Identifying Vulnerabilities in Multilayer Spatial Networks: A Case Study of Food Flows in The United States

Bin Zhang[a] and Caglar Koylu[a*]

[a] Department of Geographical and Sustainability Sciences,
University of Iowa, Iowa City, IA, USA
* caglar-koylu@uiowa.edu

Keywords: food flows, multilayer spatial networks, network measures, network vulnerability

Introduction

Flows of people, commodities and information form spatial networks in which a node represents a location or area, and an edge represents flows between a pair of locations. Understanding the patterns of connections between locations is essential for identifying the vulnerabilities of locations and the complex system. Various indices such as degree, betweenness and closeness centrality (Freeman 1977), entropy (Li et al. 2008, Koylu and Guo 2013), clustering coefficient (Saramäki et al. 2007), GINI coefficient (Lin, Dang, and Konar 2014) have been used to assess network characteristics in complex systems. However, these measures are often inadequate to identify vulnerabilities in spatial networks because there are often multiple types or categories of interactions between each pair of locations (Kivelä et al. 2014). For example, state-to-state food flows include several different types of food such as cereal grains and livestock. A state that is important for sustaining a connected supply-chain network for a certain type of food such as cereal grains may be insignificant for the distribution of another type of food such as meat and livestock. In this study, we model the food flows between Freight Analysis Framework (FAF) regions in the U.S. as a multilayer spatial network in which layers represent seven different types of food. We adopt a series of network measures to identify network characteristics and potential vulnerabilities of locations in the food network using the multilayer network approach. By calculating the Kendall's Tau coefficients, we perform two evaluations to identify the level of similarity between (1) the different layers (food flow category) for each measure; and (2) each pair of network measures for each layer.

Case Study and Data

Food supply and security have been under threat by global catastrophes as a result of climate change, natural disasters, drought and wars (Godfray et al. 2010, Dilley and Boudreau 2001). The 2020 coronavirus crisis have already shown the potential impact of pandemic spread on global economies, and food supply-chain network. It is critical to improve the measurement of food security and vulnerability to solve the lack of sufficient dietary energy and achieve a more equitable access to food products (Barrett 2010). Previous work in food security used network metrics such as high clustering of key nodes distribution (Lin, Dang, and Konar 2014), node betweenness (Ercsey-Ravasz et al. 2012), and attack measures (Holme et al. 2002) to reveal insights about food supply-chain, while they failed to consider shifting vulnerabilities based on different types of food.

In this study, we used food flow data in the US from 2007 to 2012, which is generated by the Commodity Flow Survey (CFS) and modelled the food flows as a multilayer network of multiple food categories: (1) live animal/fish, (2) cereal grains, (3) other agriculture products, (4) animal feed (5) meat/seafood, (6) milled grain products and (7) other foodstuffs. To illustrate the data, we first transformed the raw flows into modularity flows and then employed multivariate clustering (Guo 2009) and visualized the seven categories of food flows on top of the percentage of within flows to all flows (Figure 1). Within flow percentages of most regions are lower than 20%, which is a consequence of regional and global food trade connections. Generally, FAF regions where the economy is dominated by agriculture (e.g., MidWest) tend to have higher within flow percentage values. Flows with red colors are more dominant on cereal grains (SCTG2), while flows with blue colors are more dominant on other agriculture products (SCTG3) and animal feed (SCTG4). While the multivariate clustering reveals distinct flow patterns for each category, distinguishing locational characteristics require overlaying of flows with network characteristics.

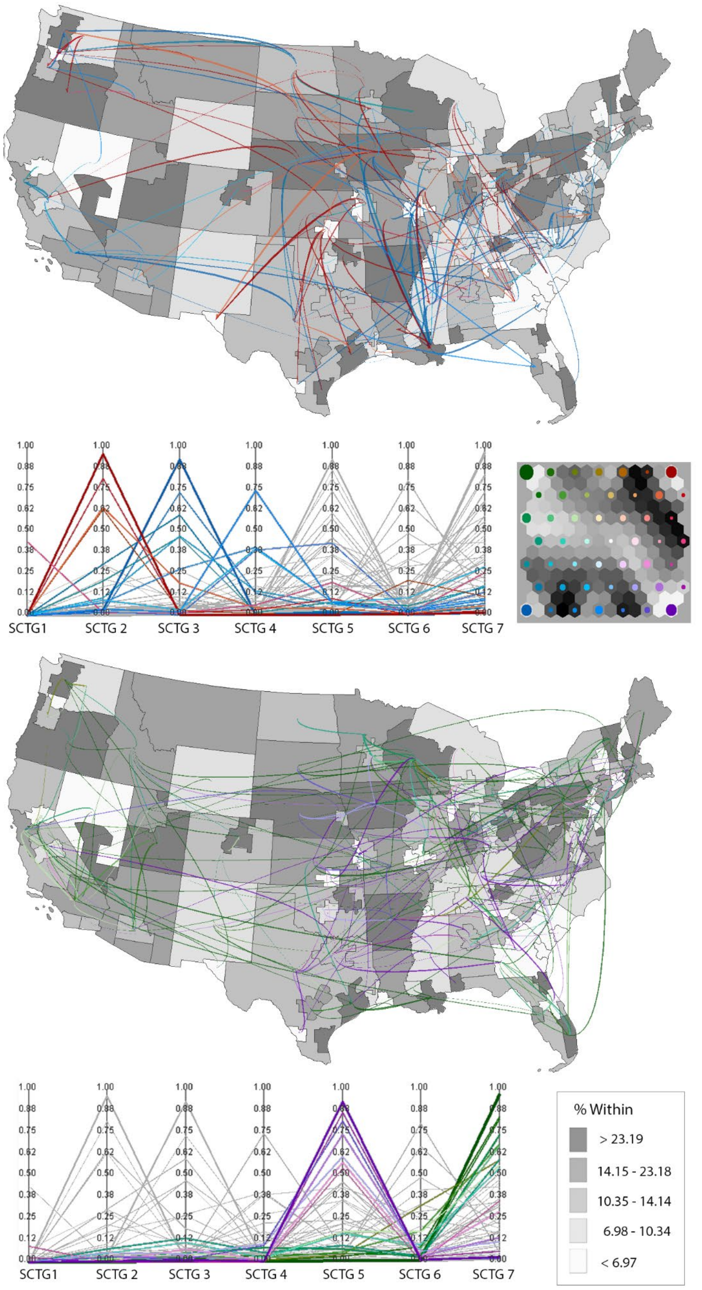


Figure 1: The multivariate clustering of food flows by category (>50% of all flows). The choropleth map shows the percentage of within flows for FAF regions.

Method

We adopted the following network measures to assess the network characteristics of locations (FAF regions): total flow, within flow, netflow ratio, GINI Coefficient, betweenness centrality and entropy. Netflow ratio is the division of the difference between inflow and outflow by the summation of the inflow and outflow. GINI coefficient is a measurement for equality, where the index varies from 0 to 1 or equality to inequality (Lin, Dang, and Konar 2014). Betweenness centrality measures the number of shortest paths between pairs of nodes (locations) and illustrates importance of nodes in connecting other nodes together (Freeman 1977). We calculate weighted betweenness centrality by minimizing the sum of the weights of the edges that each path passess through. Entropy illustrates the variation of flow connections and volumes for a location, and can provide important insights about the structure of the network and the characteristic of the location (Li et al. 2008). Entropy value ranges between 0 and 1. The entropy value is measured by the following equation:

$$E_i = -\sum_{i=1}^{j} \frac{x_{ij} * \log(x_{ij})}{\log(j-1)}$$

where E is the entropy of location i, j is the total number of connection in the network that location i has, x_{ij} is the proportion of volume between i and current connected location in total volume that i has. The total number of non-zero connections node i has is j-1 (possible outcomes). If the flow volumes are equally distributed across those j-1 number of connections, we would have this maximum entropy value. We normalize the entropy by N to scale it to be between 0 and 1. A small entropy value is usually generated when a location is connected to only a few locations or even a single location. In addition, when a location has large deviation of volume on its connections, entropy tends to be smaller. On the other hand, a large entropy value indicates that a location connects to other locations in about equal volume, which is interpreted as disordered and unpredictable. By calculating the Kendall’s tau coefficients (Sen 1968), we performed two evaluations to identify the level of similarity (1) between different layers (food flow category) for each measure; and (2) between each pair of network measures for each layer. Kendall’s Tau is a measure of ordinal association between a pair of variables. The measure varies from –1 to 1 and reveals the correspondence between two rankings.

Results

We calculated the pairwise similarity (Kendall’s Tau) between the six measures using all food flows (Figure 2). While red hues with increasing darkness represent positive rank correlation, blue hues with increasing darkness represent negative correlation indicating reverse rank correlation. Lighter colors close to white show smaller correlation. Total flow, within flow and GINI produce very similar ranking because they are related to the total volume or capacity of the location to produce flows. Entropy is slightly correlated with volume measures: total (0.25) and within flow (0.31), however, these correlations are not as strong as GINI (0.62, 0.51). On the other hand, netflow ratio and betweenness produce distinct results, but in opposite ranking with the volume measures and GINI. Overall, entropy produces the most distinct results, which we further explore to compare the differences of entropy between different layers.

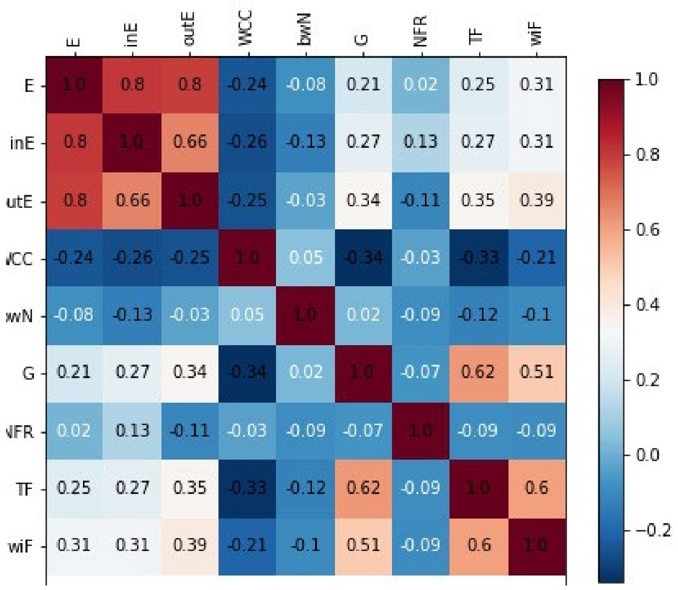


Figure 2: Pairwise Kendall's Tau coefficients of: Entropy (E), Inflow Entropy (inE), Outflow Entropy (outE), Weighted Cluster Coefficient (WCC), Betweenness Centrality (bwN), Gini Coefficient (G), Net Flow Ratio (NFR), Total Flow (TF), and Within Flow (wiF).

We calculated the pairwise similarity between the entropy measures of the seven layers (Figure 3). Other agriculture products (SCTG3) has relatively strong disagreements with other categories. Live animal/fish (SCTG1) is slightly correlated with meat/seafood (SCTG5), while cereal grains (SCTG2) is positively correlated with SCTG4 (animal feeds). Layers from SCTG4 to SCTG7 have stronger correlations with each other, however these correlations do not exceed 0.37. This result itself clearly shows how distinct entropy measures are for different network layers.

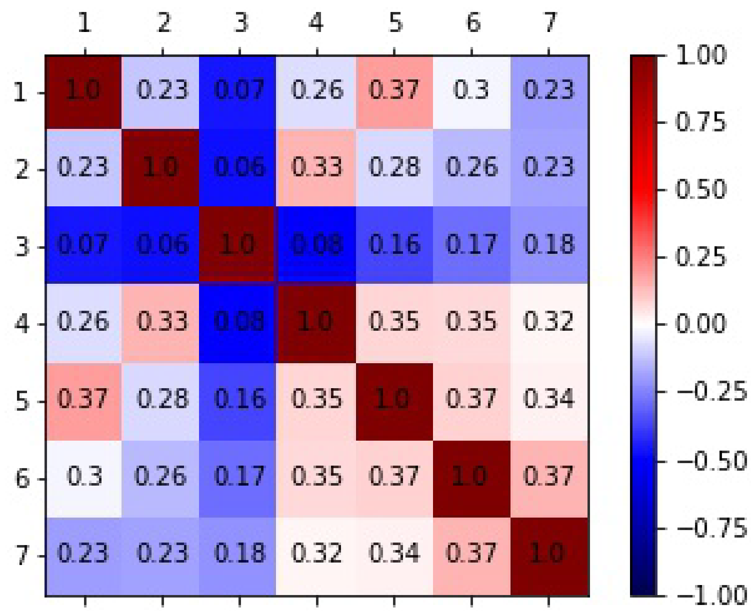
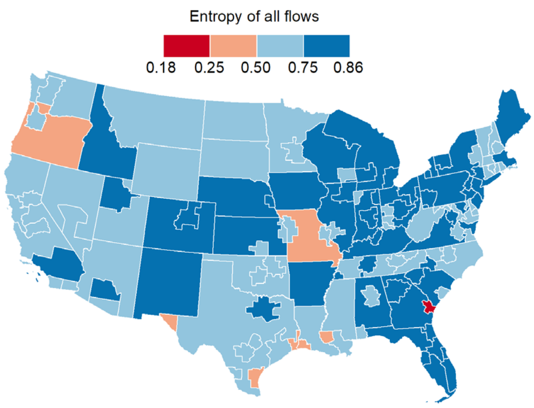


Figure 3: Pairwise Kendall's Tau coefficients of entropy measure for the seven layers: (1) live animal/fish, (2) cereal grains, (3) other agriculture products, (4) animal feed (5) meat/seafood, (6) milled grain products and (7) other foodstuffs.

Measure Mapping and Interpretation

It is important to distinguish the measurement of vulnerability in spatial networks into the concepts of system vulnerability and location vulnerability. For example, low values of entropy indicate that a location is sending and receiving flows from only a few locations, while high values of entropy indicates that the location is receiving and sending flows from diverse locations with similar volumes. Low values of entropy may indicate high vulnerability for a location in the case of disruptions such as natural disasters effecting those few connections that the location has. In contrast, while high entropy means that at location is self-sufficient and robust, this may pose a potential threat to the systems vulnerability because the system becomes highly vulnerable if the location has any disruptions in its distribution of flows.

Figure 4 illustrates the entropy calculate by all flows and other agricultural products. These two measure maps illustrate distinct results. While red colors illustrate low entropy (higher locational vulnerability), blue colors represent high entropy (lower locational but higher system vulnerability. While entropy values of all flows range between 0.18 and 0.86, entropy values for SCTG2 are mostly below 0.5 with a few values of 0 entropy meaning that those locations have only one connection in the network.



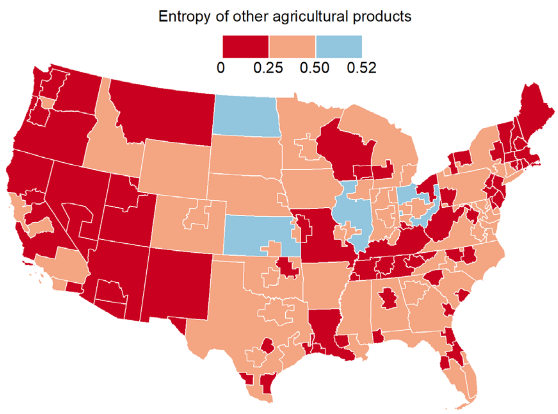


Figure 4: Entropy calculated by all flows (Left) and other agricultural products (Right).

Conclusion and Future Directions

We conducted two major evaluations: between measures and between layers. We discussed how vulnerabilities can differ based on selected measure, layers of a multilayer network, and the concepts of locational vulnerability versus system vulnerability. For future work, we propose a multilayer and multi-measure ranking system that will allow us to systematically evaluate every layer and every measure for the identification of vulnerabilities. We further plan to develop an indicator based on the average rankings of measures. This would ultimately carry information from each measure, which reflects a different characteristic of for the locations and the network. While developing such an indicator, we will remove redundant measures (e.g., measures that highly correlate with each other) to reduce noise and redundancy. Also, measure results are often dependent on the scale of networks. We used FAF regions to identify location characteristics, however, we plan to apply our methodology to finer scale data sets (e.g., simulated county-to-county flows) and state-to-state food flows.

Our prelimineray analysis included undirected network measures. We plan to further expand our analysis to include directed network measures such as inflow and outflow entropy and GINI coefficient. These measures would reveal directional imbalances in flows. While network measures provide important insight into potential locational vulnerabilities, the interpretation of measure results are limited unless flows are also overlayed to illustrate connections for each location. Connections together with locational characteristics would help us better understand the complex system and its vulnerabilities. We plan to analyze the relationship between food flows and the factors such as food production and consumption, population and demographics, and transport hubs in our future studies.

References

Barrett, Christopher B. 2010. "Measuring food insecurity." *Science* 327 (5967):825-828.

Dilley, Maxx, and Tanya E Boudreau. 2001. "Coming to terms with vulnerability: a critique of the food security definition." *Food policy* 26 (3):229-247.

Ercsey-Ravasz, Mária, Zoltán Toroczkai, Zoltán Lakner, and József Baranyi. 2012. "Complexity of the international agro-food trade network and its impact on food safety." *PloS one* 7 (5):e37810.

Freeman, L. C. 1977. "Set of Measures of Centrality Based on Betweenness." *Sociometry* 40 (1):35-41.

Godfray, H Charles J, John R Beddington, Ian R Crute, Lawrence Haddad, David Lawrence, James F Muir, Jules Pretty, Sherman Robinson, Sandy M Thomas, and Camilla Toulmin. 2010. "Food security: the challenge of feeding 9 billion people." *science* 327 (5967):812-818.

Guo, D. 2009. Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. IEEE Transactions on Visualization and Computer Graphics: IEEE.

Holme, Petter, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. 2002. "Attack vulnerability of complex networks." *Physical review E* 65 (5):056109.

Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. 2014. "Multilayer networks." *Journal of complex networks* 2 (3):203-271.

Koylu, Caglar, and Diansheng Guo. 2013. "Smoothing locational measures in spatial interaction networks." *Computers Environment and Urban Systems* 41:12-25. doi: 10.1016/j.compenvurbsys.2013.03.001.

Li, J., B. H. Wang, W. X. Wang, and T. Zhou. 2008. "Network Entropy Based on Topology Configuration and Its Computation to Random Networks." *Chinese Physics Letters* 25 (11):4177-4180.

Lin, Xiaowen, Qian Dang, and Megan Konar. 2014. "A network analysis of food flows within the United States of America." *Environmental science & technology* 48 (10):5439-5447.

Saramäki, Jari, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and Janos Kertesz. 2007. "Generalizations of the clustering coefficient to weighted complex networks." *Physical Review E* 75 (2):027105.

Sen, Pranab Kumar. 1968. "Estimates of the regression coefficient based on Kendall's tau." *Journal of the American statistical association* 63 (324):1379-1389.