Entrepreneurial Clustering on the Mechanisms of Evolutionary Change: A Means to Identify Region-Specific Opportunities for Innovative Growth.

Abstract: How does regional entrepreneurship evolve? This research focuses on the drivers of regional sectoral change through new business activity to more accurately identify industry clusters for greatest returns to targeted entrepreneurship policy making. This research applies a matrix of co-specialization probabilities of industries in 366 U.S. metropolitan areas to estimate industry “relatedness”, and develops a clustering technique that is highly correlated with metropolitan entrepreneurship rates. Entrepreneurship is a critical driver of economic development and represents an incentive target for regional policy makers. However, identifying the right industry clusters for policy intervention has always proved more challenging than not. This is because orthodox clustering techniques often face limitations in identifying cross-sector interactions that underscore innovative exchanges leading to new knowledge creation. The method developed here directly emphasizes the co-evolution of regional cross-sector interactions, and thus is centered on innovative growth and structural transformation. This clustering technique has three features of direct relevance to policy studies of entrepreneurship and innovation. 1. The clusters are metropolitan specific and reflect the economic histories of regions. 2. With its focus on industry specializations, this method provides a better identification of entrepreneurial clusters in metropolitan areas, and their labor demand. 3. The “relatedness” measure provides an adjacency matrix that reveals a network of interdependence across metropolitan economic activity. This allows us to study the network properties of regional clusters, for instance, in terms of regional resilience to exogenous shocks, including technological shocks.

1. Introduction
As cities grow, their wealth typically increases at a faster rate than their population (fig. 1). This is because cities grow by solving problems and increasing their efficiencies. In the U.S., they do so through the process of entrepreneurship, and what really differentiates one city from another is not just their size, but rather how well their residents are able to convert ideas into marketable products (Bettencourt, Lobo, Strumsky, & West, 2010). The growth and sustainability of cities is thus intrinsically tied into their entrepreneurship and the institutions and actors that enable innovation and economic growth.

Thus, understanding the sources and constraints of entrepreneurial opportunity in metropolitan economic ecosystems is of core interest to policy makers, even though identifying these system-level features is a difficult task as entrepreneurialism results from complex interactions within the structure of economic ecosystems. Nonetheless, policy makers have developed a variety of

---

1 In this paper, I use the terms city, metropolitan, metropolitan area, and metropolitan statistical area (MSA) interchangeably. In the analysis presented in the report I use the 2010 MSA definitions issued by the Office of Management Budget (OMB).
techniques that range from cluster analysis to qualitative case studies of regional economic activity, to target the appropriate economic communities for policy action and to design programs that inculcate a culture of entrepreneurialism and innovation.

This paper contributes to such policy efforts by developing an empirical approach to identify clusters of entrepreneurial industries that cut across industry sectors and reveal interesting new patterns of agglomeration. The goal is to assess a region’s mix of entrepreneurial capabilities that comprise the substrate from which new economic knowledge is created, a process leading to entrepreneurial innovation.

The novel contribution of this empirical method is the identification of the U.S. Industry Space and individual Metropolitan Entrepreneurship Spaces within it. These are abstracted graphs of the entrepreneurial ‘relatedness’ between 4-digit NAICS industries, and they reveal how clusters of industries concentrate together in metropolitan areas.

Three outcomes of this research are worth previewing; First, metropolitan areas with greater ‘Related density’ in their entrepreneurial clusters correlate with higher start-up activity, independent of size and wealth. Second, using ‘Related density’ a metropolitan ranking of entrepreneurialism can be constructed that better reflects a metro’s ability to convert ideas into market innovations. Third, each industry can be evaluated for the structural role it plays in the ecosystem. For instance, four new categories, or roles, that different industries play in an MSA’s entrepreneurial ecosystem are identified;

Drivers – these industries share the most related agglomerations in the economy;

Core capabilities – these industries represent the economy’s historic economic capabilities;
Hubs – these industries connect sub-communities of economic capabilities and are important for policy interventions targeting technology or knowledge diffusion, or broader adoption of standards;

Influencers – these industries are important for their relation to the most related industries, and are important for policies of contagion to shocks, and increasing resilience.

Furthermore, by identifying the metropolitan entrepreneurship space, the key industries that drive entrepreneurial knowledge creation are singled out and grouped for cluster analysis. An assessment of this entrepreneurial cluster’s current occupational and skill composition provides unique new insight and reference for comparing entrepreneurial ecosystems to one another. For instance, many regional policy evaluations will compare MSA’s by size, however studying labor demand data of an MSAs entrepreneurial cluster, allows us to identify other MSAs that show competing demand for the same occupation and skill groups, this provides for a much more specific cross-MSAs comparison based on economic capabilities of regions.

The paper is structured as follows; Section 2 provides the theoretical background for the ‘relatedness’ measure used here based on inter-sectoral interdependencies. Section 3 outlines the methodology and estimation of industry ‘relatedness’, and ‘Related density’ for metropolitan areas. Section 4 provides the main discussion of the paper by first developing a stronger empirical intuition of how industry concentrations are distributed across U.S. metropolitan areas in terms of the ubiquity, commonality, or rarity as an economic capability. Next, I return to the concept of relatedness to frame it in terms of network statistics by evaluating the graph theoretic properties of the relatedness of industries, and discussing how at the metropolitan level, related density, provides us with a ranking of metropolitan areas based on their entrepreneurial potential. I then “map” the U.S. industry space, which is a visual representation of the relatedness of industries,
and discuss specific properties of industries based on their location on the network. Section 5 concludes with policy implications and directions for future research.

2. Theoretical background

“When an industry has ... chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighborhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously” - Alfred Marshall, 1920

Industries are related to one another across resources (Farjoun, 1994), knowledge (Breschi, Lissoni, & Malerba, 2003), capabilities (Fai & von Tunzelmann, 2001), technologies (Enkel & Gassmann, 2010), and share interdependencies either deliberately or by some association. The deliberate interdependencies are more immediately observed and thus empirical measures of industry relatedness are most commonly observed in studies of supply chain interdependencies (Scherer, 1982), and corporate diversification strategies (Teece, 1994; Cefis & Rigamonti, 2013).

Most popular of such empirical formulations are input-output models stemming from and the work of Francois Quesnay, Leon Walrus, and Wassily Leontief.

However, even in situations where firms in different industries compete for the same resources, they may choose to collaboratively invest in developing new capabilities and technologies (Fukugawa, 2006). Or firms across different industries may collaboratively share the same resource without competition. Even in such situations of shared interdependencies by association, relatedness plays an important role in studies of regional branching (Boschma & Frenken, 2012; Essletzbichler, 2015), and models of geographic clustering of industries (Delgado, Porter, & Stern, 2010). As such, the relatedness of industries is a commonly used concept to study agglomerations, linkages, and commonalities within economies, as well as a tool to examine mechanisms for spillovers of knowledge (Gassmann, Daiber, & Enkel, 2011), productivity, and innovation across economic agents (Enkel & Gassmann, 2010).
More recently, this question of relatedness has generated considerable interest from macro studies investigating regional branching (Boschma & Frenken, 2012), and evaluations of locational clustering of industries (Delgado, Porter, & Stern, 2010), prominently in the field of economic geography. This is because relatedness measures yield well for evolutionary studies of path-dependence in corporate diversification strategies, and are thus adopted as key ‘distance’ measures in empirical studies of industry evolution (Neffke, Henning, & Boschma, 2011; Neffke & Henning, 2013; Essletzbichler, 2015). It is then no surprise that industry relatedness is a broadly adopted measure across economics, business, and management studies, albeit using different definitions of “relatedness”.

Similarly, the concept of relatedness also appears in studies assessing the various mechanisms of organizational change based on theories of the firm (Hodgson, 1998; Silverman, 1999), or evaluations of the inter-dependencies of production processes based on input-output (IO) based assessments also popularly use empirical measures of industry relatedness (Fan & Lang, 2000). The management literature also draws on concepts of industry relatedness as core to understanding corporate diversification strategy and organizational performance (Lien & Klein, 2008; Hussinger, 2010). Overall, the concept of industry relatedness that emerged from the literature on sectoral interdependencies is today popularized in studies on technological path-dependence, industry evolution, and has come to play a central role in discussions on innovation and strategic diversification (Kogan, Papanikolaou, Seru, & Stoffman, 2017), and consequently in investment strategies (Keil, Maula, Schildt, & Zahra, 2008).

In sum, the concept of industry relatedness has a home in traditional approaches in economics and the management sciences, but also in more heterodox evolutionary theories of the firm. Empirically, inter-industry relatedness has been measured and defined along various dimensions
using, buyer-supplier relationships, technological cohesion proxied by co-occurrence of firms in patent classification fields, or technology fields, as well as the production portfolios of different establishments within large firms. In all the studies, regardless of the dimension or measure of relatedness, the estimated relationships have provided explanatory power for how firms ‘relate’ to one another across different industry groups, and how these relationships affect their decisions and production processes.

Developing from this theoretical basis, the empirical measure developed in this paper takes the broadest view of ‘relatedness’ and estimates the degree of interdependence between two industries based on their co-specialization probabilities in U.S. metropolitan statistical areas. The following section discusses the measure and its estimation in more detail.

3. Method

Two measures are estimated in this paper and provide the basis for the analysis pursued here. The first is a relatedness measure that is estimated at the industry level, while the second is an index of ‘related density’ at the metropolitan level, which allows us to rank metros according to the relatedness of their industrial mix. It is estimated as the co-specialization probabilities of industries in metropolitan areas in the U.S. Specifically, two industries are said to be ‘related’ if they both have a high probability of being specialized in the same metropolitan area. In effect, relatedness (ζ) between industry specializations was estimated for every pair-wise combination for 305 4-digit NAICS industries at the national level using the ratio of their conditional probability to their independent probabilities. This estimation strategy of inter-industry relatedness is similar

---

2 Industry specializations refer to higher concentrations of business activity for a given industry sector in the MSA than the national average. Since this study is focused on metropolitan entrepreneurial outcomes, an establishment location quotient (LQ) > 1 was used to define industry specialization at the 4-Digit NAICS level of aggregation, where an LQ = 1 suggest the same concentration of a particular industry activity as the national.
to estimation of the interdependencies measure between occupations followed in Muneepeerakul, et. al. (2013). The intuition follows that if the specialization of some industry i is partly determined by the specialization of some other industry j, their conditional probability would be different from the product of their independent probabilities. Accordingly, relatedness (ζ) was estimated as:

\[
\zeta_{ij} = \frac{P[LQ_i^M > 1, LQ_j^M > 1]}{P[LQ_i^{M'} > 1]P[LQ_j^{M'} > 1]} - 1
\]

Where M, M', and M'' are random metropolitan areas and LQ is the establishment location quotient for industries i and j. Subtracting by 1 allows for better interpretability such that when, ζ = 0 it suggests industries i and j are independent, and a positive interdependence suggests the two industries are more likely to co-specialize in metros than if they were independently distributed across metropolitan areas. Negative values of ζ with the minimum being -1 by construction conversely indicate pairs of industries that are very unlikely to co-specialize in the same metropolitan area. Intuitively, positive values are considered indicative of economic complementarities while negative values suggest a disincentive to co-specialization. Data for this estimation comes from the Quarterly Census of Employment and Wages (QCEW) for the years 2013 to 2015. Figure 2 below provides some descriptive statistics for the relatedness measure.

**Figure 2. Descriptive Statistics of ‘relatedness’ in the U.S**

Source: Author’s estimates.
Examples of the pair-wise relatedness measures are provided in Table 1 below.

<table>
<thead>
<tr>
<th>High Positive Relatedness ($\zeta$)</th>
<th>Industry i</th>
<th>Industry j</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion picture and video industries</td>
<td>Independent artists, writers, and performers</td>
<td></td>
<td>6.4</td>
</tr>
<tr>
<td>Oil and gas extraction</td>
<td>Support activities for mining</td>
<td></td>
<td>4.1</td>
</tr>
<tr>
<td>Computer systems design &amp; related services</td>
<td>Electronic markets and agents and brokers</td>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td>Fishing</td>
<td>Seafood product preparation and packaging</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>Apparel knitting mills</td>
<td>Apparel &amp; piece goods merchant wholesalers</td>
<td></td>
<td>2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High Negative Relatedness ($\zeta$)</th>
<th>Industry i</th>
<th>Industry j</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apparel &amp; piece goods merchant wholesalers</td>
<td>Coal mining</td>
<td></td>
<td>-1.0</td>
</tr>
<tr>
<td>Independent artists, writers, and performers</td>
<td>Oilseed and grain farming</td>
<td></td>
<td>-1.0</td>
</tr>
<tr>
<td>Cut and sew apparel manufacturing</td>
<td>Electronic markets and agents and brokers</td>
<td></td>
<td>-1.0</td>
</tr>
<tr>
<td>Management &amp; technical consulting services</td>
<td>Farm product raw material merch. whls.</td>
<td></td>
<td>-0.9</td>
</tr>
<tr>
<td>Individual and family services</td>
<td>Business support services</td>
<td></td>
<td>-0.8</td>
</tr>
</tbody>
</table>

Source: Author’s estimates from the Quarterly Census of Employment and Wages (QCEW) for year 2014.

As noted earlier, this relatedness measure is estimated at the level of industries, and provides a ‘distance’ measure of the pair-wise relatedness of industries. A high threshold of 0.75 is applied to distinguish between related or not related industries. This high threshold provides an interpretation that given industry i is specialized in a metropolitan area, then industry j is at 75 percent likely to also specialize in that same metropolitan area.

After applying this threshold it becomes easier to create a metropolitan-level index of ‘related density’ by aggregating the count of related industry pairs within a metropolitan area and dividing by the total number of possible relations if all industry specialization in the metro were related to one another. This ‘related density’ can be thought of as measuring the extent of correlation of economic activity within an MSA’s industry specializations. Empirically, for some metropolitan area M, it is estimated as the proportion of realized vs possible count of interdependencies in each of its Set of Specializations (SS).
Related Density$^M = \frac{\text{(Realized count of related pairs)}^M_{SS}}{\text{(Possible count of all related pairs)}^M_{SS}}$

Related density is then a metropolitan-level measure of combinatorial advantage derived from the interdependencies present between each metropolitan’s economic specializations. In the following sections I provide a detailed discussion on the distribution of industrial activity in metropolitan areas and develop a much broader understanding of inter-industry relatedness in the data.

4. Discussion

Where do industries concentrate in their production activities in U.S. metropolitan statistical areas (MSAs)? Is an MSA’s mix of industry concentrations associated with its entrepreneurial potential? These two questions develop the core of this research which is to empirically identify the sources and constraints of entrepreneurial opportunity embedded in metropolitan’s economic ecosystem. The answer to both these questions is evident in the structure of the ‘U.S. Industry Space’ - a revealed network representation of the ‘relatedness’ of economic activity between industries in U.S. metropolitan areas.

I discuss the industry space in more detail below, but in short, graphing the U.S. Industry Space reveals that most MSAs have diverse and broad sets of industry concentrations, and the density of how related these industries are to each other is strongly associated with the MSA’s entrepreneurship rate. In the debate on whether metropolitan areas should specialize or diversify their economic capabilities, the analysis presented here suggests policy makers should aim to develop a ‘related-diversity’ in their region’s economic capabilities. Such organization of an MSA’s economic capabilities allows for greater efficiency in its ecosystems, facilitating innovation and entrepreneurship. Policy implications are to look beyond traditional measures of entrepreneurial output and consider the organization of an ecosystem’s industry mix as a primary source of entrepreneurial opportunity.
The following section begins with a discussion on how industrial activity is distributed across metropolitan areas using a simple categorization of how rarely or commonly industries concentrate in metropolitan areas. This provides the basis for why certain industries may cluster in proximity to one another, and helps develop the intuition for understanding the ‘relatedness’ between industries. Some empirical evidence is provided for the association between ‘relatedness’ in a metropolitan’s economic structure and its entrepreneurship rate. I then ‘map’ the U.S. Industry Space and assess some graphical properties of the network and discuss its implications for the location of industries on the network space.

**Rare, Common, and Ubiquitous Industry Concentrations in U.S. Metropolitan Areas**

In the U.S. some industries are ubiquitously present in most all metropolitan areas, while others concentrate only in a handful of MSAs. Examples of such ubiquitous industries include grocery stores, department stores, automobile dealers, restaurants, building equipment contractors, and gasoline stations, among others. Their ubiquity can be explained by there being a local demand for their products and services in all metropolitan areas. On the other hand, rare industries such as, motion picture industries, mining, seafood packaging, sheep and goat farming, and apparel and piece goods merchant wholesalers concentrate only in a few MSAs because they benefit from some locational advantages like access to natural resources, labor pools of skilled workers, or even proximity to other industries and markets. This idea of the ubiquity or rarity of industries across the country provides us with some insight on how productive activity in the U.S. is dispersed across MSAs.

For instance, as represented in Figure 1., categorizing 4-Digit NAICS industries by how rarely (less than 20 percent is US MSAs); commonly (between 20-80 percent of MSAs); or ubiquitously
(more than 80 percent of MSAs) they concentrate in U.S. MSAs, and grouping them by their Super Sector3 of activity, reveals some interesting information about the U.S. Industry Space;

**Figure 1: Ubiquitous, Common, and Rare Industry Specializations by NAICS Super Sectors**

The Information sector has the highest concentration of rare 4-Digit NAICS codes. Out of 358 MSAs in the data, only 28 MSAs have business concentrations in the technical industries of ‘Sound recording industries’, and 59 MSAs in ‘Motion picture and video industries’. Similarly, less than 70 MSAs have any concentrations in the high-technology industries of ‘Data processing, hosting, and related services’ and ‘Software publishers’. ‘Radio and television broadcasting’ is a ubiquitous industry, while most telecommunications industries concentrate in at least 120 MSAs.

In the Natural Resources and Mining sector, most farming and agricultural activities are common industries that concentrate in well over a 100 MSAs. The rarest industry in this sector is ‘Coal

---

3 NAICS super sectors have different number of 4-Digit NAICS industry codes under each hierarchy. For example, the Manufacturing sector has 86 different 4-Digit NAICS codes while the Information sector and Construction each have 12 and 10 4-Digit NAICS sub-codes respectively.
‘Mining’ concentrating in only 37 MSAs, followed by ‘Metal ore mining’ in 41 MSAs, ‘Fishing’ in 58 MSAs, and ‘Oil and gas extraction’ in 62 MSAs.

In the Leisure and Hospitality Sector, two rare industry concentrations stand out, ‘Independent artists, writers, and performers’ and ‘Agents and managers for public figures’, present in only a handful 30 MSAs.

The Trade, Transportation, and Utilities sector has the rarest of industry concentrations with ‘Apparel and piece goods merchant wholesalers’ found in only 24 MSAs. Other rare concentrations include, ‘Electronic markets and agents and brokers’ found in 38 MSAs, and the wholesaler industries of, ‘Furniture and furnishings merchants’, and ‘Druggists’ good merchants’ found in 66 and 61 MSAs respectively.

The Professional and Business Services sector has the second highest concentration of common 4-Digit NAICS codes. The two high-technology sectors of ‘Management and technical consulting services’ and ‘Computer and systems design and related services’ are rare concentrations in 67 and 54 MSAs respectively.

The Manufacturing sector has the highest concentration of common 4-Digit NAICS codes. All high-technology manufacturing concentrates commonly in well over 100 MSAs, with the exceptions of ‘Computer and peripheral equipment manufacturing’, and ‘Magnetic media manufacturing and reproducing’, which concentrate in 91 and 86 MSAs respectively. The rarest manufacturing concentrations are, ‘Cut and sew apparel manufacturing’, ‘Apparel knitting mills’, which concentrate in 35 and 42 MSAs respectively.
The Education and Health Services and Construction super sectors are almost entirely commonly or ubiquitously found industries. The one exception being the ‘Individual and family services’ which concentrates in 68 MSAs.

High-technology 4-Digit NAICS industries\textsuperscript{4} are common for the most part. Out of 46 high-tech industry codes, 31 industries are commonly found (mostly manufacturing), 8 are rarely found (distributed across the mining; trade; information; and, business services sectors), and 7 are ubiquitously found (entirely in healthcare related activities).

The implication of these observations is that industries may concentrate in some metropolitan areas over others for different reasons, and the rarest industry concentrations are not only those that depend on natural resources, but also those that rely heavily on the local presence of other related economic activities. Furthermore, skilled capabilities such as those in high-tech industries are commonly or ubiquitously found in MSAs, so their presence alone is not likely elucidating the uniqueness of a metropolitan ecosystem’s economic capabilities. This is important because policies broadly targeting high-tech industries for entrepreneurship development may not be capturing the true sources of competitive advantage for in those metropolitan areas.

This brings us to the next question, how then does an MSA’s mix of industry concentrations contribute toward its wealth and entrepreneurial capacity? What really makes an Entrepreneurial Ecosystem’s industry mix unique?

\textbf{Diversity of Industries in U.S. Metropolitan Areas}

How many industry concentrations do MSAs typically have? Using location quotients (LQ) to reveal metropolitan industry concentrations shows that most MSAs already have a high number of

\textsuperscript{4} High-technology industries were defined similarly as (Hecker, 2005) from the Bureau of Labor Statistics.
diverse industry concentrations. The average MSA in the U.S. boasts an establishment LQ greater than 1 in 119 out of 304 private sector 4-digit NAICS industry codes. The ‘Sacramento-Roseville-Arden-Arcade, CA’ MSA has the lowest number of concentrations at 14 industries, while the ‘Evansville, IN-KY’ MSA has the highest with 163 industry concentrations. Of the 358 MSAs considered here, half of them have over 124 industry concentrations. Figure 2 below shows a histogram of this left skewed distribution.

**Figure 2: Distribution of Industry Specializations Across MSAs in the US, 2015.**

![Histogram of Industry Specializations Across MSAs in the US, 2015.](image)

Source: Author’s estimates based on QCEW 2015 data.

Considering the data a little more carefully, there is no statistically significant relationship between the number of industry concentrations a MSA has and its wealth, even after controlling for MSA size. Furthermore, there is no statistically identifiable relationship between the number of industry concentrations in an MSA and the MSA’s establishment entry rate, nor its exit rates. In fact, once controlling for the size of the MSA, the number of industry concentrations does become statistically significant at the 99 percent level, but is negatively related to entrepreneurship in the

---

5 Location quotients (LQ) are commonly used in regional economics to show the concentration of industrial activity in a region relative to that of the nation. An LQ<1 indicated a lower concentration of that industrial activity relative to the national, while an LQ>1 indicated a concentration greater than the nation, suggestive of some regional competitive advantage in that industry.
MSAs, although the R-squared is only 0.25. This suggests that just considering how many industry concentrations an MSA has does not provide much information relevant to understanding sources of entrepreneurial opportunity in the metropolitan’s ecosystem. Having too many industry concentrations may in fact even hamper metropolitan entrepreneurial activity.

What then is driving the relationship between an MSA’s industry mix and its entrepreneurial activity? In the next sections, I introduce the idea of industry ‘relatedness’ to assess the organization of industry concentrations within an MSA’s economic structure, and provide some empirical arguments for its relevance to entrepreneurial capacity in metropolitan economic ecosystems.

**The Relatedness of Industries**

To evaluate how production activity is organized within an MSA, I introduce a measure of industry ‘relatedness.’ It is estimated as the pair-wise probability of two industries concentrating together in the same metropolitan area and it can be thought of as capturing the overlap of the regional competitive advantages favoring local industries. This overlap in competitive advantages may exist because the two industries share similar access to natural resources, labor pools of skilled workers, proximity to each other, or to common markets. As the measure is derived from a probability, I set a high cut off to identify when this co-specialization of industries is likely to occur. Two industries are said to share a high degree of ‘relatedness’ if given on industry

---

6 Industry relatedness here is similar in its conception as occupational ‘interdependence’ in Munnerpeerakul, et. al. (2013), and is a notion developed from the seminal work on the ‘Product Space’ of nations formalized by Hidalgo, et. al (2007).

7 More details on this estimation are provided in the Methods Appendix C.
concentration in the MSA, the second industry is at least 75 percent more likely to also concentrate in the same metropolitan area.

An empirical evaluation of this pair-wise relatedness between industries in the U.S. reveals that only 44 percent, or 135 4-digit NAICS industries out of 304 industry codes, share a high degree of Relatedness with at least one other industry. Figure 3 below shows this distribution of relatedness across industries.

Figure 3: Frequency Distribution of High Industry Relatedness in the US for 2015

Even of the 44 percent that are highly related to more than one other industry, less than half are related to at least 5 other industries. Majority of industries have few related industry partners while only a handful of industries have many related partners. This suggests that some industries play a more connected role in economic ecosystems, while others may behave more independently.

Within a network framework, this also suggests strong clustering of industries in the U.S. Industry Space – suggestive of underlying patterns of organization.

Most importantly, the density of related industries that a metropolitan area has in its profile of industry concentrations is statistically and significantly associated with its entrepreneurship rate, and wealth. Thus, ‘Related Density’ - the proportion of an MSA’s industry concentrations that are
related to one another, reveals much about entrepreneurial activity in U.S. metropolitan areas, and can provide empirical insight relevant to policy making for Entrepreneurial Ecosystems.

As a start, a 1 percentage point increase in the relatedness density of a metropolitan area is associated with a 1.02 percentage point increase in the metropolitan’s rate of establishing new businesses, after controlling for the size and wealth of the metropolitan areas. Furthermore, worker reallocation rates are also statistically correlated with Related Density, such that, a 1 percent increase in Related Density is associated with a 0.8 percent increase in the cluster’s worker reallocation rate after controlling for employment. This reflects that workers find more options for job switching in industry clusters that share greater industry relatedness. Figure 4 below presents a graph of 358 U.S. metropolitan areas plotted along their startup density (new establishment entry rates) and their Related Density, indexed to logarithmic scale.

It is worth noting that this is superlinear relationship. That means, as the related density of MSAs increases the entrepreneurship rate grows at an even greater rate. A more nuanced understanding of this relationship will benefit greatly from further empirical investigation, but for now, it can suffice to say that economies with greater related densities exhibit greater efficiencies in their entrepreneurship capabilities due to greater harmony between their related industries.
**Figure 4: Metropolitan entrepreneurship and ‘Related Density’**

![Graph showing the relationship between metropolitan entrepreneurship spaces and metropolitan rankings.](image)

**Metropolitan entrepreneurship spaces and metropolitan rankings**

The Metropolitan Entrepreneurship Space refers to a metropolitan area’s cluster of entrepreneurial industries that are also related in their agglomeration dynamics (referred to in this report as “relatedness”). These clusters play an essential role in the co-evolutionary dynamics of a metropolitan’s industry structure, and thus provide vital insight into a metropolitan area’s growth.
trajectory, relevant for identifying opportunities for targeted policy making for entrepreneurship and economic development.

As discussed earlier, the ‘related density’ of metropolitan areas correlates strongly and positively with its startup density (fig. 4). This relationship provides the empirical basis for studying the networked structure of relatedness to identify sources of entrepreneurial opportunity in U.S. metropolitan areas. In fact, the data identify that an increase in the density of related industries by 1 percent for an ecosystem is associated with an increase of more than 1 percent in its startup density.

Whereas more empirical research on this topic is needed to develop a nuanced understanding of the co-evolution of industries in cities, I provide some interpretation of related density here in the following sections. First, in section 3.1, I present a ranking of metropolitan areas in the U.S. according to their related density in 2015. Next, in section 3.2, I also present two short anecdotes to hypothesize how MSAs grow, or decline, in their related density over time.

**Metropolitan rankings of related density**

Table 4 below provides a ranking of the top 10 US Metropolitan Statistical Areas (MSAs) according to the Related Density of their entrepreneurial clusters. Here, an entrepreneurial cluster refers to groups of industries the MSA that demonstrate significant competitive advantage and is identified as the group of industries with concentrations of business activity at least 1.5 times that of the nation.

Compared across 357 MSAs in the US for 2015, metropolitan areas with greater Related Density tend to exhibit higher entrepreneurship rates. Related Density accordingly is a measure of how synchronous entrepreneurial activity is within an MSAs industry structure.
Table 4: Metropolitan Rankings – Top 25 US MSAs with highest Relatedness Density in 2015.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Related Density (%)</th>
<th>Startup Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Los Angeles-Long Beach-Anaheim, CA MSA</td>
<td>15.1</td>
<td>12.1</td>
</tr>
<tr>
<td>2</td>
<td>San Francisco-Oakland-Hayward, CA MSA</td>
<td>11.8</td>
<td>11.4</td>
</tr>
<tr>
<td>3</td>
<td>San Jose-Sunnyvale-Santa Clara, CA MSA</td>
<td>11.7</td>
<td>11.4</td>
</tr>
<tr>
<td>4</td>
<td>Boulder, CO MSA</td>
<td>10.5</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Denver-Aurora-Lakewood, CO MSA</td>
<td>8.7</td>
<td>12.4</td>
</tr>
<tr>
<td>6</td>
<td>El Centro, CA MSA</td>
<td>8.5</td>
<td>11.2</td>
</tr>
<tr>
<td>7</td>
<td>Bridgeport-Stamford-Norwalk, CT MSA</td>
<td>8.0</td>
<td>9.2</td>
</tr>
<tr>
<td>8</td>
<td>Austin-Round Rock, TX MSA</td>
<td>7.5</td>
<td>13.1</td>
</tr>
<tr>
<td>9</td>
<td>Washington-Arlington-Alexandria, DC-VA-MD-WV MSA</td>
<td>7.0</td>
<td>10.7</td>
</tr>
<tr>
<td>10</td>
<td>Las Vegas-Henderson-Paradise, NV MSA</td>
<td>7.0</td>
<td>14.4</td>
</tr>
<tr>
<td>11</td>
<td>Atlanta-Sandy Springs-Roswell, GA MSA</td>
<td>5.8</td>
<td>11.9</td>
</tr>
<tr>
<td>12</td>
<td>San Diego-Carlsbad, CA MSA</td>
<td>5.8</td>
<td>12.4</td>
</tr>
<tr>
<td>13</td>
<td>New York-Newark-Jersey City, NY-NJ-PA MSA</td>
<td>5.6</td>
<td>11.4</td>
</tr>
<tr>
<td>14</td>
<td>Provo-Orem, UT MSA</td>
<td>5.4</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>Oxnard-Thousand Oaks-Ventura, CA MSA</td>
<td>5.2</td>
<td>10.7</td>
</tr>
<tr>
<td>16</td>
<td>Seattle-Tacoma-Bellevue, WA MSA</td>
<td>5.0</td>
<td>11.1</td>
</tr>
<tr>
<td>17</td>
<td>Fresno, CA MSA</td>
<td>4.9</td>
<td>9.9</td>
</tr>
<tr>
<td>18</td>
<td>Phoenix-Mesa-Scottsdale, AZ MSA</td>
<td>4.7</td>
<td>11.9</td>
</tr>
<tr>
<td>19</td>
<td>Durham-Chapel Hill, NC MSA</td>
<td>4.6</td>
<td>10.1</td>
</tr>
<tr>
<td>20</td>
<td>Salt Lake City, UT MSA</td>
<td>4.5</td>
<td>11.2</td>
</tr>
<tr>
<td>21</td>
<td>Miami-Fort Lauderdale-West Palm Beach, FL MSA</td>
<td>4.5</td>
<td>14.2</td>
</tr>
<tr>
<td>22</td>
<td>Colorado Springs, CO MSA</td>
<td>4.4</td>
<td>12.3</td>
</tr>
<tr>
<td>23</td>
<td>Ann Arbor, MI MSA</td>
<td>4.2</td>
<td>9.5</td>
</tr>
<tr>
<td>24</td>
<td>Boston-Cambridge-Newton, MA-NH MSA</td>
<td>4.2</td>
<td>9.1</td>
</tr>
<tr>
<td>25</td>
<td>Portland-Vancouver-Hillsboro, OR-WA MSA</td>
<td>3.9</td>
<td>10.9</td>
</tr>
</tbody>
</table>

*A complete list of the MSA rankings is provided in the appendix.
Source: Author’s estimates based on QCEW data for 2015.

Mapping the U.S. Industry Space

The ‘relatedness’ measure discussed above is estimated pair-wise for all 305 4-Digit NAICS industries and thus can be used to produce an adjacency matrix of industry relatedness. Using a high cutoff of 75 percent conditional probability of the two industries concentrating in the same MSA, I produce a network representation of production activity in the U.S. Figure 5 below is this graph representation of the relatedness of industries in the nation.
In the network map above, 304 private sector 4-Digit NAICS industry codes are represented by individual bubbles, or nodes. The nodes are colored according to their NAICS super sectors, while their size corresponds to the number of MSAs in which they concentrate. Smaller nodes indicate rarer industries, while larger nodes represent more ubiquitous industry concentrations. A few features are noticeable immediately;

Most of the industries do not have any connections to any other industries at this high cutoff of relatedness. Only 44 percent of industry codes share a high relatedness to one another. Yet, because some industries have many connections we see clear clustering in the network space.

The smaller nodes tend to concentrate in the center of the network and are more likely to have connections to other industries. This reflects the importance of rare industry specializations in a
metropolitan’s industry mix. They contribute to the uniqueness of an MSAs industry mix, not simply because of their rarity, but also for their connectedness to many other industries.

There is significant mixing of colors in the network connections suggesting economic relatedness extends between industries in different NAICS hierarchies. This suggests that NAICS sectors are not identifying many of the various dimensions of economic relatedness industries share with one another.

Strong clusters of related activity are very evident, forming a “core” to the network. The densest cluster contains overlapping and inter-related sub-clusters of industries in the ‘Professional and Business Services’ sector, the ‘Trade and Transportation’ sector, and ‘Arts and Entertainment’ activity in the ‘Leisure and Hospitality’ sector. It is discussed in more detail later, but high-tech industries are overrepresented in the core.

Although connected blue nodes of manufacturing are visible distributed across different sections of the network map, manufacturing is underrepresented in the core of the U.S. Industry Space, whereas service industries dominate the core.

Smaller clusters of related activity in the ‘Natural resources and mining’ and ‘Trade and transportation’ sectors are relegated to the periphery of the network. Particularly, clusters of water transportation related activities, oil and gas extraction, and farming form connected components that surround the large core component in the network space.

**Topological Properties of the U.S. Industry Space**
The network properties of the U.S. industry space are stable over the short term. Table 1 below reports some basic graph statistics of the network for the most recent 3 years of data available. Noticeably, the changes in the statistics are relatively small, but still warrant a discussion on their interpretation as interesting characteristics of the industry space are revealed. Furthermore, as the U.S. industry space provides a baseline for comparatively assessing the properties of the metropolitan entrepreneurship spaces later in this report makes it relevant to understand some basic features of the structure of the network.

Table 1: Network Statistics of the U.S. Industry Space

<table>
<thead>
<tr>
<th>Network Statistic</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of connected nodes</td>
<td>162</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td>Count of edges &gt;0.75</td>
<td>729</td>
<td>642</td>
<td>615</td>
</tr>
<tr>
<td>Average degree</td>
<td>9.000</td>
<td>9.372</td>
<td>8.978</td>
</tr>
<tr>
<td>Avg. Weighted Degree</td>
<td>10.84</td>
<td>11.076</td>
<td>10.745</td>
</tr>
<tr>
<td>Diameter</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Avg. Path Length</td>
<td>2.964</td>
<td>2.964</td>
<td>3.146</td>
</tr>
<tr>
<td>Modularity</td>
<td>0.371</td>
<td>0.394</td>
<td>0.392</td>
</tr>
<tr>
<td>Avg. Clustering Coefficient</td>
<td>0.667</td>
<td>0.624</td>
<td>0.677</td>
</tr>
</tbody>
</table>

Source: Author’s estimates of 2013 to 2015 QCEW annual averages.

Count of related nodes and edges: The ‘relatedness’ threshold selected in this report is 0.75 for graphing the networks. At this level, the number of nodes in the network decreased from 162 industry codes in 2013 to 137 in following years. Similarly, the number of related connections also decreased from 729 in 2013 to 642 and 615 in the following 2 years. A longer-term study of these statistics is warranted to make any empirical conclusions about the evolution of the industry space. For now, it suffices to show that despite industry codes joining and dropping from the network, the main statistics of network structure are relatively stable.

8 The analysis presented here used annual QCEW data for years 2013 to 2015. Long term dynamics of the network are a focus of future research as creating a long enough panel faces different data challenges.
**Average degree**: This statistic provides a measure of graph density and calculates the average number of neighbors that nodes have in this network. The intuition here is that industry spaces with higher average degree have denser connections. Between 2013 and 2015 this statistic for the U.S. industry space has stayed about the same, albeit it increased in 2014 before decreasing again. On the other hand, the average weighted degree which takes into account the strength of ‘relatedness’ between industry codes, shows a similar trend.

**Network diameter**: This statistic measures the length of the shortest path between the two most distant nodes in the network. For 2013 and 2015 this statistic tells us that each industry code is related to any other industry code within 8 connections, a distance that dropped to 7 in 2014. When considering the network spaces for individual metropolitan areas in 2015, 8 is thus the upper limit of this statistic. Metropolitan entrepreneurship spaces with diameters closer to 7 indicate a broader set of related economic capabilities, while those with smaller network diameters will indicate more clustered and specialized economic capabilities.

**Average path length**: This is a measure of efficiency and tells us the average number of shortest paths between any two nodes in the network. The shorter the average path length would indicate an industry space that is more closely related in its mix of industrial activities, while a longer path length suggests a more diverse but related set. Between 2013 and 2015 the average path length increases from 2.964 connections to 3.146, a small increase suggestive of increased related diversity, but more importantly it provides an important reference for discussing the metropolitan spaces later.

**Modularity**: A measure of clustering within a network, higher modularity indicates stronger communities within the industry space. This statistic does not change much in the three years observed and remains relatively low at 0.392 in 2015. This is because cluster or communities of
related industries in the national industry space are overlap one another with many interconnections.

**Average clustering coefficient:** The average clustering coefficient provides information on how connected each node’s immediate neighborhood is on average and the statistic ranges from 0 to 1. At 0.677 in 2015 the average clustering coefficient for the U.S. industry space is significantly higher than if this network were randomly formed.

**Graph properties of industries in the U.S. industry space**

The structure of the industry space derives from the networked position of individual industries in relation to one another. Some industries are much more connected within the network space while others are relegated towards the periphery of connected communities. As connections in this network space are defined by the ‘relatedness’ of economic activity, how many related connections an industry has and with whom, reveals valuable information for entrepreneurship policy making.

Four network measures of the industry nodes are evaluated here, each providing unique information on the importance of individual industries in the U.S. industry space. Ranking of industries according to these measures is presented in Tables 2 and 3, and some interpretations for these properties is also below.

**Most related** – These industries share the most relatedness with other industries in the entrepreneurial ecosystem. The ‘degree’ measure identifies the industries with most ‘related’ neighbors, revealing their importance as agents of change in the ecosystem. In the U.S. industry space, this measure identifies the industries that share the greatest economic commonalities with other industries, and from a policy perspective interventions targeted at these industries are likely to have the greatest potential for immediate spillover effects to other related industries.
High-tech industries, especially in information technology sector, comprise a third of the top 15 economic driver industries, as shown in table 2 below. However, the top spots on the list belong to the entertainment sector reflecting the diversity of related industries across most NAICS super sectors, specifically, “Independent artists, writers, & performers’ and ‘Motion picture & video industries’, are each related to 48 and 44 industries respectively.

**Core capabilities** – These industries boast the entrepreneurial ecosystem’s unique mix of core capabilities. Using the ‘closeness centrality’ statistic here helps identify the relative position of an industry to all other industries in the network. In the U.S. industry space, this measure tells us about the proximity of an industry’s economic activity from all other activities represented in the network. From a policy perspective, an ecosystem’s core capabilities are its most established and entrenched industries where intervention is likely to face more resistance to change.

The U.S. industry space is comprised of two connected components, or two distinct networks, at the high threshold of 0.75 selected in this report. The larger component is made up of 128 ‘related’ industry codes, while the smaller component is made up of just 9 industry codes, all belonging to automotive related manufacturing. These nine manufacturing industries most often locate in geographical proximity to one another and thus identify as a commonly occurring cluster of the top 9 core capabilities in the U.S. industry space (table 2). If these are excluded the list looks similar to the list of the most related industries.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Economic Drivers</th>
<th>Core Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Independent artists, writers, &amp; performers</td>
<td>Metalworking machinery manufacturing</td>
</tr>
<tr>
<td>2</td>
<td>Motion picture &amp; video industries</td>
<td>Machine shops &amp; threaded product mfg.</td>
</tr>
<tr>
<td>3</td>
<td>Apparel &amp; piece goods merchant wholesalers</td>
<td>Coating, engraving, &amp; heat-treating metals</td>
</tr>
<tr>
<td>4</td>
<td>Computer systems design &amp; related services</td>
<td>Motor vehicle parts manufacturing</td>
</tr>
</tbody>
</table>
Management & technical consulting services | Forging & stamping
---|---
Data processing, hosting & related services | Industrial machinery manufacturing
Specialized design services | Other general-purpose machinery manufacturing
Advertising, PR, & related services | Plastics product manufacturing
Electronic markets & agents & brokers | Steel product mfg. from purchased steel
Commercial equip. merchant wholesalers | Independent artists, writers, & performers
Securities & commodity exchanges | Apparel & piece goods merchant wholesalers
Agents & managers for public figures | Motion picture & video industries
Software publishers | Computer systems design & related services
Furniture & furnishing merchant wholesalers | Management & technical consulting services
Other financial investment activities | Advertising, PR, & related services

Note: High-technology industries are in bold
Source: Author’s estimates of QCEW 2015 data.

**Hubs** – These industries connect different sub-clusters within the entrepreneurial ecosystem. Using the ‘betweenness centrality’ measure identifies the industries that are most often found on the shortest ‘related’ paths that connect any two industries in the network, revealing the production activities that connect otherwise diverse groups of economic specializations. These industries are relevant for policy interventions seeking a diffuse effect across different sub-clusters of industries within the ecosystem.

Hub industries are not often high-tech industries and the top 15 industries in this category are representative across a diverse set of super sectors (table 3). Industries such as ‘Independent artists, writers, and performers’ and ‘Apparel and piece goods merchant wholesalers’ have high betweenness centrality because they share many related connections with diverse groups of industries. Yet, industries such as ‘Fruit and tree nut farming’, ‘Cut and sew apparel manufacturing’, and ‘Pipeline transportation of crude oil’ have relatively few related neighbors, only 10, 8, and 7 respectively. Yet, they rank high because they are related to industries that connect them to much larger networks within the industry space.
**Influencers** – These industries have the most value in the network. The ‘eigenvector centrality’ measure used here provides a ranking of an industry’s importance based on the ‘relatedness’ of its related neighbors. Traditional measures such as Location Quotients (LQ) that are used pervasively in regional science assess the relative concentration as a proxy for the importance of an industry in a regional economy. On the other hand, eigenvector centrality as applied here reveals an industry’s importance as function of who it is related to.

As the topology of the U.S. industry space is characterized by a few industries having many related partners and most industries having only a few, by construction in the national space the list of high ranking dominant industries approximately corresponds to the rankings of economic driver industries. However, at the regional level this measure helps differentiate between industries that dominate smaller sub-clusters versus those in larger related communities.

The following Table 3 below provides a list of the top 15 “Hub” and “Driver” industries for the U.S. industry space according to their betweenness centrality and their eigenvector centrality respectively.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Economic Hubs</th>
<th>Dominant Industries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Independent artists, writers, &amp; performers</td>
<td>Computer systems design &amp; related services</td>
</tr>
<tr>
<td>2</td>
<td>Apparel &amp; piece goods merchant wholesalers</td>
<td>Motion picture &amp; video industries</td>
</tr>
<tr>
<td>3</td>
<td>Fruit &amp; tree nut farming</td>
<td>Independent artists, writers, &amp; performers</td>
</tr>
<tr>
<td>4</td>
<td>Unclassified</td>
<td>Data processing, hosting &amp; related services</td>
</tr>
<tr>
<td>5</td>
<td>Cut &amp; sew apparel manufacturing</td>
<td>Management &amp; technical consulting services</td>
</tr>
<tr>
<td>6</td>
<td>Pipeline transportation of crude oil</td>
<td>Advertising, PR, &amp; related services</td>
</tr>
<tr>
<td>7</td>
<td>Support activities for mining</td>
<td>Specialized design services</td>
</tr>
<tr>
<td>8</td>
<td>Other information services</td>
<td>Apparel &amp; piece goods merchant wholesalers</td>
</tr>
<tr>
<td>9</td>
<td>Motion picture &amp; video industries</td>
<td>Commercial equip. merchant wholesalers</td>
</tr>
<tr>
<td>10</td>
<td>Sea, coastal, &amp; great lakes transportation</td>
<td>Software publishers</td>
</tr>
<tr>
<td>11</td>
<td>Religious organizations</td>
<td>Electronic markets &amp; agents &amp; brokers</td>
</tr>
<tr>
<td>12</td>
<td>Office administrative services</td>
<td>Securities &amp; commodity exchanges</td>
</tr>
<tr>
<td>13</td>
<td>Forest nursery &amp; gathering forest products</td>
<td>Agents &amp; managers for public figures</td>
</tr>
<tr>
<td>14</td>
<td>Electronic markets &amp; agents &amp; brokers</td>
<td>Furniture &amp; furnishing merchant wholesalers</td>
</tr>
<tr>
<td>15</td>
<td>Sheep &amp; goat farming</td>
<td>Other financial investment activities</td>
</tr>
</tbody>
</table>

Note: High-technology industries are in bold

Source: Author’s estimates of QCEW 2015 data.

These measures can be applied to metropolitan entrepreneurship spaces as well and help extricate the unique entrepreneurial capabilities.

5. Conclusion

The measures discussed in the prior section and developed from network statistics provide a new perspective on the dynamism and interdependence of economic activity within entrepreneurial economies and allows for differentiated policy making for entrepreneurship that targets different industry clusters with specific programs, while maintaining a holistic perspective on the entire economy. These characteristics help policy makers differentiate between industries that have more relevance for policy targeting technology adoption or diffusion, or industries that are important to the economy’s resilience.

Future research extends this analysis for metropolitan areas to develop a comparative approach to evaluating entrepreneurial opportunity embedded in the industrial mix of metropolitan areas. The network measures provided above are indicators of the different network characteristics of industries and evaluating their importance to the resilience and growth of entrepreneurial ecosystems would prove interesting.
Bibliography


