Network Optimization of Food Flows in the U.S.

Caleb Robinson*, Arezoo Shirazi†, Mengmeng Liu‡, and Bistra Dilkina*
*School of Computational Science and Engineering
†School of Building Construction
‡School of Civil and Environmental Engineering
Georgia Institute of Technology
Email: dcrobins,arezoo.shirazi,mliu301,bdilkina6@gatech.edu

***Abstract*—The world food system is a globalized and interconnected network, and prior research has shown the US is one of the most important countries within this global network. Considering this, any affects on the US food network might have far reaching implications. The objective of this paper is to show how the efficiency of the US food network can be optimized and to examine the tradeoffs between sustainability, efficiency, and resiliency in a network science based approach. To do this we use the 2012 US Commodity Flow Survey (CFS) data focusing on origin-destination flows for 5 categories of food. We interpret these origin-destination flows as a directed weighted network and use a linear programming (LP) based approach to reconfigure the flows in order to minimize the "food miles" in the network. We extend our LP to a multi-objective optimization problem in order to consider partially optimized solutions that are closer to the current food network. Finally, we calculate characteristics of the optimized networks and compare them to the original network. Our results show that the US food network has the potential for more than a 50% reduction in "food miles", which will result in a food system that is not only more economically efficient but also with a smaller sustainability impact by reducing the corresponding transportation GHG emissions.**

I. Introduction and related works

Food security is becoming more and more important from the economic and political perspectives due to population and consumption growth, as well as climate change [1]. In particular, sustainability in food supply and demand is playing a crucial role in the world's economic development plans. That is why it is highly important to understand the structure, efficiency, sustainability and resilience of food trade systems at both national and international scale.

One way to do this is to take a network science approach and measure key system characteristics such as efficiency, vulnerability, resiliency, connectivity, etc. using appropriate network metrics. Network analysis is one of the powerful methods to analyze and quantify complex systems. More specifically, its flexible nature enables it to represent many real world domains such as transportation systems [2], combined energy systems [3], trade networks [4], vehicular accessibility [5] and many other integrated platforms.

A previous study on the complexity of the International agro-Food Trade Network (IFTN) recognized the world food system as a globalized and interconnected network and the US as one of its dominant nodes. Based on this study, the US was identified as one of the top seven hotspots in the IFTN, with high values of betweenness centrality in the world trade network, which indicates that changes in its status would result in large global impacts [6]. Following this study, other investigators analyzed the US domestic food flow network using the 2007 Commodity Flow Survey (CFS) data and compared how its properties are different from the global food trade network [7].

In this paper, we study the US domestic food flow network by taking advantage of the recently released 2012 CFS dataset [8]. We represent the CFS data as food flow networks where CFS areas are nodes and exports of specific food classes from one area to another correspond directed weighted edges. We then optimize the food flows to minimize the total ton-miles traveled by food in the network. Minimizing ton-miles increases efficiency by saving transportation costs, as well as improves sustainability by reducing the corresponding GHG emissions. The results of our study demonstrate the potential of a 50% more efficient US food flow network in terms of "food miles", with 28-74% savings depending on the food class. Previous research in diverse applications has identified important tradeoffs between efficiency and resilience in networks [9]. To this end, we formulate the problem of optimizing food networks as a multi-objective optimization problem in order to produce solutions of varying qualities, and our results show that these tradeoffs do exist in the food system setting. In particular, we compare the efficiency and resiliency of our optimized networks to the original baseline network using common network properties.

Based on the 2015 estimated US energy consumption flow chart [10], around 30% of US energy consumption goes to the transportation sector. In addition, the 2014 US Environmental Protection Agency (EPA) report shows that Greenhouse Gas (GHG) emissions from transportation accounted for about 26% of total U.S. GHG emissions [11]. Another study on the US food-mile climate impacts (which was conducted based on 2007 CFS data) showed that food transportation in the US is responsible for more than 15% of GHG emissions [12]. Their analysis also showed that the total food weight accounted for in the US has almost doubled within the past 5 years (from 2007 to 2012). Looking through all the provided statistics and information, and considering the increasing demand of food, any reduction in food miles incurred for transportation and distribution would also decrease the US GHG emissions and any other climate impacts associated with it.

Our results can be expanded to consider directly several

978-1-4673-9005-7/16/$31.00 ©2016 IEEE

other conditions and criteria, such as transportation modes, cost and emission analysis, temperature controlled shipments, etc. to make the model more realistic and useful to practitioners. As an example, previous studies developed an inventory/transportation model specifically for cold items [13], which can particularly be applied to agro-food shipments as well. In their model, the authors considered not only the cost of holding and shipping the cold items, but the emission from each sector, to be able to study the trade-off between the cost and emission objective functions. This was further expanded, to consider multiple types of products [14]. Consideration of alternate scenarios in studies such as this one can help transportation planners make more informed decisions.

We believe that the results of this study, when combined with the recent sustainable inventory models and supply chain logistics, can contribute to the design of a more economically and environmentally sustainable food mile distribution systems within the US food flow network. More broadly, this work is an example of the use of data-driven and algorithmic approaches to inform decisions and policies related to sustainability, which is at the heart of the growing new research field of Computational Sustainability [15].

II. Methodology

A. Data

The data used in this study was obtained from the 2012 Commodity Flow Survey (CFS) Public Use Micro-data (PUM) file, which is published every five years as part of the Economic Census by the US Census Bureau [8]. The dataset includes information about all the different commodities transported within the US. The following five categories of food commodities, based on the Standard Classification of Transported Goods (SCTG), were selected for further analysis in this study: (02) Cereal Grains (includes seed), (03) Agricultural Products (excludes Animal Feed, Cereal Grains, and Forage Products), (04) Animal Feed, Eggs, Honey, and Other Products of Animal Origin, (05) Meat, Poultry, Fish, Seafood, and Their Preparations, and (07) Other prepared foodstuffs, and fats and oils. These five categories are also aligned with the previous US food network analysis conducted on the 2007 CFS data [7].

The origins and destinations of commodity flows reported in the PUM file are defined on the state level, CFS area level, and metro area level in the PUM file. For the purpose of this study, CFS areas were selected because they are the smallest geographic areas out of the three. Based on the 2012 CFS data user guide [8], 132 CFS areas were defined in 2012, an increase from the 123 CFS areas reported earlier in 2007. Detailed geographical representation of the 2012 CFS areas are shown in Figure 1. In addition, the comparison between 2007 and 2012 food network characteristics are presented in Table I. The differences between the 2007 and 2012 data are largely due to the better quality and completeness of data in the more recent dataset.

The PUM file includes different variables for all usable shipment records (rows of data) collected by the CFS. The particular variables used in this study and their detailed descriptions are presented in Table II.

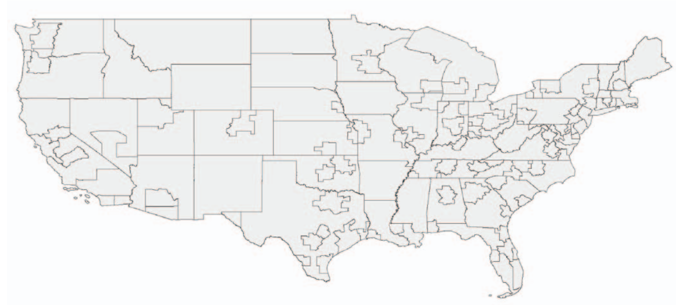


Fig. 1. Map of the CFS Areas in the continental US

TABLE I
FOOD NETWORK COMPARISON OF 2007 VS 2012 CFS DATA

Network Properties	2007 Data	2012 Data
Number of nodes $\lvert V\rvert$	123	132
Number of links $\lvert E\rvert$	4,198	11,649
Graph Density (directed)	0.28	0.674
Average Degree $< k >$	34.1	176.5

TABLE II
VARIABLE DESCRIPTIONS

Variable	Unit	Description
i	-	index of CFS area of shipment origin
j	-	index of CFS area of shipment destination
k	-	index of SCTG food category/class
m	-	index of shipment type between two CFS areas
$f^k_{i,j,m}$	pounds	weight of shipment type m from i to j for class k
$w^k_{i,j,m}$	scalar	an estimate of the total number of shipments of type m from i to j for class k
$d^k_{i,j,m}$	miles	routing distance of shipment type m from i to j for class k
$D^k_{i,j}$	miles	average routing distance for each ton shipped from i to j for class k
$F^k_{i,j}$	tons	weight of total shipments from i to j for class k
$G^k_{i,j}$	tons$*$miles	ton miles of goods shipped from i to j for class k

Considering the five SCTG food categories, there are 370,223 shipment records where each shipment record gives the indicated variables between an origin and destination for a particular category of food. For some pairs of CFS areas there will be multiple entries, each representing a different type of shipment. We aggregate all different types of shipments involving a given food class k between two CFS areas i and j following Equation 1, and obtain a matrix of flows F^k for each food class. An entry in this matrix represents the total weight of shipments from one area to another. Note that in Equation 1 the weight is divided by 2000 to convert the unit of the variable to ton for further analysis. Similar to Equation 1, we calculate the average number of miles traveled by each ton of food in Equation 2, to obtain a matrix of routing distances D^k. Finally, the element-wise multiplication of F^k and D^k, shown in Equation 3, will give us a matrix G^k where entries represent the number of "ton miles" or "food miles" incurred by the food flow from one CFS area to another. Our analysis in the remainder of the paper will be in terms of these food miles.

$$F_{i,j}^k = \sum_m \frac{f_{i,j,m}^k}{2000} w_{i,j,m}^k \tag{1}$$

$$D_{i,j}^k = \frac{\sum_m d_{i,j,m}^k * w_{i,j,m}^k * \frac{f_{i,j,m}^k}{2000}}{F_{i,j}^k} \tag{2}$$

$$G_{i,j}^k = F_{i,j}^k * D_{i,j}^k \tag{3}$$

B. Network Optimization

The k matrices, G^k, that we have calculated in Equation 3 can be treated as adjacency matrices and interpreted as networks, where nodes are CFS areas, and weighted directed edges are the number of food miles traveled between areas. Our goal is to redistribute *where* the goods in the food network are shipped, thus to minimize the number of food miles in the resulting network. In other words, we want to maintain the same number of imports and exports for each node (in terms of tons), while reorganizing *where* those imports and exports come and go from, in the most efficient manner (in terms of ton-miles).

We can frame this optimization problem as a linear program and solve it efficiently using off-the-shelf solvers. Formally, our optimization problem will look for a new food weight adjacency matrix, F^{k*}, that minimizes ton miles, under the constraints that F^{k*} has to have the same row and column sums as the original food weight adjacency matrix, F^k. Our objective function in this case is given in Equation 4, with the constraints in Equations 5, 6, and 7. In this formulation, each entry of the optimized matrix, $F_{i,j}^{k*}$, is represented by a variable, $x_{i,j}^k$.

$$\min_{\mathbf{x}^k} \sum_{i,j} x_{i,j}^k * D_{i,j}^k \tag{4}$$

s.t.

$$\sum_i x_{i,j}^k = \sum_i F_{i,j}^k \quad \forall j \in 1 \dots n, \tag{5}$$

$$\sum_j x_{i,j}^k = \sum_j F_{i,j}^k \quad \forall i \in 1 \dots n \tag{6}$$

$$x_{i,j}^k \geq 0 \quad \forall i \in 1 \dots n, \forall j \in 1 \dots n \tag{7}$$

C. Multi-Objective Network Optimization

The result of the optimization problem defined in the previous section is a re-distribution of the shipments of goods such that the food miles traveled by all goods is minimized. As shown in Section II-D, the optimization results in a very sparse yet efficient network. This extremely optimized network can result in several "real world" problems. First, the sparsity of the network can mean the network will be more vulnerable to node failure (e.g. a hurricane disrupting food supply from a southeastern state). Secondly, a fully optimized (and hence largely consolidated) network might not be able to adapt easily to a change in demands for certain goods in the future. Finally, moving from the current *status quo* to the fully optimized network will likely not be feasible to implement, as it would involve a collaboration with many entities.

All of these problems point towards needing a range of "partially optimized" solutions. To this end we can reformulate our optimization problem into a multi-objective optimization that allows us to interpolate between the original and optimal distributions of goods. For the multi-objective optimization we keep the same objective function as before, however introduce an ϵ-constraint that forces the L_1 distance between the original and optimized solution to be below a parameterized threshold. By changing the value of the parameter, ϵ, we can get solutions of varying quality.

Equations 8 through 14 show the formulation of our multi-objective optimization problem. In the additional constraints 12 - 14, the u variables are added to linearize the L_1 distance between optimized food flows $\mathbf{x}^k$ and the original food flows F^k, i.e. $u_{i,j} = |F_{i,j} - x_{i,j}|$ for a given i and j.

$$\min_{\mathbf{x}^k, \mathbf{u}} \sum_{i,j} x_{i,j}^k * D_{i,j}^k \tag{8}$$

s.t.

$$\sum_i x_{i,j}^k = \sum_i F_{i,j}^k \quad \forall j \in 1 \dots n, \tag{9}$$

$$\sum_j x_{i,j}^k = \sum_j F_{i,j}^k \quad \forall i \in 1 \dots n \tag{10}$$

$$x_{i,j}^k \geq 0 \quad \forall i \in 1 \dots n, \forall j \in 1 \dots n \tag{11}$$

$$\sum_{i,j} u_{i,j} \leq \frac{100 - \epsilon}{100} * \sum_{i,j} F_{i,j}^k \tag{12}$$

$$u_{i,j} \geq F_{i,j}^k - x_{i,j}^k \quad \forall i \in 1 \dots n, \forall j \in 1 \dots n \tag{13}$$

$$u_{i,j} \geq x_{i,j}^k - F_{i,j}^k \quad \forall i \in 1 \dots n, \forall j \in 1 \dots n \tag{14}$$

For the multi-objective optimization we use 100 evenly spaced values of ϵ in the range from 0 to 100, effectively letting the optimized solutions vary from a fully optimized configuration to the original configuration.

D. Network Analysis

The result of the single objective optimization problem is a network for each of the k classes of goods, denoted as F^{k^*} in our optimization equation. The results of the multi-objective optimization step are similar networks for each of the k classes of goods, for each value of ϵ that we consider. Intuitively, both groups of these networks are re-configurations of the tons of goods shipped between CFS areas that minimize total ton miles.

In order to analyze the effects that improving the "food mile" has on the food network efficiency as a whole we: 1) aggregate the individual networks for each food class into a summed network, for both the *optimized* and *original* configurations, then 2) compare the properties of the aggregate *optimized* network to the aggregate *original* network.

We consider the weighted directed network for each food class k, where edge weights $G_{i,j}^{k*} = F_{i,j}^{k*} D_{i,j}^{k}$ correspond to food miles traveled by food class k from CFS i to j. We aggregate the five optimized networks into an *aggregate optimized food network* with edge weights corresponding to the total food miles across all food types, $G_{i,j}^{*} = \sum_k (F_{i,j}^{k^*} D_{i,j}^{k})$. We perform a similar transformation for the original food flows considering the per class food miles original networks G^k and the *aggregate original network* G. In these aggregate networks, nodes represent CFS areas, and directed edges have weights equal to food miles of combined food flow between areas. This representation lets us comment on the wider effects (in terms of reduction of environmental emissions) that the decrease in shipping tons might bring about.

For the network comparison, we have chosen to examine the following network properties: number of nodes, number of edges, average degree and average weighted degree, density, number of strongly connected components (SCCs), number of weakly connected components (WCCs), transitivity, reciprocity, bipartite betweenness centrality, and bipartite average clustering. These properties allow us to quantitatively comment on the differences between the two networks. It should be noted that we do *not* examine centrality measures that use shortest path calculations, as the CFS data only describes shipments in terms of 'single-hops'; there is no data to suggest that a shipment from area i to j will carry on to a further area k. To calculate the bipartite betweenness centrality and bipartite average clustering measures we convert the network into a two-mode format using the following procedure. Given an adjacency matrix, A, of size $n \times n$, we create a new adjacency matrix, $A' = [0, A; A, 0]$, of size $2n \times 2n$. This is akin to creating a new node u' for each node u in the original network, then replacing each edge, (u, v), in the original network, with edges (u, v') and (u', v). A formal definition of the properties that we examined are defined as follows:

Percent Improvement for this network is defined as the percentage of improvement in reducing the total ton-miles of flow in the US. It is calculated as shown in Equation 15.

$$\alpha = \left(\frac{ton * mile_{original} - ton * mile_{optimized}}{ton * mile_{original}}\right) * 100 \tag{15}$$

Average Degree $< k >$ is the average number of neighbors of a node in a network. This property gives us an idea of how many other areas a particular CFS area interacts with.

Average Weighted Degree$< k^w >$ is the number of edges connected to each node weighted by the edge weights of those edges. As an example, the weighted-in node degree (k^w_{in}), measures the total ton miles of imports to the node, and the weighted-out node degree (k^w_{out}), measures the total ton miles of exports from the node. In general, the average weighted in degree and average weighted out degree of a directed network the same. Therefore, the average weighted node degree for the entire network is the average of the weighted (either in or out) node degrees.

Network Density quantifies how interconnected the network nodes are by measuring the ratio of the number of edges $|E|$ to the number of possible edges, as shown in Equation 16 for a directed network.

$$\phi = \frac{E}{V(V-1)} \tag{16}$$

Connectedness measures the way in which a network is connected. Two classification of connectedness in networks are defined as:

- Weakly Connected Component (WCC): A collection of nodes in which there exists a path from any node to any other, regardless of the direction of the edges.
- Strongly Connected Component (SCC): A collection of nodes in which there exists a directed path from any node to any other.

Transitivity is a measure of the fraction of all the possible triangles that are present in a graph. Possible triangles are identified by the number of "triads" (two edges with a shared nodes). Formally this is:

$$T = \frac{3 * \text{Number of triangles}}{\text{Number of connected triplets of vertices}} \tag{17}$$

Reciprocity is defined on directed graphs as the fraction of the number of edges that are present in both directions between a pair of nodes over the total number of edges.

$$r = \frac{|\{(u,v)|\exists(u,v) \in E, \exists(v,u) \in E\}|}{|E|} \tag{18}$$

(Bipartite) Average Betweennesss Centrality $< b_{avg} >$ measures the number of times a node k acts as a bridge along the shortest paths between two other nodes i, j. This number is then normalized by the maximum possible value which for bipartite graphs is limited by the relative size of the two node sets [16]. To this end, the b_k should be divided by either $b_{v_1 max}$ or $b_{v_2 max}$ as shown in Equations 19 and 20, depending on whether node i belongs to node set V_1 or V_2 (the two sets of nodes making up the bipartite network). The bipartite average betweenness centrality for the entire network is the average of the b_k of all the nodes.

$$b_{v_1 max} = \frac{1}{2}[n_2^2(s+1)^2 + n_2(s+1)(2t-s-1) - t(2s-t+3)], \tag{19}$$

where

$$s = (n_1 - 1)/n_2,$$
$$t = (n_1 - 1) \bmod n_2$$

$$b_{v_2max} = \frac{1}{2}[n_1^2(p+1)^2 + n_1(p+1)(2r-p-1) - r(2p-r+3)], \quad (20)$$

where

$$p = (n_2 - 1)/n_1,$$
$$r = (n_2 - 1) \bmod n_1$$

(Bipartite) Average Clustering Coefficient $< c_{avg} >$ measures the ratio of existing links connecting a node's neighbors to each other to the maximum possible number of such links. However, there can be no triangle in a bipartite graph. Hence, a clustering coefficient metric is defined for pairs of nodes (in the same set V_1 or V_2) as shown in Equation 21 which captures the overlap between neighborhoods of nodes. Consequently, the clustering coefficient of one node at the average of its clustering coefficients with other nodes is defined as represented in Equation 22 where $N(N(u))$ are the second order neighbors of u in the bipartite graph excluding u [17]. Finally, The bipartite average clustering coefficient for the entire network is the average of the c_u of all the nodes.

$$c_{uv} = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (21)$$

$$c_u = \frac{\sum_{v \in N(N(u))} c_{uv}}{|N(N(u))|} \quad (22)$$

III. Results

In this section, we present our optimization results, the corresponding network properties analysis, and discussion on the resiliency of the resulting networks. Table III summarizes the key results in terms of optimizing the food miles in all of the networks. The efficiency of the food network in terms of the total food miles can be increased for all of the five food categories, and can be increased more than 50 percent in the aggregate optimized network over the original network. This indicates that the optimized network has 50 percent less ton miles of food traveling within it, which could result in huge savings in both transportation costs and CO2 emissions, and consequently in a more economically and environmentally sustainable food network system within the US.

To conduct a simple calculation based on the EPAs regulations and standards, a typical heavy duty tractor's (class-8 day-cab with high roof) baseline CO2 emission is around 98 grams of CO2 per ton-mile [18]. Hence, switching to our fully optimized US food network, results in a huge reduction of around 0.26 megatons of CO2 emissions per year in the US. Based on 2007 CFS data, it was previously shown that that food transportation in the US is responsible for more than 15% of GHG emissions [12], hence such reductions constitute a significant reduction in the overall US emissions footprint.

TABLE III
Results of network optimization on food ton miles

Class	Original Ton Miles	Optimized Ton Miles	% Improvement
2	174,156,490,610	124,849,999,605	28.31%
3	97,442,455,171	33,196,525,701	65.93%
4	54,461,071,730	24,677,048,286	54.69%
5	43,438,581,969	15,731,447,048	63.78%
7	196,462,265,497	49,157,552,077	74.98%
All	565,960,864,977	247,612,572,716	56.25%

The results from the network analysis of the original and optimized networks can be seen in Table IV and Figures 2, 3 and 4. In Table IV we show how the network properties differ between the original and fully optimized networks over each of the different classes of goods.

From Table IV, we observe that the number of edges and consequently the network density decreased dramatically in the optimized networks, which means that higher volumes of food can be transported together across the shortest distances to decrease the total miles traveled in the network. The average weighted outgoing degree of nodes also decreases in the optimized networks, showing that inefficiencies, where the same food class is both imported and exported by a given CFS area, are reduced. In addition, we can see that the original aggregate network combining the food flows of all 5 food classes considered, G, has a single strongly connected component (SCC), i.e. there is a directed food-dependency path from any CFS area to any other and vice versa. In the optimized network, G^*, there is still a single weakly connected component (WCC), however the number of strongly connected components increases from 1 to 38. We map the strongly connected components in Figure 2. This map shows that the aggregate US food network becomes segmented geographically. This spatial clustering result is to be expected when minimizing food miles as demands will be satisfied by the closest available supply in all cases. It also suggests that the demands from different geographic areas can be satisfied by short range trade, rather than cross continental shipments.

Table IV also shows that both transitivity and reciprocity factors are decreased after the optimization for each class' network. These measures indicate a key structural difference between the original and optimized networks. Transitivity values near zero in the optimized network suggest that there aren't many triangles in the network, which means that it is unlikely that a node's neighbors will interact with each other. Reciprocity values near zero in the optimized network mean that there aren't many pairs of nodes where shipments happen in both directions. Both the transitivity and reciprocity results indicate that the optimized network is *less* resilient to events that could disrupt the supply from a given node. In the original network if a random node is eliminated, the node's neighbors will most likely have other connections from which they can import a fraction of their demands. This is not the case in the

TABLE IV
NETWORK PROPERTIES OF ORIGINAL AND OPTIMIZED FOOD MILE NETWORKS

Name	$\lvert V \rvert$	$\lvert E \rvert$	$< k >$	$< k^w >$	Density	SCCs	WCCs	Transitivity	Reciprocity	Betweenness	Clustering
Aggregate Original Network	132	11,649	176.50	4,287,582,310	0.67	1	1	0.75	0.76	0.01	0.54
Aggregate Optimal Network	132	736	11.15	1,875,852,824	0.04	38	1	0.15	0.17	0.02	0.13
(2) Cereal Grains, Original	132	970	14.70	1,319,367,353	0.06	40	7	0.15	0.25	0.01	0.06
(2) Cereal Grains, Optimal	132	223	3.38	945,833,330	0.01	132	7	0.02	0	0.06	0.00
(3) Agricultural Products, Original	132	4,469	67.71	738,200,418	0.26	8	1	0.35	0.41	0.01	0.13
(3) Agricultural Products, Optimal	132	258	3.91	251,488,831	0.01	128	1	0.03	0.01	0.00	0.19
(4) Products of Animal Origin, Original	132	3,278	49.67	412,583,877	0.19	21	1	0.29	0.01	0.01	0.10
(4) Products of Animal Origin, Optimal	132	248	3.76	186,947,335	0.01	131	1	0.02	0.01	0.00	0.16
(5) Meat, Poultry, etc., Original	132	4,899	74.23	329,080,166	0.28	10	1	0.41	0.47	0.01	0.16
(5) Meat, Poultry, etc., Optimal	132	256	3.88	119,177,629	0.01	131	1	0.03	0.01	0.00	0.20
(7) Other prepared foodstuffs, Original	132	9,427	142.83	1,488,350,496	0.55	1	1	0.66	0.67	0.01	0.41
(7) Other prepared foodstuffs, Optimal	132	263	3.98	372,405,698	0.02	132	1	0.02	0	0.00	0.18

optimized network, where most nodes will get their demand from a few nearby nodes which themselves are not connected to each other. It follows that if a random node is eliminated in the optimized network, then there will be a much higher chance of a neighboring node not receiving a large fraction of its demand. In general, if nodes rely mainly on small numbers of shipping connections to fulfill their demand, as in the optimized networks, they will be vulnerable to shortages if any of their suppliers are disrupted. From another resiliency perspective, as shown in Figure 2, the potential cascading effects of disruptions will likely be more localized in the optimized network which has 38 SCC clusters of nodes versus the single cluster in the almost completely connected original network. The difference in number of components is even more stark when considering single food classes, going from 1 in the original to 132 in the optimized network for food class 7. So in the optimized networks one could expect that node disruptions will result in larger and harder to recover from cascading effects, but fewer and more localized neighboring nodes or areas will be affected. Note that considering resilience with respect to elimination of specific links is less realistic as food shipments travel over roads and any disruption to a road link could likely be corrected by taking an alternative route.

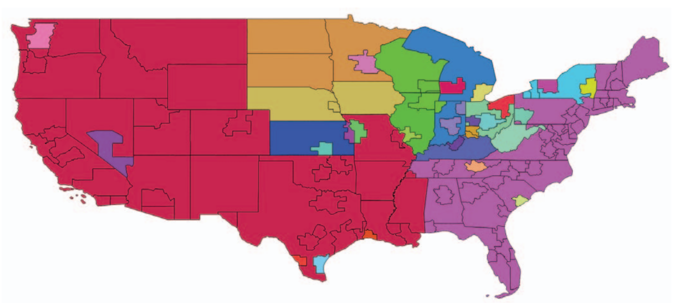


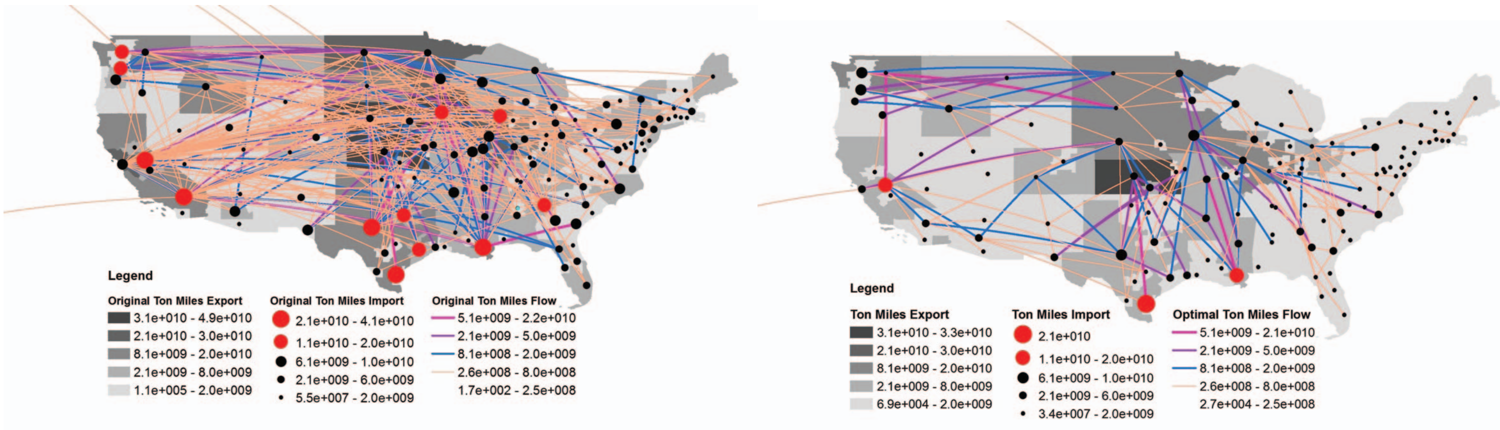
Fig. 2. Visualization of the strongly connected components in the *optimized* network G^*.

In Figures 3 and 4, we visualize the spatially-explicit structural differences between the original and optimized networks when edge weights are considered as either food miles (Fig. 3) or as food weight (Fig. 4). The food miles networks (Fig. 3) show the effect that ton mile optimization has on food mile edges. The optimized networks are much sparser than the original networks with larger fraction of heavy weight edges. The size of dots in the network maps represents the amount of food miles incurred by food imported by the corresponding CFS area. We can see that most CFS areas reduced their import food miles in the optimized network. Additionally, the number of CFS areas with very large food miles imports, which are represented by red dots on the map, are dramatically reduced. This indicates that our optimization network improves a lot, and that there are relatively more improvements among the CFS areas with larger original food miles imports. The shading of the CFS areas in the food miles networks represents the total exports (in terms of ton-miles). From the optimized map we can see that the export food mile footprint of an area generally increases the closer it is to the center of the US. This is not the case in the original network, as California has a relatively large export food mile footprint. In the food weight networks (Fig. 4), we observe that in the optimized network there are more edges with high food amounts but that there are much fewer long edges with significant food flows.

As discussed earlier in Section II-C, there are several problems with using a fully optimized network. This has led us to use a multi-objective optimization approach to generate partially optimized networks that minimize food miles.

In Figure 5 we examine the results of our multi-objective optimization for the class "Other prepared foodstuffs, and fats and oils". Although we are showing results for just this class of goods, we have observed similar results with the other classes. The five graphs show how the properties of partially optimized networks change as the optimized food network is constrained to be between 0% and 100% the same as the original. With no constraints, the optimized network will be the same as the network from the single-objective optimization result, and with a 100% constraint the resulting network will be the same as the original network.

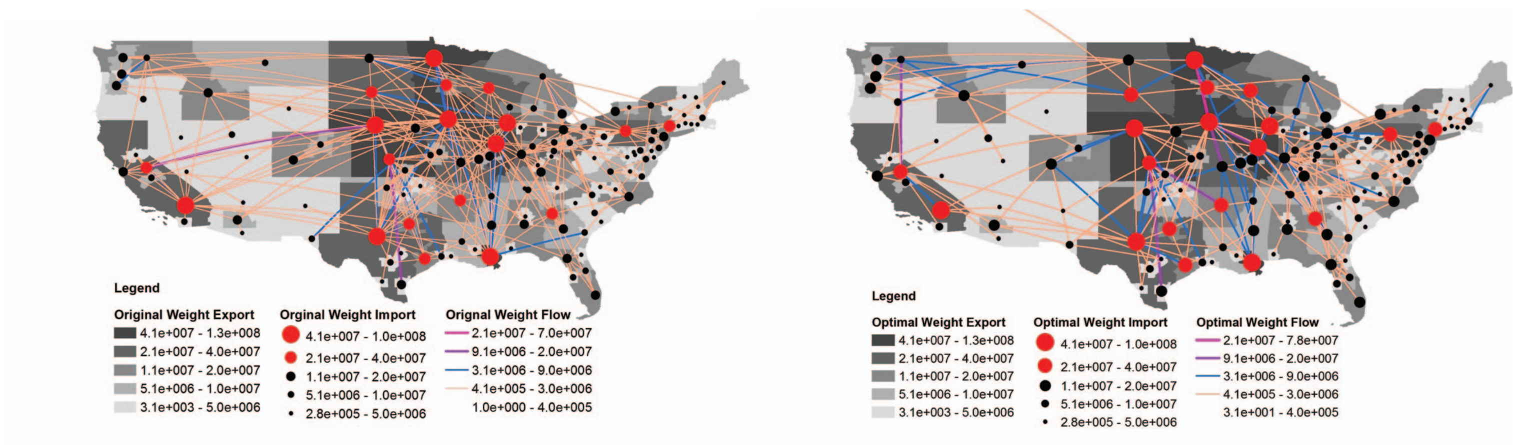
The top graph shows how the percent efficiency improvement (in food miles) achieved by the optimization step decreases as the problem becomes more constrained to the original. We also observe that the number of edges in the network increases as the problem becomes more constrained,



(a) **Original** ton miles network

(b) **Optimized** ton miles network

Fig. 3. Comparison of the original and optimized networks in terms of food mile flows. **(a)** Shows the original food miles network. **(b)** Shows the optimized food miles network. The edges in the smallest category are not shown for all maps in order to make the 'original' networks easier to interpret.



(a) **Original** food weight network

(b) **Optimized** food weight network

Fig. 4. Comparison of the original and optimized networks in terms of food weight. **(a)** Shows the original food weight network. **(b)** Shows the optimized food weight network. The edges in the smallest category are not shown for all maps in order to make the 'original' networks easier to interpret.

while the number of strongly connected components drop. This behaviour is expected considering the results of the single-objective optimization. In the more optimized scenarios the network is highly fragmented with almost every node constituting its own strongly connected component, with very few superfluous edges. From the "Export Food Weight' plot we can see that the more optimized networks have larger maximum weight edges, which means that a single link failure could cause larger food flow disruption than in less optimized ones. Such food networks would be very vulnerable to natural disasters, drought, or other extreme events. The loss of a single node or edge would mean, with high probability, that an entire area would be cut off from supply.

To present an alternate configuration to the unconstrained (i.e. fully optimized) scenario, we can examine the network that has been constrained to be 60% similar to the original. In the 60% constrained case, the percent improvement over the original network is still high, at $\approx 58\%$ (with a max of $\approx 75\%$ in the fully unconstrained case), however the resulting network will be much more resilient than the fully optimized one. There are less than 5 strongly connected components in the 60% constrained case, and an order of magnitude more edges than the fully optimized network. An increased number of edges means that the neighbors of a node will be more likely to have neighbors, preventing them from becoming fully cutoff with the loss of the original node. In general, these network properties suggest that the partially constrained result would be more capable of surviving some sort of extreme event without any node becoming completely cut off. Finally, this 60% constrained network is a reasonable alternate network to consider because of the clear "jump" in the "Max Export Food Miles per Edge" line in the "Export Food Miles' plot. As the constraint value changes from 50% to 70% there is a dramatic increase in the largest link in terms of food miles. By intelligently considering a constraint value that avoids this large increase, we are able to prevent a costly shipment route.

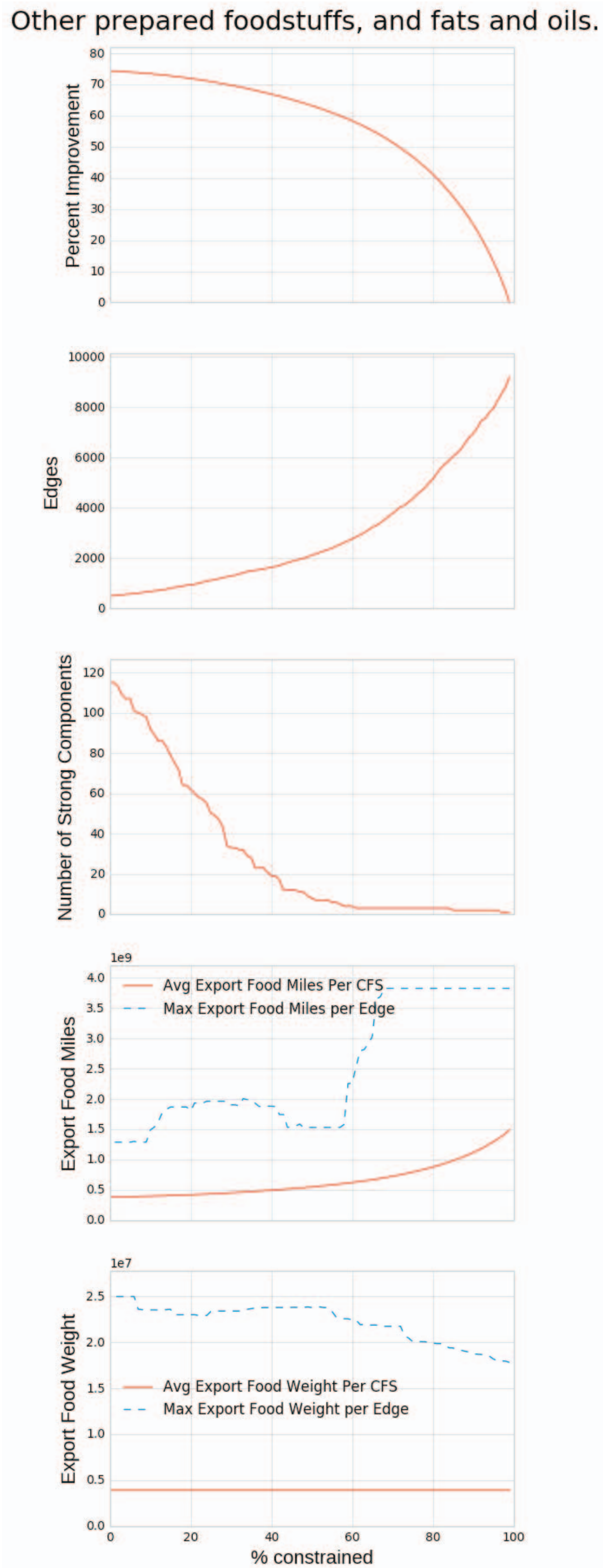


Fig. 5. Multi-objective optimization results for the category “Other prepared foodstuffs, and fats and oils” measuring different network properties, as the flows $\mathbf{x}$ are constrained to match between 0% and 100% of the original food network flows.

IV. Conclusion and Future Work

A. Conclusion

The current US food network has already been extensively analyzed in terms of its network properties and also within the larger context of the global food trade network [6], [7]. The goal of this study was not only to reanalyze the US food network using more recent data sets, but also to demonstrate possible ways of improving this network from the sustainability perspective. To this end, optimization strategies were adopted to improve the efficiency and sustainability of the system, while maintaining network resiliency. The results of the analysis indicate that by keeping around the 60% of our current network, assuming that the current network is a good representative of a resilient network, it is still possible to improve network’s efficiency and sustainability by around 60% over the original 2012 food network.

B. Future Work

There are many different future directions that this work can take. We have restricted our optimization and analysis to the 2012 CFS data, which could easily be extended to include 2007 data. The CFS data also has more information about shipments, including mode of transportation. We could further segment the 2012 dataset to achieve fine-grained results for shipments using each mode of transportation. Additionally, our methods could be applied to international trade and shipping data to get a better idea of how the global food network can be improved under similar assumptions to this study. On the methodological front, we can change our optimization formulation in several different ways including: using different objective functions, using different distance metrics, and relaxing the constraints that each area must have exactly the same amount of imports and exports as the original network (to approximate the redistribution of production and demand). Lastly, studies can be done that examine the impacts of climate change on the current, and optimized food networks.

Acknowledgment

This work is supported by the NSF grants CCF-1522054 (COMPUSTNET: Expanding Horizons of Computational Sustainability), 1300918 and 1441208.

References

[1] H. C. J. Godfray et al., ”Food security: The challenge of feeding 9 Billion people,” Science, vol. 327, no. 5967, pp. 812818, Jan. 2010.

[2] M. Ouyang, L. Zhao, L. Hong, and Z. Pan, ”Comparisons of complex network based models and real train flow model to analyze Chinese railway vulnerability,” Reliability Engineering and System Safety, vol. 123, pp. 3846, Mar. 2014.

[3] A. Quelhas, E. Gil, J. D. McCalley, and S. M. Ryan, ”A Multiperiod generalized network flow model of the U.S. Integrated energy system: Part IModel description,” IEEE Transactions on Power Systems, vol. 22, no. 2, pp. 829836, May 2007.

[4] S. T. Shutters and R. Muneepeerakul, ”Agricultural trade networks and patterns of economic development,” PLoS ONE, vol. 7, no. 7, p. e39756, Jul. 2012.

[5] S. Porta, P. Crucitti, and V. Latora, ”Multiple centrality assessment in Parma: A network analysis of paths and open spaces,” URBAN DESIGN International, vol. 13, no. 1, pp. 4150, Mar. 2008.

[6] M. Ercsey-Ravasz, Z. Toroczkai, Z. Lakner, and J. Baranyi, "Complexity of the international Agro-Food trade network and its impact on food safety," PLoS ONE, vol. 7, no. 5, p. e37810, May 2012.

[7] X. Lin, Q. Dang, and M. Konar, "A network analysis of food flows within the United States of America," Environmental Science and Technology, vol. 48, no. 10, pp. 54395447, May 2014.

[8] "US census bureau commodity flow survey main page," 2010. [Online]. Available: https://www.census.gov/econ/cfs/. Accessed: Sep. 17, 2016.

[9] M. Brede and B. J. M. de Vries, "Networks that optimize a trade-off between efficiency and dynamical resilience," Physics Letters A, vol. 373, no. 43, pp. 39103914, Oct. 2009.

[10] "Lawrence Livermore National Laboratory Energy Flow Charts," 2009. [Online]. Available: https://flowcharts.llnl.gov/. Accessed: Oct. 15, 2016.

[11] "US EPA Regulations and standards: Heavy-duty," 2015. [Online]. Available: https://www3.epa.gov/otaq/climate/regs-heavy-duty.htm. Accessed: Oct. 15, 2016.

[12] C. L. Weber and H. S. Matthews, "Food-miles and the relative climate impacts of food choices in the United States," Environmental Science and Technology, vol. 42, no. 10, pp. 35083513, May 2008.

[13] A. Bozorgi, J. Pazour, and D. Nazzal, "A new inventory model for cold items that considers costs and emissions," International Journal of Production Economics, vol. 155, pp. 114125, Sep. 2014.

[14] A. Bozorgi, "Multi-product inventory model for cold items with cost and emission consideration," International Journal of Production Economics, vol. 176, pp. 123142, Jun. 2016.

[15] C. P. Gomes, "Computational sustainability: Computational methods for a sustainable environment, economy, and society," The Bridge, vol. 39, no. 4, pp. 513, 2009.

[16] S. P. Borgatti, M. G. Everett, and J. C. Johnson, Analyzing social networks. London, United Kingdom: SAGE Publications, 2013.

[17] M. Latapy, C. Magnien, and N. D. Vecchio, "Basic notions for the analysis of large two-mode networks," Social Networks, vol. 30, no. 1, pp. 3148, Jan. 2008.

[18] "US EPA Emission factor fact sheets," 2014. [Online]. Available: https://www3.epa.gov/otaq/consumer.htm. Accessed: Oct. 15, 2016.