bias to the bottom (-205.64 ± 434.12 and -75.1 ± 364.49 respectively). Even more, we can see that both distributions reach negative values because of the larger expected losses, indicating higher risk involved in those treatment plans as they do not prepare the landscape to disrupt future wildfire events. Values of $\lambda \geq 0.5$ lead to more compact distributions – still with negative results – by sacrificing valuable outcomes but assuring the future protection of relevant flammable areas detected by the DPV layer, mitigating the expected losses. This is reflected in the distribution of the discounted utility when $\lambda = 1$, with an expected discounted utility of 211.05 an a standard deviation of 103.45, contrasting the results above with $\lambda \in \{0, 0.25\}$.

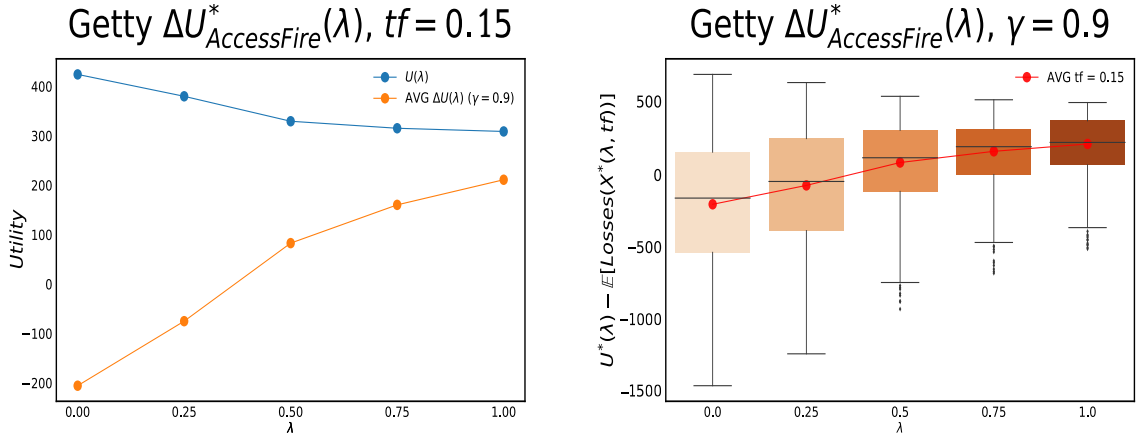


Figure 5.9: (a) Getty center instance raw utility (blue) $U(\lambda)$ with accessibility as the dominant category and average discounted utility (orange) $\Delta_{tf}U(\lambda) = U^*(\lambda) - \mathbb{E}\left[Losses(X^*(\lambda, tf))\right]$ including future expected losses due to wildfire events as a function of λ. Treatment fraction is set to 15%. (b) Distribution of the optimal utility discounted by future expected wildfire losses ($\gamma = 0.9$) for different λ levels when protecting 15% of landscape.

5.4 Conclusions

Decision-makers usually face multiple objectives when making decisions intersecting landscape and fire management. In our original approach, the different cells within the landscape were given equal importance. We extend these previous analyses to include multiple objectives into the optimization model and analyze the trade-off between current weighted objectives and future protection against wildfire risk. Experiments were performed on three distinct regions using multiple utility functions obtained from the combination of demographic and environmental raster layers, highlighting different priorities of the decision-maker.

The study areas are selected to illustrate the framework when dealing with different fuel compositions and utility distributions. These differences are reflected in the Napa instance with much higher proportions of flammable fuel types having a more expanded DPV matrix compared to the other two instances presenting area-focused DPV because they are composed of higher fractions of non-flammable landscape types. The DPV matrix provides insight into high-risk areas within the landscape and informs actions to reduce expected wildfire impacts. Regions with more compact matrices could suggest that fewer resources are needed to mitigate the effects of future wildfires in these areas. As conjectured from the calculated DPV, the expected area burned and expected losses for all utility functions as a percentage of the total instance area and the total utility available (heatmaps) was highest in the Napa instance. In addition, we observe that the fuel types commonly involved in fire propagation patterns consist mainly of some combinations of grass, woody shrubs, and shrub litter across all three instances.

In our analysis, the trade-off between present utility and the future protection of the landscape based on the DPV is evaluated. As expected, in general, the present utility tends to decrease as λ value rises as the decision-maker increasingly focuses more resources on areas with higher fire spread risk. However, present utility values are overestimated when λ drops, as the decision-maker emphasizes more the impact on the current utility function and less the future wildfire risk. Despite these common trends, certain utility layers present different patterns depending on the instance. For example, we can find that global maximum ΔU values can be obtained with $\lambda = 0$ (e.g., the forest utility layer in Napa) when the most likely propagation patterns are not associated with the highest values of the utility heatmap, reducing the risk at a very high cost of present utility. In other cases, a balanced $\lambda = 0.5$ accounts for a good trade-off between minimizing expected losses and maximizing the present utility as in the Napa U_{Equal} scenario. In addition, some scenarios including certain combinations of utility layers and instances characteristics are harder to balance with future expected losses, mainly due to the fragmentation of the instance and the distribution of the utility layer over the landscape, which can sometimes be too focused on a single point or significantly sparse, focusing the treatment plan on areas that are not likely to match the most relevant propagation patterns of the expected fires.

The results presented in this work are conditional upon the occurrence of fires within specific time frames, meaning that the evaluation of the effectiveness of the fuel treatment plans assumes that new wildfires occur during the period of time in which treatments are

highly effective. This may not be the case in practice, where the temporal dimension (when to apply the treatment) is crucial to implement effective plans.

Chapter 6

Conclusion

Overall, the work presented in this dissertation has followed a data-driven methodology for quantifying and mitigating risk focusing on food and forestry supply chains. The was done by first developing a stochastic global allocation model to understand how to effectively allocate food resources on the global scale. This model allows stakeholders to incorporate this risk when analyzing potential solutions for decision making. In the future, several extensions of this framework can be explored. A risk-averse version can be formulated by explicitly including the risk of experiencing "bad" scenarios (i.e., with a strong impact on the performance of the network). For this, risk-aware stochastic formulations including terms such as the known conditional value-at-risk (CVAR, [101]) will be explored in future iterations of the framework. In addition, the framework can be naturally extended to incorporate multiple products simultaneously, analyzing the performance of the whole agricultural supply chain instead of focusing on one particular commodity. More so, we would like to better model and include factors that influence trade relationships between countries into the cost function and/or incentives between edges. Some possible implementations include incorporating quotas or minimum supply/demand constraints between countries and the addition of stochasticity to random variables such as Beta that control the minimum inventory levels at the end of each period.

The simplified representation of supply disruption in the previous chapter led us to design and incrementally updated a data-driven measure that quantifies the network risk levels including downstream propagation effects in chapter 3. Our current risk measure, DSRM solves two challenges around including propagation effects of supply disruption and normalizing between nodes and across networks. Future work can still be done to improve the normalization process as the distribution of DSRM values range from 0 to values much smaller than 1 due to repeated normalization. In addition, when creating significance tree, selecting when to cut out repeated nodes is not trivial as sometimes the may or may not be its most significant/ impactful occurrence in the tree level.

In chapter 4, we present a methodology to support the development of scalable policies by aggregating the different varieties of food supply chains into a manageable subset of groups based on the supply chain structure and characteristics. In the future, we want to

investigate changes in network structure as they shift between groups. In addition, we want to expand the product quantity and diversity. Additionally, we are interested in predicting risk levels given network structure: graph embeddings are used to capture the structure of the network and can become more representative features for prediction (Graph ML) both at the network level and at the node scale. More so, this framework can allow the inclusion of socio-environmental metrics such as food access equity and emission levels into the clustering algorithm. The addition of these measures could enable policy makers to evaluate the trade-off when designing policies for groups based on different metrics.

Finally, we quantify the socio-environmental costs of supply chain systems facing risk by extending these previous analyses to include multiple objectives into the optimization model and analyze the trade-off between current weighted objectives and future protection against wildfire risk. The results suggest that utility functions need to be carefully generated and analyzed by the decision-maker in order to represent his/her expectations and concerns about the landscape because they play a crucial role in generating adequate treatment plans. Some extensions to this work would involve solving a multi-stage version integrating the current framework with forest growth and the inclusion of an explicit multi-criteria optimization model including the generated utility functions as the main objectives, analyzing the trade-off of multiple feasible plans via useful techniques such as Pareto frontiers. In addition, we can incorporate wildfire severity and frequency prediction model into the framework, developing and end-to-end solution for the decision-makers. Another future direction can involve using future projected layer values (e.g., population) as model inputs to consider providing solutions that include projected changes in different demographic and environmental factors of interest.

Bibliography

[1] John T Abatzoglou and A Park Williams. "Impact of anthropogenic climate change on wildfire across western US forests". In: *Proceedings of the National Academy of Sciences* 113.42 (2016), pp. 11770–11775.

[2] Access. *Study: 40 percent of businesses fail to reopen after a disaster*. Oct. 2021. URL: https://www.accesscorp.com/press-coverage/study-40-percent-businesses-fail-reopen-disaster/.

[3] Serpil Aday and Mehmet Seckin Aday. "Impact of COVID-19 on the food supply chain". In: *Food Quality and Safety* 4.4 (2020), pp. 167–180.

[4] Belarmino Adenso-Dıaz, Julio Mar-Ortiz, and Sebastián Lozano. "Assessing supply chain robustness to links failure". In: *International Journal of Production Research* 56.15 (2018), pp. 5104–5117.

[5] James K Agee and Carl N Skinner. "Basic principles of forest fuel reduction treatments". In: *Forest ecology and management* 211.1-2 (2005), pp. 83–96.

[6] Fermın J Alcasena et al. "Optimizing prescribed fire allocation for managing fire risk in central Catalonia". In: *Science of the total environment* 621 (2018), pp. 872–885.

[7] Thankgod Chukwumeka Aleruchi. "Strategies to Minimize Perishable Food Loss in the Retail Grocery Business". PhD thesis. Walden University, 2019.

[8] Hagar H Amer, Noha M Galal, and Khaled S El-Kilany. "A simulation study of sustainable agri-food supply chain". In: *Proceedings of the International Conference on Industrial Engineering and Operations Management Paris, France*. 2018, pp. 2264–2275.

[9] Erin Baldassari. "Camp Fire death toll grows to 29, matching 1933 blaze as state's deadliest". In: *East Bay Times (November 12, 2018)* (2018).

[10] World Bank. "World integrated trade solution". In: 2014.

[11] Kevin Barnett et al. "Beyond fuel treatment effectiveness: Characterizing Interactions between fire and treatments in the US". In: *Forests* 7.10 (2016), p. 237.

[12] Opher Baron, Oded Berman, and David Perry. "Continuous review inventory models for perishable items ordered in batches". In: *Mathematical Methods of Operations Research* 72.2 (2010), pp. 217–247.

[13] Opher Baron, Oded Berman, and David Perry. "Continuous review inventory models for perishable items with leadtimes". In: *Probability in the Engineering and Informational Sciences* (2017), pp. 1–26.

[14] G Behzadi et al. "Allocation flexibility for agribusiness supply chains under market demand disruption". In: *International Journal of Production Research* 56.10 (2018), pp. 3524–3546.

[15] Golnar Behzadi, Michael Justin O'Sullivan, and Tava Lennon Olsen. "On Metrics for Supply Chain Resilience". In: *European Journal of Operational Research* (2020).

[16] Golnar Behzadi et al. "Agribusiness supply chain risk management: A review of quantitative decision models". In: *Omega* 79 (2018), pp. 21–42.

[17] Golnar Behzadi et al. "Robust and resilient strategies for managing supply disruptions in an agribusiness supply chain". In: *International Journal of Production Economics* 191 (2017), pp. 207–220.

[18] Sherine Beshara, Khaled S El-Kilany, and Noha M Galal. "Simulation of agri-food supply chains". In: *International Journal of Industrial and Manufacturing Engineering* 6.5 (2012), pp. 899–904.

[19] Richard A Birdsey. *Carbon storage and accumulation in United States forest ecosystems*. Vol. 59. US Department of Agriculture, Forest Service, 1992.

[20] Alfredo Braunstein et al. "Network dismantling". In: *Proceedings of the National Academy of Sciences* 113.44 (2016), pp. 12368–12373.

[21] Robert E Burgan. *Behave: fire behavior prediction and fuel modeling system, fuel subsystem*. Vol. 167. Intermountain Forest and Range Experiment Station, Forest Service, US Department of Agriculture, 1984.

[22] Maria Calviño-Cancela et al. "Wildfire risk associated with different vegetation types within and outside wildland-urban interfaces". In: *Forest Ecology and Management* 372 (2016), pp. 1–9.

[23] B Carducci et al. "Food systems, diets and nutrition in the wake of COVID-19". In: *Nature Food* 2.2 (2021), pp. 68–70.

[24] Jaime Carrasco et al. "Adjusting Rate of Spread Factors through Derivative-Free Optimization: A New Methodology to Improve the Performance of Forest Fire Simulators". In: *arXiv preprint arXiv:1909.05949* (2019).

[25] Sarah Cascone. "The Getty Museum Closes Its Doors as Wildfires Rage in California, But Says the Art Is Safe Inside Its Fireproof Building". In: *Artnet News* (Oct. 2019). URL: https://news.artnet.com/art-world/getty-museum-closed-wildfires-1691749.

[26] Bureau Census. "U.S. Census Bureau QuickFacts: Napa County, California". In: *Census Bureau QuickFacts* (2019). URL: https://www.census.gov/quickfacts/napacountycalifornia.

[27] Egemen K Çetinkaya et al. "Modelling communication network challenges for future internet resilience, survivability, and disruption tolerance: A simulation-based approach". In: *Telecommunication Systems* 52.2 (2013), pp. 751–766.

[28] Vaibhav Chaudhary, Rakhee Kulshrestha, and Srikanta Routroy. "State-of-the-art literature review on inventory models for perishable products". In: *Journal of Advances in Management Research* (2018).

[29] Hernán Chávez et al. "Simulation-based multi-objective model for supply chains with disruptions in transportation". In: *Robotics and Computer-Integrated Manufacturing* 43 (2017), pp. 39–49.

[30] Md Maruf Hossan Chowdhury and Mohammed A Quaddus. "A multiple objective optimization based QFD approach for efficient resilient strategies to mitigate supply chain vulnerabilities: The case of garment industry of Bangladesh". In: *Omega* 57 (2015), pp. 5–21.

[31] Martin Christopher and Hau Lee. "Mitigating supply chain risk through improved confidence". In: *International journal of physical distribution & logistics management* (2004).

[32] Martin Christopher and Helen Peck. "Building the resilient supply chain." In: (2004).

[33] Jaekwon Chung and Dong Li. "The prospective impact of a multi-period pricing strategy on consumer perceptions for perishable foods". In: *British Food Journal* (2013).

[34] Woodam Chung. "Optimizing fuel treatments to reduce wildland fire risk". In: *Current Forestry Reports* 1.1 (2015), pp. 44–51.

[35] Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to derivative-free optimization*. Vol. 8. Siam, 2009.

[36] Thomas Curt, Thibaut Fréjaville, and Sébastien Lahaye. "Modelling the spatial patterns of ignition causes and fire regime features in southern France: implications for fire prevention policy". In: *International Journal of Wildland Fire* 25.7 (2016), pp. 785–796.

[37] Scott J DeWald, Scott John Josiah, and Rebecca Erdkamp. *Heating with wood: Producing, harvesting and processing firewood*. Cooperative Extension, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln, 2005.

[38] Yongrui Duan et al. "Inventory models for perishable items with inventory level dependent demand rate". In: *Applied Mathematical Modelling* 36.10 (2012), pp. 5015–5028.

[39] David K Duvenaud et al. "Convolutional networks on graphs for learning molecular fingerprints". In: *Advances in neural information processing systems* 28 (2015).

[40] William P Edwards. "The New Normal: Living with Wildland Fire". In: *Natural Resources & Environment* 33.3 (2019), pp. 30–33.

[41] Pelagie Elimbi Moudio, Cristobal Pais, and Zuo-Jun Max Shen. "Quantifying the impact of ecosystem services for landscape management under wildfire hazard". In: *Natural Hazards* 106.1 (2021), pp. 531–560.

[42] M. A. Finney. "FARSITE: fire area simulator — model development and evaluation". In: *Rev. ed. US Dept Agric., ForServ. Res. Pap. RMRS-RP-4* (2004), p. 48.

[43] Mark A Finney. "Design of regular landscape fuel treatment patterns for modifying fire growth and behavior". In: *Forest Science* 47.2 (2001), pp. 219–228.

[44] Mike D Flannigan et al. "Implications of changing climate for global wildland fire". In: *International journal of wildland fire* 18.5 (2009), pp. 483–507.

[45] Félix-Antoine Fortin et al. "DEAP: Evolutionary algorithms made easy". In: *Journal of Machine Learning Research* 13.Jul (2012), pp. 2171–2175.

[46] Shodor Education Foundation. "How to weight a tree". In: *Project SUCCEED-HI: For Students!* (2002). URL: http://www.shodor.org/succeedhi/succeedhi/weightree/teacher/activities.html.

[47] Thomas Fuller. *Three Weeks After Fire, Official Search for Dead Is Completed*. Nov. 2018. URL: https://www.nytimes.com/2018/11/29/us/victims-california-fires-missing.html.

[48] Franziska Gaupp. "Extreme events in a globalized food system". In: *One Earth* 2.6 (2020), pp. 518–521.

[49] Clive George. "Environment and Regional Trade Agreements". In: (2014). DOI: https://doi.org/https://doi.org/10.1787/5jz0v4q45g6h-en. URL: https://www.oecd-ilibrary.org/content/paper/5jz0v4q45g6h-en.

[50] Rossitza Goleva and Seferin Mirtchev. "Traffic Modelling in Disruption-tolerant Networks". In: *Annual Seminar of the PTT College "Modelling and Control of Information Processes", CTP, Sofia*. 2010, pp. 6–20.

[51] Noel Gorelick et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote Sensing of Environment* (2017). DOI: 10.1016/j.rse.2017.06.031. URL: https://doi.org/10.1016/j.rse.2017.06.031.

[52] Suresh Kumar Goyal and Bibhas Chandra Giri. "Recent trends in modeling of deteriorating inventory". In: *European Journal of operational research* 134.1 (2001), pp. 1–16.

[53] Aric Hagberg et al. "Networkx. High productivity software for complex networks". In: *Webová strá nka https://networkx. lanl. gov/wiki* (2013).

[54] William E Hart et al. *Pyomo-optimization modeling in python*. Vol. 67. Springer, 2017.

[55] R Hausmann, J Hwang, and D Rodrick. ""What you export matters", NBER Working paper No. 11905, National Bureau of Economic Research, Cambridge, MA". In: (2005).

[56] Ricardo Hausmann and Bailey Klinger. "The structure of the product space and the evolution of comparative advantage". In: *CID Working Paper Series* (2007).

[57] Kelvin Hirsch et al. "Fire-smart forest management: a pragmatic approach to sustainable forest management in fire-dominated ecosystems". In: *The Forestry Chronicle* 77.2 (2001), pp. 357–363.

[58] Seyedmohsen Hosseini, Dmitry Ivanov, and Alexandre Dolgui. "Review of quantitative methods for supply chain resilience analysis". In: *Transportation Research Part E: Logistics and Transportation Review* 125 (2019), pp. 285–307.

[59] kristina Houck and Patch Staff. *Top 20 most destructive California wildfires*. 2018. URL: https://patch.com/california/san-diego/top-20-deadliest-most-destructive-wildfires-ca-history.

[60] HI Hsiao, JGAJ Van der Vorst, SWF Omta, et al. "Logistics outsourcing in food supply chain networks: Theory and practices". In: *International agri-food chains and networks: management and organization* (2006), pp. 135–150.

[61] Matthew O Jackson and Asher Wolinsky. "A strategic model of social and economic networks". In: *Journal of economic theory* 71.1 (1996), pp. 44–74.

[62] Steven Jaffee, Paul Siegel, and Colin Andrews. "Rapid agricultural supply chain risk assessment: A conceptual framework". In: *Agriculture and rural development discussion paper* 47.1 (2010), pp. 1–64.

[63] David S Johnson. "The NP-completeness column: an ongoing guide". In: *Journal of Algorithms* 6.3 (1985), pp. 434–451.

[64] Steven G Johnson. *The NLopt nonlinear-optimization package [Software]*. 2014.

[65] W Matt Jolly et al. "Climate-induced variations in global wildfire danger from 1979 to 2013". In: *Nature communications* 6 (2015), p. 7537.

[66] Masoud Kamalahmadi and Mahour Mellat-Parast. "Developing a resilient supply chain through supplier flexibility and reliability assessment". In: *International Journal of Production Research* 54.1 (2016), pp. 302–321.

[67] K Katsaliaki, P Galetsi, and S Kumar. "Supply chain disruptions and resilience: a major review and future research agenda". In: *Annals of Operations Research* (2021), pp. 1–38.

[68] Yusoon Kim, Yi-Su Chen, and Kevin Linderman. "Supply network disruption and resilience: A network structural perspective". In: *Journal of operations Management* 33 (2015), pp. 43–59.

[69] Thomas N Kipf and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks". In: *arXiv preprint arXiv:1609.02907* (2016).

[70] Jack PC Kleijnen and JGAJ Van der Vorst. "Designing robust and sustainable fresh-food supply chains: Improved simulation methodology for reducing waste". In: *Department of Information Systems & Management/Center for Economic Research, Tilburg University* (2005).

[71] Priya Krishnakumar. *We mapped every wine country fire. They're larger and more destructive than ever*. Nov. 2019. URL: https://www.latimes.com/projects/wine-country-fires-california-sonoma-napa/.

[72] Matthai Chakko Kuruvila and Ellen Lee. *Paradise fire evacuees starting to return home*. Feb. 2012. URL: https://www.sfgate.com/bayarea/article/Paradise-fire-evacuees-starting-to-return-home-3209813.php.

[73] David Laborde et al. "COVID-19 risks to global food security". In: *Science* 369.6503 (2020), pp. 500–502.

[74] Magdalena Leithner and Christian Fikar. "Simulating fresh food supply chains by integrating product quality". In: *Operations research proceedings 2017*. Springer, 2018, pp. 671–676.

[75] Ruxian Li, Hongjie Lan, John R Mawhinney, et al. "A review on deteriorating inventory study". In: *Journal of Service Science and Management* 3.01 (2010), p. 117.

[76] Radojko Lukic, Dragana Vojteski Kljenak, and Dragica Jovancevic. "RETAIL FOOD WASTE MANAGEMENT." In: *Management Research & Practice* 6.4 (2014).

[77] John Manners-Bell. *Supply chain risk: Understanding emerging threats to global supply chains*. Kogan Page Publishers, 2014.

[78] Vlad Manole et al. "WITS–World Integrated Trade Solution". In: *World Scientific Book Chapters* (2005), pp. 541–548.

[79] Dmytro Matsypura, Oleg A Prokopyev, and Aizat Zahar. "Wildfire fuel management: network-based models and optimization of prescribed burning". In: *European Journal of Operational Research* 264.2 (2018), pp. 774–796.

[80] William G Moseley and Jane Battersby. "The vulnerability and resilience of African food systems, food security, and nutrition in the context of the COVID-19 pandemic". In: *African Studies Review* 63.3 (2020), pp. 449–461.

[81] Marie Pelagie Elimbi Moudio, Cristobal Pais, and Zuo-Jun Max Shen. "Data-driven Planning in the Face of Supply Disruption in Global Agricultural Supply Chains". In: *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE. 2021, pp. 238–242.

[82] Jafar Namdar et al. "Supply chain resilience for single and multiple sourcing in the presence of disruption risks". In: *International Journal of Production Research* 56.6 (2018), pp. 2339–2360.

[83] Fabrizio Natale et al. "Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread". In: *Preventive veterinary medicine* 92.4 (2009), pp. 341–350.

[84] NOAA. *Billion-dollar weather and climate disasters*. 2022. URL: https://www.ncei.noaa.gov/access/billions/.

[85] Ritesh Ojha et al. "Bayesian network modelling for supply chain risk propagation". In: *International Journal of Production Research* 56.17 (2018), pp. 5795–5819.

[86] Philip N Omi. "Theory and practice of wildland fuels management". In: *Current Forestry Reports* 1.2 (2015), pp. 100–117.

[87] Bhakti Stephan Onggo et al. "Agri-food supply chains with stochastic demands: A multi-period inventory routing problem with perishable products". In: *Simulation Modelling Practice and Theory* 97 (2019), p. 101970.

[88] Javier Arturo Orjuela-Castro, Lizeth Andrea Sanabria-Coronado, and Andrés Mauricio Peralta-Lozano. "Coupling facility location models in the supply chain of perishable fruits". In: *Research in transportation business & management* 24 (2017), pp. 73–80.

[89] G Padmanabhan and Prem Vrat. "EOQ models for perishable items under stock dependent selling rate". In: *European Journal of Operational Research* 86.2 (1995), pp. 281–292.

[90] Cristobal Pais et al. "Cell2Fire: A Cell Based Forest Fire Growth Model". In: *arXiv preprint arXiv:1905.09317* (2019).

[91] Cristobal Pais et al. "Downstream Protection Value: Detecting critical zones for effective fuel-treatment under wildfire risk". In: *Under R1 in Computers & Operations Research* (2020).

[92] Cristobal Pais et al. "Downstream Protection Value: Detecting critical zones for effective fuel-treatment under wildfire risk". In: *Computers & Operations Research* 131 (2021), p. 105252.

[93] Gillian Pais, Kartik Jayaram, and Arend van Wamelen. "Safeguarding Africa's food systems through and beyond the crisis". In: *Retrieved February* 26 (2020), p. 2021.

[94] P Pakdeenarong and T Hengsadeekul. "Supply chain risk management of organic rice in Thailand". In: *Uncertain Supply Chain Management* 8.1 (2020), pp. 165–174.

[95] Supun Perera, Michael Bell, and Michiel Bliemer. "Network Science approach to Modelling Emergence and Topological Robustness of Supply Networks: A Review and Perspective". In: *arXiv preprint arXiv:1803.09913* (2018).

[96] Supun S Perera et al. "Topological structure of manufacturing industry supply chain networks". In: *Complexity* 2018 (2018).

[97] Michael J Puma et al. "Assessing the evolving fragility of the global food system". In: *Environmental Research Letters* 10.2 (2015), p. 024007.

[98] Fred Raafat. "Survey of literature on continuously deteriorating inventory models". In: *Journal of the Operational Research society* 42.1 (1991), pp. 27–37.

[99] Joaqun Ramırez, Santiago Monedero, and David Buckley. "New approaches in fire simulations analysis with wildfire analyst". In: *7th International Conference on Forest Fire Research*. 2011.

[100] Bob Ritchie and Clare Brindley. "An emergent framework for supply chain risk management and performance measurement". In: *Journal of the Operational Research Society* 58.11 (2007), pp. 1398–1411.

[101] R Tyrrell Rockafellar, Stanislav Uryasev, et al. "Optimization of conditional value-at-risk". In: *Journal of risk* 2 (2000), pp. 21–42.

[102] Matthew G Rollins. "LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment". In: *International Journal of Wildland Fire* 18.3 (2009), pp. 235–249.

[103] S. M. Running. "Is global warming causing more large wildfires?" In: *Science* 313 (2006), pp. 927–928. DOI: 10.1126/science.

[104] L Russo et al. "Complex network statistics to the design of fire breaks for the control of fire spreading". In: *Chemical Engineering Transactions* (2015).

[105] Michele Salis et al. "Evaluating alternative fuel treatment strategies to reduce wildfire losses in a Mediterranean area". In: *Forest Ecology and Management* 368 (2016), pp. 207–221.

[106] Boris Schäling. *The boost C++ libraries*. Boris Schäling, 2011.

[107] Joe H Scott. *Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2005.

[108] Winnie Septiani et al. "Method and approach mapping for agri-food supply chain risk management: A literature review". In: *International Journal of Supply Chain Management* 5.2 (2016), pp. 51–64.

[109] Sarbjit Singh. "An inventory model for perishable items having constant demand with time dependent holding cost". In: *Mathematics and Statistics* 4.2 (2016), pp. 58–61.

[110] Umang Soni, Vipul Jain, and Sameer Kumar. "Measuring supply chain resilience using a deterministic modeling approach". In: *Computers & Industrial Engineering* 74 (2014), pp. 11–25.

[111] K Sun et al. "Wildfire spread in Wildland Urban Interface (WUI) communities in California: Introducing the urban fuel". In: *AGUFM* 2019 (2019), NH43C–0950.

[112] Christopher Tang and Brian Tomlin. "The power of flexibility for mitigating supply chain risks". In: *International journal of production economics* 116.1 (2008), pp. 12–27.

[113] Xinmin Tian et al. "Intel® OpenMP C++/Fortran Compiler for Hyper-Threading Technology: Implementation and Performance." In: *Intel Technology Journal* 6.1 (2002).

[114] Benjamin R Tukamuhabwa et al. "Supply chain resilience: definition, review and theoretical foundations for further study". In: *International Journal of Production Research* 53.18 (2015), pp. 5592–5623.

[115] C. Tymstra et al. "Development and structure of Prometheus: the Canadian Wildland Fire Growth Simulation Model". In: *Information Report NOR-X-Edmonton (AB): Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre* 417 (2010), p. 102.

[116] Cor N Verdouw et al. "Virtualization of food supply chains with the internet of things". In: *Journal of Food Engineering* 176 (2016), pp. 128–136.

[117] Jelena V Vlajic, EMT Hendrix, and JGAJ van der Vorst. *Food supply chain network robustness: a literature review and research agenda*. Tech. rep. Wageningen University, 2008.

[118] Thomas L Vollrath, Mark J Gehlhar, and Charles B Hallahan. "Bilateral import protection, free trade agreements, and other factors influencing trade flows in agriculture and clothing". In: *Journal of Agricultural Economics* 60.2 (2009), pp. 298–317.

[119] Dale R Waddell. "Estimating load weights with Huber's Cubic Volume formula: a field trial." In: *Res. Note. PNW-RN-484. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 12 p* 484 (1989).

[120] Xiaojun Wang and Dong Li. "A dynamic product quality evaluation based pricing model for perishable food supply chains". In: *Omega* 40.6 (2012), pp. 906–917.

[121] Hui-Ming Wee. "Economic production lot size model for deteriorating items with part ial back-ordering". In: *Computers & Industrial Engineering* 24.3 (1993), pp. 449–458.

[122] A Park Williams and John T Abatzoglou. "Recent advances and remaining uncertainties in resolving past and future climate effects on global fire activity". In: *Current Climate Change Reports* 2.1 (2016), pp. 1–14.

[123] Jonathan Woetzel, Dickon Pinner, and Hamid Samandari. *Climate Risk and response*. Tech. rep. McKinsey Global Institute, 2020.

[124] Stephen A Wood et al. "Trade and the equitability of global food nutrient distribution". In: *Nature Sustainability* 1.1 (2018), pp. 34–37.

[125] Group World Bank. *Food security and covid-19*. Dec. 2021. URL: https://www.worldbank.org/en/topic/agriculture/brief/food-security-and-covid-19.

[126] Bo Yan, Ping Shi, and FL Wang. "Risk assessment and control of agricultural supply chain based on CVaR". In: *Soft Sci.* 27.10 (2013), pp. 111–115.

[127] Nyamah Edmond Yeboah et al. "Agricultural supply chain risk identification-a case finding from Ghana". In: *Journal of management and strategy* 5.2 (2014), p. 31.

[128] Kang Zhao et al. "Supply chain network robustness against disruptions: Topological analysis, measurement, and optimization". In: *IEEE Transactions on Engineering Management* 66.1 (2018), pp. 127–139.

Appendix A

Global Allocation Model Extended

A.1 Appendix: Two-stage Problem Extended Results

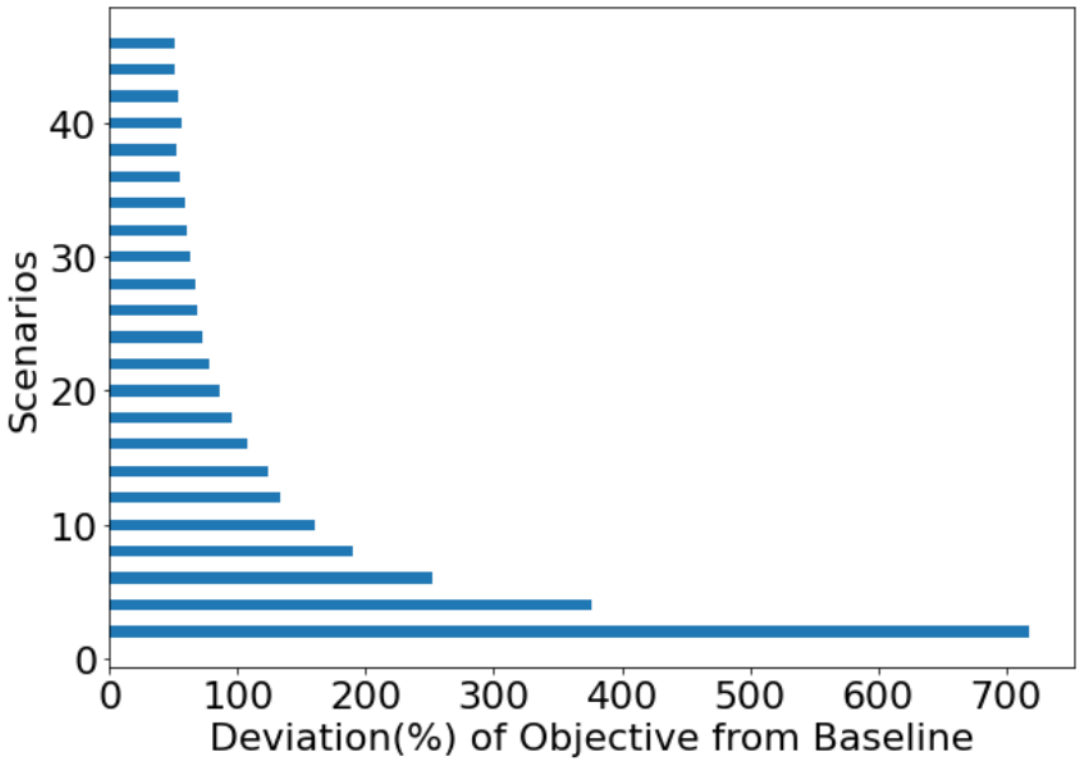


Figure A.1: Bar plot shows the percentage (%) deviation of Objective from deterministic baseline. The baseline is set to be the case in which no country experiences a capacity disruption, representing the most basic planning situation.

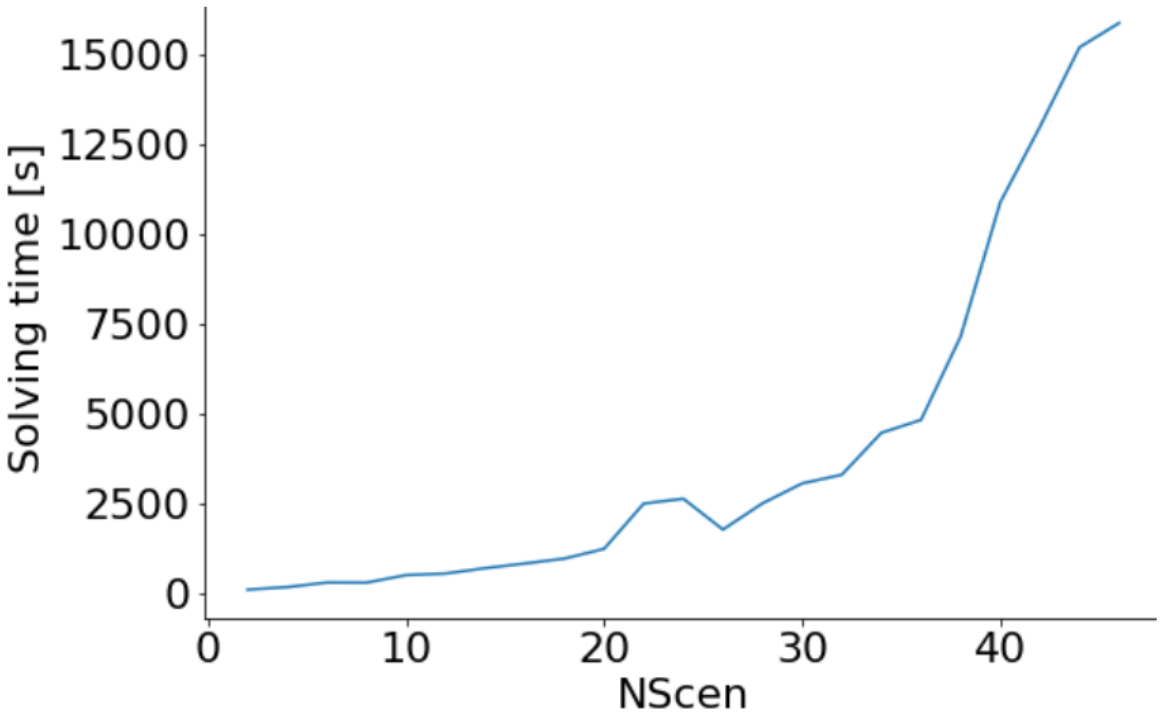


Figure A.2: Evolution of Risk Neutral Solving time versus the number of scenarios. Non-linear relationship between solving time and number of scenarios with an explosion after incorporating 36 scenarios.

Table A.1: presents for single country disruption scenarios the Objective values, relative inventory for the disrupted values and the average inventory for the top 20 suppliers when that country is disrupted.

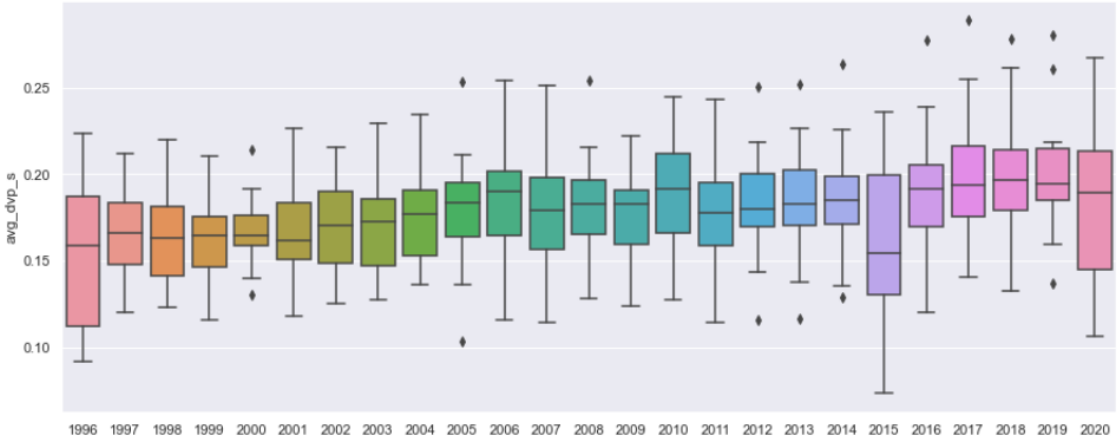
Country	Objective	Relative Inventory	Average Inventory
Brazil	1839.660	0.021	0.035
Colombia	810.800	0.027	0.034
Indonesia	509.959	0.022	0.034
Germany	402.228	0.023	0.034
Guatemala	350.534	0.023	0.034
Honduras	326.664	0.023	0.034
Peru	301.183	0.023	0.034
India	287.834	0.023	0.034
Italy	282.844	0.022	0.034
Ethiopia	266.253	0.022	0.034
Uganda	261.363	0.022	0.034
Mexico	254.843	0.022	0.034
Belgium	207.708	0.021	0.034
Costa Rica	207.708	0.021	0.034
Côte d'Ivoire	167.646	0.017	0.034
Nicaragua	192.284	0.020	0.034
El Salvador	182.976	0.019	0.034
United States	172.836	0.015	0.034
Papua New Guinea	168.476	0.014	0.034

Appendix B

Agricultural Network Health Measure Extended

B.1 Distribution of risk measures (DPV) through time across all products.

Figure B.1: Box-plot showing the distribution of Normalized DPV values across all Products between 1996 and 2020. There is a slight positive trend towards risk with time potentially signally overall increasing vulnerability of networks globally.



B.2 DSRM Pseudocode

Algorithm 1 DSRM Single node Pseudo-code

```
1: procedure
2: Given a subtree, with root
3:     for node ∈ leafNodes do:
4:         metric[node]=0
5:         while len(parents)>0 and some nodes unexplored: do
6:             for parent ∈ parents do:
7:                 if parent = root then:
8:                     continue
9:                     temp = 0
10:                    for child ∈ children[parent] do:
11:                        if child not in metric then:
12:                            break
13:                            temp+=tree.edges[parent,child]['delta]
14:                            *tree.edges[parent,child]['delta weight']
15:                            grandparent = getparents([parent],subtree)
16:                            if grandparent is not root then:
17:                                temp = 0.5*(temp +
18:                                tree.edges[grandparent[0],parent]['delta'])
19:                                metric[parent] = temp
20:                                leafNodes = list(parents)
21:                                parents = getparents(leafNodes,subtree)
22:
```

Algorithm 2 DSRM Whole Network Proceduce Pseudo-code

```
1: procedure
2: Given a Digraph, G (V, E), with vertices, V and Edges, E
3:     for node ∈ G.nodes do:
4:         subtree = genSubtree(G, source=node, depthlimit=100)
5:         subtree = addAttr(subtree, G,node)
6:         subtree = treeEdgePrune(subtree, epsilon = 0.05, verbose = verbose)
7:         subtree = addparents(subtree, node)
8:         DSRM = dsrm(subtree,verbose = verbose)
9:         DSRMs[node] = DSRM
10:        if node ∈ DSRM then:
11:            dsrmMaster[node] = DSRM[node]
12:        else
13:            dsrmMaster[node] =0
14:            for node ∈ G.nodes do:
15:                G.nodes[node]['dsrm'] = dsrmMaster[node]
16:
```

Appendix C

Network Clustering Extended Analysis

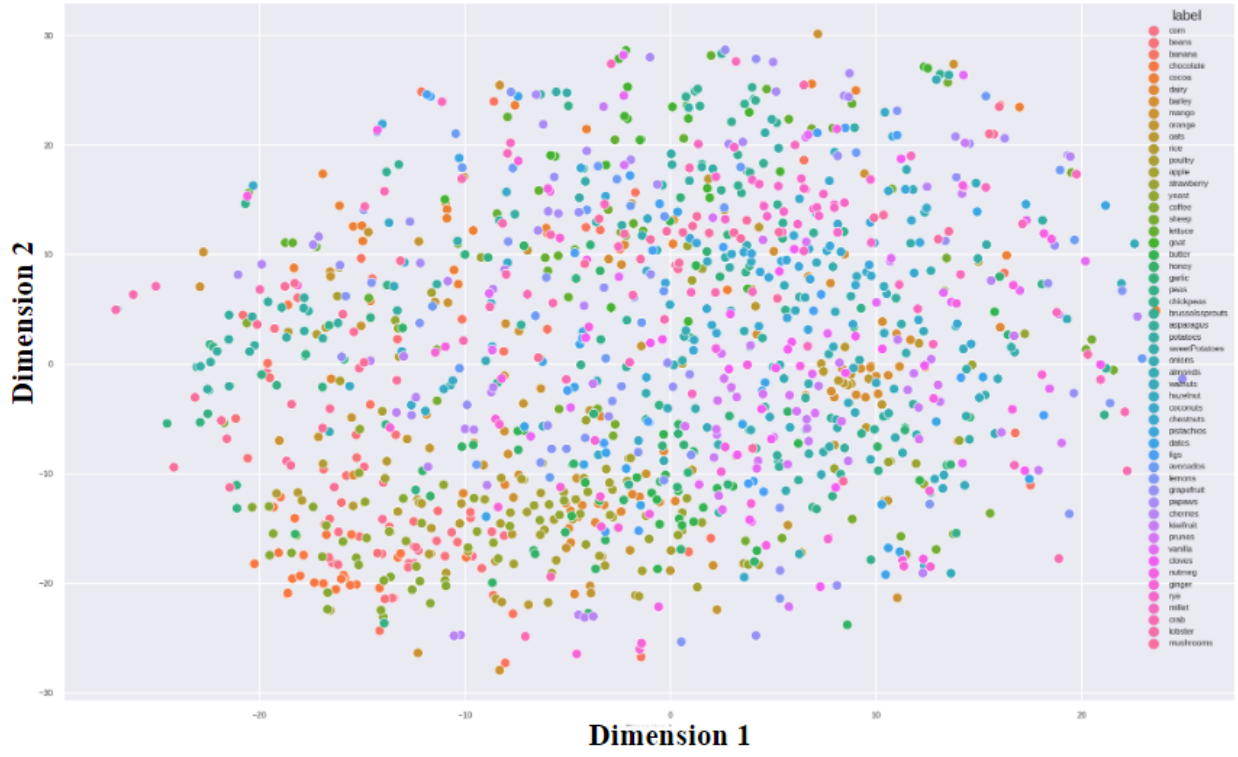
C.1 Clustering Networks based on structure - Summary

Table C.1: Summary table - Combinations of Hyperparameters for clustering framework. The cells highlighted in green indicate models (predominantly GCN models) that are expressive enough to identify clustering. The cells in red used the Graph2Vec model which was not expressive enough to capture distinctive clustering behavior in 2 dimensions.

Model	**Dimensions**	**Attributed vs not**	**Clustering algorithms**	**Number of clusters**	**Dimension reduction**
GCN	64	Attributed	Community detection	∼11	TSNE & UMAP
GCN	128	Attributed	Community detection	11	TSNE & UMAP
GCN	128	Attributed	k-means	10-30	TSNE & UMAP
GCN	256	Attributed	Community detection	–	TSNE & UMAP
Graph2Vec	64	Attributed	Community detection	n/a	n/a
Graph2Vec	64	not Attributed	Community detection	n/a	n/a
Graph2Vec	128	Attributed	Community detection	n/a	n/a
Graph2Vec	128	not Attributed	Community detection	n/a	n/a

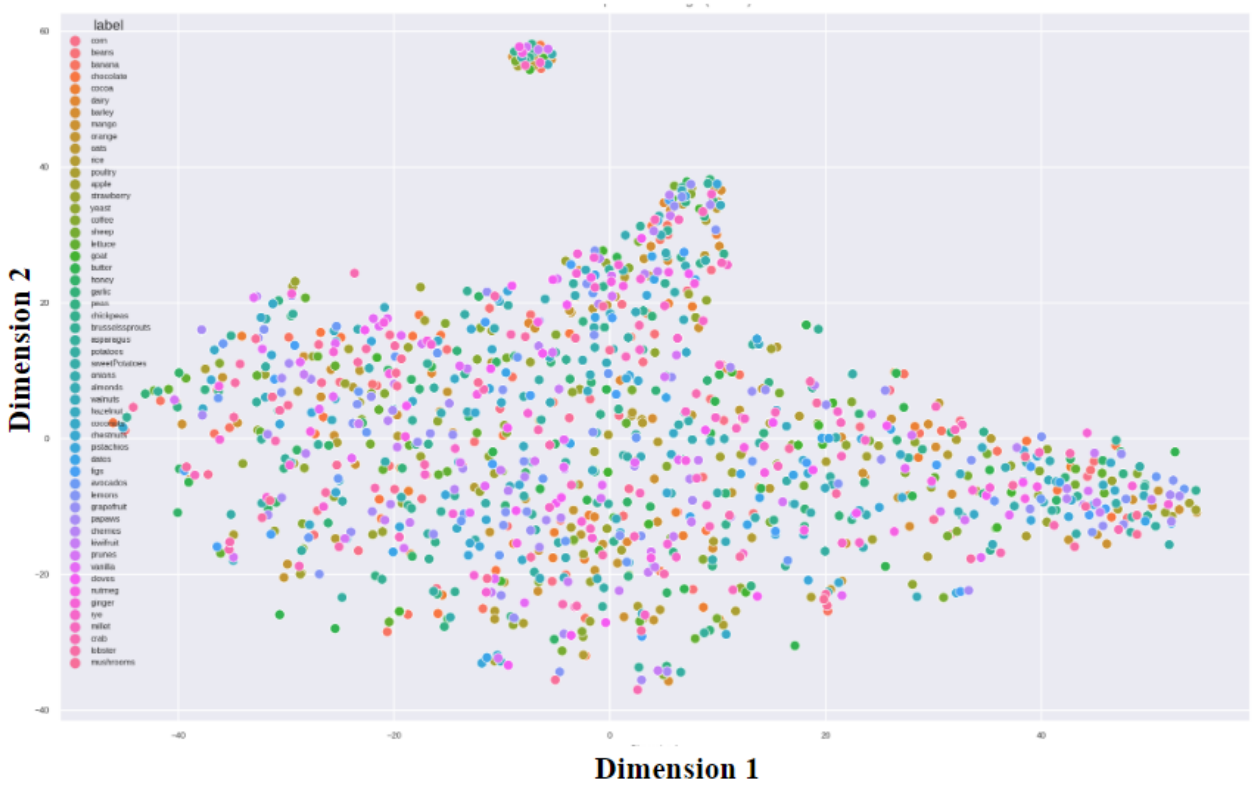
C.2 Graph2Vec Extension

Figure C.1: Two dimensional projection with TSNE of 64 dimensional Graph2Vec embeddings (non-attributed). This 2D projection does not reveal clear separations in data



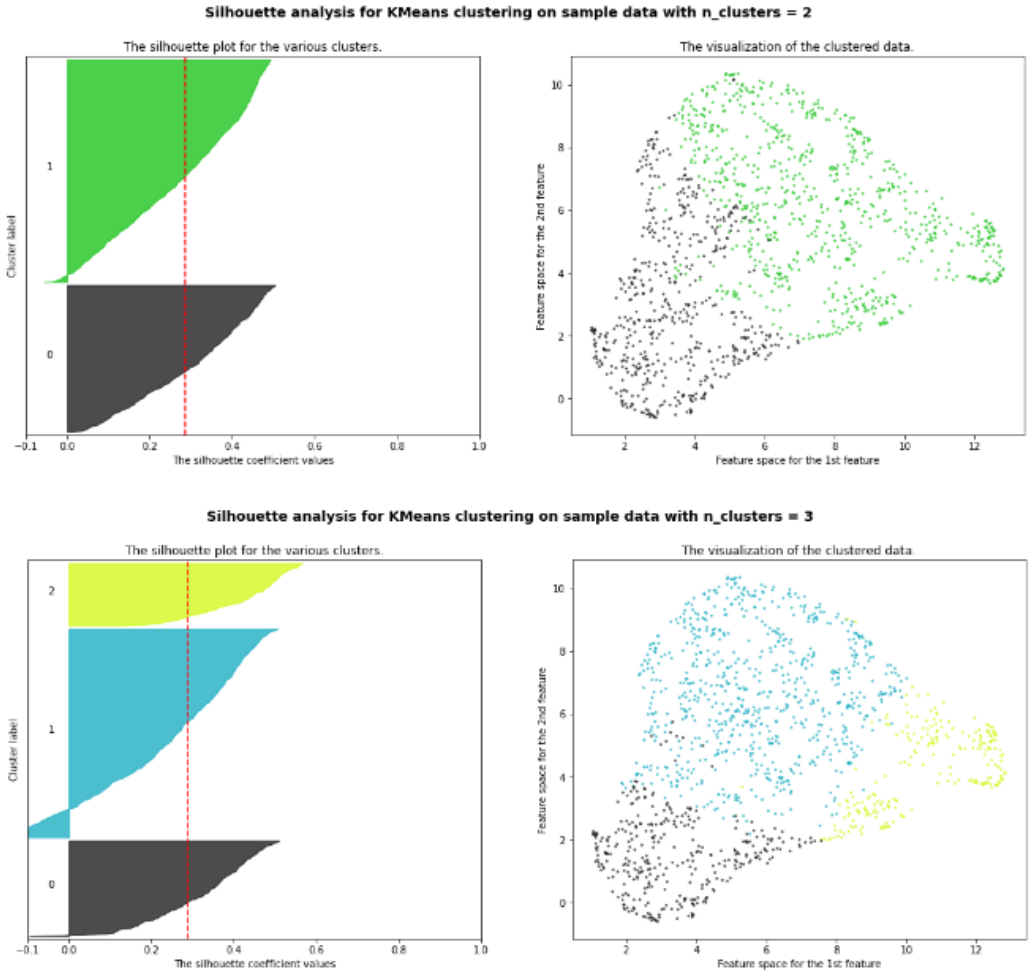
C.3 Selecting Number of Clusters

Figure C.2: Two dimensional projection with TSNE of 64 dimensional Graph2Vec embeddings (attributed). This 2D projection does not reveal clear separations in data



C.4 Silhouette Plots for number of clusters = 6 and 7

Figure C.3: silhouette plots for number of clusters ranging between 2 and 3 and their corresponding 2D cluster visualization on the top and bottom respectively. The clusters are generated obtained from spectral clustering.



C.5 Using dendrogram for Selecting Number of Clusters

.

Figure C.4: silhouette plots for number of clusters ranging between 6 and 7 and their corresponding 2D visualization on the top and bottom respectively. The clusters are generated obtained from spectral clustering.

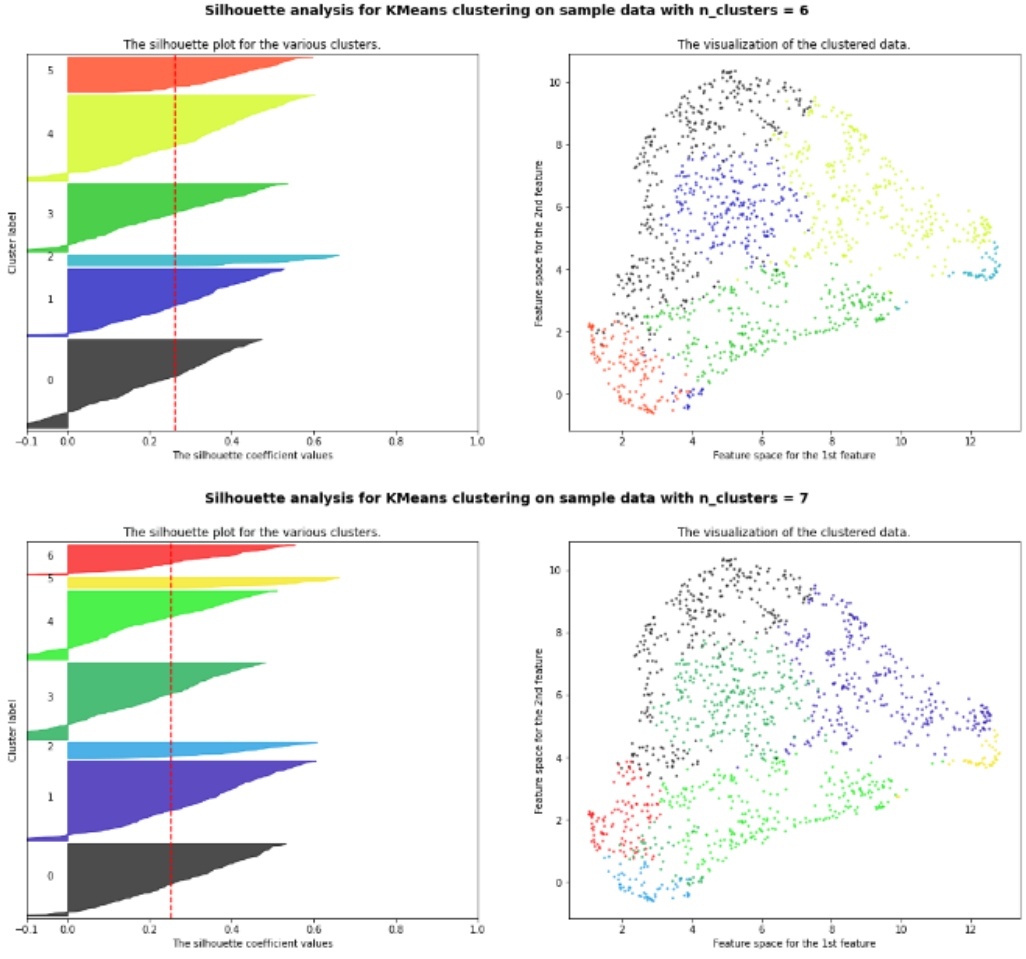
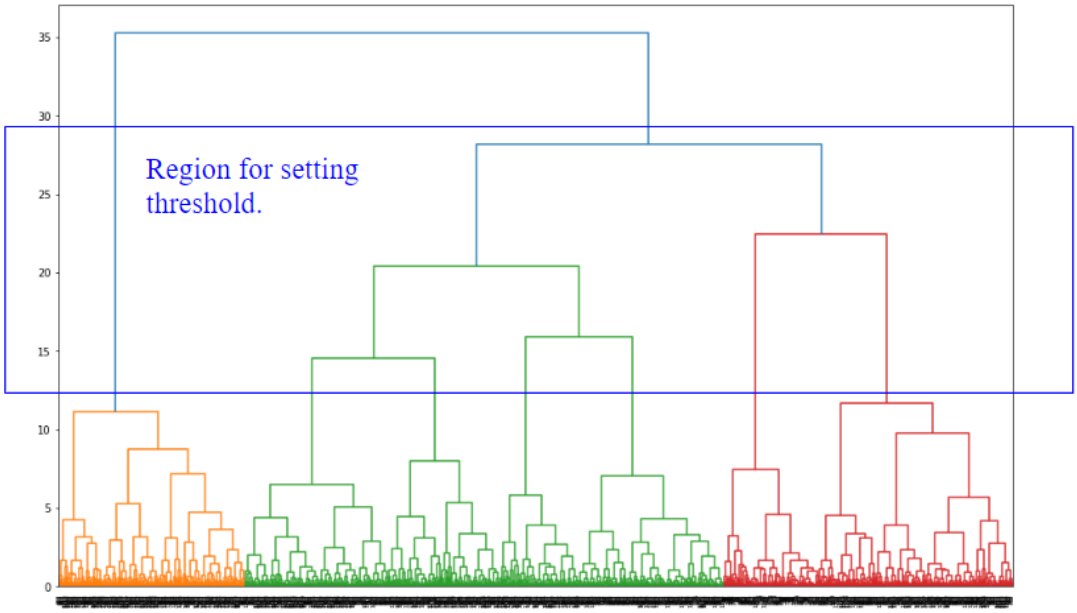


Figure C.5: Observing the dendrogram from the top to bottom, we note that the big difference between clusters is between the components of the orange cluster versus the red and green clusters as the vertical height (blue line) is longer for the former



Appendix D

Detailed Results of Extended Wildfire Framework

D.1 Mathematical formulation

Based on the model applied in conrad2012wildlife,pais2020DPV, the $\overline{\lambda}$-connected version of the optimal resource allocation problem ($ORAP_{\overline{\lambda}}$) can be formulated as the following Mixed-integer programming (MIP) model:

$$
\begin{aligned}
(ORAP_{\overline{\lambda}}) \quad \max \sum_{i \in V} U_i(\overline{\lambda}) &= \sum_{i \in V} \left(\overline{\lambda} DPV(i) + (1-\overline{\lambda}) NV_i\right) x_i && \text{(D.1)}\\
s.t. \quad z + \sum_{j \in V} y_{(s,j)} &= Q && \text{(D.2)}\\
\sum_{i \in V} \alpha_i &\leq 1 && \text{(D.3)}\\
y_{(s,j)} &\leq Q\alpha_j \qquad \forall j \in V && \text{(D.4)}\\
y_{(i,j)} &\leq Q x_j \qquad \forall\, (i,j) \in E_s && \text{(D.5)}\\
\sum_{i \in V:(i,j) \in E_s} y_{(i,j)} &= \sum_{l \in V:(j,l) \in \tilde{E}} y_{(j,l)} + x_j \qquad \forall j \in V_s && \text{(D.6)}\\
\sum_{i \in V} b_i x_i &= B && \text{(D.7)}\\
\sum_{i \in V} c_i x_i &= C && \text{(D.8)}\\
x_i, \alpha_i &\in \{0,1\} \quad \forall i \in V && \text{(D.9)}\\
y_{(i,j)} &\in \mathbb{R}^+ \qquad \forall (i,j) \in E_s && \text{(D.10)}\\
z &\in \mathbb{R}^+ && \text{(D.11)}
\end{aligned}
$$

In this formulation, an extra cell s acts as the source of all flow of the network obtaining

a new set of nodes $V_s = V \cup \{s\}$. The source is then connected to each node inside V by a set of directed edges $E_s = \tilde{E} \cup \{(s,i) \mid i \in V\}$. Flow from the source s is sent to the original network or absorbed by an auxiliary variable $z \in [0, Q]$ with Q the maximum number of nodes to be selected (eq. (4)). Adjacency constraints are enforced by eqs. (5)-(8). In eq. (5), at most one cell $i \in V$ is acting as a link between the source s and the original network using the binary variable α_i, equal to 1 if the cell i gets the flow from the source. To force the connectivity of the solution, the remaining cells do not get any flow from the source, setting those flow variables to zero (eq. (6)). Constraint (7) ensures that cells not included as part of the solution cannot receive flow from any adjacent node $i \in V$. Flow conservation is modeled by eq. (8), where each node $i \in V$ with a positive flow consumes one unit of flow and any remaining one is sent to adjacent cells l if an edge (j, l) exists.

Equations (9)-(10) keep track of the total economic benefit (B) and cost (C) of the optimal resource allocation plan, if provided, to account for extra constraints such as budget restrictions. We force the optimal solution to include the desired amount of cells Q to be selected by adding the following equation:

$$\sum_{i \in V} x_i = Q \tag{D.12}$$

In this way, the optimal connected subgraph S will include exactly Q cells.

Experiments Pseudo-code

Algorithm 3 Experiments Pseudo-code

procedure EXPERIMENTS
  for $dom \in \{forest, access, carbon, population, equal\}$ **do**
    if $dom == equal$ **then**
      `Calculate` $NV_{dom} = 0.7L_{dom} + \sum_{k \in K: k \neq dom} 0.1L_k$
    else
      `Calculate` $NV_{dom} = 0.25 \sum_{k \in K} L_k$
      `Calculate` $DPV(i) = \alpha_i \sum_{j \in T_i} NV(j),\ \forall i \in V$
      for $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$ **do**
        `Calculate utility` $U_i(\lambda) = \lambda DPV(i) + (1-\lambda)NV(i),\ \forall i \in V$
        for $tf \in \{0.15, 0.25, 0.5\}$ **do**
          Get optimal solution $X^*(\lambda, tf) \leftarrow$ Solve$(ORAP_\lambda(tf))$
          `Calculate` $\mathbb{E}\left[Losses(X^*(\lambda, tf))\right] \leftarrow Simulation(X^*(\lambda, tf), R = 100)$
          `Calculate` $\Delta_{tf} U(X^*(\lambda, tf)) := U(X^*(\lambda, tf)) - \gamma \mathbb{E}\left[Losses(X^*(\lambda, tf))\right]$

Detailed results

Table D.1: Average discounted utility results for Napa instance evaluated from 100 simulations. Results for all λ combinations between the DPV heatmap and NV layers are presented by dominating utility category (column 1) and treatment fraction level (column 2).

Utility	tf	$\Delta U(0)$	$\Delta U(0.25)$	$\Delta U(0.5)$	$\Delta U(0.75)$	$\Delta U(1)$
Access	0.05	-630.36 ± 697.47	-719.95 ± 733.02	-212.01 ± 343.35	-2.29 ± 217.35	168.94 ± 123.13
	0.10	-204.23 ± 680.14	-222.66 ± 685.27	275.49 ± 260.19	444.88 ± 165.37	567.69 ± 86.51
	0.15	76.93 ± 725.53	482.08 ± 431.61	731.61 ± 220.05	800.39 ± 151.74	903.39 ± 65.42
	0.20	747.88 ± 539.28	939.52 ± 371.53	1143.22 ± 213.7	1134.43 ± 147.12	1216.88 ± 57.02
	0.25	1218.59 ± 453.58	1373.17 ± 344.61	1533.92 ± 213.11	1552.99 ± 132.53	1508.5 ± 53.91
	0.5	3298.82 ± 198.32	3324.25 ± 169.89	3423.14 ± 99.82	3352.41 ± 42.74	3097.83 ± 13.78
Carbon	0.05	-98.88 ± 264.17	-236.28 ± 369.97	-388.85 ± 298.23	-217.91 ± 196.83	-119.62 ± 146.07
	0.10	126.69 ± 258.7	15.7 ± 339.87	-155.04 ± 217.61	-8.59 ± 137.88	60.61 ± 103.34
	0.15	330.78 ± 231.39	345.71 ± 214.09	172.81 ± 178.55	154.35 ± 118.93	209.04 ± 75.63
	0.20	496.06 ± 199.97	520.36 ± 183.79	438.15 ± 161.36	404.96 ± 109.03	339.83 ± 59.88
	0.25	759.32 ± 167.12	822.89 ± 151.53	715.74 ± 139.45	608.36 ± 79.46	494.92 ± 46.97
	0.5	1661.86 ± 113.59	1623.56 ± 116.32	1559.84 ± 132.09	1533.92 ± 78.67	1170.72 ± 25.24
Equal	0.05	-393.71 ± 441.85	-471.49 ± 519.2	-346.28 ± 325.8	-139.49 ± 204.77	-1.67 ± 132.82
	0.10	-49.04 ± 430.45	-100.32 ± 480.92	19.05 ± 243.76	154.46 ± 151.34	262.15 ± 94.0
	0.15	272.19 ± 389.96	347.36 ± 302.84	356.9 ± 203.05	403.82 ± 135.65	491.04 ± 70.41
	0.20	563.05 ± 337.88	626.91 ± 260.67	692.6 ± 191.55	635.5 ± 129.05	694.99 ± 59.39
	0.25	827.96 ± 284.05	966.14 ± 236.93	1017.39 ± 186.06	958.71 ± 108.22	934.82 ± 52.73
	0.5	2373.07 ± 197.32	2251.69 ± 195.28	2237.53 ± 174.00	2288.82 ± 56.59	1980.16 ± 18.02
Forest	0.05	-29.83 ± 222.46	-151.18 ± 304.25	-367.75 ± 267.31	-267.42 ± 203.46	-181.82 ± 163.4
	0.10	196.82 ± 198.16	51.23 ± 275.67	-198.23 ± 206.55	-71.48 ± 134.11	-11.37 ± 95.93
	0.15	355.85 ± 184.95	231.59 ± 249.27	-38.34 ± 177.71	53.45 ± 107.85	105.97 ± 72.67
	0.20	523.46 ± 165.59	482.76 ± 179.07	242.87 ± 164.76	307.56 ± 107.06	229.7 ± 50.59
	0.25	756.43 ± 156.15	751.86 ± 162.06	573.52 ± 156.47	493.94 ± 78.28	350.95 ± 43.45
	0.5	1437.75 ± 119.00	1485.83 ± 83.09	1366.29 ± 102.55	1244.37 ± 69.88	957.45 ± 25.99
Population	0.05	-54.52 ± 145.67	-122.11 ± 201.67	-215.18 ± 282.34	-202.21 ± 166.13	-154.92 ± 139.03
	0.10	47.22 ± 142.73	10.4 ± 147.42	-7.44 ± 149.98	-82.36 ± 139.5	-40.6 ± 110.42
	0.15	126.16 ± 136.79	129.71 ± 112.21	99.94 ± 121.06	7.84 ± 116.75	48.0 ± 84.91
	0.20	213.52 ± 121.23	213.94 ± 107.48	172.36 ± 125.59	150.56 ± 103.08	185.98 ± 70.19
	0.25	288.26 ± 105.29	285.93 ± 118.84	241.63 ± 125.67	250.5 ± 99.98	273.0 ± 52.72
	0.5	681.34 ± 103.15	645.12 ± 127.37	856.13 ± 30.42	817.33 ± 26.46	748.70 ± 17.05

Table D.2: Average discounted utility results for the Paradise instance evaluated from 100 simulations. Results for all λ value combinations between the DPV heatmap and the NV layer are presented by dominating utility category (column 1) and treatment fraction level (column 2).

Utility	tf	$\Delta U(0)$	$\Delta U(0.25)$	$\Delta U(0.5)$	$\Delta U(0.75)$	$\Delta U(1)$
	0.05	178.43 ± 318.39	219.03 ± 277.09	271.39 ± 171.98	314.29 ± 106.98	351.16 ± 69.99
	0.10	709.31 ± 257.46	713.05 ± 247.5	772.21 ± 135.73	763.92 ± 94.98	770.0 ± 65.49
Access	0.15	1170.32 ± 257.5	1182.01 ± 235.67	1236.91 ± 133.02	1201.7 ± 94.7	1188.84 ± 66.45
	0.20	1628.9 ± 254.39	1647.97 ± 215.35	1666.09 ± 131.14	1639.13 ± 95.41	1585.71 ± 72.64
	0.25	2074.37 ± 243.06	2096.7 ± 200.28	2079.68 ± 128.94	2046.8 ± 89.94	1981.67 ± 74.27
	0.5	4107.38 ± 142.12	4121.74 ± 152.96	4139.48 ± 51.61	3974.30 ± 15.38	3440.25 ± 0.93
	0.05	170.0 ± 167.37	119.33 ± 212.13	112.99 ± 202.85	79.86 ± 169.74	92.99 ± 144.35
	0.10	501.83 ± 154.96	395.71 ± 203.75	406.94 ± 205.52	377.67 ± 163.68	326.32 ± 146.97
Carbon	0.15	880.92 ± 119.85	829.95 ± 143.02	749.78 ± 139.05	621.85 ± 66.66	594.98 ± 62.37
	0.20	1136.25 ± 115.12	1111.82 ± 140.39	1000.25 ± 82.06	851.37 ± 35.68	770.77 ± 30.33
	0.25	1366.23 ± 118.15	1322.46 ± 141.34	1267.54 ± 79.92	1079.41 ± 23.32	930.57 ± 13.21
	0.5	2325.86 ± 75.56	2263.15 ± 73.76	2242.70 ± 17.81	2128.67 ± 8.69	1917.48 ± 1.24
	0.05	146.99 ± 253.0	96.19 ± 298.17	126.65 ± 192.1	139.03 ± 133.54	163.79 ± 96.69
	0.10	517.53 ± 254.44	463.82 ± 299.69	454.8 ± 188.2	430.92 ± 131.2	402.6 ± 97.28
Equal	0.15	869.89 ± 250.2	814.78 ± 293.29	846.11 ± 114.27	716.38 ± 133.36	658.84 ± 98.7
	0.20	1338.95 ± 210.91	1296.81 ± 237.16	1148.0 ± 85.11	1026.82 ± 31.18	906.99 ± 103.73
	0.25	1740.49 ± 207.27	1744.59 ± 192.15	1516.46 ± 71.49	1302.24 ± 21.32	1154.24 ± 107.06
	0.5	3295.83 ± 103.08	3286.24 ± 104.47	3190.06 ± 45.89	2953.84 ± 13.94	2390.28 ± 1.29
	0.05	216.81 ± 202.38	249.91 ± 204.67	213.87 ± 184.86	198.49 ± 157.21	190.93 ± 134.61
	0.10	683.54 ± 154.58	694.02 ± 161.88	624.48 ± 165.31	545.56 ± 150.06	479.31 ± 137.23
Forest	0.15	1024.93 ± 150.04	1017.27 ± 141.46	990.61 ± 110.4	848.03 ± 62.04	778.14 ± 56.25
	0.20	1298.57 ± 148.71	1296.31 ± 142.27	1250.72 ± 69.61	1115.27 ± 37.33	991.0 ± 28.6
	0.25	1528.34 ± 140.52	1562.9 ± 106.9	1500.39 ± 60.08	1395.67 ± 31.94	1192.01 ± 12.08
	0.5	2598.65 ± 64.54	2566.28 ± 63.29	2551.14 ± 17.53	2452.89 ± 8.03	2296.49 ± 0.90
	0.05	324.29 ± 70.44	268.57 ± 125.4	-26.16 ± 106.26	-12.03 ± 91.99	-0.25 ± 86.53
	0.10	695.32 ± 71.5	630.53 ± 129.06	354.47 ± 104.33	91.32 ± 87.87	93.99 ± 82.53
Population	0.15	955.1 ± 73.52	889.47 ± 127.92	707.54 ± 105.29	183.65 ± 88.69	173.49 ± 85.17
	0.20	1164.93 ± 72.59	1101.32 ± 123.66	928.23 ± 102.43	418.05 ± 93.16	241.01 ± 90.85
	0.25	1333.8 ± 69.21	1279.45 ± 69.42	1106.89 ± 103.3	811.5 ± 93.61	319.53 ± 93.13
	0.5	2007.73 ± 35.38	1958.83 ± 62.23	1852.77 ± 114.98	1741.65 ± 78.41	708.24 ± 3.74

Table D.3: Average discounted utility results for the Getty center instance evaluated from 100 simulations. Results for all λ value combinations between the DPV heatmap and the NV layer are presented by dominating utility category (column 1) and treatment fraction level (column 2).

Utility	tf	$\Delta U(0)$	$\Delta U(0.25)$	$\Delta U(0.5)$	$\Delta U(0.75)$	$\Delta U(1)$
	0.05	-472.96 ± 453.7	-392.1 ± 417.92	-262.43 ± 303.32	-139.59 ± 230.58	-80.92 ± 201.81
	0.10	-353.71 ± 443.07	-227.01 ± 375.65	-72.71 ± 266.06	4.14 ± 189.54	69.61 ± 150.55
Access	0.15	-205.64 ± 434.12	-75.1 ± 364.49	82.81 ± 213.83	160.22 ± 136.52	211.05 ± 103.45
	0.20	-64.41 ± 431.02	82.69 ± 350.25	225.15 ± 184.66	284.95 ± 120.86	330.84 ± 74.25
	0.25	125.28 ± 407.03	211.91 ± 338.75	355.09 ± 159.94	406.88 ± 97.7	445.96 ± 51.63
	0.5	905.18 ± 291.65	949.64 ± 256.68	971.89 ± 138.03	967.81 ± 70.83	988.54 ± 24.95
	0.05	-205.02 ± 215.85	-252.35 ± 283.41	-250.66 ± 255.97	-177.37 ± 220.13	-151.03 ± 213.97
	0.10	-95.77 ± 205.28	-131.89 ± 259.1	-85.9 ± 175.91	-83.15 ± 171.28	-47.19 ± 142.25
Carbon	0.15	10.27 ± 191.77	-48.71 ± 245.27	7.7 ± 146.14	33.27 ± 111.25	50.12 ± 103.33
	0.20	106.72 ± 176.78	51.74 ± 222.78	112.5 ± 140.3	104.59 ± 82.4	114.97 ± 82.65
	0.25	183.74 ± 174.64	121.02 ± 217.3	187.24 ± 118.38	170.12 ± 69.47	178.04 ± 51.78
	0.5	578.29 ± 149.24	569.52 ± 136.31	514.78 ± 94.54	492.50 ± 48.50	467.99 ± 28.17
	0.05	-291.81 ± 270.13	-340.11 ± 348.64	-267.66 ± 274.53	-156.46 ± 214.39	-121.35 ± 200.53
	0.10	-182.42 ± 264.34	-238.24 ± 329.64	-127.19 ± 229.6	-52.0 ± 168.74	-18.22 ± 153.08
Equal	0.15	-77.81 ± 262.55	-137.06 ± 321.14	10.99 ± 165.64	64.96 ± 117.05	82.66 ± 107.17
	0.20	11.86 ± 255.07	-37.58 ± 311.16	96.22 ± 150.43	146.98 ± 100.09	170.95 ± 82.78
	0.25	96.76 ± 246.77	31.2 ± 307.0	203.37 ± 136.78	225.95 ± 82.37	244.5 ± 52.99
	0.5	636.70 ± 186.57	633.69 ± 165.57	589.46 ± 112.89	600.49 ± 54.09	599.59 ± 26.80
	0.05	-154.77 ± 168.04	-216.76 ± 232.17	-285.57 ± 288.36	-226.0 ± 244.0	-164.97 ± 214.63
	0.10	-56.82 ± 156.05	-114.55 ± 208.5	-133.43 ± 205.9	-115.09 ± 178.6	-72.47 ± 144.47
Forest	0.15	19.91 ± 144.58	-26.92 ± 189.24	-45.63 ± 158.45	-11.54 ± 119.47	15.09 ± 103.73
	0.20	96.95 ± 133.05	36.43 ± 176.55	38.2 ± 128.73	48.1 ± 96.83	67.37 ± 77.59
	0.25	147.82 ± 131.6	110.52 ± 165.47	115.68 ± 120.95	108.12 ± 77.0	125.16 ± 53.59
	0.5	474.79 ± 99.54	474.01 ± 102.52	390.05 ± 94.46	369.58 ± 54.05	359.10 ± 28.73
	0.05	-70.86 ± 65.58	-112.05 ± 109.7	-193.76 ± 204.4	-173.36 ± 196.66	-170.81 ± 198.96
	0.10	-43.46 ± 64.26	-70.31 ± 89.42	-136.47 ± 160.56	-115.5 ± 150.06	-112.4 ± 151.19
Population	0.15	-16.44 ± 64.65	-27.52 ± 66.72	-75.13 ± 113.62	-59.41 ± 105.36	-56.56 ± 105.75
	0.20	7.38 ± 61.85	-4.0 ± 59.09	-38.04 ± 93.7	-21.94 ± 85.87	-19.29 ± 84.26
	0.25	35.15 ± 59.66	22.34 ± 55.45	-2.93 ± 74.14	14.77 ± 55.57	18.24 ± 53.49
	0.5	158.55 ± 43.87	123.64 ± 43.91	106.78 ± 48.59	119.83 ± 32.95	123.32 ± 28.41

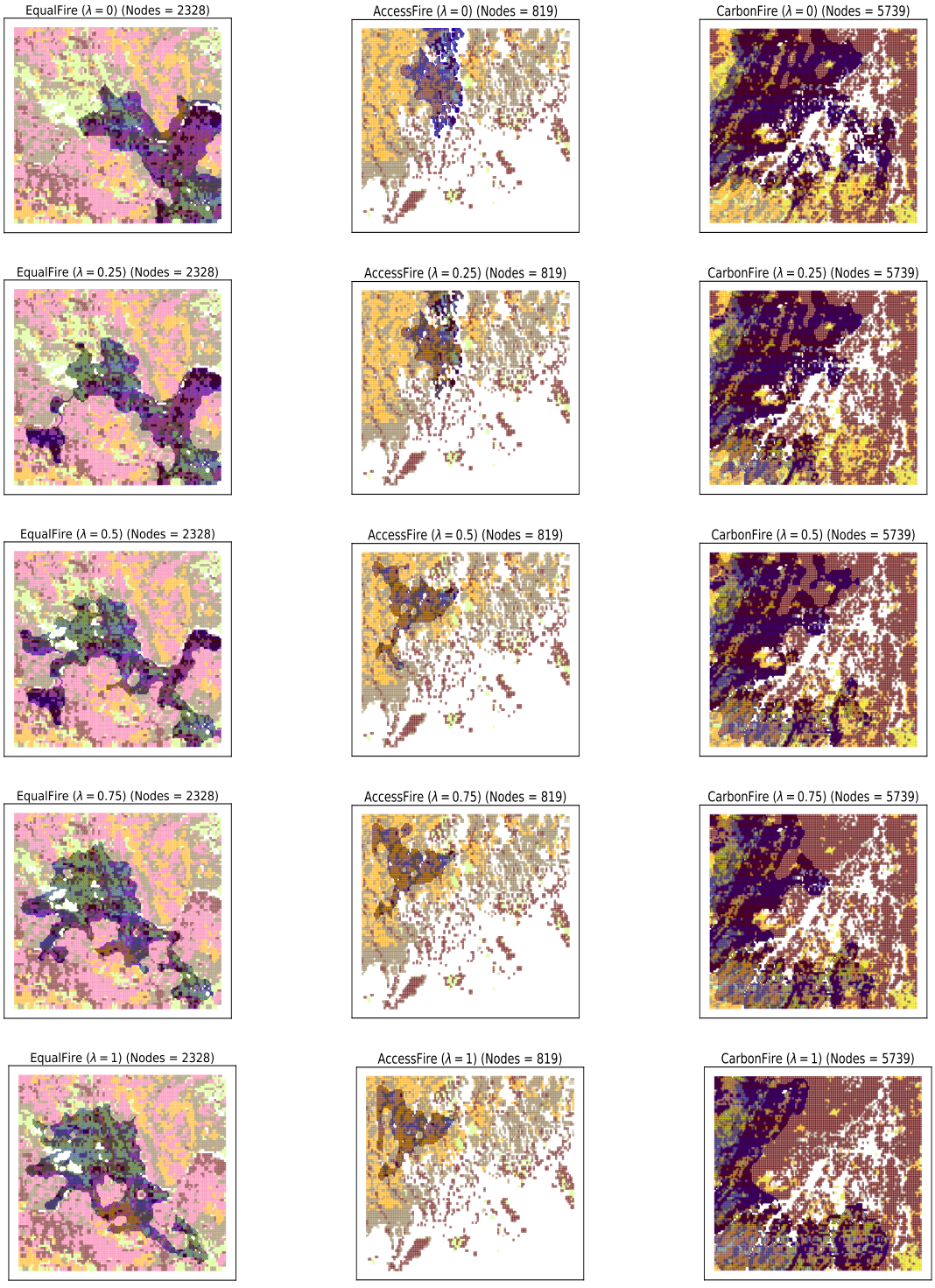


Figure D.1: Resource allocation sample plans for Napa valley, Getty center, and Paradise (columns) instances for different λ weights (rows) to account for the expected losses due to future wildfires, at a specific treatment fraction tf. Significant variations in the optimal plans can be observed as the λ values are modified to include future wildfire risk into the objective function. Original land cover colors have been modified for better contrast and non-flammable nodes have been removed (white space).