WILEY | Hindawi

Research Article

Urban Economic Structures as Multidimensional Networks: A Complex Systems Framework for Analyzing Economic Development Trajectories

Shade T. Shutters [1,2]

[1]*School of Complex Adaptive Systems, Arizona State University, P.O. Box 872701, Tempe 85282, AZ, USA*
[2]*Global Climate Forum, Neue Promenade 6, Berlin 10178, Germany*

Correspondence should be addressed to Shade T. Shutters; shade.shutters@asu.edu

Urban and regional systems often face the difficulty and necessity of structural transitions. These transitions, which can be imposed by external circumstances or initiated by a city itself, include energy transitions, transitions to a circular economy, transitions following a pandemic or natural disaster, or intentional policies meant to "move" an urban economy toward a desired state. However, what does economic structure mean in these cases? Traditional notions of economic structure are ambiguous and simplistic and typically consist of simple distributions, such as number of workers per industry. Yet to better understand, guide, or respond to system transitions, planners must move beyond these nebulous notions toward a theoretically grounded, quantifiable definition of economic structure. A recent trend emerging from the nexus of complexity science and urban science has been to operationalize urban economic structures as networks of interacting economic components. Typically based on colocation patterns of some type of entity, these networks have previously been constructed using economic entities such as products, occupations, or labor skills. Yet different types of entities also exhibit colocation patterns with each other, such as patent technology classes and industries. Here, those cross-entity colocation patterns are used to merge multiple types of entities into a single network representation of urban economies, offering a granularity not possible using a single node type. Occupations, industries, college degrees, and patent technology codes are merged into one multidimensional or multinodal network. As in previous studies, a dense core of highly connected entities emerges in this network. The network locations of individual cities are contrasted, and community detection algorithms are used to identify clusters of highly connected economic entities, showing that the densely connected network core is associated with science, technology, and business-related economic entities. Proximities between individual cities within the network are also measured revealing that many cities that are close to each other in the network are also close to each other in physical space. This framework offers potential applications including the ability to quantify structural change over time in response to a shock or to assess the relative difficulty of future desirable trajectories. More broadly, this framework might be applied to the study of structural change in other complex adaptive systems from human institutions to ecosystems.

1. Introduction

Urban and regional systems face the difficulty and necessity of structural transitions. These transitions, which can be imposed by external circumstances or initiated by a city itself, include energy transitions (e.g., decarbonization), transitions to a circular economy, transitions following a pandemic or other shock (building back better), or any

number of policies intended to "move" an urban economy toward an aspirational state. Yet a region's current economic structure constrains its potential pathways for future transitions [1–4], and policy-makers can benefit from understanding the constraints—and possibilities—applicable to their regions.

But what does structure mean in these cases? For most people, the term structure evokes an image of something that

can be touched and seen, like a building or a bridge. Unlike physical structures, the economic structure of a city is not tangible, and yet it directly impacts the lives of every resident of the city. While a person managing a building has the luxury of blueprints and maps that enable detailed planning and analysis, urban planners typically equate economic structure with simple distributions such as workers per occupation, GDP per industry, or exports per product.

To better understand developmental trajectories of urban systems, researchers have begun to view urban economies as networks of interacting economic entities. These networks—or spaces—capture the internal heterogeneity of economies by quantifying the interdependencies or relatedness between economic components such as labor occupations. This approach was brought to prominence in 2007 by Hidalgo et al. [5] who sought to understand how a nation's product space constrains its future development pathways. Variations of the methodology are now used frequently to compare cities, to explain past urban transitions, or to explore possible future trajectories of urban economies. Economic relatedness, which is central to this approach, has been calculated based on products produced [5], similarity of processes and outputs [6, 7], patterns of geographical colocation [2], supply-chain linkages in an input-output matrix [8], and even patterns of co-occurrence within other economic units, such as how skills co-occur within occupations [9]. These measures, which capture the magnitude of interdependence between pairs of entities, have been used to construct various parts of a city's economic structure, including its industry structure [10–12], occupational structure [2, 13, 14], labor skills structure [9, 15, 16], technology structure [17, 18], and scientific research structure [19]. These studies typically represent economic structures as networks for which some measure of relatedness or interaction is used to weight the links between parts.

These various spaces interact strongly with and influence each other yet are almost always studied in isolation. However, when these spaces are operationalized as networks, they may be merged and studied as a single integrated system. This merger can be accomplished either by introducing multiple node types within a single network or by linking networks in a multilayered network. Combining spaces can reveal previously undiscovered patterns of co-occurrence and relatedness and may offer novel insights regarding the constraints and developmental possibilities of urban economies. One example of this approach, at the country level, merges three types of economic entities based on colocation patterns among countries [20].

Here, this approach is applied at a subnational level by analyzing colocation patterns for four entity types across nearly 400 U.S. metropolitan areas. Multiple economic datasets are integrated to examine several instantiations of this previously unexplored multidimensional network level. Colocation patterns are analyzed not only of various industry pairs, occupation pairs, etc., but also for industry-occupation pairs, occupation-degree pairs, technology-industry pairs, and every other possible combination. The

result is a non-Euclidian map of the U.S.–a multinodal network in which individual cities can be located and their complex economic structures characterized in finer detail than has previously been possible. Whereas two cities may look almost identical in, for instance, industry space, they are likely distinct in a merged occupation-industry-technology-degree space. With such detail one might ask:

(1) Why are these cities similar in one economic dimension but not in others?

(2) What combinations of industries, skills, and technologies are most likely to grow quality jobs?

(3) What skills are absent or deficient and thus preventing a city from moving to a new area of the network?

(4) Which parts of the network are most vulnerable to shock or most enhance a city's economic resilience?

One key to addressing these questions is the ability to quantify the notion of similarity between two economic structures. Note that the strength of interaction between individual nodes translates to a measure of closeness or proximity between those nodes. Two nodes that have a high interdependence value will have high proximity in a network. Past studies have taken advantage of this concept to generalize across nodes and create an aggregate measure of proximity between two economic structures [3, 4]. Here, this method is applied to a series of multidimensional networks to calculate the proximity between every pair of MSAs at different levels of network granularity.

While this is similar to measuring physical proximities between cities, there are important differences between measuring proximity in physical space and in a network. First, this study measures proximities as opposed to distance. Thus, when two cities occupy the same place they will have a distance of 0, but a proximity of 1. Second, depending on the method one uses to aggregate proximity between two sets of nodes, proximity may not be symmetrical in a network. In other words, city A may be highly proximate to city B in a network, but it does not necessarily mean that city B is highly proximate to city A. Such a case could occur if, for instance, most of the economic entities present in city B were also present in city A, but few of the entities present in city A were present in city B.

2. Data and Methods

2.1. Geographical Units. The spatial units of analysis used in this study are U.S. Metropolitan Statistical Areas (MSAs). For each MSA collected, four categories of economic data were collected as described below. These various datasets are generated by different agencies and do not typically cover the exact same set of MSAs. Therefore, a process of harmonizing areas across datasets was required. The resulting list standardizes MSA codes and names across the four datasets and excludes MSAs that were not present in all four datasets. The final harmonized list contains 387 MSAs, including those in Puerto Rico, for which all four data sources were available.

2.2. *Data*. Data for each of four economic categories occupations, industries, degrees, and technologies, were procured or derived at the level of MSA from publicly available sources. Details of each dataset are described below. A summary of the number of entities present at each hierarchical level for each data type is presented in Table 1. In all cases, 2019 data are used unless otherwise noted.

2.2.1. *Employment per Occupation*. Occupational employment data are published annually for MSAs in the U.S. Bureau of Labor Statistics' Occupational Employment and Wage Statistics (OEWS) [21]. Data are published by Standard Occupation Classification (SOC) code at 2 levels of aggregation, "broad" (2-digit SOC) and "detailed" (6-digit SOC). By simple truncation of codes, two further levels of aggregation were created that were previously reported in the OEWS but ceased to be after 2017, "major" (3-digit SOC) and "minor" (5-digit SOC). Occupations having less than 10 employees within a given MSA are suppressed in the publicly available data, and for simplicity, these cases are taken to be 0 employment.

One idiosyncrasy of the OEWS data is that, despite being labeled as employment by MSA, data on urban areas in the six states of New England are published not for MSAs but for an alternative geographical unit known as New England City and Town Areas (NECTAs). NECTA boundaries are similar to New England MSA boundaries but are not identical. Thus, there exists both a Boston MSA and Boston NECTA, each with a different boundary. To harmonize OEWS data with remaining datasets, NECTAs were assumed to correspond to MSAs of the same or similar name. For example, occupation data are published for NECTA 74950, Manchester, NH, while industry, degree, and patent data are published for MSA 31700, Manchester-Nashua, NH. These spatial units are assumed to be equivalent in this study. Because this study's analytic methodology is based on proportions within each MSA and not on absolute numbers, this approach is justified (see the MSA Harmonization Table in the accompanying data repository for further details).

2.2.2. *Employment per Industry*. Industry employee data are published both quarterly and annually for MSAs in the U.S. Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) [22]. Here, the annual file is used. Data are reported by North American Industry Classification System (NAICS) code at three aggregation levels, 2-digit NAICS, 3-digit NAICS, and 4-digit NAICS. While further granularity is available (5- and 6-digit NAICS), these data are typically too sparse for the needs of this analysis and are not used in this study.

2.2.3. *Employment per College Degree*. Degree data for employed workers are taken from U.S. Census Bureau's American Community Survey (ACS) 1-year dataset [23]. The Census Bureau publishes a sample of this survey known as the Public Use Microdata Sample (PUMS), which covers approximately 1% of the U.S. population and which assigns one of 174 possible college degree codes to each worker for

TABLE 1: Summary of curated data and their hierarchical structures.

MSA data type	Hierarchy level	Unique codes
Occupations (O)	Major	22
	Minor	94
	Broad	450
	Detailed	781
Industries (I)	2-digit NAICS	20
	3-digit NAICS	92
	4-digit NAICS	304
College degrees (D)	acs1	6
	acs2	16
	cip2	38
	Detailed	174
Patent technology codes (T)	Section	10
	Class	129
	Subclass	665

both a first and second (if applicable) college degree. One of the possible degree values is "N/A or less than a bachelor's." PUMS data are aggregated to geographical units known as a Public Use Microdata Areas (PUMAs) which do not correspond to any other generally used spatial unit for collection of regional statistics in the U.S. Therefore, a crosswalk, published by iPUMS, is used to allocate PUMA population characteristics to MSAs [24]. Degrees are tabulated at four levels of aggregation: detailed, which is the 4-digit code used in the raw PUMS data, and both 3-digit and 1-digit ACS categories are defined in [25], each of which is an aggregation of detailed degree codes. The fourth aggregation uses an alternative categorization based on the 2-digit Classification of Instructional Programs (CIP) codes defined by the U.S. Department of Education [26]. Each 2-digit CIP code is an aggregation of 4-digit ACS codes. For further detail, refer to the degree code crosswalk included in the data repository accompanying this paper.

2.2.4. *Technologies per Patent*. The U.S. Patents and Trademark Office (USPTO) publishes data on each patent it grants including the county of residence of each inventor and a list of each cooperative patent classification (CPC) code included on the patent [27]. CPC codes represent the specific technologies present in a patent based on common subject matter, and a single patent may include several CPC codes. The number of times each CPC code was used on a patent is then tallied for each county in a given year. Using the 2020 official mapping of counties to MSAs from the U.S. Office of Management and Budget [28], county totals are then aggregated to MSA totals. CPC codes on patents having multiple inventors are assigned to the county of each inventor unless the inventors reside in the same county. In cases where multiple inventors live in the same county, each CPC code used on a patent is tabulated only once for that county. Likewise, if a patent has inventors in multiple counties, but those counties are all within the same MSA, the patent's CPC codes are counted only once for the corresponding MSA. CPC code usage is tabulated at three aggregation levels, "section" (1-digit CPC), "class" (3-digit CPC), and "subclass" (4-digit CPC). Because patenting rates

can exhibit high variability from year to year, the time window is expanded to include patents for the years 2015–2019 for each MSA. Finally, only utility patents are considered in this study (thus ignoring design patents and plant patents).

2.2.5. Data Used for Networks. For each network constructed, only one hierarchical level of data for each of the four data types shown in Table 1 is selected. Using hierarchical levels with a low number of entities (codes) does not provide enough heterogeneity for meaningful networks. On the other hand, using hierarchical levels with a large number of entities increases the granularity of the analyses but it also exponentially increases computation time. Furthermore, as the number of entities increases the data become more sparse and therefore less useful.. Thus, this study is focused on intermediate hierarchical levels of each data type. More precisely, of 144 possible combinations of data slices, four networks were constructed and analyzed using slices as shown in Table 2. Results and visualizations in this paper are primarily derived from networks 1 and 3.

2.3. Calculating Interdependence. After selecting a hierarchical level at which to aggregate each of the four data types in Table 1, interdependence value between each pair of economic entities is calculated using the method described in [2]. This method requires that raw data first be recast as presence-absence data so that a matrix of employees by occupation of MSA, for instance, becomes a matrix of 1's and 0's (Figure 1). An economic entity is determined to be present in an MSA if the location quotient (LQ) for that entity is greater than one in the selected MSA. The LQ of entity e in MSA m is defined as follows:

$$\text{LQ}_{e,m} = \frac{\left(c_{e,m} / \sum_e c_{e,m} \right)}{\left(\sum_m c_{e,m} / \sum_m \sum_e c_{e,m} \right)}, \qquad (1)$$

where c is the count of entity e in MSA m such as the number of employees in a particular occupation in m. An entity is determined to be present in m if $\text{LQ}_{e,m} \geq 1$ and absent if $\text{LQ}_{e,m} < 1$.

Location quotients and the determinations of present or absent are calculated for each data type separately before being merged into a single list of presence-absence for all entities. Using this master list of presence-absence data, the method of [2] is applied to MSA co-occurrence patterns to calculate an interdependence value x between every pair of entities i and j as follows:

$$x_{i,j} = \frac{P\left[LQ_{i,m} > 1, LQ_{j,m} > 1 \right]}{P\left[LQ_{i,m'} > 1 \right] P\left[LQ_{j,m''} > 1 \right]} - 1, \qquad (2)$$

where m, m', and m'' denote randomly selected MSAs. Thus, $x > 0$ when two entities co-occur in the same MSAs more often than expected by chance, and $x < 0$ when they co-occur less often than expected.

The resulting entity × entity matrix of interdependence values is used to create a network in which each economic entity is a node and the interdependence values between entities are the weights of links between nodes. Because

study data cover four types of economic entities, the resulting networks have four node types, each of which is interdependent with and, therefore, connected to every other node. Thus, the resulting networks are complete, weighted, nondirected, and multinodal networks.

2.4. Calculating Proximities between MSAs. Having calculated an interdependence value x between every pair of economic entities i and j, an aggregate measure of proximity between any two MSAs p and q can then be calculated.

To begin, let S_p be the set of all economic entities present in MSA p. Consider an entity i not present in S_p, $i \notin S_p$. For each such entity i, an aggregate value of proximity is calculated between i and all the members of S_p. This measure, known as the transition potential of i [2], is defined as follows:

$$V_i\left(S_p\right) = 1 - \prod_{j \in S_p} \left(1 - c\left(x_{ij} + 1 \right) P\left[\text{LQ}_{i,p} > 1 \right] \right), \qquad (3)$$

where c is simply a tuning parameter chosen to result in a useful range of values of V. Thus, the transition potential of entity i, where $i \notin S_p$, is the aggregate proximity of i to all the entities $j \in S_p$.

Finally, this value is calculated for every member of S_q and aggregated to a measure of proximity R from MSA p to MSA q:

$$\underset{p \longrightarrow q}{R} = \frac{1}{N_q} \sum_{i \in S_q} V_i\left(S_p\right)^{1-\delta_i}, \qquad (4)$$

where N_q is the number of entities in S_q and δ_i is an indicator function: it is 1 if i is already a member of S_p and 0 if it is not [3, 4]. Thus, an entity i has $V_i = 1$ if it is already present in MSA p and the value determined by function (3) if it is not. In the case that all entities in S_q are already present in S_p, the proximity of S_p to S_q would be 1, the maximum possible proximity.

It is critical to note that proximity R is not symmetrical. That is, it is not required that $R_{p \longrightarrow q} = R_{q \longrightarrow p}$. To illustrate, consider the case where all entities present in MSA p are present in MSA q, but q also has many more entities present that are not present in p. In this case, the proximity of p to $q R_{p \longrightarrow q}$ would be relatively low, while $R_{q \longrightarrow p} = 1$.

2.5. Network Visualizations. Networks were visualized using the igraph package for R [29]. Pairwise interdependence values are used as link weights, and link weights less than 0 are dropped before rendering. The igraph package's multidimensional scaling layout algorithm is used to spatially arrange nodes, and only links having a weight above a certain threshold value are shown to highlight those pairwise relationships having the highest interdependence values.

2.6. Community Detection. Clusters of related economic entities within each network were identified using igraph's walktrap community detection algorithm [29]. The walktrap algorithm uses random walks to identify densely connected

TABLE 2: Description of data slices used in the four networks constructed for this study.

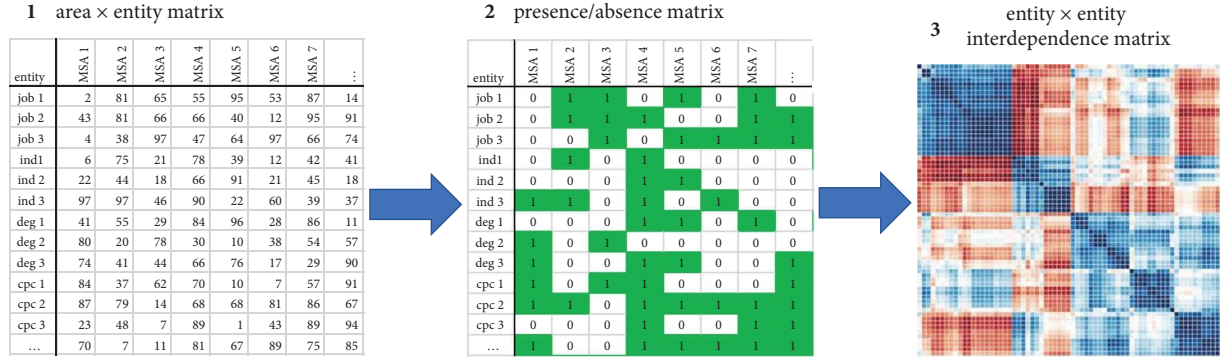|	Network 1	Network 2	Network 3	Network 4
Occupation level	Major ($O = 22$)	Major ($O = 22$)	Minor ($O = 94$)	Broad ($O = 450$)
Industry level	2-digit ($I = 20$)	2-digit ($I = 20$)	3-digit ($I = 92$)	4-digit ($I = 304$)
Degree level	acs2 ($D = 16$)	acs2 ($D = 16$)	cip2 ($D = 38$)	cip2 ($D = 38$)
Technology level	Section ($T = 10$)	Class ($T = 129$)	Class ($T = 129$)	Class ($T = 129$)
Total nodes (N)	68	187	353	921

FIGURE 1: Extracting additional information from distributions to build a network. Starting with distributions across spatial units (1), raw counts are converted to presence-absence data (2), and then compared to null models of colocation to quantify the interaction or interdependence between every pair of economic entities (3). Those interdependencies can be positive or negative and are used to construct a network in which nodes are the economic entities of interest and pairwise interdependencies are the link weights between nodes.

subnetwork communities within a larger network and requires that nonpositive link weights be dropped. Because walktrap is a hierarchical detection method, users may select the number of communities to be detected. Output includes both a dendrogram and community membership list of every node.

3. Results and Discussion

Networks were constructed for various combinations of data hierarchy levels as shown in Table 2. A rendering of network 3, with number of nodes $N = 353$, is shown in Figure 2. Links with an interdependence value less than 2 are not shown so that only pairs of entity pairs with relatively high interdependence are highlighted.

As with previous work using only one type of economic entity, Figure 2 displays a dense core of highly connected entities. All networks created using the data slices in Table 2 displayed a similar dense core. Previous studies have shown this dense core to be associated with high-wage, knowledge-intensive dimensions of an economy, such as so-called "creative" jobs [2, 3]. Here, for the first time, a multidimensional space reveals that this dense network core includes not only certain occupations but also certain industries, college degrees, and technologies.

Note that although patents are generally associated with creativity and innovation, many CPC codes do not appear in the dense network core. However, studies have demonstrated that some technology codes are more associated with innovation than others [30], and only a small fraction of patents lead to innovations that affect markets [31]. Thus, perhaps, it should not be unexpected that many CPC codes
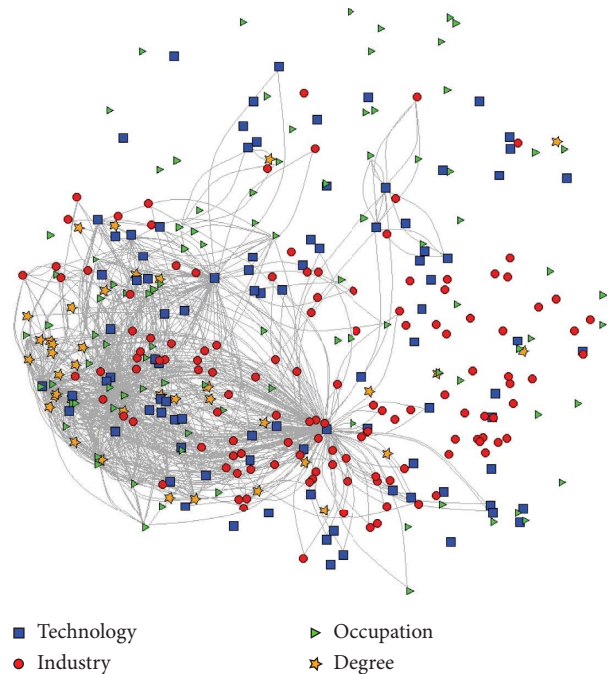

FIGURE 2: Visualizing the 2019 U.S. multimodal economic network structure using network 3 ($N = 353$). Links with value less than 2 are not shown. Note the characteristic dense core of highly connected entities which has been shown in other studies to be associated with high-wage and/or creative activities. Networks constructed with other hierarchical levels of data aggregation display a similar dense core.

appear outside of the core, and the network approach developed here may offer a new method of assessing the innovative potential of each CPC technology code.

The concentration of college degrees in the core is even more pronounced, with nearly all degrees falling within the dense core. Only a small number of relatively ubiquitous degrees, such as education degrees, fall outside the core. Thus, one interpretation of this study's network approach reinforces the notion that advanced education is critical for a knowledge-based, innovation economy.

3.1. Locating MSAs in the Network.

As with related studies that used only a single dimension, this study finds that the network location of individual MSAs varies considerably. Figure 3 compares two economically distinct MSAs, revealing that San Francisco is primarily located in the creative core while Dalton, Georgia, is not. However, Figure 3 also shows that San Francisco is almost exclusively within the creative core which could indicate that San Francisco's creative economy may be less sustainable than those of other knowledge intensive regions because it lacks the ancillary economic activities required to support a high-wage, high-tech sector. In contrast, other MSAs associated with creativity, including Boston and San Jose, do occupy several nodes outside the dense core.

While researchers have previously mapped cities within network composed of single entity types, merging spaces provide an unprecedented level of granularity in the underlying network map and an ability to clearly distinguish between various economic structures. To demonstrate this advance, consider two MSAs, Blacksburg, Virginia (13980) and Muskegon, Michigan (34740). Data for 2019 show that these MSAs have the same presence/absence pattern for industries. Thus, when located in a network space composed of industries only, these cities would be indistinguishable and have a proximity of 1. Their locations are identical.

However, when located in the merged, multidimensional networks created in this study, the cities are distinct. Even in network 1 ($N = 68$), the coarsest multidimensional network created in this study, substantial differences emerge. Figure 4 highlights only those economic entities that are unique to each MSA. Note that Blacksburg has several college degrees present, primarily in the network's creative core, while Muskegon has none. Each MSA also specializes in different technologies. In quantitative terms, the correlation between the presence/absence vectors of industries—taken as a vector of 1's and 0's—for Blacksburg and Muskegon is $R = 1.00$, while the correlation between the presence/absence vectors of all economic entities in network 1 is $R = 0.35$.

Thus, a key advantage of multidimensional economic networks, in this context, is the ability to identify and quantify differences between cities that may be invisible within a single dimension. While this example uses the least detailed multidimensional network, networks with higher granularity should only enhance the ability to identify subtle and unique components of each city's economic structure.

3.2. Identifying Economic Communities.

It is first important to note that the membership and nature of communities detected will be different depending on the network used, and on the number of communities that the algorithm is instructed to detect. Results of the community detection algorithm produced intuitive results even at the coarsest level of data. In network 1, which is constructed with only 68 nodes, five economic communities were detected (Figure 5), and were subjectively named based on the members of each (Table 3). These communities include manufacturing, business and services, construction, STEAM (science, technology, engineering, arts, and mathematics), and general commercial economic communities.

Note that all college degrees other than education fields are found in the STEAM and business/services communities. Education is the lone degree within the general commercial community, which likely corresponds to the ubiquity of educational activities. The degree code for "less than a bachelor's" falls within the manufacturing community which may indicate limited requirements for advanced education in manufacturing relative to other industry sectors. On the other hand, the manufacturing community contains many patent technology codes, suggesting that the nature of U.S. manufacturing has evolved to focus more on technologies than labor skills.

3.3. Determining Marshallian Channels of Agglomeration.

A long-standing goal of regional economics is to understand the causes of observed industry agglomerations [32–35]. Researchers generally focus on three so-called Marshallian channels or economic forces that lead certain industries to colocate more frequently than expected, and research has sought to disentangle which channel is the primary driver of different industry agglomerations [12, 13, 36–40]. These channels, which are assumed to infer different cost benefits, include the followings:

(1) Labor access—industries that share similar labor requirements (e.g., similar skill sets) tend to colocate because they need access to the same pool of workers

(2) Industry linkages—industries that are in a customer-supplier relationship tend to colocate because they can lower transportation costs by being near each other

(3) Knowledge spillovers—industries that can gain technological knowledge from each other by being in close proximity (e.g., where employees can mingle) tend to colocate

The use of multidimensional networks can contribute to this discourse by explicitly determining the types of economic entities that are associated with closely linked industry pairs. Industry agglomerations are first identified by selecting industry pairs (or groups) with a relatively high interdependence value, indicating that the pair of industries is generally both present and absent in an area. To explain what forces may underly a high interdependence value, one would then identify other economic entities to which both industries are highly linked. For instance, if a pair of agglomerating industries are both highly linked to several occupations, it is likely the industries coagglomerate because of shared labor
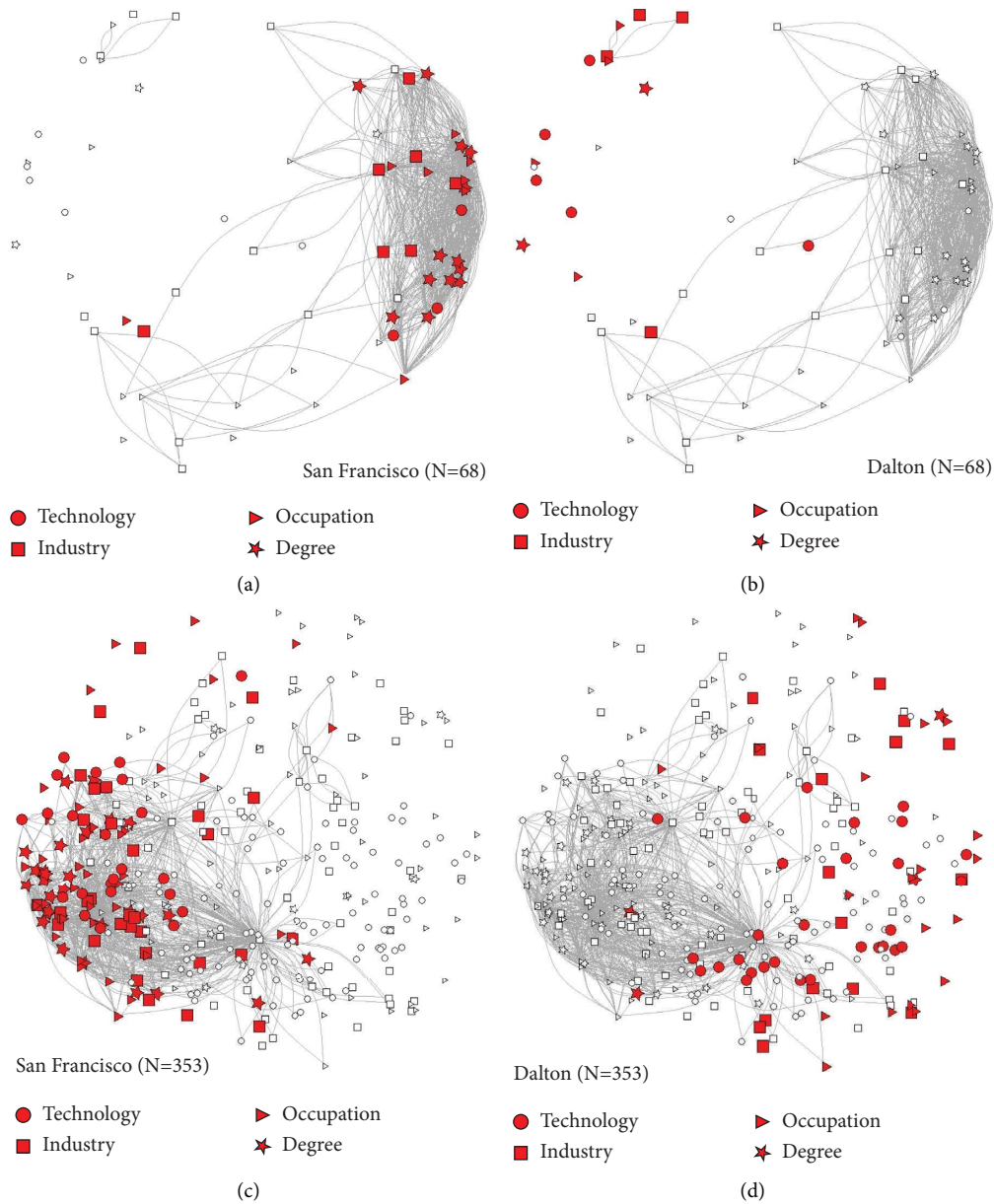
Figure 3: Comparing the network locations of individual MSAs. When only the entities present in an MSA are highlighted, the network location of that MSA is revealed. Here, the locations of San Francisco (a and c) and Dalton, Georgia (b and d), are compared first in network 1, with $N = 68$ (a and b) and second in network 3, with $N = 353$ (c and d). MSA locations are distinctly different with San Francisco appearing almost exclusively within a dense network core and Dalton almost exclusively outside that core. Differences in network locations can be quantified to give a proximity between any two economic structures. Links with value less than 0.4 are not shown in panels (a) and (b) while links with value less than 2 are not shown in panels (c) and (d).

needs. If, on the other hand, a pair of agglomerating industries are both highly linked primarily to technologies, it is more likely they coagglomerate because of knowledge spillovers. Industry linkages are likely the driving force in cases where two agglomerating industries are most highly linked with other industries. Cases in which agglomerating pairs are mostly closely linked with college degrees present the tantalizing possibility that not only is labor access likely the driving agglomeration force but also it is access to particular types of cognitive skills, as opposed to physical skills, that is driving agglomeration.

The high-level communities identified in Table 3 (and shown in Figure 5) already point to the ability of multidimensional networks to distinguish different drivers of industry agglomeration. Notice that cluster 1 (STEAM) is primarily composed of college degrees along with a single industry, "Professional and Technical Services." The fact that this sector appears alone with several degree entities suggests that firms in this sector coagglomerate not only because of shared requirements for labor but also for highly skilled labor. On the other hand, cluster 2 (manufacturing), which also has a single industry, is dominated not by college
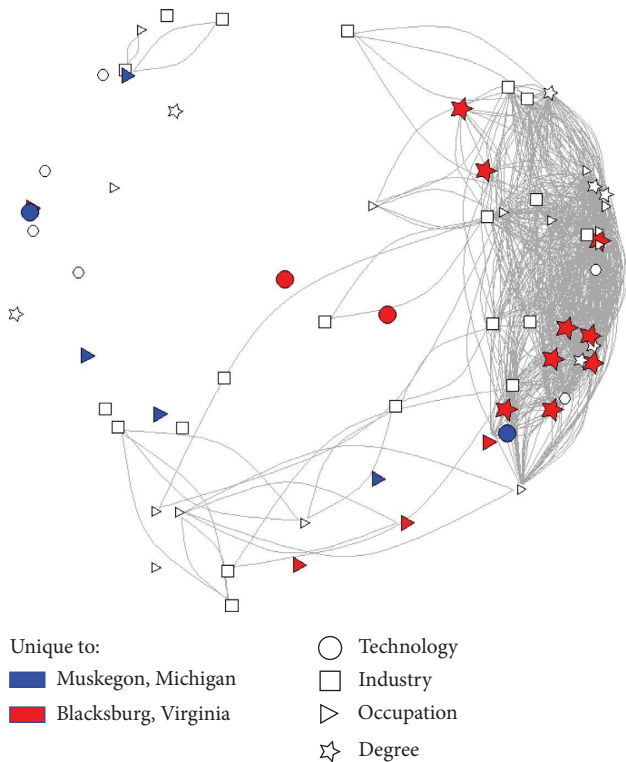
FIGURE 4: Juxtaposition of two MSAs highlighting only differences between the two. Using network 1 ($N = 68$), Muskegon, Michigan and Blacksburg, Virginia are identical in terms of industry structure, meaning that they share the exact same set of industries present. Yet in a multidimensional space, substantial differences between the two emerge, particularly in terms of college degrees. Note also that Blacksburg has many more unique entities within the dense network core. Links with value less than 0.4 are not shown.

degrees but by technologies, suggesting that firms in this sector coagglomerate more because of knowledge spillovers. Note that the communities shown in Table 3 are detected only in the coarsest network of this study (network 1) and more detailed networks are likely to reveal communities within industry sectors that agglomerate for different reasons.

Regardless of the network used, the fact that all pairwise relationships are quantified means that the relative contribution of each Marshallian channels to observed agglomerations should also be quantifiable. There are likely several ways to construct such a measure, but such an endeavor is beyond the scope of the current study, and it is left to future studies to undertake this promising challenge.

3.4. Proximities between MSAs.
Using network 3 ($N = 387$), proximities were calculated between every MSA pair. A selection of the highest and lowest proximity values between MSAs is shown in Table 4. Recall that proximity ranges from 0 to 1, with 1 being the highest possible proximity. One striking feature of these values is the fact that many of the cities having high economic proximity also have high physical proximity. Of the five highest economic proximity values, three MSA pairs—Chattanooga-Dalton, Riverside-El
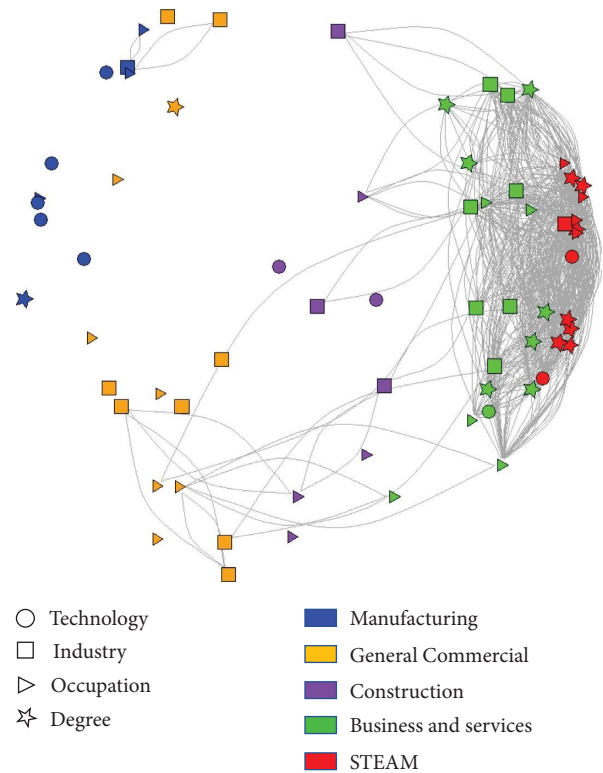

FIGURE 5: Location of five economic communities detected in network 1. Results are from the walktrap community detection algorithm using network 1 (nodes $N = 68$). The number of communities was chosen a priori. Subjective community names were assigned by the author. Note that the dense core is composed of STEAM and business and services entities, see Table 3 for a detailed list of members of each community. Links with value less than 0.4 are not shown.

Centro, and Charleston-Sumter—are physically adjacent to each other, while a fourth pair, North Port-Sebring, is separated by just one county. Thus, spatial context and physical embeddedness likely play key roles in the evolution of urban economic structures. This phenomenon may also indicate that the definitions of metropolitan statistical areas do not adequately capture regional units in the U.S.

On the other hand, the fourth highest proximity is between Guayama, Puerto Rico, and Oklahoma City, demonstrating that structural similarity can also emerge in places far apart physically.

3.5. Identifying Transition Gaps.
While the previous section focused on the network proximity between two MSAs, more generally, the network proximity between any two economic structures can be measured. There is no requirement that those structures represent existing cities. A structure might instead be a desired future economy. Thus, proximity can be calculated, for instance, between every city and the communities detected above.

To demonstrate, Table 5 presents the 10 closest MSAs, in terms of proximity in network 1, to the STEAM and business & services communities detailed in Table 3. Such measures not only give urban planners a quantified metric of the

TABLE 3: Economic communities (5) detected in network 1 ($N = 68$), see Figure 5 for visualization.

Community 1: STEAM

Occupations
(i) Management
(ii) Business and financial operations
(iii) Computer and mathematical
(iv) Arts, design, entertainment, sports, and media
Industries
(i) Professional and technical services
Degrees
(i) Computers, mathematics, and statistical
(ii) Physical and related science
(iii) Social science
(iv) Engineering
(v) Literature and language
(vi) Visual and performing arts
(vii) Communications
Technologies
(i) Physics
(ii) Electricity

Community 2: manufacturing

Occupations
(i) Installation, maintenance, and repair
(ii) Production
(iii) Transportation and material moving
Industries
(i) Manufacturing
Degrees
(i) N/A (less than bachelor's degree)
Technologies
(i) Performing operations; transporting
(ii) Textiles; paper
(iii) Fixed constructions
(iv) Mechanical engineering; lighting; heating; weapons
(v) General tagging of new technological developments

Community 3: construction

Occupations
(i) Community and social service
(ii) Building and grounds cleaning and maintenance
(iii) Office and administrative support
(iv) Construction and extraction
Industries
(i) Construction
(ii) Administrative and waste services
(iii) Unclassified
Degrees
(i) (None)
Technologies
(i) Human necessities
(ii) Chemistry and metallurgy

TABLE 3: Continued.

Community 4: general commercial

Occupations
(i) Healthcare practitioners and technical
(ii) Healthcare support
(iii) Protective service
(iv) Food preparation and serving related
(v) Sales and related
(vi) Farming, fishing, and forestry
Industries
(i) Agriculture, forestry, fishing, and hunting
(ii) Mining, quarrying, and oil and gas extraction
(iii) Utilities
(iv) Wholesale trade
(v) Retail trade
(vi) Transportation and warehousing
(vii) Health care and social assistance
(viii) Accommodation and food services
Degrees
(i) Education
Technologies
(i) (None)

Community 5: business and services

Occupations
(i) Architecture and engineering
(ii) Life, physical, and social science
(iii) Legal
(iv) Educational instruction and library
(v) Personal care and service
Industries
(i) Information
(ii) Finance and insurance
(iii) Real estate and rental and leasing
(iv) Management of companies and enterprises
(v) Educational services
(vi) Arts, entertainment, and recreation
(vii) Other services, except public administration
Degrees
(i) Biological, agricultural, and environmental sciences
(ii) Psychology
(iii) Multidisciplinary studies
(iv) Science- and engineering-related
(v) Business
(vi) Liberal arts and history
(vii) Other fields
Technologies
(i) Other—no section assigned

Note. Occupation level: major (2-digit), industry level: 2-digit NAICS, degree level: acs2, and technology level: cpc section (1-digit).

TABLE 4: Five highest and a selection from the 20 lowest MSA-to-MSA proximity values using network 3.

MSA a	MSA b	$R_{a \longrightarrow b}$
5 highest values of R		
Chattanooga, TN-GA (16860)	Dalton, GA (19140)	0.823
Riverside-San Bernardino-Ontario, CA (40140)	El Centro, CA (20940)	0.816
North Port-Sarasota-Bradenton, FL (35840)	Sebring, FL (42700)	0.808
Oklahoma city, OK (36420)	Guayama, PR (25020)	0.803
Charleston-North Charleston, SC (16700)	Sumter, SC (44940)	0.791
5 of the 20 lowest values of R		
Dalton, GA (19140)	San Francisco-Oakland-Hayward, CA (41860)	0.073
Bloomsburg-Berwick, PA (14100)	San Jose-Sunnyvale-Santa Clara, CA (41940)	0.105
Odessa, TX (36220)	New York-Newark-Jersey city, NY-NJ-PA (35620)	0.108
Dalton, GA (19140)	San Diego-Carlsbad, CA (41740)	0.108
Hinesville, GA (25980)	Milwaukee-Waukesha-West Allis, WI (33340)	0.123

TABLE 5: MSAs closest to the STEAM and business & services communities on network 1, detailed in Table 3.

| Closest to STEAM community	Closest to business & services community
San Francisco-Oakland-Hayward, CA (41860)	Boston-Cambridge-Newton, MA-NH (14460)
Austin-Round Rock, TX (12420)	Raleigh, NC (39580)
Washington-Arlington…, DC-VA-MD-WV (47900)	Philadelphia-Camden…, PA-NJ-DE-MD (37980)
Seattle-Tacoma-Bellevue, WA (42660)	Baltimore-Columbia-Towson, MD (12580)
Durham-Chapel Hill, NC (20500)	New York-Newark-Jersey city, NY-NJ-PA (35620)
Portland-Vancouver-Hillsboro, OR-WA (38900)	San Francisco-Oakland-Hayward, CA (41860)
New York-Newark-Jersey city, NY-NJ-PA (35620)	Denver-Aurora-Lakewood, CO (19740)
Boulder, CO (14500)	Minneapolis-St. Paul-Bloomington, MN-WI (33460)
San Diego-Carlsbad, CA (41740)	Madison, WI (31540)
Bridgeport-Stamford-Norwalk, CT (14860)	San Diego-Carlsbad, CA (41740)
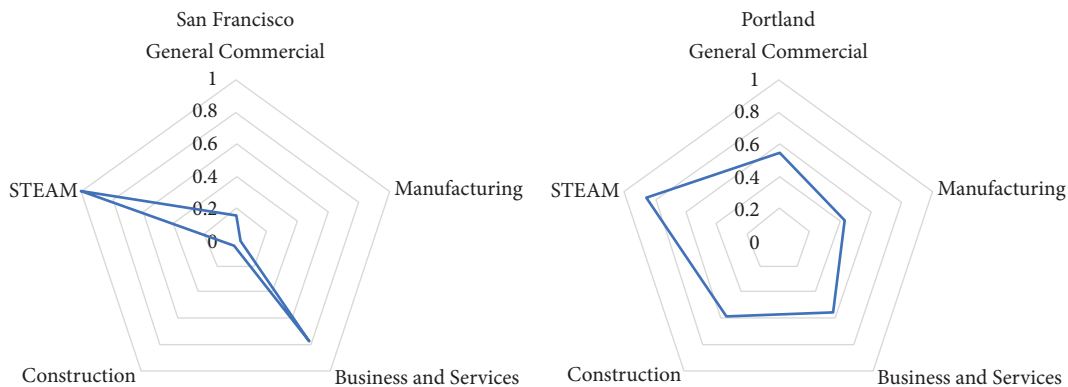
FIGURE 6: Radar diagrams showing proximities of two MSAs to each of the five communities, detailed in Table 3. While Portland shows a diversity of high-to-moderate proximities among communities, San Francisco is high in proximity to two communities and close to 0 for all others. This could indicate a lower diversity of economic activities in San Francisco compared to Portland and is an area for future research. Proximities were calculated using network 1 ($N = 68$).

difficulty of transitioning to a desired economic structure but also enable planners to identify the economic entities missing in their local economy that must be grown or acquired. For instance, an MSA seeking to grow a STEAM economy as defined in Table 3 may find that it is missing certain degrees or technologies. Planners could use this information to develop strategies for acquiring those missing economic elements, for instance, by working with local community colleges. Coupled with network proximity measures, planners could conceivably outline a long-term

plan including time-phased intermediate steps for transitioning to a desirable future economy in a manner that enhances returns on investment.

Proximities to various communities can also be compared within an MSA to assess a degree of diversity for each region. Such assessments are visualized as radar diagrams for two MSAs in Figure 6. Note that, while San Francisco has high proximity to the STEAM and business & services communities, it has very low proximity to others. In contrast, Portland has a moderate-to-high proximity to all

communities. This could indicate that Portland has a more diverse and balanced economy than San Francisco, though more research is needed to determine the nature and interpretation of this type of analysis.

4. Conclusion and Future Directions

Here, an increasingly popular framework for viewing urban economies as networks has been extended by merging four separate economic networks into a single multidimensional space. In these networks, nodes represent one of four components of an urban economy, such as industries or occupations, while link weights quantify the interdependence between pairs of economic components and are derived from the co-occurrence patterns of components across a country's urbanized areas. While researchers have previously attempted to merge networks at the national level, this study implements this technique at the level of metropolitan statistical areas in the United States. The resulting network enables a refined characterization of individual regional economies not possible when using only a single network. This network serves as a non-Euclidian map in which cities can be located and compared, and their developmental trajectories analyzed over time.

The key benefit of this network approach to representing urban economies is that it provides a method of visualizing and quantifying economic structure in a manner more sophisticated and useful than mere frequency distributions. Being represented as a network makes urban economies amenable to a suite of computational methods and metrics from network science and graph theory. The key contribution of this study is that it expands this network framework into multiple economic dimensions, or node types, offering a large improvement in the granularity with which an individual urban economy can be characterized.

There are several potential policy applications for this framework, and many ways the framework can be improved, some of which are outlined below.

4.1. Refining the Methodology. In addition to numerous possible applications that have not been discussed, there remains much fundamental research to do in this domain. The methodology used in this study requires multiple steps, including the following:

(1) Determine whether an entity is present or absent in an area

(2) Quantify the interaction between any two entities

(3) Measure the proximity or distance between two subnetworks within a network

Yet for each of these stages depicted in Figure 1, there exists no method that is generally accepted in the relevant literature as superior to others. For instance, several measures have been proposed for quantifying interactions between entities based on co-occurrence patterns, though there has been little work to determine which is optimal. Future research should seek to compare existing measures, to explore new measures, and to determine which measures offer the most meaningful outputs.

4.2. Dealing with Negative Interactions. An obstinate issue in working with network spaces like those developed in this study is how to treat negative interaction values. While negative interactions are a hallmark of the ecological thinking that inspired much of this methodology, they have been difficult to deal with in economic networks. Indeed, in the foundational work by Hidalgo and Hausmann [5], which most studies in this area acknowledge as inspiration, negative interactions are not permitted at all. Even in studies where they are, negative interactions introduce difficulties in network renderings and community detection and so typically end up being ignored, such as was required in this study.

Yet such interactions carry important information. In ecosystems, species interacting negatively may competitively exclude one another so that only one species is generally present in an area. In economic networks, such considerations become important if policy-makers were to apply this methodology, for example, to identify potential targets for job growth. Such targets may be less desirable if they interact negatively with many of a city's existing economic strengths.

Consider also the previous discussion on determining agglomeration forces. One possibility that is seemingly not addressed in relevant literature is that two industries agglomerate not because they are drawn together but because they are both driven away from a common set of other industries. Even in cases where industries are drawn to each other, it is likely that negative interactions play a moderating role in each channel.

Such issues related to negative economic interactions should be more thoroughly investigated and where needed, such as in community detection, new methods should be developed that integrate such relationships instead of ignoring them. This is likely an excellent opportunity for cross disciplinary collaboration between ecosystem theorists and regional economists.

4.3. Next Steps. Another aspect of this study that demands further examination is the relationship between network-based results and spatial patterns. For instance, do MSAs that fall primarily into one network community cluster spatially across the U.S.? Or does MSA population size better explain the network location and community of individual MSAs? These important questions comprise a fruitful research agenda at the nexus of regional science and complexity science.

Finally, the method used in this study is applicable to any country for which data are available for multiple economic entities at a meaningful level of geography (generally unified labor market areas). This could be extended to groups of countries (e.g., European Union), provided they share a common coding schema and data structure. Thus, this domain of inquiry has potential for a broad and diverse suite of future applications and research directions with global appeal.

Data Availability

Disclosure

An early version of this manuscript was posted as a preprint at [41].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Florida, S. T. Shutters, and G. Spencer, "Pathways to Ontario's Knowledge Economy: a system for identifying existing regional strengths and future prospects," 2016, https://rogerlmartin.com/mpi/wp-content/uploads/2016/2008/2016-MPIWP-2001_Pathways-to-Ontarios-Knowlede-Economy_Florida-Shutters-Spencer.pdf.

[2] R. Muneepeerakul, J. Lobo, S. T. Shutters, A. Goméz-Liévano, and M. R. Qubbaj, "Urban economies and occupation space: can they get there from here?" *PLoS One*, vol. 8, no. 9, Article ID e73676, 2013.

[3] S. T. Shutters, R. Muneepeerakul, and J. Lobo, "Constrained pathways to a creative urban economy," *Urban Studies*, vol. 53, no. 16, pp. 3439–3454, 2016.

[4] S. T. Shutters, R. Muneepeerakul, and J. Lobo, "How hard is it for urban economies to become green?" *Environment and Planning B: Planning and Design*, vol. 43, no. 1, pp. 198–209, 2016.

[5] C. A. Hidalgo, B. Klinger, A. L. Barabási, and R. Hausmann, "The product space conditions the development of nations," *Science*, vol. 317, no. 5837, pp. 482–487, 2007.

[6] J. Essletzbichler, "Relatedness, industrial branching and technological cohesion in US metropolitan areas," *Regional Studies*, vol. 49, no. 5, pp. 752–766, 2015.

[7] K. Frenken, F. Van Oort, and T. Verburg, "Related variety, unrelated variety and regional economic growth," *Regional Studies*, vol. 41, no. 5, pp. 685–697, 2007.

[8] Y. Han and S. J. Goetz, "Predicting US county economic resilience from industry input-output accounts," *Applied Economics*, vol. 51, no. 19, pp. 2019–2028, 2019.

[9] A. Alabdulkareem, M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, and I. Rahwan, "Unpacking the polarization of workplace skills," *Science Advances*, vol. 4, no. 7, Article ID eaao6030, 2018.

[10] C. Jara-Figueroa, B. Jun, E. L. Glaeser, and C. A. Hidalgo, "The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms," *Proceedings of the National Academy of Sciences*, vol. 115, no. 50, pp. 12646–12653, 2018.

[11] F. Neffke, M. Henning, and R. Boschma, "How do regions diversify over time? Industry relatedness and the development of new growth paths in regions," *Economic Geography*, vol. 87, no. 3, pp. 237–265, 2011.

[12] N. O'Clery, S. Heroy, F. Hulot, and M. Beguerisse-Diaz, "Unravelling the Forces Underlying Urban Industrial Agglomeration," 2019, https://arxiv.org/abs/1903.09279.

[13] T. Farinha, P.-A. Balland, A. Morrison, and R. Boschma, "What drives the geography of jobs in the US? Unpacking relatedness," *Industry & Innovation*, vol. 26, no. 9, pp. 988–1022, 2019.

[14] S. T. Shutters, J. Lobo, R. Muneepeerakul et al., "Urban occupational structures as information networks: the effect on network density of increasing number of occupations," *PLoS One*, vol. 13, no. 5, Article ID e0196915, 2018.

[15] S. Kok and B. T. Weel, "Cities, tasks, and skills," *Journal of Regional Science*, vol. 54, no. 5, pp. 856–892, 2014.

[16] S. T. Shutters and K. Waters, "Inferring networks of interdependent labor skills to illuminate urban economic structure," *Entropy*, vol. 22, no. 10, p. 1078, 2020.

[17] R. Boschma, P.-A. Balland, and D. F. Kogler, "Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010," *Industrial and Corporate Change*, vol. 24, no. 1, pp. 223–250, 2015.

[18] G. Tóth, Z. Elekes, A. Whittle, C. Lee, and D. F. Kogler, "Technology network structure conditions the economic resilience of regions," *Economic Geography*, vol. 98, no. 4, pp. 355–378, 2022.

[19] P. Catalán, C. Navarrete, and F. Figueroa, "The scientific and technological cross-space: is technological diversification driven by scientific endogenous capacity?" *Research Policy*, vol. 51, no. 8, Article ID 104016, 2022.

[20] E. Pugliese, G. Cimini, A. Patelli, A. Zaccaria, L. Pietronero, and A. Gabrielli, "Unfolding the innovation system for the development of countries: coevolution of Science, Technology and Production," *Scientific Reports*, vol. 9, no. 1, Article ID 16440, 2019.

[21] U.S. Bureau of Labor Statistics, "Occupational employment and wage statistics," 2020, http://www.bls.gov/oes/.

[22] U.S. Bureau of Labor Statistics, "2019 quarterly Census of employment and wages," 2020, https://www.bls.gov/cew/.

[23] U.S. Census Bureau, "American community survey (ACS), one-year public use Microdata sample (PUMS), 2005-2019," 2020, https://www.census.gov/programs-surveys/acs/microdata/access.html.

[24] iPUMS, "Crosswalk between 2013 MSAs and 2010 PUMAs," 2019, https://usa.ipums.org/usa/resources/volii/MSA2013_PUMA2010_crosswalk.xls.

[25] J. Siebens and C. L. Ryan, *Field of Bachelor's Degree in the United States: 2009. Ameican Community Survery Reports*, US Census Bureau, Washington, DC, USA, 2012.

[26] National Center for Education Statistics, *The Classification of Instructional Programs*, U.S. Department of Education, Washington, DC, USA, 2022.

[27] U.S. Patent and Trademark Office, "USPTO public bulk data release," 2022, https://patentsview.org/download/data-download-tables.

[28] U.S. Office of Management and Budget, *OMB bulletin No. 20-01: revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas*, White House, Washington, DC, USA, 2020.

[29] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *International journal of complex systems*, vol. 1695, pp. 1–9, 2006.

[30] R. Capello and C. Lenzi, "4.0 Technologies and the rise of new islands of innovation in European regions," *Regional Studies*, vol. 55, no. 10-11, pp. 1724–1737, 2021.

[31] I. Getz and A. G. Robinson, "Innovate or die: is that a fact?" *Creativity and Innovation Management*, vol. 12, no. 3, pp. 130–136, 2003.

[32] G. Duranton and D. Puga, "Micro-foundations of urban agglomeration economies," in *Handbook of Regional and Urban Economics*, J. V. Henderson and J.-F. Thisse, Eds., pp. 2063–2117, Elsevier, Amsterdam, Netherlands, 2004.

[33] E. L. Glaeser, H. D. Kallal, J. A. Scheinkman, and A. Shleifer, "Growth in cities," *Journal of Political Economy*, vol. 100, no. 6, pp. 1126–1152, 1992.

[34] A. Marshall, *Principles of Economics*, Macmillan, London, UK, 1920.

[35] A. Marshall and M. P. Marshall, *The Economics of Industry*, Macmillan, London, UK, 1879.

[36] K. Behrens, "Agglomeration and clusters: tools and insights from coagglomeration patterns," *Canadian Journal of Economics/Revue canadienne d'économique*, vol. 49, no. 4, pp. 1293–1339, 2016.

[37] D. Diodato, F. Neffke, and N. O'Clery, "Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time," *Journal of Urban Economics*, vol. 106, pp. 1–26, 2018.

[38] G. Ellison, E. L. Glaeser, and W. R. Kerr, "What causes industry agglomeration? Evidence from coagglomeration patterns," *The American Economic Review*, vol. 100, no. 3, pp. 1195–1213, 2010.

[39] G. Faggio, O. Silva, and W. C. Strange, "Heterogeneous agglomeration," *The Review of Economics and Statistics*, vol. 99, no. 1, pp. 80–94, 2017.

[40] E. Howard, C. Newman, and F. Tarp, "Measuring industry coagglomeration and identifying the driving forces," *Journal of Economic Geography*, vol. 16, no. 5, pp. 1055–1078, 2015.

[41] S. T. Shutters, *Using Multidimensional Networks to Better Understand Constraints and Possibilities of Urban Development*, Arizona State University, Tempe, Arizona, 2023.