

# Complex Network Analysis of Imported and Exported Goods in the United States of America

Konstadinos G. Goulias(✉) and Hui Shi

University of California, Santa Barbara, CA 93106, USA
kgoulias@ucsb.edu

**Abstract.** In this paper the network structure of imported and exported freight flows in United States in 2022 is explored. The data used are from the Freight Analysis Framework (FAF) that represents freight movement among states and major metropolitan areas by all modes of transportation. Using tools from network science hubs and central nodes making the network of freight flow vulnerable are identified and communities of nodes based on shipments of goods in value of tonnage flows are extracted to distinguish different functional regions. Different network topology in 2022 is found analyzing imported goods and exported goods. The networks are graphs with 132 nodes that are spatial subdivisions based on major Metropolitan Statistical Areas (MSAs) and the non-MSA parts of their States. The links are the directional movement of freight flows among the nodes. The imports graph explored here has 10739 directional links among the 132 nodes. The graph of international exports is smaller with 9722 links among the 132 nodes and displays different freight flow behavior and community clustering than the much larger 2022 domestic freight flow network that has 17356 directional links. The analysis here finds some similarities in the location of vulnerable nodes. Similarities are found in functional regions formed by data-driven clusters of nodes dividing the US into Northeast, Southeast, Southcentral, West, and Midwest with differences among the domestic, import and export networks. The number of distinct clusters in the imports and exports networks is higher than the domestic network clusters and in a few clusters node membership violates regional contiguity and proximity. This is due to the infrastructure connectedness and economic complementarity (e.g., nodes that exchange goods for auto manufacturing with Detroit, Michigan, and Laredo in Texas that imports goods from Mexico).

**Keywords:** Network science · Functional region · Community · Freight flow · Freight value · Exports · Imports

## 1 Introduction

Regions around the world are interdependent and often form trade alliances with a high degree of international economic and political influence [1, 2]. The impact of these agreements and alliances is different among localities in a country for a variety of regional and historical development reasons but also local policies create spatially

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14819, pp. 346–365, 2024.
https://doi.org/10.1007/978-3-031-65282-0_23

differentiated impacts such as found by Hayward when exploring the impact of the European Community formation and resulting trade with the US [3]. Local influence of international trade could be examined at the level of a State or Province as was done by [4, 5] for the United States and Canada but a more detailed geography may reveal key differences within states. This motivates our analysis here using a more fine-grained geographic subdivision in the Freight Analysis Framework (FAF) maintained by United States agencies. FAF combines data from the US Census Bureau Commodity Flow Survey (CFS) with a variety of other sources to develop trade flow data among 132 places (metropolitan areas and non-metropolitan areas in each of the 50 states and Washington DC). This captures goods movement in agriculture, resource extraction, utility, construction, retail, services, and other sectors. The data used here are presented in a later section and they are from FAF5 that has been recently completed [6]. This paper focus is on the international freight flows among the 132 FAF5 places in the US and the worldwide regions. Domestic monetarized flows are used as benchmark. Network science tools are used to detect and describe regions in the FAF5 data for 2022, which is the closest year to the writing of this paper and has been validated with observed data. FAF provides three metrics of freight flow that are tonnage of goods, value of this tonnage, and ton-miles. The focus here only on the value of tonnage flows in USD 2022.

Network science views origins, destinations, and the movement of goods between them as a graph of nodes (vertices) and links (edges). Importance of edges is based on the size of freight flows moving from one node to another, which are commonly called weights of the edges (links). Node importance is based on the number of connections with other nodes and the total amount of commodities moving into a node and/or moving out of a node. In this way one can identify nodes that function as major predominantly producers, major predominantly consumers and as hubs of freight flow movement. Network statistics also allow to identify node hierarchy within a network based on the node degree of influence exerted to the entire network. In this way vulnerability of a network can be studied in terms of loss when eliminating an influential node. Another outcome of network analysis is the identification of groups of nodes (in this paper they are metropolitan places or entire states) that have higher levels of freight flows within their group. These places create functional regions derived from relationships among nodes and may be a factor of resiliency for a group of places and for the entire freight network as a whole. Making these distinctions among nodes and groups of nodes requires a specialistic analysis that is network science and reviewed briefly later in the paper. The key research questions are:

1. Are there topological differences between import and export freight networks?
2. Are there critical nodes of production, consumption, and intermediaries that make the import and/or export networks vulnerable to threats?
3. Are there groups of nodes that belong together in a functional region and are they geographically clustered?
4. Does membership in these functional regions differ between imports and exports?

## 2 Data Used Here

The freight flow data analysis here is modeled using a multimodal and intermodal network of facilities that bring produced goods (e.g., goods by agricultural fields, extraction of crude oil in wells, mining of iron ore) to facilities that transform these goods (e.g., preparing food and packaging, refining crude oil to gasoline, manufacturing cars) that in turn are transported to the end consumer (e.g., grocery stores, retail gasoline pumps, car buyers at salons or residences). Many commodities have distinct infrastructure components underlining their movement geography such as petroleum crude. Its supply chain includes extraction technologies at the origins where production happens (e.g., oil wells), pipelines, railways, highways and trucks, barges, tanks for storage, refineries, and a distribution system to the ultimate customers (e.g., power generating plants, industrial complexes, businesses, and residential customers). All these form spatial networks that take a different form impacting the nature of node and links. In this paper no distinction among the different goods, or the infrastructure used to transport them are made. Instead, network science is used to identify clusters of nodes and links that represent groups of regions that participate in international trade of goods with higher "intensity" among them than with other regions in the U.S.

The core data and the spatial organization of origins and destinations are from the Commodity Flow Survey (CFS) conducted by the Bureau of Transportation Statistics (BTS), the US Department of Transportation, the US Census Bureau, and the US Department of Commerce to provide data on domestic freight shipments in a variety of industries with the latest version in 2017 (https://www.census.gov/programs-surveys/cfs/about.html). The nodes are metropolitan areas in the United States and places in a state that are not metropolitan areas. The data assembled and combined produce the Freight Analysis Framework (FAF) using the CFS regions as origins and destinations (see Appendix A). To distinguish among different states in a major metropolitan area CFS divides an area by state. For example, the area with code 408 is divided in a Connecticut part, a New Jersey part, a New York part, and a Pennsylvania part. These regional economic centers aim to balance geographic representation and good quality of data. This delineation includes Metropolitan Areas, state remainders that are not in the Metropolitan Areas, and entire states such as North Dakota. FAF in essence is a major fusion of data from different agencies and includes CFS to provide estimates of flow of goods in the form of a matrix of origins and destinations (the 132 zones that CFS is also using). Appendix A shows the map of the current (2017 and later) FAF zones and their list. The freight flow is classified into categories of the Standard Classification of Transported Goods (SCTG) and FAF uses the two-digit classification from SCTG to develop a 43 class list that spans from the commodities of agricultural products to the precisions machinery and fuels [6]. The value in dollars is based on shipments in CFS and other available data and it is the *net selling value* and excludes freight charges and excise taxes [6, 7]. The current version of FAF5 (FAF5.2) provides estimates for tonnage (unit: thousand tons), value (unit: million dollars), and a combined weight and distance (unit: million ton-miles) by origin-destination pair of FAF regions, commodity type, and mode for the base year (2017), predicted values and validated/verified with observed data for 2018, 2019, 2020, 2021, and 2022 and forecast year estimates for other years. The metrics provided differentiate among imports, exports, domestic, and total amounts.

## 3 Method

In the context of flows a network is composed of two major elements that are the places (nodes or vertices) where commodities are originating and/or destined and the relations between these places represented as nodes of a network called links or edges. A relation can be simply the connection between two nodes and in our analysis are directed indicating that the link from node A to B is different than from node B to A in terms of the amount of goods moved. This allows the possibility of asymmetry in the movement between origins and destinations (e.g., goods produced at one place are transported to another place without equal reciprocity). Goods are not only produced at nodes but they within but are also transshipments from international origins or to international destinations and other domestic nodes. Movement of goods happens only along the links. The combination of nodes and links form a network with a number of indicators (metrics) that are used to describe the properties of this network. Many of these indicators originated in graph theory and the terminology used in this analysis follows this formalization of a network as a graph. This section uses ideas and metrics based on methods developed first in social networks analysis [8, 9] and later developed further to explore other networks with complex structure by physics-oriented approaches [10–13].

A graph indicated as G = (V,E) is a combination of nodes/vertices (V) and links/edges (E). Let's assume the number of nodes is N and the number of links is L. The N nodes form an adjacency matrix A of dimensions N by N with elements $a_{ij}$ (with i = 1, 2,…,N and j = 1,2,…,N) taking the value 1 if there is a link between the i node and the j node. To account for the size of commodities flowing from node i to node j we define a weights matrix W with elements $w_{ij}$. In the commodities flow this matrix is not symmetrical which means $w_{ij} \neq w_{ji}$. In commodities we also have material that originates in one place and is destined in the same place. The weights matrix in this case has non-zero diagonal elements. In networks language this creates a loop link (called loops herein). In one network here we study the tonnage network in which we can have tons of goods in a place in which local manufacturers transform raw material in other commodities that are also consumed within the node. This means that commodities (goods) produced in a node are transformed, stored, consumed or shipped to other nodes.

*Average degree* is the ratio between edges and nodes. The word *degree* is always used to describe the number of edges connected to a node. The *in-degree* is the number of edges pointing to a node (in terms of commodities this means the number of connections carrying commodities to a destination). The *out-degree* is the number of edges leaving a node representing the commodities transported to other nodes. Each link can have a *weight* that represents the physical weight or value of all commodities moved along the edge/link. In this way the *weighted degree* is the total sum of weights for all the edges in the in-degree and the opposite for the weighted out-degree.

In networks we look for *connectivity*, *centrality*, and *community*. Connectivity is represented by the weighted and unweighted *in-degree* and *out-degree*. The higher these are and the more connected the nodes are. A super-connected node would be the one that has the highest degree. This may also be a *weakness* of the network if there are no alternatives to bypass the highest degree node that functions as intermediary if commodities need to be rerouted around it.

Another metric of network connectivity is its *diameter*, which is the topological distance between the two furthest away nodes. We can also enumerate all the paths with topological distance lower than a threshold number. A network with better connectivity is the network that has more short topological paths and a smaller diameter. Often random graphs are better connected than real world networks. *Graph density* is the number of connected edges over the total number of possibly connected edges and it is between 0 and 1 with the value of 1 for the network with the highest connectivity with no holes in the graph. Graphs that have density much lower than 1 are sparse network types. *Average path length* is the number of intermediaries between any two nodes along the shortest paths between them and it is an indicator of network overall efficiency. Average path length together with the clustering coefficient are used to judge if a network is a small-world network (i.e., few degrees of separation among the nodes).

Of interest in analyzing networks is to find out if nodes form groups based on their characteristics and relationships. One way to examine if nodes belong to a close-knit community is to enumerate how many nodes of a node's connections are connected to each other in the network. When many of these nodes are connected to each other, the node belongs to a highly clustered network. The clustering coefficient is the number of neighbors that are connected to each other over the highest number of possible connections. The word clustering in this context has been challenged because in reality it indicates *transitivity* (i.e., "…this means the presence of heightened numbers of triangles in the graph" see Newman [9], page 183). The calculation counts the number of triangles (or triplets) over the potential number of triangles (triplets) that the neighbors of a node can have. In this case we account for the edge direction. The average clustering coefficient is the average of the clustering coefficients of the nodes in the network. A high average clustering coefficient for a network and a low average path length show that we have a tightly connected network in which we can reach from one node all the other nodes with a few steps. This is called a *small-world network* and often compared to a random graph that is usually having higher clustering coefficients all else equal. Small-world networks are more *robust* than random networks because removal of a node in a small-world will have much smaller repercussions for the rest of the nodes than the removal of a node in a random graph. However, it is also important to identify the relative location of a node within the structure of a network. For example, removal of a central node will divide the entire network in multiple independent structures making this network vulnerable in its functioning just by the removal of one node. This is where centrality indicators are important to identify hubs or super connected nodes. Related to this is also a scale-free network in which the degree distribution follows a power law, which means we have a few nodes with extremely high number of connections and then a very large number of nodes with an extremely low degree (few connections to other nodes). There are different centrality indicators for nodes. The betweenness centrality that measures how important a node is for other nodes moving commodities from one node to another is used here. This indicator shows how often a node lies on the shortest path between two other nodes.

From the social network analysis literature, a community is a subgraph of a network with nodes that are more connected with each other than with any other nodes outside the subgraph (e.g., also called a clique or cohesive group). Since randomly wired networks lack a community structure a metric that compares a random network's wiring with the same number of nodes to the wiring of an observed network can be used. This metric is called modularity (Q) because in essence divides the network graph into modules (subgraphs) with each module containing nodes that are more tightly connected than nodes belonging to other modules [14]. This was first developed for undirected and unweighted graphs and then extended to directed weighted graphs [15]. The computation and allocation of nodes is done by a series of steps that assign nodes to groups and computing the modularity until a maximum is found corresponding to an optimal separation of nodes into groups (see the method named the Louvain algorithm in [16] that is used in this paper as implemented in Gephi). Computation of the modularity metric becomes more complex when we have a large origin-destination directed network with weights (the freight amount and/or value moving from an origin node to a destination node as explored in this paper). In this sense the modularity metric is maximized like a fit metric. At the end of the process the modularity algorithm assigns nodes into an aggregated group or cluster based on shared characteristics. One interpretation of modularity is as the difference between the fraction of edges within an observed community and the number of edges of the same nodes that a random connectivity would produce. The higher the modularity then the more systematic the presence of edges in a group becomes.

Methods based on the network science techniques have been used in the analysis of aviation [17], in maritime transportation as in the review of more than 120 papers in [18], in high-speed rail [19], and urban travel demand verification and validation [20]. They have also been extended and used in emergencies and crises [21, 22], but also to develop a variety of methods improving community detection [23, 24, and 25].

## 4 Analysis and Results

There are two examples in this section to explore the role played by a few major hubs in the US and are based on international imports and international exports and a benchmark network of domestic flows. US is the largest goods importer in the world and its top 5 suppliers are China, Mexico, Canada, Japan, and Germany. US is also the second largest exporter of goods in the world with the top five countries receiving the US exports ranked as Canada, Mexico, China, Japan, and United Kingdom (www.ustr.gov). The ratio of imported goods in value over exported goods is approximately 1.6 (US Census, 2023). Our interest in this analysis is to first identify the hubs to which imports arrive in the US and then if there is a community that exchanges these imports with each of these major hubs with higher intensity than others. Identifying the hubs of exit of exported goods and their community of nodes that provide exported goods is also an interesting question in this paper. As shown later not all places in the US participate in imports and/or exports. Table 1 shows the statistics of these two networks and the domestic network that is by far the densest network and shows a clear propensity to have clusters heavily influenced by contiguity, adjacency, and closeness forming six

functional regions that are the NorthEast, SouthEast, SouthCentral, West, MidWest, and NorthCentral. Major hubs in this network are as expected major metropolitan areas in New York, Florida, Georgia, Texas, California, Illinois, and Michigan. Figure 1 shows a map of these functional regions together with the hubs in each region with the size of the circles proportional to the weighted degrees of the nodes.

Figure 2 shows international imported goods in USD by port of entry in the US. The communities shown by different colors are defined based on final destination of imported goods in USD 2022. The size of each node is proportional to weighted out-degree in USD 2022 to depict the role each node plays in distributing imported goods to the rest of the 131 places and to itself. For example, goods arriving in Los Angeles are valued in 2022 USD at approximately 370576 million and are destined to itself and other places (imported by Los Angeles and staying in Los Angeles are goods worth 114290 million 2022 USD) making this node the largest importing node in value in the United States. The lions share (approximately 62.8%) of these imported goods are from Eastern Asia and this includes China, Hong Kong, Japan, Macao, Mongolia, North Korea, South Korea, and Taiwan. This node receives most goods through its major ports of entry that are Port of Long Beach, Port of Los Angeles, and Los Angeles International Airport forming a major vulnerable location for the entire United States.

This node not only transfers to other nodes goods but also receives large value of goods from other places valued at approximately 82353 million USD of 2022 from within California such as San Francisco, San Diego, Sacramento, Fresno, and Rest of California (i.e., international goods imported at San Francisco or San Diego are transported to Los Angeles) but also other nodes further out forming a community (based on the modularity) shown with the same color in Fig. 1 and containing nodes that are far from California. This community contains Louisville, KY, that is transferring higher value of goods to Los Angeles than Los Angeles to Louisville while another node of the same community, Sacramento, does the opposite. This implies the international import communities in Fig. 2 sometimes violate the geographic contiguity and closeness relationships found in the domestic networks. However, as shown in Fig. 2 many nodes of this community are also from places contiguous to California (e.g., nodes in Arizona).

Figure 3 shows international export by port of exit and the communities displayed by different colors are defined based on origin of the exported goods and the point of departure to international destinations (i.e., the size of nodes is proportional to weighted in-degree in USD 2022). It should be noted the communities in Imports22 and Exports22 are different (see the bottom of Table 1 illustrating differences and commonalities between the two network community structures). Some nodes however play major roles in freight flow for imports and exports as explained below but also domestic goods flows. In fact, Los Angles is a hub for both Imports22 and Exports22 and dominates the community labeled here as West and SouthWest with 12.88% of nodes in Imports22 and 20.45% of nodes in Exports22.

Table 1 shows the network metrics of the domestic, import, and export networks and communities identified using modularity. Both international networks have lower connectivity than the domestic network. Domestic22 network has all its nodes connected to all other nodes with lower network diameter, highest node density, and an order of magnitude higher average weighted degree is a robust network with many options for

hubs and routes. Also shows the domestic production and associated consumption are by far higher than exports and imports. The international import and export networks also show more communities with hubs distributed throughout the US and communities with multiple hubs spatially distributed within the communities. They also show spatial interactions that can span the entire nation. For example, the community in Imports22 called SouthCentral-Texas includes major Texas hubs such as Laredo and Houston and Detroit in Michigan. Similar occurrence in Exports22 is for the community called NorthCentral-Detroit that includes a hub in Laredo, Texas. This example is a reflection of manufacturing corridors heavily influenced by the role Laredo plays as exporting and importing goods to and from Mexico and then used in manufacturing.

The presence of more communities is an indication that some hubs can function as a redundant system of US ports of entry or exit to other systems (e.g., one community of nodes supporting another community of nodes in emergencies). This is heavily dependent on the supporting infrastructure surrounding them (e.g., highways, pipelines, railways, and regional airports) that can function as alternate routes in emergencies and/or disasters at other hubs and communities. The eastern seaboard of the US has many nodes that are endowed places with ports that could function as substitutes for each other in emergencies. In contrast, the western seaboard of the US has Los Angeles dominating the entire region with supporting roles by San Francisco and Seattle.

Tables 2 and 3 illustrate the different roles played by the top 20 nodes in imports and exports. In Table 2 when the ratio of weighted out-degree over the weighted in-degree is greater than 1 we have a node that functions as a net distributor of goods to other nodes. In Table 3 the ratio of weighted in-degree to the weighted out-degree is an index of the net collector role each node plays. In essence, other nodes and itself provide goods to be exported by the collector node. However, the same node plays a different role in different communities depending on analysis of imports or exports. Notable in this role is Laredo in TX that has a ratio of 56 in Imports22 and 80.40 for Exports22. This makes Laredo the largest distributor and collector in the US. In fact, Laredo is a hub with major infrastructure components to transport goods from and to Mexico that includes major vehicle and rail bridges and an inland port. Goods moved to and from this node are automobiles, automobile parts, fuels, and electronics (www.laredoedc.org). Another node that emerges in this analysis is Savannah, GA, as both a distributor (ratio 15.02) and collector (ratio 11.07) and this is mainly due to its major international containers port. North Dakota is another node of importance for imports and exports from and to Canada for fossil fuel and agricultural machinery ( https://www.usglc.org/state-facts/north-dakota/). The minimum weighted in-degree in Imports22 is 520 million for the node Rest of Hawaii and the minimum in Exports22 weighted out-degree is 96 million USD also for the Rest of Hawaii node.

Figure 4 illustrates the difference among zones that specialize in exports and zones that are specializing in imports using the total weighted degree of imports and exports for each node of the two networks Imorts22 and Exports22. The smooth line separating these is one way to draw a border in the two different data of imports and exports and captured the non-linear correlation between imports and exports. Notable is again the role of major metropolitan areas and the diversity among them (e.g. New York in the NY portion vs the NJ portion).

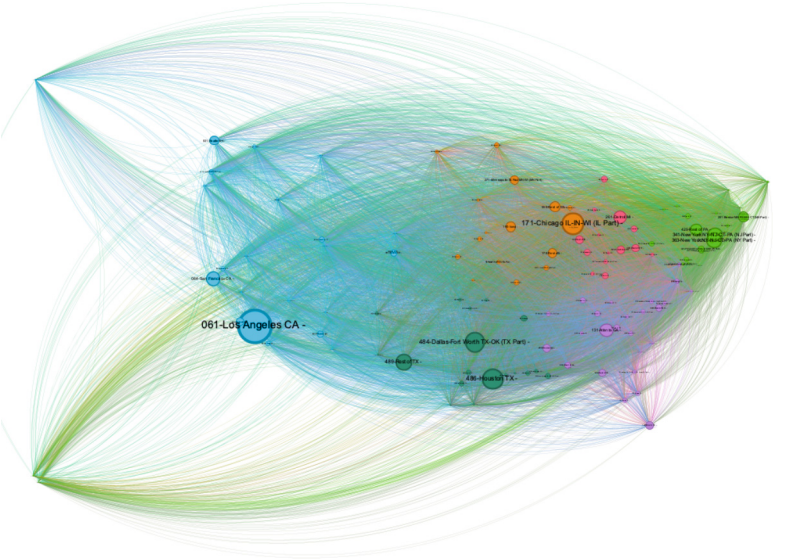


**Fig. 1.** Domestic network defined with size of nodes proportional to weighted degree in USD 2022 (Domestic22)

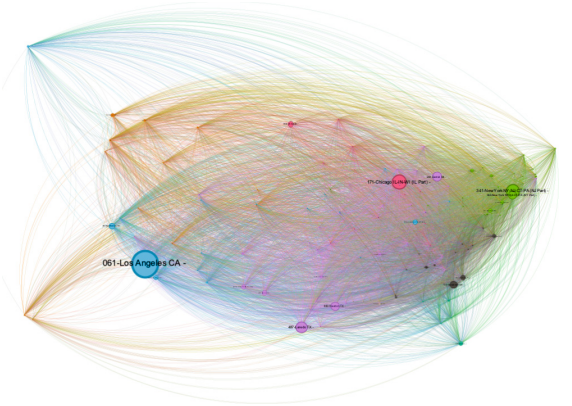


**Fig. 2.** International imports by port of entry and communities defined based on final destination (size of nodes is proportional to weighted in-degree in USD 2022) – Imports22

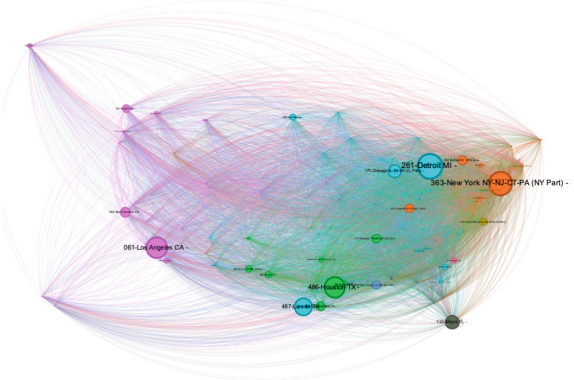


**Fig. 3.** International Export by Port of Exit and communities defined based on export origin (size of nodes is proportional to weighted in-degree in USD 2022) – Exports22

**Table 1.** Freight Flow US Network 2022 for Domestic and International Imports and Exports

| | Imports22 | Exports22 | Domestic22 |
|---|---|---|---|
| Average Degree | 81.356 | 73.652 | 131.485 |
| Number of Nodes | 132 | 132 | 132 |
| Number of Edges | 10739 | 9722 | 17356 |
| Average Weighted Degree | 18477 | 11534 | 112408 |
| Network Diameter | 3 | 3 | 2 |
| Graph Density | 0.621 | 0.562 | 1.0 |
| Average Path Length | 1.3 | 1.4 | 1.0 |
| Average Clustering Coefficient | 0.722 | 0.692 | 0.996 |
| Highest Betweenness Centrality | Many nodes | Many nodes | Many nodes |
| **Community Detection** | | | |

(*continued*)

**Table 1.** (*continued*)

| | Imports22 | | Exports22 | | | Domestic22 | |
|---|---|---|---|---|---|---|---|
| Modularity | 0.396 (8 classes) | | 0.428 (8 classes) | | | 0.35 (6 classes) | |
| | %nodes | Major Hubs | %nodes | Major Hubs | | % nodes | Major Hubs |
| Northeast | 22.73 | New York NY & NJ parts | 22.73 | New York, NJ part | | 21.21 | New York NY & NJ parts |
| MidAtlantic – Virginia | In NorthWest & SouthEast | | 3.03 | Virginia Beach- Norfolk, VA | | In SouthEast | |
| SouthEast | 11.36 | Savannah, GA and Charleston, NC | 2.27 | Charleston, NC | | 21.21 | Atlanta, GA, Miami, FL |
| Southeast – Florida | 3.79 | Miami, FL | 4.55 | Miami, FL | | In SouthEast | |
| SouthCentral - | 27.27 | Texas many hubs with Mexico | 9.09 | Houston, TX | | 13.64 | Dallas-Ft Worth & Houston, TX |
| SouthCentral – Louisiana | 3.03 | New Orleans, LA | 3.03 | New Orleans, LA | | In SouthCentral | |
| West and SouthWest | 12.88 | Los Angeles, CA | 20.45 | | Los Angeles, CA | 18.18 | Los Angeles, CA |
| NorthWest | 9.85 | Seattle, WA & Portland, OR & high dispersion | In SouthWest | | | In SouthWest | |
| MidWest | 9.09 | Chicago, IL | In NorthCentral | | | 14.39 | Chicago, IL |
| NorthCentral – Detroit | In SouthCentral | | 34.85% | Detroit, Mi, Laredo, TX, & Chicago, IL | | 11.36 | Detroit, MI |

**Table 2.** Top 20 Distributing Nodes of International Imports in 2022 (Imports22)

| Top 20 Distributors for Imports | | | | |
|---|---|---|---|---|
| | Weighted in-degree (IN) in USD | Weighted out-degree (OUT) in USD | Ratio OUT/IN | Modularity Class |
| 061-Los Angeles CA | 196643 | 370576 | 1.88 | SouthWest |
| 171-Chicago IL-IN-WI (IL Part) | 131724 | 193744 | 1.47 | MidWest |
| 341-New York NY-NJ-CT-PA (NJ Part) | 97463 | 176604 | 1.81 | NorthEast |
| 487-Laredo TX | 2631 | 147317 | 55.99 | SouthCentral |
| 261-Detroit MI | 67678 | 118055 | 1.74 | SouthCentral |
| 486-Houston TX | 78312 | 110624 | 1.41 | SouthCentral |
| 363-New York NY-NJ-CT-PA (NY Part) | 92601 | 107306 | 1.16 | NorthEast |
| 132-Savannah GA | 6231 | 93590 | 15.02 | SouthEast |
| 064-San Francisco CA | 90893 | 71598 | 0.79 | SouthWest |
| 380-North Dakota | 3132 | 66727 | 21.31 | MidWest |
| 451-Charleston SC | 6723 | 58299 | 8.67 | SouthEast |
| 489-Rest of TX | 64934 | 56383 | 0.87 | SouthCentral |
| 212-Louisville KY-IN (KY Part) | 14675 | 51806 | 3.53 | SouthWest |
| 485-El Paso TX-NM (TX Part) | 8086 | 50888 | 6.29 | SouthCentral |
| 063-San Diego CA | 32641 | 49916 | 1.53 | SouthWest |
| 122-Miami FL | 27665 | 46645 | 1.69 | SouthEast-Florida |
| 241-Baltimore MD | 17504 | 45140 | 2.58 | NorthEaest |
| 531-Seattle WA | 34357 | 44704 | 1.30 | NorthWest |

(*continued*)

**Table 2.** (*continued*)

| Top 20 Distributors for Imports | | | | |
|---|---|---|---|---|
| 512-Virginia Beach-Norfolk VA-NC (VA Part) | 6326 | 40440 | 6.39 | SouthEast |
| 421-Philadelphia PA-NJ-DE-MD (PA Part) | 31876 | 36667 | 1.15 | NorthEast |

**Table 3.** Top 20 Collector Nodes of Exported Goods in 2022 (Exports22)

| Top 20 Collectors for Exports | Weighted in-degree | Weighted out-Degree | Ratio In/out Collector if > 1 | Modularity Class |
|---|---|---|---|---|
| 261-Detroit MI | 127842 | 39411 | 3.24 | NorthCentral |
| 363-New York NY-NJ-CT-PA (NY Part) | 121573 | 54327 | 2.24 | NorthEast |
| 061-Los Angeles CA | 107068 | 76493 | 1.40 | SouthWest |
| 486-Houston TX | 106616 | 75833 | 1.41 | SouthCentral |
| 487-Laredo TX | 88947 | 1106 | 80.40 | NorthCentral |
| 122-Miami FL | 68857 | 21977 | 3.13 | SouthEast - Florida |
| 171-Chicago IL-IN-WI (IL Part) | 64541 | 52719 | 1.22 | NorthCentral |
| 064-San Francisco CA | 49736 | 51439 | 0.97 | SouthWest |
| 362-Buffalo NY CFS Area | 45713 | 7756 | 5.89 | NorthEast |
| 483-Corpus Christi TX | 44234 | 5051 | 8.76 | SouthCentral |
| 471-Memphis TN-MS-AR (TN Part) | 42869 | 4891 | 8.77 | SouthCentral |
| 212-Louisville KY-IN (KY Part) | 42047 | 9161 | 4.59 | NorthEast |
| 223-New Orleans LA-MS (LA Part) | 39623 | 24456 | 1.62 | SouthCentral |

(*continued*)

**Table 3.** (*continued*)

| Top 20 Collectors for Exports | Weighted in-degree | Weighted out-Degree | Ratio In/out Collector if > 1 | Modularity Class |
|---|---|---|---|---|
| 531-Seattle WA | 36497 | 34052 | 1.07 | SouthWest |
| 485-El Paso TX-NM (TX Part) | 33562 | 4200 | 7.99 | SouthCentral |
| 512-Virginia Beach-Norfolk VA-NC (VA Part) | 33267 | 3922 | 8.48 | MidAtlantic |
| 489-Rest of TX | 32397 | 104127 | 0.31 | SouthCentral |
| 482-Beaumont TX | 30819 | 5864 | 5.26 | SouthCentral |
| 380-North Dakota | 29652 | 4146 | 7.15 | NorthCentral |
| 132-Savannah GA | 26323 | 2378 | 11.07 | NorthCentral |

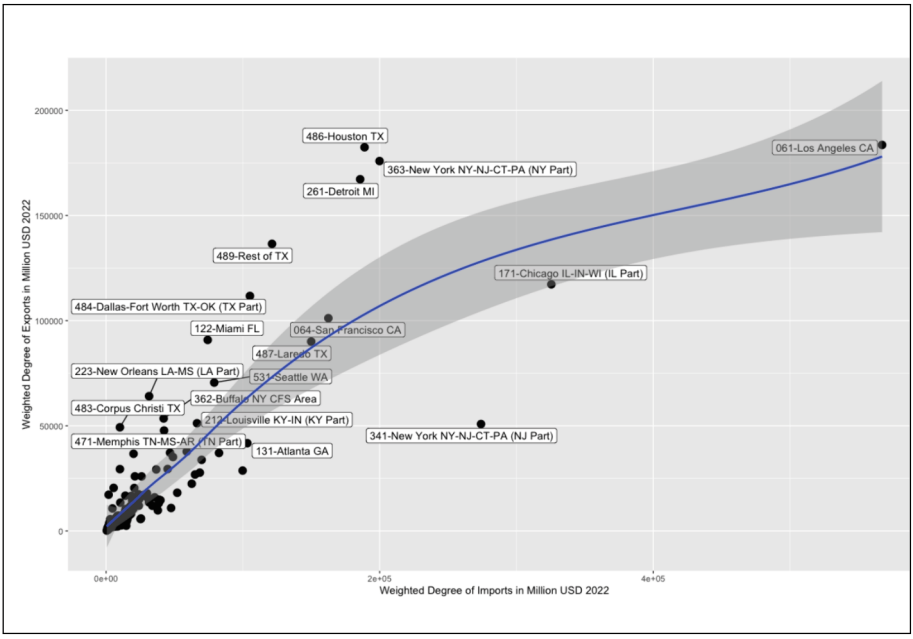


**Fig. 4.** Weighted degree in USD 2022 of international exports versus international imports

## 5 Summary and Conclusion

In this paper network science is used to explore freight flows among the 132 CFS and FAF areas used to collect and provide domestic and international freight origins and destinations as well as directional flows. The analysis is done with networks that are based on the value of freight flows. Freight flows are viewed as produced by a node, attracted to a node, and flowing onto directional links among all nodes considered here. Each link between two nodes is weighted by the metric used to quantify freight flow in US dollars of 2022. In this analysis network science statistics are computed that show formation of communities of nodes and their geographic location in the U.S., the importance of each of the 132 places (nodes) analyzed as hubs for the entire country and as hubs within their community of places, and the relationship of all these patterns with distance and contiguity. NAFTA and the border effect with Mexico and Canada are clearly exhibited by the patterns in the networks analyzed here. There are also internal to the U.S. borders among the communities identified here but these are different depending on domestic, import, and export goods.

With respect to the research questions in the introduction major topological differences are present among the three networks analyzed here and particularly between imports and exports. The topology of international imports and exports is not only different than the total freight flow networks but also shows the existence of trade corridors crossing the entire country with directional bias from North to South and the opposite.

We also find critical nodes making the networks here vulnerable and they are often major metropolises endowed with transportation infrastructure and large population centers. For international trade, hubs are the portals of imports or exports such as Laredo in Texas and Detroit in Michigan. The megaregion/community structure and membership of each node in a community has similarities to more traditional division of the US into functional regions but with key differences that network science tools enable us to identify. The modularity analysis for all the networks considered here identifies geographical subdivisions with strong relationships among places within a group (community or functional region). This multilateral analysis of freight flow reveals the formation of communities of freight flow that resemble in shape typical megaregions in the US (e.g., the Northwest, SouthEast, SouthCentral, MidWest, West). However, many places in the U.S. show marked differences in their membership to communities altering topological aspects of community formation and hub emergence depending on the role played by specific places explored in international import and export network topologies.

Throughout this analysis, however, the place endowment in major infrastructure investments in ports and airports and also border crossing portals such as Laredo, TX, and the role played by Los Angeles, New York (both portions of NY and NJ), Chicago, and Houston signify subregional economic roles that are different depending on the type of economic activity examined. The existence of communities that are different among domestic goods, imports and exports challenge the oversimplifying modeling approaches such as the shift-share type of models or new economic geography. However, alternative techniques are still to be developed using network science.

Network science is an efficient computational technique to identify through pattern recognition and visualization the topology of exchange networks and pinpoint many aspects of the movement of goods (major and minor hubs, different communities that depend on the type of metric and the type of freight transported). However, these tools are not designed to detect qualitative differences among nodes (e.g., Houston vs Rest of Texas). These are defined by CFS as origins and destinations and they can be very different geographies and regional subdivisions that are treated computationally as equivalent (i.e., computing in-degree and out-degree using the same definition). Similar issue is also the handling of some States as one node (e.g., Iowa, North Dakota, and so forth).

There are many additional next steps emerging from our experience with the network models here. *First*, an attempt could be made to increase the spatial resolution using the 3,244 counties and county equivalent subdivisions as nodes. For transportation planning applications the FAF origin and destinations have been disaggregated to a finer level of geographical subdivision in Florida to predict the impact of these flows on the transportation infrastructure [26]. Similar exercises are found in other geographies [27] and for emission modeling national level [28] and could be expanded and not only cover the entire country but possibly include Canadian and Mexican fine resolution subdivisions. *Second*, the node-to-node links do not distinguish the movements by different modes (e.g., air, rail, highway, pipeline) and this blurs the connectivity among nodes and the vulnerability of nodes and links. Developing methods to bring the physical aspects of mode -specific networks enhances our ability to view hubs as major intermodal facilities and assess their robustness using methods as in [29]. *Third*, the modularity method used here is found to be an effective way to group nodes together in communities based on the exchange of goods. There are competing approaches to detect communities that one could experiment with in future analyses as in city interaction [30].

## Appendix A

See Figure A.1 and Table A.1.




**Fig. A.1.** Maps of the 2017 Commodity Flow Survey Zones

**Table A.1** United States and their two-letter abbreviations

| | | | |
|---|---|---|---|
| ALABAMA | AL | MONTANA | MT |
| ALASKA | AK | NEBRASKA | NE |
| ARIZONA | AZ | NEVADA | NV |
| ARKANSAS | AR | NEW HAMPSHIRE | NH |
| CALIFORNIA | CA | NEW JERSEY | NJ |
| COLORADO | CO | NEW MEXICO | NM |
| CONNECTICUT | CT | NEW YORK | NY |
| DELAWARE | DE | NORTH CAROLINA | NC |
| FLORIDA | FL | NORTH DAKOTA | ND |
| GEORGIA | GA | OHIO | OH |
| HAWAII | HI | OKLAHOMA | OK |
| IDAHO | ID | OREGON | OR |
| ILLINOIS | IL | PENNSYLVANIA | PA |
| INDIANA | IN | RHODE ISLAND | RI |
| IOWA | IA | SOUTH CAROLINA | SC |
| KANSAS | KS | SOUTH DAKOTA | SD |
| KENTUCKY | KY | TENNESSEE | TN |
| LOUISIANA | LA | TEXAS | TX |
| MAINE | ME | UTAH | UT |
| MARYLAND | MD | VERMONT | VT |
| MASSACHUSETTS | MA | VIRGINIA | VA |
| MICHIGAN | MI | WASHINGTON | WA |
| MINNESOTA | MN | WEST VIRGINIA | WV |
| MISSISSIPPI | MS | WISCONSIN | WI |
| MISSOURI | MO | WYOMING | WY |
| DISTRICT OF COLUMBIA | DC | | |

Porto Rico, Guam, Northern Mariana Island, and Virgin Islands are not included in the analysis here.\

## References

1. Erickson, R.A., Hayward, D.J.: The international flows of industrial exports from US regions. Ann. Assoc. Am. Geogr. **81**(3), 371–390 (1991)
2. Hayward, D.J., Erickson, R.A.: The North American trade of US states: a comparative analysis of industrial shipments, 1983–91. Int. Reg. Sci. Rev. **18**(1), 1–31 (1995)
3. Hayward, D.J.: International trade and regional economies: the impacts of European integration on the United States. Routledge (2019)

4. Warf, B., Cox, J.: The US-Canada free trade agreement and commodity transportation services among US States. Growth Chang. **24**(3), 341–364 (1993)
5. Andresen, M.A.: Trade specialization and reciprocal trading relationships in Canada and the United States, 1989 and 2001. Ann. Assoc. Am. Geogr. **99**(1), 163–183 (2009)
6. Hwang, H.L., et al.: Freight Analysis Framework Version 5 (FAF5) Base Year 2017 Data Development Technical Report(No. ORNL/TM-2021/2154). Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States) (2021)
7. Bureau of Transportation Statistics: Commodity Flow Survey (CFS) (2023). https://www.bts.gov/cfs. Accessed Nov 2023
8. Wasserman, S., Faust, K.: Social network analysis: methods and applications. Cambridge University Press (1994)
9. Watts, D.J.: Small Worlds: The Dynamics of Networks Between Order and Randomness. Princeton University Press, Princeton (1999)
10. Barabasi, A.L.: Linked: How Everything is Connected to Everything else and What it Means for Business, Science, and Everyday Life. Basic Books (2014)
11. Barabasi, A.L.: Network Science. Cambridge University Press, Cambridge (2016)
12. Newman, M.E.: Random graphs as models of networks. Handbook Graph. Netw. **1**, 35–68 (2003)
13. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
14. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)
15. Leicht, E.A., Newman, M.E.: Community structure in directed networks. Phys. Rev. Lett. **100**(11), 118703 (2008)
16. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. **2008**(10), P10008 (2008)
17. Wang, J., Mo, H., Wang, F., Jin, F.: Exploring the network structure and nodal centrality of China's air transport network: a complex network approach. J. Transp. Geogr. **19**(4), 712–721 (2011)
18. Álvarez, N.G., Adenso-Díaz, B., Calzada-Infante, L.: Maritime traffic as a complex network: a systematic review. Netw. Spat. Econ. **21**(2), 387–417 (2021)
19. Xu, W., Zhou, J., Qiu, G.: China's high-speed rail network construction and planning over time: a network analysis. J. Transp. Geogr. **70**, 40–54 (2018)
20. Saberi, M., Rashidi, T.H., Ghasri, M., Ewe, K.: A complex network methodology for travel demand model evaluation and validation. Netw. Spat. Econ. **18**, 1051–1073 (2018)
21. Göbel, M., Araújo, T.: A network structure analysis of economic crises. In: Complex Networks and Their Applications VIII: Volume 2 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019 8 (pp. 547–560). Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-36683-4_44
22. Garcia, S., Rajtmajer, S., Grady, C., Mohammadpour, P., Mejia, A.:. Performance of a multi-layer commodity flow network in the United States under disturbance. In: Complex Networks and Their Applications VIII: Volume 2 Proceedings of the Eighth International Conference on Complex Networks and Their Applications Complex Networks 2019 8 (pp. 645–655). Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-36683-4_52
23. Belyi, A., Sobolevsky, S.: Network size reduction preserving optimal modularity and clique partition. In: International Conference on Computational Science and Its Applications, pp. 19–33. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-10522-7_2

24. Gouvêa, A.M., Vega-Oliveros, D.A., Cotacallapa, M., Ferreira, L.N., Macau, E.E., Quiles, M.G.: Dynamic community detection into analyzing of wildfires events. In: International Conference on Computational Science and Its Applications, pp. 1032–1047. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58799-4_74
25. Wang, X., Xue, F., Lu, S., Jiang, L., Wu, Q.: Functional community detection in power grids. In: International Conference on Complex Networks and Their Applications, pp. 883–894. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-36683-4_70
26. Viswanathan, K., Beagan, D., Mysore, V., Srinivasan, N.: Disaggregating freight analysis framework version 2 data for Florida: methodology and results. Transp. Res. Rec. **2049**(1), 167–175 (2008)
27. Rodriguez-Roman, D., Masoud, N., Jeong, K., Ritchie, S.G.: Goal programming approach to allocate freight analysis framework mode flow data. Transp. Res. Rec. **2411**(1), 82–89 (2014)
28. Auld, J., Rashidi, T.H., Lin, J.: Analysis of national county-level heavy-duty freight truck emissions. In: Transportation, Land Use, Planning, and Air Quality: Selected Papers of the Transportation, Land Use, Planning, and Air Quality Conference 2009, pp. 1–10 (2009)
29. Auerbach, J., Kim, H.: Measuring robustness and coverage of transportation networks with multiple routes and hubs. Ann. Am. Assoc. Geogr. **112**(6), 1741–1760 (2022)
30. Zhang, W., Thill, J.C.: Mesoscale structures in world city networks. Ann. Am. Assoc. Geogr. **109**(3), 887–908 (2019)