Social capital, agglomeration, and entrepreneurship

Jae Beum Cho
Indiana Business Research Center
Kelley School of Business
Indiana University
Bloomington IN, USA
jaebcho@iu.edu

Abstract

In this study, I argue that social capital and agglomeration are inherently interlinked. Utilizing a structural approach, I construct novel measures of regional social capital using the Ellison-Glaeser (EG) index of coagglomeration (Ellison et al., 2010) in the context of nonprofit organizations. Also constructing metrics for Marshall's agglomeration mechanisms, I use panel data from US city-industries for 2004-2013 to estimate the count of new establishment births as a function of social proximity, industry agglomeration, and their interactions. I find that 'related variety', or the balance of proximity across different dimensions, is key to promoting entrepreneurship across all industries. In particular, I find that the interplay between social proximity and own industry concentration is particularly strong, while labor market pooling is most related with social proximity out of the Marshallian forces. However, I find that these relationships vary significantly across industries, highlighting the need for nuanced approaches that consider proximity's various dimensions in the context of entrepreneurship that are sensitive to industry and region characteristics.

Keywords: Entrepreneurship, social capital, agglomeration, related variety, nonprofit organizations
JEL classifications: L26, L31, M13, R12

1. Introduction

The observed spatial variation in entrepreneurship rates has garnered much attention in the field of economic geography due to its positive effects on job creation and economic growth (Acs and Armington, 2006; Delgado et al., 2010; Glaeser et al., 2010). Naturally, this has sparked discussion regarding the causes for such disparity. Notably in the economic geography and urban economics fields, regional variations in entrepreneurship have been linked with agglomeration economies (Marshall, 1890), where among others, proximity to suppliers and customers, the thickness of labor markets, and the spillover of knowledge have been argued to facilitate positive externalities that benefit new firms (Ghani et al., 2014; Glaeser and Kerr, 2009; Rosenthal and Strange, 2003).

In tandem, we have learned that social capital is instrumental in determining regional outcomes (Iyer et al., 2005; Malecki, 2012). Broadly, the view that economic outcomes are in part driven by social forces is certainly not new (Granovetter, 2005). A classic example is Saxenian's (1996) documentation of Silicon Valley, where interactions made possible through horizontal integration aided entrepreneurs in acquiring new knowledge and well suited hires. Safford (2009) documents how two rustbelt communities affected by the offshoring of steel manufacturing experienced drastically different economic trajectories due to contrasting social structures. Economic geography has coined terms such as embeddedness, buzz, local norms, network capital, and untraded interdependencies to identify key social components of regional success (Bürker and Minerva, 2014; Porter, 1998b; Storper, 1995; Storper and Venables, 2004), with some even referring to social interactions as the 'magic' that determines regions' competitiveness (Ioannides, 2013). Social capital is also closely linked to entrepreneurship. At the individual level, it has been theorized to promote both the discovery of promising opportunities and the effective mobilization of resources in exploiting those opportunities (Davidsson and Honig, 2003; Stuart and Sorenson, 2005). Similarly at the regional level, social capital aids entrepreneurs by enabling access to exclusive community resources as well as reducing transaction costs (Audia et al., 2006; Dahl and Sorenson, 2009; Sorenson and Audia, 2000; Westlund et al., 2014).

Such studies have much advanced our understanding of the individual benefits that social capital and agglomeration bring to entrepreneurs. Less understood however, is the complex interplay that is possible between these two forces. Certainly, Marshall's mechanisms of agglomeration embody not only the economic gains, but also the social benefits of co-location. Knowledge spillovers, for example, are facilitated through proximal social linkages which catalyze information exchange (Storper and Venables, 2004). A plausible hypothesis that follows is that social capital and knowledge spillovers complement each other in promoting entrepreneurship. Similarly, proximity to customers and suppliers reduces the transport costs of goods, but also facilitates stronger social connections through repeated transactions which increase trust and access to exclusive resources. Finally, labor market pooling provides entrepreneurs with an abundant supply of better suited workers, but is also indicative of homophily, implying greater social familiarity and lower communication costs (Caragliu and Nijkamp, 2016). These interlinkages are not confined to Marshall's microfoundations. Currid and Williams (2010) document how dense social interactions benefit the cultural industry by providing conventions, establishing 'taste', and facilitating access to gatekeepers.

Evaluating the effects of these interlinkages in determining entrepreneurs' location decisions is the main contribution of this article. I argue that acknowledging this interplay is crucial to better understanding the entrepreneurial ecosystem and formulating educated policy responses sensitive to region and industry circumstances. Consider the proximity paradox, which refers to how geographic proximity's benefits can be hindered by superfluous proximity in the social, institutional, or cognitive dimensions (Broekel and Boschma, 2012). In such cases, 'related variety', or the state in which regional entities are neither too proximal nor distant, is more beneficial than raw concentration alone (Boschma et al., 2012; Caragliu and Nijkamp, 2016). Such observations are mirrored in the social capital literature, where the negative effects of excessive social proximity are referred to as 'the dark side of social capital' (Portes, 1998). For entrepreneurship, the benefits of agglomeration (i.e. g`eographic proximity) can be thus hindered or augmented by social structure, depending on the density and composition of social interactions within a region. The relationship can work both ways, with the benefits of social capital also depending on the degree of agglomeration. To my knowledge, this paper contains one of the first empirical attempts to evaluate these intricacies.

It is noteworthy that social capital research has been criticized extensively in the past due to issues concerning its ambiguous definition and resulting inconsistent measurement (Arrow, 2000; Durlauf, 2002; Manski, 1993). Social capital has traditionally been defined in two different ways. The structural approach (Burt, 2005; Coleman, 1988; Granovetter, 1973) emphasizes the characteristics of social networks themselves, such as density, the types of links, and embedded hierarchies, while the instrumental approach (Bourdieu, 1986; Lin, 2001) views social capital as the social resources *procured through* network links. In this study, I adopt the structural view and define regional social capital as 'the characteristics of local social interactions that facilitate the identification, creation, and mobilization of socioeconomic resources'. This approach is chosen for two reasons. First, considering social capital as a resource (such as trust or knowledge) effectively renders it nearly indistinguishable from outcomes due to other common factors including education, institutions, and agglomeration (Sobel, 2002). Second, equating social capital with resources leads to circular arguments where positive outcomes are due to social capital, but evidence for social capital are those outcomes themselves (Portes, 1998). A structural definition mitigates these concerns by distinguishing social interaction characteristics with the resources procured through these interactions, while also allowing for the much needed examination of how regional social structures influence agglomeration and entrepreneurship.

Even with concrete definitions, the proper measurement of social capital has proven prohibitively difficult due to challenges in collecting social data that is adequate both in its depth and breadth (Rupasingha et al., 2006). The common approach has been to utilize readily available proxies such as trust, crime rates, or voter turnout, which have been met with extensive criticism regarding their validity (for a review, see Durlauf, 2002; and Sobel, 2002). Addressing these concerns, I follow the civic engagement and associational activity literature (Putnam et al., 1993; Putnam, 2001; Woolcock, 2001) and construct social capital measures based on a unique dataset containing organization level panel data on the near entirety of nonprofit organizations in the United States. The use of this extensive dataset along with an approach that captures structural characteristics allows the small number problem to be sidestepped while also addressing validity issues. Nonetheless, I do not claim this approach to yield an all-encompassing measure of regional social capital; indeed, more general – and possibly compelling – conceptualizations exist (Malecki, 2012; Storper, 2005). Rather, I argue that while still a proxy, my measure captures particularly well the social context as it relates to associational activity, which at the regional scale has shown to be a key determinant of community structure across a variety of settings, both national and subnational (Alesina and La Ferrara, 2000; Samila and Sorenson, 2017; Woolcock and Narayan, 2000). I add that a 'circus tent' approach to social capital including all possible social constructs carries the danger of diluting the currency of the concept by divorcing it from concrete theoretic roots (de Souza Briggs, 2004; Lin, 2001). My choice responds to these concerns as well.

Another shortcoming of many existing entrepreneurship studies is limited industry coverage, with the common focus being on manufacturing industries (Ellison et al., 2010; Rosenthal and Strange, 2001), traded industries (Delgado et al., 2014; Porter, 2003), or specific industry groups (Rosenthal and Strange, 2003).[1] This is concerning especially when considering social constructs as entrepreneurship determinants. The effects of social capital are likely to differ for the services sector as opposed to, for example, the mining sector. This paper studies nearly all relevant industries. The empirical results suggest that the variability across industries is substantial, highlighting the need for a more systematic approach towards assessing spatial variations in entrepreneurship that considers region and industry specific characteristics.

The remainder of the article is organized as follows. Section 2 examines the literature on social capital and explains the data and methods used to measure it in this study. Section 3 presents the different agglomeration theories, as well as the specific data and metrics used to capture these forces. Section 4

[1] Notable exceptions include Ghani et al. (2014), Glaeser et al. (Glaeser et al., 2014), and Nyström (Nyström, 2007). However, none consider social capital effects.

discusses the empirical framework, while Section 5 presents the results. Section 6 concludes.

2. Regional social capital

2.1. Social interactions and social capital: theories and concepts

Economic geography has developed a wide-ranging literature studying social interactions from a spatial perspective. Related constructs include i) 'untraded interdependencies', the local relations and conventions which enable communication and coordination between economic agents (Storper, 1997), ii) 'buzz', the face-to-face interactions within space that facilitate knowledge exchange and build trust (Storper and Venables, 2004), and iii) the 'social milieu', the social agglomerations that enable mechanisms necessary for industry growth (Currid and Williams, 2010). Such concepts are common in that they include both a social and spatial aspect, and that they emphasize proximity. However, less attention is given to the specific social structures which facilitate meaningful interactions among the social and geographic realms. This black box view of social interactions has been reexamined in recent years, especially for studies of endogenous growth and knowledge spillovers (Tura and Harmaakorpi, 2005). Scholarly debate has increasingly focused on more complex definitions of proximity across different dimensions (Capello, 2009), acknowledging the link between geographic proximity and various dimensions of familiarity (Caragliu and Nijkamp, 2016). For example, lock-in occurs when industries or regions are too proximal in the technological or cognitive dimensions, leading to stifled innovation (Boschma et al., 2012). This alludes to the vital role structures of social interaction play, where in many cases *less* proximity and more bridging of unconnected networks benefits regional outcomes.

Social capital studies have also become increasingly sensitive to such nuances. Emphasis has been placed on the possible dysfunctionality of social capital, where some argue that high levels of social capital hinder development by promoting excessive community closure and restricting receptiveness to new ideas (Portes, 1998; Woolcock and Narayan, 2000). Note how such arguments mirror the debate regarding tensions between different proximity dimensions. Indeed, while excessive social proximity hinders economic outcomes, social proximity does not necessarily equate to social capital. In as much as social phenomena is multifaceted, tie sparsity together with the presence of social connections that traverse groups may very well be another valid form of social capital.

I address these concerns by defining social capital structurally as the characteristics of social interactions that are related to the procurement and mobilization of social resources. Acknowledging the critical tension between social proximity and social distance, I operationalize the concept by further distinguishing between *bonding* and *bridging* social capital, concepts popularized by Putnam (1993; 2001) and Woolcock (2000), but with roots tracing back to the strong (Coleman, 1988; Fukuyama, 1995) and weak (Burt, 2005; Granovetter, 1973) ties literature within sociology. Bonding social capital is characteristic of strong, repeated interactions among those that share similar characteristics and interests, and is commonly related to increased trust, reciprocity, and the enforcement of social norms. Bridging social capital on the other hand refers to weak ties connecting socially distant others that aids individuals in getting ahead by facilitating access to non-redundant information and unique insights. The critical distinction between bonding and bridging social capital is that the former relates to social proximity, while the latter emphasizes social distance. In the context of entrepreneurship, bonding social capital may aid entrepreneurs by, for example, providing easier access to exclusive resources, reducing transaction costs, or fostering entrepreneurial norms. A classic example is immigrant entrepreneurship within ethnic enclaves, where entrepreneurs benefit extensively from close ties with others of the same ethnicity (Wilson and Portes, 1980). Bridging social capital could benefit entrepreneurs seeking information on promising business opportunities or innovative knowledge, a phenomenon well documented in knowledge intensive industries (Agrawal et al., 2006; Stuart and Sorenson, 2003).

Such a distinction is important in that the bonding and bridging social capital dichotomy closely relates to the aforementioned tensions of proximity. At the regional level, both types of social capital consider social interactions manifested within spatial proximity. However, within space, the two distinguish themselves in terms of *social* proximity, with bonding social capital characterizing proximity through homophily and bridging social capital representing heterophily through social distance. Thus the argument for ‘related variety’ (Broekel and Boschma, 2012), at least within the social dimension, can be thought of as a different way of framing the tension between bonding and bridging social capital. This tension can also manifest itself across different proximity definitions that span the social, spatial, and economic realms. Relatedly, Glückler (2007) identifies various typologies of networks within space that consider these forces, suggesting that a combination of strong clusters and bridging ties benefit regional outcomes. The aim of this article is to empirically test such theories, with the main hypothesis being that *related variety across proximity dimensions is key in promoting entrepreneurship*.

2.2. Regional social capital: data and measurement

Regional social capital by its name encompasses both the social and spatial dimensions, with some even arguing that social capital is inherently space-specific (Caragliu and Nijkamp, 2016). Thus a natural starting point to consider these two dimensions in tandem is to link social proximity with the agglomeration of social activity. Zipf’s (1949) famous notion of the ‘principle of least effort’ reasons that propinquity influences interactions because distant ties require greater effort to maintain. Remarkably, this simple notion has stood the test of time even in the internet era, where even online social connections are observed to be geographically bound (Bailey et al., 2018). Considering that homophily is the single most dominant force in social relations (McPherson et al., 2001), it can be reasoned that spatial and social proximity are tightly related at the regional scale (Glückler, 2007). Buzz for example, reinforces social proximity through local face-to-face interactions that build trust, reinforce norms, and establish regional cultures (Currid and Williams, 2010; Storper and Venables, 2004).

Consistent with this reasoning, I consider the agglomerative forces among nonprofit organizations within a region to measure social proximity, with the critical assumption that the greater the agglomerative forces, the greater the social proximity. My examination of nonprofit activity follows the popular literature on civic engagement and associations, where the formation of groups and other forms of collective civic activity have been shown to strongly determine social characteristics at the national and regional scale (Alesina and La Ferrara, 2000; Knack and Keefer, 1995; Putnam et al., 1993; Rupasingha et al., 2006; Woolcock, 2001). It also adheres to the recognition of "community" as an indispensable component of the regional economy (Marshall, 1919; Storper, 2005).

Specifically, I use data published by the National Center for Charitable Statistics (NCCS) that encompasses all tax exempt organizations in the United States documented by the Internal Revenue Service (IRS), which I collect for years 2004-2012. Tax exempt status is granted by the IRS to organizations operating exclusively for societal or mutual benefit and whose earnings are not inured to private shareholders or individuals. In this sense, the data provide the most comprehensive account of associational activity possible given practical data restrictions, and in 2012 this amounted to roughly 1.5 million documented nonprofits. The online appendix provides additional details regarding the dataset, its limitations, and the construction of the social metrics.

Critically, the data contain information on organization level main activities classified based on the National Taxonomy of Exempt Entities (NTEE), a classification scheme developed jointly by the NCCS and IRS which encompasses a broad variety of organization types ranging from arts organizations to religious congregations. This amounts to over 600 mutually exclusive categories, determined from program descriptions on tax forms and nonprofit directories. Excluding irrelevant categories such as cemeteries, public utilities, and K-12 education, I divide the remainder into "civic engagement" (C-type) and "knowledge seeking" (K-type) groups (see Table 1). The former includes organizations that generally relate to civic activity, such as arts, civil rights, and community improvement, while the latter includes organizations relating to education, employment, and scientific research. This distinction is made based on social capital theory, which suggests that different motives govern participation in these two realms. Civic engagement is mostly linked to the enforcement of norms, building of trust, and volunteerism and is dominated by homophilous ties, while knowledge seeking activities mainly constitute heterophilous relationships that concern the creation and diffusion of knowledge that can be used to 'get ahead' (Malecki, 2012). Admittedly, the distinction is not perfect: civic activity also occurs in knowledge seeking groups, and vice versa. Nonetheless, as the proceeding analysis shows, social interactions taking

Table 1. NTEE major groups and classifications

NTEE major group	Description	Classification
A	Arts, Culture & Humanities	Civic (C-type)
B	Education	Knowledge (K-type)
C	Environmental Quality, Protection, and Beautification	Civic (C-type)
D	Animal-Related	Civic (C-type)
E	Health Care	Civic (C-type)
F	Mental Health & Crisis Intervention	Civic (C-type)
G	Voluntary Health Associations & Medical Disciplines	Civic (C-type)
H	Medical Research	Knowledge (K-type)
I	Crime & Legal-Related	Civic (C-type)
J	Employment, job related	Knowledge (K-type)
K	Food, Agriculture & Nutrition	Civic (C-type)
L	Housing & Shelter	Civic (C-type)
M	Public Safety, Disaster Preparedness & Relief	Civic (C-type)
N	Recreation, Sports, Leisure, & Athletics	Civic (C-type)
O	Youth Development	Civic (C-type)
P	Human Services	Civic (C-type)
Q	International, Foreign Affairs & National Security	Civic (C-type)
R	Civil Rights, Social Action & Advocacy	Civic (C-type)
S	Community Improvement & Capacity Building	Civic (C-type)
T	Philanthropy, Voluntarism & Grantmaking Foundations	Civic (C-type)
U	Science and Technology Research Institutes	Knowledge (K-type)
V	Social Science Research Institutes	Knowledge (K-type)
W	Public & Societal Benefit	Civic (C-type)
X	Religion Related, Spiritual Development	Civic (C-type)
Y	Mutual & Membership Benefit Organizations	Civic (C-type)

Notes: Certain subcategories within each NTEE major group are classified differently. Not all subcategories are included in the analysis. The online appendix provides further details regarding C and K-type groupings.

place within these realms exhibit drastically different associations with entrepreneurship, which is likely in part due to underlying differences between these two social dimensions.

As a first step, I measure the agglomerative forces between nonprofit categories by calculating the Ellison-Glaeser (EG) index of coagglomeration (Ellison et al., 2010) for pairwise combinations of NTEE codes, which is done separately for the C and K-type organizations. The index is defined as

$$EG_{ab} = \frac{\sum_m (s_{ma} - x_m)(s_{mb} - x_m)}{1 - \sum_m x_m^2} ,$$

where s_{ma} refers to the share of NTEE code a's employment contained in region m, and x_m is region m's share of total nonprofit employment. The index is calculated at the Metropolitan Statistical Area (MSA) level to maintain consistency with the empirical analyses, and robustness checks are conducted based on county level calculations. Panels are pooled and median employment values are taken to reduce noise due to coding errors and intermittent reporting. It can be seen that the index is closely related to the covariance of employment shares across regions, and Ellison and Glaeser (1997) and Ellison et al. (2010) provide mathematical justification of how the index measures the strength of agglomerative forces in a particular model of location choice. The index is quite general in that it places little restrictions on location decisions, allowing it to be generalizable to the nonprofit context. Higher values of EG_{ab} indicate stronger attractive forces due to factors such as shared interests, inputs, labor requirements, or natural advantages.

Table 2 presents the 10 most coagglomerated pairs for the C and K-type groups. For C-type organizations, coagglomeration is generally strong among nonprofits engaged in international affairs, while for K-type organizations, social science research categories are the most coagglomerated. These activities are generally concentrated in large cities such as Washington DC, New York, and Boston. Compared to estimates reported for pairwise manufacturing industries in Ellison et al. (2010), the index values for nonprofits are right skewed, reflecting that some nonprofit activities exhibit very strong tendencies to coagglomerate. The lowest value of the index for C-type organizations was -0.077 for Public automotive safety (M42) and In-home assistance (P44), while for the K-types it was -0.040 for Economics & behavioral science research (V22-V24) and Graduate & professional schools (B50).

The EG coagglomeration index is a global measure of the agglomerative forces between groups. From this global measure, a region specific metric of social proximity is calculated as

$$\text{Social proximity}_{rt} = \frac{\sum_a \sum_{b>a} W_{art} \cdot W_{brt} \cdot EG_{ab}}{\sum_a \sum_{b>a} W_{art} \cdot W_{brt}} \;, \qquad where$$

$$W_{art} = Count_{art} \times Size_a \,.$$

Here, $Count_{art}$ is the count of NTEE code a organizations in region r for year t, while $Size_a$ is average employment for NTEE code a organizations.[2] These two components constitute the weights (W), and both weights for categories a and b enter multiplicatively for each EG_{ab} term. The denominator normalizes the metric, thus allowing it to be interpreted as the weighted average of EG_{ab} values. As for

[2] The product of count and average size is used in lieu of actual employment for two reasons. First, as opposed to counts, employment data are not available for all panel years, and in many cases are reported intermittently. Second, small nonprofits with less than $25,000 (but greater than $5,000) in annual receipts and some religious organizations do not report employment to the IRS, but are included in counts. The weights account for size differences between, for example, universities and churches, while allowing the inclusion of all nonprofits documented in the data.

Table 2. Highest EG coagglomeration index (EG_{ab}) values for C and K-type groups

Rank	Nonprofit category a (NTEE code)	Nonprofit category b (NTEE code)	EG index
A. Civic engagement (C-type) organizations: EG coagglomeration index using MSA level employment			
1	International Agricultural & Economic Dev. (Q31-Q32)	International Democracy & Civil Society Dev. (Q35)	0.368
2	Voluntary Health Assoc., Surgical Specialties (G9B)	Disabled Persons Rights (R23)	0.344
3	International Agricultural & Economic Dev. (Q31-Q32)	International Peace & Security (Q40)	0.300
4	Human Services, In-Home Assistance (P44)	Philanthropy Orgs., Voluntarism Promotion (T40)	0.295
5	International Agricultural & Economic Dev. (Q31-Q32)	Public & Societal Benefit, Public Administration (W20)	0.293
6	Animal and Wildlife, Fisheries Resources (D33)	International Agricultural & Economic Dev. (Q31-Q32)	0.289
7	Human Services, In-Home Assistance (P44)	International Affairs Alliances & Advocacy (Q01-Q03)	0.287
8	Voluntary Health Assoc., Alzheimers disease (G83)	Public & Societal Benefit, Other (W90-W99)	0.283
9	International Affairs Alliances & Advocacy (Q01-Q03)	Philanthropy Orgs., Voluntarism Promotion (T40)	0.274
10	Human Services, In-Home Assistance (P44)	International Human Rights (Q70)	0.263
B. Knowledge seeking (K-type) organizations: EG coagglomeration index using MSA level employment			
1	Foreign Affairs & Other Misc. Research (Q,R,T,X,Y05)	Social Science Research Support Orgs. (V01-V19)	0.351
2	Social Science Research Support Orgs. (V01-V19)	Economics & Behavioral Science Research (V22-V24)	0.338
3	Foreign Affairs & Other Misc. Research (Q,R,T,X,Y05)	Economics & Behavioral Science Research (V22-V24)	0.299
4	Science & Tech. Research Support Orgs. (U01-U19)	Social Science Research Support Orgs. (V01-V19)	0.264
5	Public & Societal Benefit Research (W05)	Social Science Research Support Orgs. (V01-V19)	0.257
6	Foreign Affairs & Other Misc. Research (Q,R,T,X,Y05)	Science & Tech. Research Support Orgs. (U01-U19)	0.247
7	Employment Preparation & Procurement (J20)	Social Science Research Support Orgs. (V01-V19)	0.223
8	Medical Disciplines Research (H90)	Social Science Research Support Orgs. (V01-V19)	0.223
9	Science & Tech. Support Organizations (U01-U19)	Economics & Behavioral Science Research (V22-V24)	0.219
10	Foreign Affairs & Other Misc. Research (Q,R,T,X,Y05)	Public & Societal Benefit Research (W05)	0.217

the EG coagglomeration index, I calculate the social proximity metric separately for C and K-type groups.

This metric is in essence a simple density measure of social interactions, where *higher* values indicate greater average social proximity and thus greater *bonding* social capital, while *lower* values are indicative of greater average social distance and thus greater *bridging* social capital. Figure 1 provides a network representation of C-type organizations connected to each other by links representing EG_{ab} values. The network is superimposed with the relative differences in shares of nonprofits between the Washington-Arlington-Alexandria MSA and the Philadelphia-Camden-Wilmington MSA, two representative regions similar in size. The first noteworthy observation is that the network exhibits a clear core-periphery structure, with a dense central cluster surrounded by a weakly connected periphery. This indicates that the coagglomerative forces are strong for only a subset of activities, which mostly consist of public and societal benefit (NTEE W), international development (NTEE Q), and civil rights and social action (NTEE R) organization categories. Contrasting Washington DC with Philadelphia, it is seen that Washington dominates the core, which results in a much higher social proximity value for the region

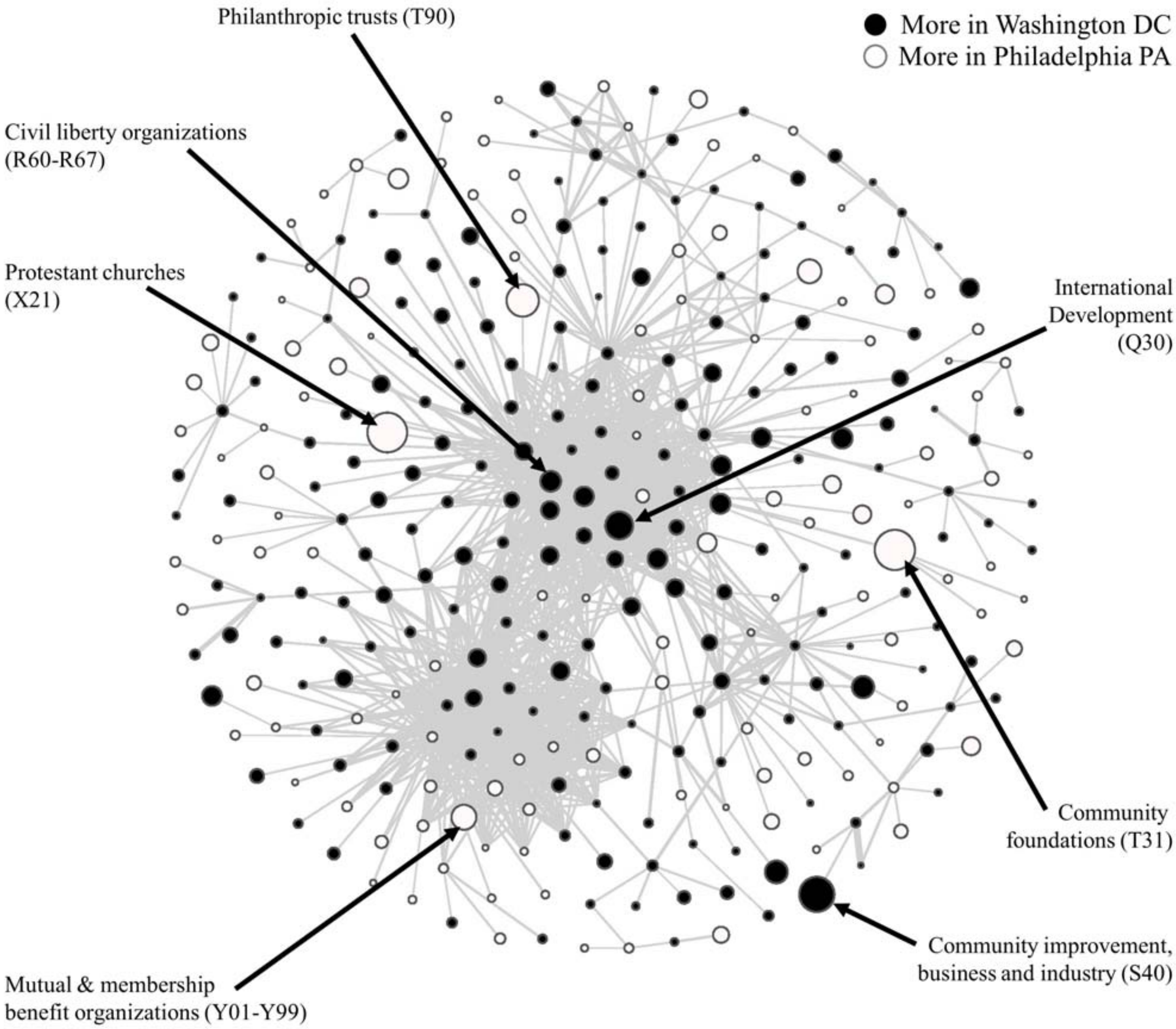

Figure 1. The coagglomeration network for civic engagement (C-type) organizations, superimposed with the absolute differences in nonprofits shares between Washington DC and Philadelphia PA.

Notes: Nodes represent NTEE codes. Edges represent EG coagglomeration index values. Node size is proportional to absolute differences in nonprofit shares. To make visualization manageable, only edges that are either part of the maximum spanning tree (Hidalgo et al., 2007) or those corresponding to EG index values of 0.01 or greater are shown. The network is visualized using the ForceAtlas2 (Jacomy et al., 2014) and Fruchterman-Reingold (1991) algorithms in succession.

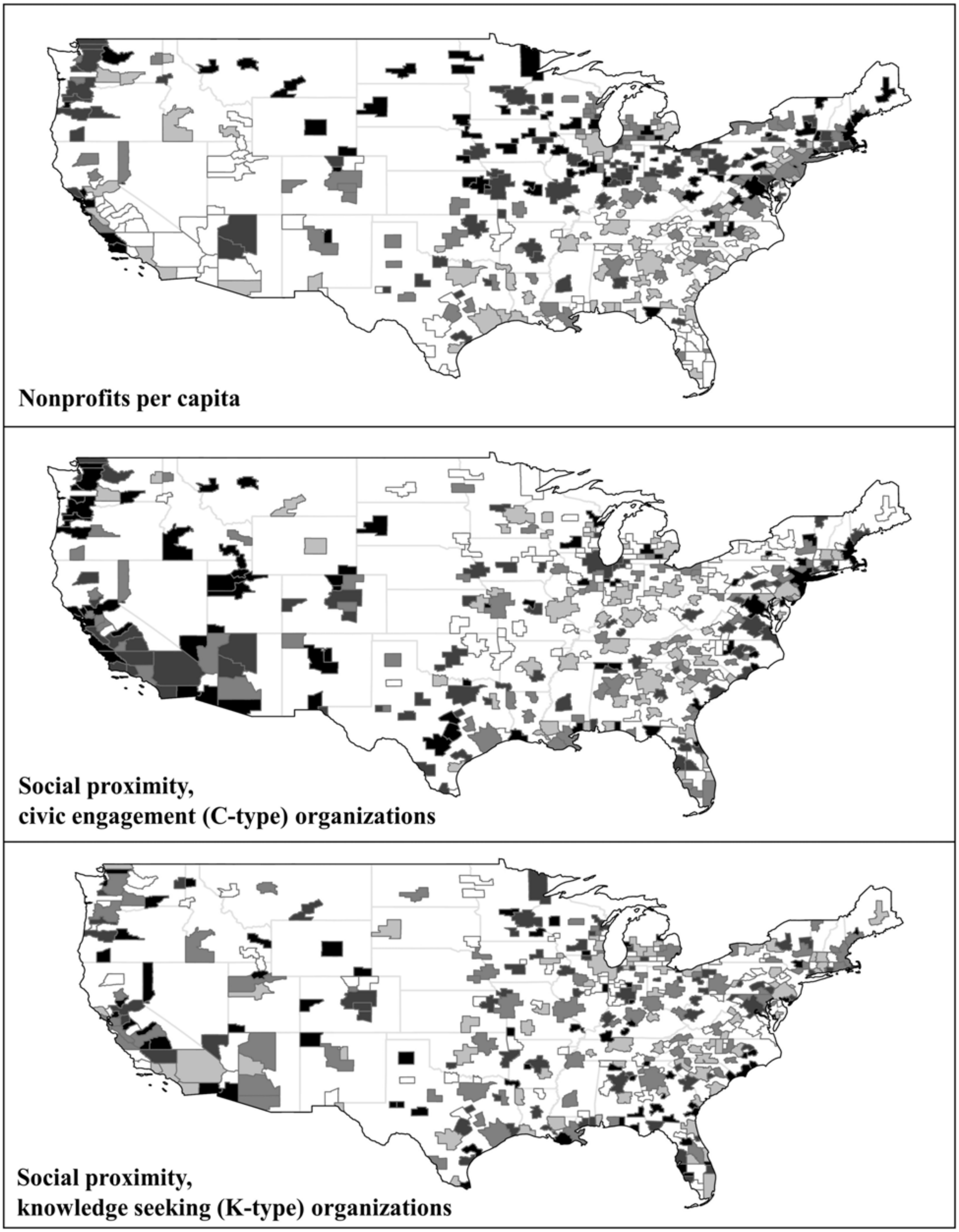

Figure 2. Nonprofits per capita and social proximity for C and K-type organizations, 2004-2012 median. Darker corresponds to higher values.

compared to Philadelphia, despite roughly similar populations. Figure 2 shows spatial patterns for the number of nonprofits per capita and the social proximity metrics. The correlation between the social proximity metrics for C and K-type groups was 0.03, while their correlations with nonprofits per capita were 0.13 and -0.19 respectively. The highest social proximity measured for C-type organizations was 0.0022 for Gulfport-Biloxi Mississippi in 2005, while for K-type organizations Hanford-Corcoran California was highest at 0.0158 in 2011.

Notably, the EG index, while appealing in its generality, is unable to differentiate between different coagglomeration mechanisms. For manufacturing industries, this results in coagglomeration due to the sharing of customers and suppliers, labor, or knowledge being indistinguishable from coagglomeration due to, for example, the need to be close to the shore (Ellison and Glaeser, 1997). Similarly for nonprofits, coagglomeration can occur due to exogenous factors. The high social proximity value for C-type organizations in Washington DC is in large part due to the concentration of foreign affairs and public policy organizations which benefit from being located in the nation's capital. I do not attempt to distinguish between these different forces, and rather consider all coagglomeration mechanisms germane in reinforcing social proximity. Indeed, foreign affairs and policy advocacies both being reliant on federal government ties constitutes another not insignificant mechanism that augments social interactions.

3. Agglomeration: concentration and inter-industry relations

Agglomeration economies refer to the mechanisms that drive the geographic concentration of employees and firms, and is a centripetal force that acts against the centrifugal forces of congestion, pollution, and higher land costs (Krugman, 1991). Fundamentally, gains from concentration come from the reduction of transportation costs, whether it be the transport of goods, labor, or ideas. My main strategy is to measure four agglomeration factors, which constitute own industry concentration and the three agglomeration economies of Marshall (1890). Own industry concentration is measured by the location quotient (LQ), where

$$LQ_{irt} = \frac{employment_{irt}/employment_{rt}}{total\ employment_{it}/total\ employment_{t}},$$

and i, r, and t denote industries, regions, and years respectively.

For all agglomeration metrics, industry employment data spanning years 2004-2012 are taken from the Wholedata Establishment and Employment Database, a dataset developed by the Upjohn Institute that provides an unsuppressed version of the County Business Patterns series published by the Census Bureau (Isserman and Westervelt, 2006). Industries are defined at the four-digit level based on the 2007 North American Industry Classification System (NAICS), and I exclude agriculture (NAICS 11), public administration (NAICS 92), and private households (NAICS 8141), along with a few industries for which reliable data were not available.[3] Regions are defined at the MSA level using 2009 definitions published by the Office of Management and Budget. The online appendix provides additional information regarding the data and methods used for constructing the metrics.

The following subsections start by discussing each of the Marshallian forces and the metrics and data used to measure them. The three measures are conceptually and mechanically similar in that they capture different dimensions of proximity for a focal industry in a region, only differing in how inter-industry linkages are defined. However, it is expected that the Marshallian mechanisms do not constitute the primary forces of agglomeration for some industries, especially the services sector. Thus I construct an additional metric of agglomeration that measures the overall strength of all agglomerative forces using the EG coagglomeration index for industries, further explained below. Readers are turned to Glaeser and Gottlieb (2009) for a more extensive review of the literature on agglomeration.

3.1. Proximity to customers and suppliers: input-output linkages

The concentration of firms brings with it reductions in the costs of transporting goods, thus increasing productivity. When inputs are bulky, producing close to raw materials is beneficial, while producing close to customers is ideal when finished products are costly to transport. Examples of this phenomenon include the refined sugar industry in nineteenth century New York (Glaeser and Kerr, 2009), or the 1960s and 70s steel industry in the U.S. northeast.

I utilize the supply and use tables of the 2007 Benchmark Input-Output Accounts published by the Bureau of Economic Analysis to measure the strength of customer-supplier relationships (Ellison et al., 2010; Jofre-Monseny et al., 2011). Supplier relationships are defined as

$$W_{i \leftarrow j}^{Input} = inputs_{i \leftarrow j} \; / \; total\ inputs_i \quad ,$$

and thus represent the share of industry i's inputs that come from industry j. Customer relationships ($W_{i \rightarrow j}^{Output}$) are defined analogously as the share of outputs that are sold to industry j. Only inter-industry relations are considered, excluding other factors such as value added or final demand. Own industry sales are also excluded in calculations. I construct a combined measure of the flow of goods between industries as $W_{ij}^{IO} = mean\,\{\, W_{i \leftarrow j}^{Input},\ W_{i \rightarrow j}^{Output} \,\}$. This metric is not symmetric due to differences in industry size (i.e. $W_{ij}^{IO} \neq W_{ji}^{IO}$). Panel A of Table 3 presents the top five industry pairs that exhibit the strongest customer-supplier relationships.

---

[3] These industries are petroleum and coal products manufacturing (NAICS 3241), rail transportation (NAICS 4821), postal services (NAICS 4911), central banks (NAICS 5211), and insurance and employee benefit funds (NAICS 5251).

Table 3. Strength of inter-industry relationships

Rank	Industry i (4-digit NAICS)	Industry j (4-digit NAICS)	Value
A. Input-output linkages, W_{ij}^{IO} *(mean of Input and Output shares, Benchmark IO accounts, 2007)*			
1	Motor Vehicle Manufacturing (3361)	Motor Vehicle Parts Manufacturing (3363)	0.619
2	Radio and Television Broadcasting (5151)	Advertising, Public Relations, and Related Services (5418)	0.568
3	Other Information Services (5191)	Advertising, Public Relations, and Related Services (5418)	0.503
4	Newspaper, Periodical, Book, Publishers (5111)	Advertising, Public Relations, and Related Services (5418)	0.468
5	Motor Vehicles, Parts & Supplies Wholesalers (4231)	Automotive Repair and Maintenance (8111)	0.445
B. Labor market pooling, W_{ij}^{labor} *(correlations, Occupational Employment Statistics, 2000 - 2012, median)*			
1	Ventilation, Heating, Air-Cond. Equipment Manufacturing (3334)	Other General Purpose Machinery Manufacturing (3339)	0.949
2	Converted Paper Product Manufacturing (3222)	Soap, Cleaning Compound & Toilet Manufacturing (3256)	0.946
3	Industrial Machinery Manufacturing (3332)	Other General Purpose Machinery Manufacturing (3339)	0.945
4	Alumina & Aluminum Production & Processing (3313)	Foundries (3315)	0.945
5	Nonferrous Metal Production and Processing (3314)	Foundries (3315)	0.943
C. Knowledge spillovers, $W_{ij}^{knowledge}$ *(patent citation shares, NBER patent database, 1976-2006)*			
1	Shoe Stores (4482)	Footwear Manufacturing (3162)	0.781
2	Other Telecommunications (5179)	Communications Equipment Manufacturing (3342)	0.689
3	Child Day Care Services (6244)	Other Miscellaneous Manufacturing (3399)	0.651
4	Jewelry, Luggage, & Leather Goods Stores (4483)	Other Leather & Allied Product Manufacturing (3169)	0.644
5	Scenic & Sightseeing Transportation, Land (4871)	Ship & Boat Building (3366)	0.639
D. EG coagglomeration index for pairwise industries, W_{ij}^{EG} *(Unsuppressed County Business Patterns, 2004-2012, median)*			
1	Cut and Sew Apparel Manufacturing (3152)	Motion Picture and Video Industries (5121)	0.144
2	Cut and Sew Apparel Manufacturing (3152)	Agents & Managers for Artists & Other Public Figures (7114)	0.142
3	Cut and Sew Apparel Manufacturing (3152)	Independent Artists, Writers, and Performers (7115)	0.120
4	Cut and Sew Apparel Manufacturing (3152)	Sound Recording Industries (5122)	0.106
5	Motion Picture and Video Industries (5121)	Agents & Managers for Artists & Other Public Figures (7114)	0.104

Using these relationships, I measure the proximity to customers and suppliers for a focal industry i in region r, year t as

$$IO\ proximity_{irt} = \sum_{j \neq i} \left(\frac{employment_{jrt}}{total\ employment_{rt}} \times W_{ij}^{IO} \right) ,$$

which in essence is the employment weighted average of customer-supplier linkages to all other industries. The highest observed Input-Output proximity was 0.1849 for Motor Vehicle Manufacturing (NAICS 3361) in Kokomo Indiana in 2005, while the lowest was 0.0003 for Nonscheduled Air

Transportation (NAICS 4812) in Morristown Tennessee in 2004.

3.2. Proximity to workers: labor market pooling

Firms also agglomerate to reap the benefits of a large labor pool, which facilitates easier worker movement across firms and industries. This allows for risk sharing in the labor market, and thus maximizes productivity while reducing wage fluctuations. Agglomerations also facilitate better worker-firm matches (Helsley and Strange, 1990), with entrepreneurs tending to start firms in areas with better access to a suitable labor force (Combes and Duranton, 2006).

To capture this mechanism, I measure the extent to which two industries share similar labor requirements. I use data on occupational employment by industry taken from the Occupational Employment Statistics series published by the Bureau of Labor Statistics for years 2000-2012, which provides detailed employment patterns for roughly 800 occupations across all industries. Following Ellison et al. (2010), I calculate similarities in labor requirements between pairwise industries (W_{ij}^{labor}) by taking the median of yearly correlations of employment shares across occupations. Panel B of Table 3 provides the top five most similar industry pairs. Labor market proximity ($Labor\ proximity_{irt}$) is calculated in an analogous manner to $IO\ proximity_{irt}$, substituting W_{ij}^{IO} with W_{ij}^{labor}. The highest observed labor market proximity was 0.7115 for Hardware, and Plumbing and Heating Equipment and Supplies Merchant Wholesalers (NAICS 4237) in Dalton Georgia for 2005, while the lowest was -0.0288 for Child Daycare Services (NAICS 6244) in Elkhart-Goshen Indiana, also for year 2005.

3.3. Proximity to ideas: knowledge spillovers

The third Marshallian force is the presence of suppliers of ideas. Marshall asserted that spatial proximity eases the movement of knowledge and ideas among firms and workers, famously quoting how the mysteries of the trade were "in the air". If firms colocate to share knowledge, it should be that industries using similar knowledge cluster together (Glaeser and Kerr, 2009; Jofre-Monseny et al., 2011; Saxenian, 1996).

To test the importance of knowledge flows, I measure the extent to which two industries share technologies through patent citations (Jaffe et al., 1993). Data are taken from the NBER Patent Database, which documents all patent citations from 1976-2006, and I exclude self-citations as well as patents filed outside of the U.S. (Hall et al., 2001). I map technologies to industries using a mapping scheme based on an algorithmic approach that mines the textual content of patent abstracts, matching keywords with industry descriptions (Goldschlag et al., 2016). Analogous to $W_{i \leftarrow j}^{Input}$, I calculate the share of patents associated with industry i that cite patents associated with industry j. Like the supplier relationship metric, this measure is not symmetric ($W_{ij}^{knowledge} \neq W_{ji}^{knowledge}$). Panel C of Table 3 provides the top five industry pairs. Again, $Knowledge\ proximity_{irt}$ is calculated analogously to the IO and labor proximity metrics using $W_{ij}^{knowledge}$. The highest observed knowledge proximity was 0.2384 for Scenic and Sightseeing Transportation, Water (NAICS 4872) in Pascagoula Mississippi for 2010, and the lowest was 3.9×10^{-5} for Residential Mental Retardation, Mental Health and Substance Abuse Facilities (NAICS 6232) in Laredo Texas, also for 2010.

Despite their importance, knowledge spillovers are notoriously difficult to measure. They encompass many different concepts ranging from endogenous growth to social interactions theory. I note that rather than being a measure for all intellectual spillovers, this measure rather captures the exchange of technologies (Ellison et al., 2010), and is thus limited in its scope to knowledge transfer that is mainly pecuniary in nature. It is also important to note that knowledge spillovers occur through the other Marshallian channels, as well as through social interactions. This is especially true in this study, as the social proximity measure for K-type organizations can be regarded as another measure of knowledge flows, albeit in a different setting. As such, the effect of knowledge proximity on entrepreneurship is expected to be weaker than the other agglomeration factors.

3.4. A composite index of inter-industry relations

Not all agglomeration is due to the presence of customer-supplier linkages, labor market pooling, or knowledge spillovers. For example, the arts industries cluster in Los Angeles and New York due to social factors that include shared tastes and norms (Currid and Williams, 2010), and shared dependencies for certain natural advantages constitute another significant reason to coagglomerate. For many industries, it may be the case that such factors dominate the Marshallian forces in driving agglomeration.

Similar to the construction of the social proximity metrics, I utilize the EG coagglomeration index defined across *industries* to construct a composite index of agglomerative forces (W_{ij}^{EG}) using industry-MSA level employment taken from the unsuppressed County Business Patterns data. Due to the nature of the EG index, this metric can be thought of as measuring all agglomerative forces, including Marshall's mechanisms, social factors, and natural advantages. Panel D of Table 3 lists the top five industry pairs. Analogous to the other Marshallian metrics, I construct a composite proximity metric ($Composite\ proximity_{irt}$) utilizing W_{ij}^{EG}. I find that the highest observed composite proximity was 0.0234 for Pipeline Transportation of Crude Oil (NAICS 4861) in Midland Texas in 2012, while the lowest was -0.0164 for Securities and Commodity Exchanges (NAICS 5232) in Pascagoula Mississippi in 2010.

4. Estimation strategy

The dependent variable (B_{irt+1}) is the lagged count of new establishment births in industry i, MSA r, and year $t + 1$. I follow Glaeser et al. (2010) and define entrepreneurship as establishment births of single-unit enterprises that are not part of an existing firm, with robustness checks conducted for new establishments of existing firms (i.e. facility expansions). I focus on establishment counts rather than employment to work around critical disclosure issues that hamper granular analyses of employment data for births at the industry-region level. Nonetheless, as seen in previous studies (for example, Delgado et al., 2010), the estimates are not expected to differ significantly based on either measure mainly due to the fact that most new establishments (especially single-unit entrants) are rather small in size (Glaeser and Kerr, 2009). Establishment births data are taken from custom tabulations of the Statistics of U.S. Businesses (SUSB) yearly change series, which encompasses all employer firms[4] with an Employer Identification Number. The SUSB series is based on the Business Register, the Census Bureau's most accurate data source for U.S. business establishments that compiles data from economic censuses, business surveys, federal tax records, and other departmental and federal statistics. Descriptive statistics for the variable are provided in Table 4 and Table A1 of the appendix.

The high observed skewness and large number of zero births (58% of observations are zeros) for industry-MSA pairs present major problems in linear estimation, and thus I implement the Poisson quasi-maximum likelihood (PQML) estimator with industry, region, and year fixed effects. The PQML estimator has its advantages in i) allowing for overdispersion due to imposing no restrictions on the conditional variance (provided the use of fully robust standard errors), ii) being consistent under a weaker assumption of correctly specified conditional means compared to other nonlinear estimators such as zero-inflated models or the negative binomial that rely on additional assumptions, and iii) allowing to sidestep the 'incidental parameters' problem, where for most nonlinear estimators excluding the Poisson, a large number of fixed effects leads to inconsistent parameter estimation under fixed T, $N \to \infty$ asymptotics (Cameron and Trivedi, 2013). In all specifications, standard errors are of the robust variety, clustered at the four-digit industry by MSA level.

The main empirical specification can be represented as

$$E(B_{irt+1}) = \exp(\alpha + \beta_{SOC} X_{rt}^{SOC} + \beta_{AGG} X_{irt}^{AGG} + \beta_{INT} X_{rt}^{SOC} \cdot X_{irt}^{AGG} + \beta_{CON} Z_{irt} + I_i + R_r + T_t)$$

where X_{rt}^{SOC} refers to the set of social proximity metrics for C and K-type organizations, X_{irt}^{AGG} refers to own industry concentration and the various inter-industry agglomeration metrics, $X_{rt}^{SC} \cdot X_{irt}^{AGG}$ refers to the

[4] The series excludes data for private households, railroads, agricultural production, government entities, as well as for industries outlined in Section 3.

Table 4. Descriptive statistics

	Total (2005-2013)	Mean	SD	Min	Max
A. Industry by MSA characteristics					
B_{irt+1} **(single-unit enterprises)**	**4,462,993**	**4.879**	**32.188**	**0**	**2,911**
Mining & utilities	13,078	0.500	3.226	0	91
Construction	617,773	18.909	67.287	0	2,326
Manufacturing	130,247	0.469	3.962	0	774
Trade & transportation	850,590	3.567	19.189	0	1,702
FIRE & business services	1,416,937	11.413	52.848	0	1,821
Other services	1,434,368	6.652	39.824	0	2,911
LQ_{irt}		**1.097**	**4.371**	**0**	**664.627**
Mining & utilities		1.646	9.816	0	380.085
Construction		1.090	2.967	0	237.714
Manufacturing		1.420	6.749	0	664.627
Trade & transportation		0.997	2.314	0	260.380
FIRE & business services		0.759	1.266	0	127.066
Other services		0.921	1.495	0	154.092
$IO\ proximity_{irt}$		**0.005**	**0.003**	2.70×10^{-4}	**0.185**
Mining & utilities		0.005	0.002	8.31×10^{-4}	0.061
Construction		0.004	0.004	7.89×10^{-4}	0.034
Manufacturing		0.003	0.002	3.49×10^{-4}	0.185
Trade & transportation		0.005	0.003	2.70×10^{-4}	0.038
FIRE & business services		0.005	0.002	9.34×10^{-4}	0.052
Other services		0.006	0.003	5.88×10^{-4}	0.051
$Labor\ proximity_{irt}$		**0.433**	**0.074**	**-0.029**	**0.712**
Mining & utilities		0.402	0.060	0.126	0.637
Construction		0.413	0.048	0.138	0.646
Manufacturing		0.456	0.052	0.113	0.708
Trade & transportation		0.459	0.065	0.089	0.712
FIRE & business services		0.446	0.058	0.148	0.680
Other services		0.373	0.082	-0.029	0.672
$Knowledge\ proximity_{irt}$		**0.002**	**0.003**	3.90×10^{-5}	**0.238**
Mining & utilities		0.002	0.001	1.65×10^{-4}	0.024
Construction		0.002	0.002	2.07×10^{-4}	0.052
Manufacturing		0.002	0.002	8.80×10^{-5}	0.060
Trade & transportation		0.002	0.003	7.50×10^{-5}	0.238
FIRE & business services		0.002	0.003	2.11×10^{-4}	0.072
Other services		0.003	0.003	3.90×10^{-5}	0.174
$Composite\ proximity_{irt}$		-6.71×10^{-6}	7.59×10^{-4}	**-0.016**	**0.023**
Mining & utilities		5.30×10^{-4}	0.001	-0.003	0.017
Construction		1.70×10^{-4}	0.000	-0.001	0.006
Manufacturing		2.78×10^{-4}	0.000	-0.010	0.012
Trade & transportation		-4.09×10^{-5}	0.001	-0.010	0.023
FIRE & business services		-3.15×10^{-4}	0.001	-0.016	0.009
Other services		-2.50×10^{-4}	0.001	-0.012	0.007
$BDUM_{ir}$		**0.491**	**0.500**	**0**	**1**
$Expansions_{irt}$		**0.191**	**0.275**	**0**	**1**
B. MSA characteristics					
$Social\ proximity_{rt}^{C-type}$		2.43×10^{-5}	2.77×10^{-4}	**-0.001**	**0.002**
$Social\ proximity_{rt}^{K-type}$		-3.58×10^{-4}	**0.002**	**-0.004**	**0.016**
$HHI_{rt}^{nonprofits}$		**0.028**	**0.017**	**0.011**	**0.406**
$Nonprofits\ per\ 1000_{rt}$		**3.877**	**1.244**	**0.954**	**16.868**
$HHI_{rt}^{industries}$		**0.021**	**0.013**	**0.011**	**0.179**
$\ln population_{rt}$		**12.660**	**1.060**	**10.910**	**16.770**
$Patents\ per\ 1000_{rt}$		**0.245**	**0.420**	**0**	**6.067**

Notes: (1) 280 industries (excluding those outlined in the text) classified at the four-digit NAICS level using 2007 definitions; (2) FIRE refers to the Finance, Insurance, and Real-Estate sectors; (3) 363 MSAs within the contiguous U.S. defined using 2009 OMB definitions; (4) Total number of observations is 914,760.

interaction terms between the social proximity and agglomeration variables, Z_{irt} refers to the set of controls, and I_i, R_r, T_t refer to the industry, region, and year fixed effects respectively. All explanatory variables encompass years 2004-2012 (excluding an indicator variable for pre-existing start-up activity, further explained below), while the dependent variable spans years 2005-2013, for a total of 9 years of data. 280 industries (excluding those outlined in Section 3) are defined using 2007 NAICS definitions at the 4-digit level, which strikes a balance between granularity and noise due to concordances made with the 2002 and 2012 NAICS. 363 MSAs within the contiguous U.S. are included in the study, for a total of 914,760 (i.e.$9 \times 280 \times 363$) observations.

I include a set of control variables to account for other factors that may explain firm births. At the MSA level, I include the Hirschman-Herfindahl Index for nonprofit shares ($HHI_{rt}^{nonprofits}$) and the number of nonprofits per 1,000 ($Nonprofits\ per\ 1000_{rt}$) to control for nonprofit characteristics that may confound the estimates for the social proximity metrics. I also include the Herfindahl index for industry employment ($HHI_{rt}^{industries}$) to control for urbanization economies (Jacobs, 1969), as well as the log of population ($\ln population_{rt}$) to control for city size effects. In addition, I include the number of patents per 1,000 ($Patents\ per\ 1000_{rt}$) taken from U.S. Patent and Trade Office statistics, which controls for regional innovation as well as general levels of educational attainment. At the industry-MSA level, I include a dummy variable ($BDUM_{ir}$) equal to one for any pre-existing start-up activity in year 2003 (Delgado et al., 2010), as well as the share of incumbent establishments that experienced expansions ($Expansions_{irt}$) to control for unmodeled factors (such as city policies promoting certain industries) that may promote firm births in particular industry-MSAs (Glaeser and Kerr, 2009).

It is important to note that there may be a number of other explanations for variations in new establishment counts. The inclusion of the full range of MSA and industry fixed effects controls for the most worrisome confounders that are time invariant, such as a city's natural advantages, climate, or fixed differences in industry characteristics. It is also worthwhile to note that the MSA fixed effects generally control for social characteristics such as demographic composition or culture that are known to be relatively persistent (Andersson and Larsson, 2016), and thus mostly fixed given the timeframe. The year fixed effects control for time-specific shocks such as macroeconomic conditions or business cycles, which is particularly important for this study considering that the panel includes years that were affected by the recent global recession.

Using the count of new firms as the dependent variable also addresses simultaneity issues. Rosenthal and Strange (2003) point out that entrepreneurs are unconstrained by previous decisions and make location choices taking the existing environment as exogenously given. Thus the characteristics of cities are seen as fixed from the viewpoint of an entrepreneur. This along with the inclusion of a 1 year lag alleviates issues of reverse causality. Nonetheless, I lack enough variation to include industry by MSA fixed effects, which would exclude all industry-MSA pairs that experienced zero births during the timeframe, thus biasing the sample. While the share of incumbent establishments that experienced expansions ($Expansions_{irt}$) is expected to net out a significant portion of unmodeled time-varying factors at the industry-MSA level, nonetheless it is not a perfect solution. In addition, I am unable to fully rule out the presence of unobserved time-varying city level determinants that simultaneously affect social proximity and firm births. Overall, even with the careful selection of controls and the arsenal of fixed effects, one should interpret the results throughout the paper as partial correlations rather than causal effects.

5. Results

5.1. Baseline estimations

Table 5 presents the baseline PQML estimates. Only the variables of interest are reported for brevity, with the full model documented in the online appendix. For every specification, all explanatory variables excluding the interactions and the dummy for pre-existing startup activity ($BDUM_{ir}$) are standardized to have mean zero and unit standard deviation to aid interpretation. Variables are also standardized prior to interaction to maintain main effects. All estimation results hereon forth report non-exponentiated coefficients. Thus in column 6, the results imply that a 1 standard deviation increase in social proximity for civic activity decreases the log count of establishment births in an industry-region by .021 (i.e. decreases births by roughly 2.1%), given average levels of the location quotient and Marshallian metrics.

Across all specifications, the main effects of own industry concentration and the three Marshallian metrics are positive and significant, with labor market pooling ($Labor\ proximity_{irt}$) being the most dominant of the Marshallian forces. This mirrors results reported by Glaeser and Kerr (2009), who conduct similar analyses using data for U.S. manufacturing industry-city pairs spanning years 1977-1999. Considering their different measurement of Marshallian factors, the timeframe of studies, and the scope of included industries, the results suggest that the explanatory power of labor market pooling is particularly strong compared to customer-supplier linkages and knowledge spillovers.

The first column includes just the social proximity metrics and the fixed effects. The main effects of both social proximity metrics are negative, suggesting that greater *bridging* social capital positively associates with entrepreneurship. This association is nearly unchanged with the inclusion of the industry traits as well as the controls in columns 3 and 4. However, the main effects are positive and highly significant when excluding the MSA fixed effects and including the interaction terms (column 5), which is likely capturing persistent characteristics such as entrepreneurship culture that are positively correlated with both social proximity and entrepreneurship. Thus although regions with higher levels of bonding

Table 5. Poisson quasi-maximum likelihood (PQML) estimation results

	Dependent variable: count of single-unit establishment births in industry-region					
	(1) Social proximity only	(2) Industry traits	(3) All proximity measures	(4) Including controls	(5) No MSA fixed effects	(6) Full estimation
$Social\ proximity_{rt}^{C-type}$	-0.027***		-0.027***	-0.025***	0.062***	-0.021***
	(0.005)		(0.005)	(0.005)	(0.008)	(0.006)
$Social\ proximity_{rt}^{K-type}$	-0.009***		-0.009***	-0.007***	0.019***	-0.004
	(0.003)		(0.003)	(0.002)	(0.005)	(0.004)
LQ_{irt}		0.085***	0.085***	0.083***	0.114***	0.115***
		(0.011)	(0.011)	(0.010)	(0.007)	(0.007)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$					-0.003	-0.003
					(0.003)	(0.003)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$					-0.037***	-0.035***
					(0.004)	(0.004)
$IO\ proximity_{irt}$		0.049***	0.049***	0.051***	0.074***	0.061***
		(0.008)	(0.008)	(0.008)	(0.010)	(0.010)
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$					-0.018***	-0.006
					(0.005)	(0.005)
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$					0.014***	0.009
					(0.005)	(0.005)
$Labor\ proximity_{irt}$		0.127***	0.127***	0.123***	0.139***	0.119***
		(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$					0.023***	0.025***
					(0.008)	(0.008)
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$					-0.012**	-0.011***
					(0.005)	(0.004)
$Knowledge\ proximity_{irt}$		0.036***	0.036***	0.036***	0.027***	0.036***
		(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$					-0.011**	-0.012***
					(0.004)	(0.004)
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$					0.004	-0.000
					(0.005)	(0.004)
Controls	No	No	No	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	Yes	Yes	Yes	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudo-likelihood	-1,330,469	-1,310,702	-1,310,563	-1,304,522	-1,403,994	-1,297,296
Observations	914,760	914,760	914,760	914,760	914,760	914,760

Notes: (1) Robust standard errors clustered at the four-digit NAICS by MSA level reported in parentheses; (2) all explanatory variables (excluding interactions and $BDUM_{ir}$) are standardized to have mean zero and unit standard deviation to aid interpretation; (3) interacted variables are standardized prior to interaction to maintain main effects; (4) Constant included but not reported, and controls included unless otherwise noted. Full estimation results provided in the online appendix.
*p<0.1, **p<0.05, ***p<0.01.

social capital generally attract more entrepreneurs, increases in bonding social capital per se are not associated with higher levels of entrepreneurship. Such a finding is not insignificant considering that many past studies including social capital measures based on cross-sectional variation conclude that social proximity positively associates with entrepreneurship (for example, Andersson and Larsson, 2016; Westlund et al., 2014). The main effect of social proximity for K-type organizations does become insignificant in the preferred specification reported in column 6 however, and the coefficient for $Social\ proximity_{rt}^{C-type}$ is also smaller in magnitude. This along with the various specifications outlined in the following sections suggest that social forces mainly act through agglomeration mechanisms, rather than independently.

The interaction between $Social\ proximity_{rt}^{K-type}$ and the location quotient is the largest in magnitude and most precisely estimated out of all interaction terms. The positive association between own industry concentration and entrepreneurship is augmented by greater social distance (bridging social capital) in knowledge seeking activities. This is intuitively appealing considering that industry concentration is another form of proximity, and supports the view that related variety – a combination of closeness and distance across different proximity dimensions – is beneficial for economic outcomes (Caragliu and Nijkamp, 2016). The interaction effect between $Social\ proximity_{rt}^{C-type}$ and LQ_{irt} is also quite strong when considering sectors individually (further discussed below), but insignificant overall due to opposing forces across industries, particularly in the construction sector. Overall, this suggests a comparatively local role for social forces in primarily influencing own industry concentration rather than through inter-industry linkages.

When considering Marshallian factors, I find that the interaction effects are strongest for labor market pooling. One possible explanation is the fact that proximity to customers, suppliers, and knowledge as defined in this study are mainly based on pecuniary linkages, while labor market pooling – indicative of homophily – is fundamentally more socially aligned. The relatively large positive interaction effect observed between labor market pooling and social proximity in civic activity supports this explanation. Bonding social capital in the civic sphere and bridging social capital in knowledge seeking activity augments the benefits of thick labor markets, which is consistent with social capital theory as well as with theories of related variety. The interaction effects for IO and knowledge proximity are weaker and mostly insignificant, although a negative interaction effect between $Social\ proximity_{rt}^{C-type}$ and knowledge spillovers is identified. Upcoming estimations by sector shed more light on the contrasting interaction effects between agglomeration and social proximity across sectors.

5.2. Variations by sector

Table 6 documents estimations of the relationship between the variables of interest and entrepreneurship by sector. For each sector, the first column reports estimates including the Marshallian factors and their interactions, while the second column replaces them with the composite proximity measure derived using the EG coagglomeration index.

The main effects of social proximity are not particularly robust across sectors and specifications. Social forces mainly act through agglomeration effects, particularly in tandem with own industry concentration. The interaction effects between social proximity and the location quotient are strong and precisely estimated across all sectors excluding mining and utilities. For the mining and utilities sector, neither social proximity nor the Marshallian forces are particularly significant determinants, suggesting that new establishments in these industries most likely locate in areas with certain natural advantages. The construction sector is unique in that the interaction effects for the location quotient are both negative. Social distance augments the benefits of own industry concentration disproportionately for construction compared to other sectors. These results coincide with the construction industry's characteristics in relying on new customer ties in order to secure contracts, and reflect the industry's highly competitive nature in being dominated by small and medium sized enterprises, many of which are run as ethnic businesses (Choi and Spletzer, 2012; Walton-Roberts and Hiebert, 1997). For all other sectors, higher levels of bonding social capital in civic activities coupled with bridging social capital in knowledge seeking activity is seen to augment concentration's benefits, consistent with social capital theory.

Overall I find that the strength and significance of the Marshallian metrics' interactions vary widely by sector, which in many cases is due to the relative insignificance of the main effects themselves. Nonetheless, the interaction effects between social proximity and labor market pooling are very strong in magnitude for labor intensive industries including mining, utilities, construction, and manufacturing, which again supports the notion that labor market pooling is closely related to social interactions among workers of similar occupations. For these sectors, both bonding social capital in the civic sphere and bridging social capital in knowledge seeking activity augments the benefits of proximity to workers. The interaction effects for customer-supplier linkages and knowledge spillovers are generally less significant, excluding the FIRE & business services sector where a related variety effect is observed for both metrics.

As expected, the Marshallian factors are most strongly associated with entrepreneurship in manufacturing, while performing comparatively poorly in other sectors. Replacing these factors with the composite proximity measure allows identification of the tension between social proximity and agglomeration in sectors for which the Marshallian metrics are of lesser importance, including trade and transportation and other services. For the other sectors, it is seen that the measure's interaction terms roughly average out the individual interaction effects of each Marshallian factor. A noteworthy observation is that comparing log likelihoods, the explanatory power of composite proximity is generally

Table 6. Poisson quasi-maximum likelihood (PQML) estimation results: variations by sector

	Mining & utilities		Construction		Manufacturing		Trade & transportation		FIRE & business services		Other services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Social\ proximity_{rt}^{C-type}$	-0.053	-0.075*	-0.033**	-0.009	-0.056***	-0.022	-0.028*	-0.009	-0.003	-0.015	-0.043***	-0.018**
	(0.043)	(0.043)	(0.015)	(0.011)	(0.020)	(0.019)	(0.017)	(0.013)	(0.010)	(0.010)	(0.009)	(0.008)
$Social\ proximity_{rt}^{K-type}$	0.042	0.164***	0.001	-0.012	0.056***	0.002	0.022**	0.002	-0.008	0.005	0.001	-0.014**
	(0.056)	(0.061)	(0.009)	(0.008)	(0.019)	(0.014)	(0.011)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
LQ_{irt}	0.230***	0.204***	0.246***	0.199***	0.159***	0.124***	0.088***	0.077***	0.162***	0.150***	0.133***	0.096***
	(0.046)	(0.043)	(0.028)	(0.027)	(0.016)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.013)	(0.009)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$	0.041	0.038	-0.059***	-0.048***	0.034***	0.020***	0.074***	0.031***	0.018***	0.019***	0.025***	0.023**
	(0.039)	(0.041)	(0.008)	(0.008)	(0.007)	(0.006)	(0.018)	(0.010)	(0.006)	(0.007)	(0.009)	(0.010)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$	-0.014	0.046**	-0.044***	-0.033***	-0.063***	-0.046***	-0.028***	-0.010**	-0.065***	-0.060***	-0.096***	-0.055***
	(0.018)	(0.020)	(0.006)	(0.006)	(0.010)	(0.006)	(0.009)	(0.005)	(0.007)	(0.008)	(0.016)	(0.011)
$IO_{proximity\ irt}$	0.021		0.027**		0.136***		-0.002		0.067***		0.120***	
	(0.017)		(0.012)		(0.016)		(0.038)		(0.012)		(0.016)	
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$	-0.016		-0.027***		-0.018		0.043***		-0.012***		0.001	
	(0.019)		(0.006)		(0.016)		(0.014)		(0.004)		(0.011)	
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$	0.002		-0.006		-0.008***		-0.010		0.018***		0.012	
	(0.016)		(0.010)		(0.002)		(0.008)		(0.007)		(0.008)	
$Labor\ proximity_{irt}$	0.298*		0.118***		0.414***		0.317***		0.093***		-0.085***	
	(0.165)		(0.035)		(0.059)		(0.054)		(0.027)		(0.031)	
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$	0.134**		0.049***		0.097***		-0.029		0.025***		0.010	
	(0.054)		(0.008)		(0.027)		(0.020)		(0.006)		(0.010)	
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$	-0.115*		-0.036***		-0.103***		-0.037**		0.001		0.003	
	(0.064)		(0.009)		(0.025)		(0.016)		(0.007)		(0.008)	
$Knowledge\ proximity_{irt}$	-0.013		0.026***		0.050***		0.020***		0.036**		-0.014	
	(0.061)		(0.010)		(0.014)		(0.007)		(0.016)		(0.017)	
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$	-0.010		-0.008		-0.005		-0.024***		-0.029***		0.002	
	(0.034)		(0.009)		(0.016)		(0.005)		(0.007)		(0.009)	
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$	0.027		-0.003		0.019**		-0.001		0.030***		0.015*	
	(0.052)		(0.009)		(0.009)		(0.009)		(0.009)		(0.009)	
$Composite\ proximity_{irt}$		0.289***		0.174***		0.274***		0.300***		0.173***		0.230***
		(0.039)		(0.024)		(0.008)		(0.016)		(0.031)		(0.011)
$Social\ proximity_{rt}^{C-type} \times Composite_{proximity\ irt}$		0.017		0.038***		0.041***		-0.054***		-0.027		-0.054***
		(0.018)		(0.011)		(0.012)		(0.012)		(0.018)		(0.018)
$Social\ proximity_{rt}^{K-type} \times Composite_{proximity\ irt}$		-0.129***		-0.017		-0.006		-0.037***		0.054***		0.049***
		(0.027)		(0.017)		(0.012)		(0.012)		(0.020)		(0.013)
Log pseudo-likelihood	-13,341	-12,954	-82,466	-81,359	-143,402	-136,682	-341,111	-329,888	-251,329	-252,626	-371,850	-361,785
Observations	26,136	26,136	32,670	32,670	277,695	277,695	238,491	238,491	124,146	124,146	215,622	215,622

Notes: See Table 4. (1) All specifications include a constant and the control variables, as well as industry, MSA, and year fixed effects; (2) Mining & utilities: NAICS 21, 22; Construction: NAICS 23; Manufacturing: NAICS 31-33; Trade & transportation: NAICS 44-49; FIRE (Finance, Insurance, Real Estate) & business services: NAICS 52-56; Other services: all other NAICS industries excluding those outlined in the text.

greater than the three Marshallian metrics combined, with it being strong enough to crowd out a significant portion of the effects of even own industry concentration and its interactions. This is due to the nature of the EG coagglomeration index, which encompasses all agglomerative forces including those not explained by the Marshallian factors. Nonetheless, even with the inclusion of the composite proximity measure, the interaction effects for the location quotient remain significant in magnitude and quite precisely measured, which is indicative of the robustness of the relationship between social proximity and own industry concentration.

All in all, I find support for the main hypothesis that a *balance* of proximity across different dimensions is key in promoting entrepreneurship. In all specifications, the positive association between agglomeration and entrepreneurship is mediated by negative interaction effects that suggest the importance of social distance (i.e. *bridging* social capital) in at least one dimension. To my knowledge, this study is one of the first to identify such interplay. The results are in line with social capital theory that has long suggested that excessive bonding social capital can hinder development outcomes, often by encouraging seclusion and hampering receptiveness to new ideas (Portes, 1998; Woolcock, 2001). Another explanation for these mediating effects may be due to the fact that entrepreneurship requires innovative ideas and the identification of new opportunities, which is more effectively facilitated through bridging ties that traverse social distance (Davidsson and Honig, 2003). I also find that compared to the strong interaction effects between social proximity and own industry concentration, the interaction effects for the Marshallian forces are limited and specific to certain sectors. While the composite proximity measure helps to identify the related variety effect for sectors where Marshallian forces are less important, nonetheless I am unable to satisfactorily discern whether the relative weakness of the Marshallian metrics' interaction effects for certain sectors are due to actual sectoral differences or simply due to how these forces are measured. These characteristics persist when considering various alternative specifications highlighted in the following section. Future work will hopefully further clarify potential differences in the relationship between social proximity and Marshallian forces across sectors.

5.3. Robustness checks

Tables A2-A5 of the appendix report robustness checks relating to the sector level estimations documented in Table 6. I first assess the possibility that the results may be capturing certain aspects of entrepreneurship during the recent global recession. While the year fixed effects control for overall macroeconomic fluctuations, nonetheless the underlying associations may differ according to business cycles. Table A2 documents estimations excluding years $t = 2007$, 2008, and 2009, which excludes years 2008-2010 for the dependent variable. Table A3 considers an alternative measurement of social proximity where the EG coagglomeration index for nonprofits (EG_{ab}) is defined at the county rather than MSA level. While Ellison et al. (2010) note that the EG coagglomeration index rescales the observed covariance to be theoretically invariant to the fineness of geographic breakdown, nonetheless the calculated EG index values for the two geographic units are quite different, exhibiting correlations of 0.64 and 0.80 for C and K-type activities respectively. This may be due to the fact that MSAs do not encompass all counties, or because different agglomerative forces act at different spatial scales. Table A4 documents estimations with new establishments of existing firms (facility expansions) as the dependent variable, which assesses whether the results are robust across different establishment populations. Finally, Table A5 documents estimations utilizing a zero-inflated Poisson model to assess whether the observed overdispersion and high number of zeros affect the PQML estimates. One could argue that there might be instances where a particular industry is very unlikely to reside in a certain region (e.g. coal mining in New York City), in which case a two-step process that separately models these 'true zeros' would be more adequate. I utilize the logit link for the inflation model using all explanatory variables, excluding the interactions to aid convergence. The fixed effects are also excluded in the inflation model to address the incidental parameters problem, but included in the count model.

I find that that the main effects for both social proximity metrics continue to be generally insignificant, although fairly robust in direction and magnitude across specifications. This reinforces the observation that social proximity mainly acts through agglomerative forces. The direction and magnitude of the interaction effects between social proximity and the location quotient are also remarkably robust across specifications, excluding minor differences in significance for a handful of cases particularly when the composite proximity measure is used in lieu of individual Marshallian metrics. This is especially encouraging considering that these effects are consistently identified even when considering facility expansions. The strong interplay between social proximity and own industry concentration seems to persist regardless of establishment type.

The interaction effects for the Marshallian metrics and composite proximity metric are also quite robust, excluding facility expansions estimates. For single-unit establishments, the interactions for labor market pooling continue to be strong for labor intensive industries, and the related variety effect is consistent observed for FIRE & business services. The interaction terms for the Marshallian metrics lose most of their significance when considering facility expansions. This is especially pronounced for labor market pooling in the mining and utilities, construction, and manufacturing sectors, for which the interaction effects are strongest in the main estimations. Accordingly, the strength of the interaction effects for composite proximity are also smaller in magnitude for nearly all sectors when considering facility expansions. This suggests that in general, social forces are less of a factor in determining the location of new establishments that are part of an existing firm. This most likely reflects the fact that these new businesses mainly consist of new branches, offices, facilities, or franchise establishments of existing firms, for which location decisions are mainly made based on the central firm's needs as well as local business conditions.

6. Summary and conclusions

Explanations regarding the spatial variations in entrepreneurship are of particular importance in that the geography of entrepreneurship tends to drive the geography of growth and development (Andersson and Larsson, 2016). In this study, I modeled entrepreneurship at the industry-MSA level as a function of social capital, agglomeration, and their interactions. Critically, I utilized a structural approach to measure social capital that captured regional levels of social proximity by applying the EG coagglomeration index (Ellison et al., 2010) to the context of associational activity. Relatedly, I was able to construct a time varying metric of social capital that also sidestepped the 'small number problem' by taking advantage of a unique dataset that documents nearly all identifiable nonprofit organizations in the United States.

Most broadly, this study finds strong support for the positive association between 'related variety' (Boschma et al., 2012) and entrepreneurship. The positive association between agglomeration and entrepreneurship is made stronger in regions possessing a *blend* of bonding and bridging social capital across different social dimensions. I find that the interactions between social proximity and own industry concentration in particular are strong and remarkably robust across different specifications, even when considering facility expansions. This aligns well with theories of clusters, where one of the main benefits of own industry concentration is in the continuous monitoring and comparing that takes place among like firms which creates incentives for innovation and differentiation (Porter, 1998a). Such mechanisms are most likely strongly influenced by social forces in both civic and knowledge seeking activity. However, I find that the main effects of social proximity are neither positive nor particularly significant when including MSA fixed effects: if at all the association seems to be generally negative. This contrasts many cross-sectional studies which identify strong positive effects for social proximity, implying that one should be cautious of omitted factors when employing across-variation to measure social effects.

The interaction effects between social proximity and the Marshallian metrics are comparatively limited in scope to select industry sectors. The interaction effects for labor market pooling are quite strong, although these effects are generally confined to labor intensive industries including mining, utilities, construction, and manufacturing. For these sectors, bonding social capital in the civic sphere and bridging social capital in knowledge seeking activity augments the benefits of thick labor markets, consistent with social capital theory. The interaction effects for customer-supplier linkages and knowledge spillovers are not particularly significant, excluding the FIRE and business services sector where a related variety effect is identified. Overall, the Marshallian metrics perform poorly for non-manufacturing industries, limiting cross-sector comparisons of the interaction effects. Using the composite proximity measure – calculated using the EG coagglomeration index for industry employment – in lieu of these metrics reveals the related variety effect between social proximity and inter-industry agglomeration for sectors such as services, trade, and transportation where Marshallian forces are less significant.

Overall, the estimations identify a consistent phenomenon where agglomeration's benefits are mediated by negative interactions with social proximity, indicating the importance of social *distance* (i.e. bridging social capital). In no case are the effects of social proximity – either direct or through agglomeration effects – uniformly positive. The 'buzz-and-pipeline' model in cluster theory notes similar tensions in the optimal development of industry concentrations (Bathelt et al., 2004), while the social capital literature also refers to the downsides of excessive social proximity as the 'dark side of social capital' (Portes, 1998). Regardless of terminology, the results caution against the overzealous embracing of social proximity as an all positive force. As Portes (2014) writes, the negative effects of excessive social proximity 'are felt not only by members of the group, but by the entire society'. This study contributes to the literature by identifying this tension empirically for entrepreneurship across a broad range of industries.

One limitation of this study is the arguably narrow definition of social capital, which only encompasses social forces as they manifest in associational activity. The use of nonprofit data in this study mainly responds to data limitations. While this approach allows for the measurement of statistics that are quite close to population parameters, nonetheless organization level data most likely masks individual level social interaction mechanisms. In this regard, future research may take advantage of recent advances in big data analysis along with datasets that are increasingly more granular and broad in scope, which present promising opportunities for identifying individual level social interaction mechanisms (Bailey et al., 2018).

Another limitation is the rather crude measurement of Marshallian forces. The relatively weak effects of these metrics in non-manufacturing industries may in fact be due to how inter-industry linkages are defined. As Glaeser and Kerr (2009) also note, this is especially the case for knowledge spillovers, which are notoriously difficult to measure. The utilization of patent citations captures a very limited aspect of inter-industry knowledge flows that only consider the spillover of knowledge at the highest levels. In addition, the mapping of technologies to industries is also imperfect, relying on probabilistic schemes. Hopefully, future agglomeration theory will outline better concepts and how to best measure them empirically using new data sources.

A possible topic for further research is the empirical investigation of how the interplay between social proximity and agglomeration varies according to distance. Even with the large nonprofit dataset, the thinness of data (particularly for small rural counties) prohibits construction of reliable social proximity metrics defined at geographic delineations finer than the MSA level. Nevertheless, provided the availability of more detailed datasets and utilization of alternative social proximity metrics, a meaningful exercise would be to test to what degree the effects of social proximity – both direct and indirect – observe distance decay. This would be especially interesting considering past research identifying a sharp distance decay effect for knowledge spillovers (Figueiredo et al., 2015), as well as studies that identify differing geographic extents of agglomerative externalities across industries (Rosenthal and Strange, 2003).

Appendix

Table A1. Pairwise correlations

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) $Social\ proximity_{rt}^{C-type}$	1.000													
(2) $Social\ proximity_{rt}^{K-type}$	0.032	1.000												
(3) $HHI_{rt}^{nonprofits}$	-0.088	0.075	1.000											
(4) $Nonprofits\ per\ 1000_{rt}$	0.131	-0.192	0.048	1.000										
(5) LQ_{irt}	-0.009	0.006	0.007	-0.009	1.000									
(6) $IO\ proximity_{irt}$	0.043	-0.029	-0.037	0.025	0.040	1.000								
(7) $Labor\ proximity_{irt}$	-0.004	0.037	0.012	-0.057	0.025	-0.007	1.000							
(8) $Knowledge\ proximity_{irt}$	0.008	-0.048	-0.032	0.036	0.023	0.183	0.060	1.000						
(9) $Composite\ proximity_{irt}$	-0.013	0.013	0.010	-0.018	0.157	0.086	0.142	0.071	1.000					
(10) $HHI_{rt}^{industries}$	0.045	0.092	0.102	-0.040	0.005	-0.048	-0.059	-0.003	0.001	1.000				
(11) $\ln population_{rt}$	0.029	-0.277	-0.110	-0.092	-0.018	0.136	0.001	-0.013	-0.025	-0.252	1.000			
(12) $Patents\ per\ 1000_{rt}$	0.108	-0.146	-0.130	0.139	-0.009	0.098	-0.021	0.165	-0.022	0.061	0.193	1.000		
(13) $Expansions_{irt}$	-0.006	-0.059	-0.023	0.006	0.105	0.158	0.049	0.039	0.052	-0.058	0.165	0.031	1.000	
(14) $BDUM_{ir}$	0.017	-0.093	-0.050	-0.010	0.019	0.259	0.090	0.074	-0.035	-0.096	0.326	0.073	0.242	1.000

Table A2. Poisson quasi-maximum likelihood (PQML) estimation results: variations by sector, excluding recession years

	Mining & utilities		Construction		Manufacturing		Trade & transportation		FIRE & business services		Other services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Social\ proximity_{rt}^{C-type}$	-0.038	-0.065	-0.035**	-0.014	-0.048**	-0.017	-0.026	-0.010	-0.003	-0.015	-0.043***	-0.020**
	(0.049)	(0.050)	(0.017)	(0.014)	(0.022)	(0.021)	(0.018)	(0.015)	(0.011)	(0.012)	(0.010)	(0.009)
$Social\ proximity_{rt}^{K-type}$	0.011	0.126*	-0.008	-0.020*	0.054***	0.001	0.017	-0.005	-0.015*	-0.001	-0.003	-0.014*
	(0.064)	(0.068)	(0.011)	(0.010)	(0.021)	(0.017)	(0.011)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
LQ_{irt}	0.226***	0.207***	0.201***	0.158***	0.157***	0.122***	0.089***	0.081***	0.164***	0.152***	0.171***	0.122***
	(0.050)	(0.045)	(0.022)	(0.021)	(0.016)	(0.010)	(0.009)	(0.008)	(0.011)	(0.011)	(0.018)	(0.014)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$	0.043	0.048	-0.034***	-0.029***	0.031***	0.017***	0.075***	0.018	0.032***	0.031***	0.018**	0.014
	(0.041)	(0.046)	(0.007)	(0.006)	(0.006)	(0.006)	(0.020)	(0.012)	(0.007)	(0.008)	(0.008)	(0.009)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$	-0.011	0.039*	-0.058***	-0.033**	-0.064***	-0.047***	-0.030***	-0.003	-0.073***	-0.066***	-0.033**	-0.013
	(0.019)	(0.021)	(0.021)	(0.013)	(0.010)	(0.007)	(0.012)	(0.008)	(0.008)	(0.008)	(0.014)	(0.011)
$IO_{proximity\ irt}$	0.019		0.032**		0.136***		-0.000		0.077***		0.120***	
	(0.017)		(0.014)		(0.016)		(0.036)		(0.012)		(0.016)	
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$	-0.018		-0.026***		-0.017		0.042***		-0.015***		-0.000	
	(0.018)		(0.007)		(0.016)		(0.014)		(0.005)		(0.012)	
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$	0.000		-0.012		-0.008***		-0.011		0.020**		0.008	
	(0.018)		(0.011)		(0.002)		(0.009)		(0.008)		(0.009)	
$Labor\ proximity_{irt}$	0.311*		0.143***		0.439***		0.318***		0.099***		-0.072**	
	(0.162)		(0.036)		(0.060)		(0.053)		(0.027)		(0.030)	
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$	0.119**		0.044***		0.094***		-0.033*		0.027***		0.011	
	(0.055)		(0.008)		(0.027)		(0.019)		(0.007)		(0.010)	
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$	-0.148**		-0.028***		-0.099***		-0.035**		-0.001		0.003	
	(0.068)		(0.010)		(0.025)		(0.016)		(0.008)		(0.008)	
$Knowledge\ proximity_{irt}$	-0.007		0.021**		0.045***		0.020***		0.035**		-0.021	
	(0.063)		(0.010)		(0.016)		(0.007)		(0.016)		(0.018)	
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$	-0.010		-0.007		-0.006		-0.024***		-0.026***		0.001	
	(0.034)		(0.009)		(0.018)		(0.005)		(0.007)		(0.009)	
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$	0.044		0.000		0.019*		0.000		0.029***		0.008	
	(0.052)		(0.010)		(0.010)		(0.009)		(0.009)		(0.009)	
$Composite\ proximity_{irt}$		0.284***		0.175***		0.273***		0.296***		0.174***		0.214***
		(0.039)		(0.023)		(0.009)		(0.017)		(0.031)		(0.011)
$Social\ proximity_{rt}^{C-type} \times Composite_{proximity\ irt}$		0.014		0.036***		0.043***		-0.046***		-0.034*		-0.048***
		(0.020)		(0.011)		(0.012)		(0.012)		(0.019)		(0.017)
$Social\ proximity_{rt}^{K-type} \times Composite_{proximity\ irt}$		-0.141***		-0.012		-0.007		-0.035***		0.056***		0.033**
		(0.031)		(0.017)		(0.013)		(0.013)		(0.020)		(0.013)
Log pseudo-likelihood	-9,082	-8,815	-56,644	-55,802	-97,972	-93,257	-232,259	-224,465	-170,941	-172,219	-250,825	-244,396
Observations	17,424	17,424	21,780	21,780	185,130	185,130	158,994	158,994	82,764	82,764	143,748	143,748

Notes: See Table 5. Estimations exclude years $t = 2007$, 2008, and 2009

Table A3. Poisson quasi-maximum likelihood (PQML) estimation results: variations by sector, EG_{ab} using county employment

	Mining & utilities		Construction		Manufacturing		Trade & transportation		FIRE & business services		Other services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Social\ proximity_{rt}^{C-type}$	-0.084*	-0.074	-0.001	0.013	-0.038	-0.004	-0.011	0.004	0.009	-0.011	-0.037***	-0.013
	(0.048)	(0.047)	(0.014)	(0.013)	(0.023)	(0.019)	(0.015)	(0.014)	(0.011)	(0.011)	(0.009)	(0.010)
$Social\ proximity_{rt}^{K-type}$	0.082	0.226***	0.006	-0.011	0.065***	0.004	0.017	-0.006	-0.005	0.008	-0.003	-0.019**
	(0.060)	(0.061)	(0.010)	(0.009)	(0.020)	(0.014)	(0.012)	(0.007)	(0.008)	(0.008)	(0.006)	(0.008)
LQ_{irt}	0.234***	0.207***	0.311***	0.229***	0.172***	0.131***	0.097***	0.078***	0.162***	0.150***	0.157***	0.114***
	(0.049)	(0.043)	(0.034)	(0.034)	(0.019)	(0.009)	(0.008)	(0.008)	(0.010)	(0.010)	(0.015)	(0.010)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$	0.052	0.064*	-0.096***	-0.070***	0.031***	0.018**	0.077***	0.031*	0.018**	0.017**	0.027***	0.023**
	(0.035)	(0.035)	(0.012)	(0.012)	(0.007)	(0.007)	(0.016)	(0.017)	(0.008)	(0.009)	(0.008)	(0.009)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$	-0.024	0.038*	-0.052***	-0.034***	-0.067***	-0.047***	-0.005	-0.002	-0.081***	-0.073***	-0.066***	-0.040***
	(0.023)	(0.023)	(0.013)	(0.010)	(0.012)	(0.006)	(0.008)	(0.008)	(0.010)	(0.010)	(0.011)	(0.007)
$IO_{proximity\ irt}$	0.028		0.017		0.137***		-0.004		0.063***		0.117***	
	(0.023)		(0.012)		(0.020)		(0.042)		(0.012)		(0.017)	
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$	0.015		-0.029***		-0.007		0.026**		-0.010**		0.005	
	(0.018)		(0.007)		(0.010)		(0.012)		(0.005)		(0.008)	
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$	-0.006		0.000		-0.005		-0.010		0.021**		0.019*	
	(0.016)		(0.011)		(0.003)		(0.010)		(0.008)		(0.010)	
$Labor\ proximity_{irt}$	0.397**		0.141***		0.424***		0.316***		0.097***		-0.089***	
	(0.166)		(0.037)		(0.063)		(0.059)		(0.028)		(0.034)	
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$	0.116**		0.047***		0.083***		-0.019		0.025***		0.009	
	(0.050)		(0.009)		(0.018)		(0.017)		(0.007)		(0.008)	
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$	-0.132**		-0.040***		-0.126***		-0.041**		-0.002		0.001	
	(0.064)		(0.011)		(0.027)		(0.019)		(0.008)		(0.009)	
$Knowledge\ proximity_{irt}$	-0.031		0.023**		0.049***		0.029***		0.037**		-0.017	
	(0.065)		(0.009)		(0.013)		(0.009)		(0.016)		(0.018)	
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$	0.034		0.001		-0.001		-0.028***		-0.036***		-0.007	
	(0.038)		(0.009)		(0.012)		(0.009)		(0.008)		(0.008)	
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$	0.012		-0.002		0.016		0.005		0.035***		0.014	
	(0.048)		(0.009)		(0.013)		(0.010)		(0.009)		(0.009)	
$Composite\ proximity_{irt}$		0.325***		0.174***		0.273***		0.281***		0.167***		0.221***
		(0.041)		(0.022)		(0.009)		(0.016)		(0.025)		(0.012)
$Social\ proximity_{rt}^{C-type} \times Composite_{proximity\ irt}$		0.008		0.048***		0.055***		-0.058***		-0.014		-0.051***
		(0.018)		(0.012)		(0.021)		(0.017)		(0.018)		(0.019)
$Social\ proximity_{rt}^{K-type} \times Composite_{proximity\ irt}$		-0.155***		-0.022		-0.003		-0.021		0.074***		0.051***
		(0.025)		(0.016)		(0.012)		(0.014)		(0.020)		(0.016)
Log pseudo-likelihood	-13,335	-12,890	-82,430	-81,331	-143,306	-136,708	-342,048	-330,371	-251,278	-252,763	-370,879	-362,112
Observations	26,136	26,136	32,670	32,670	277,695	277,695	238,491	238,491	124,146	124,146	215,622	251,622

Notes: See Table 5. Estimations utilize social proximity measures calculated using the EG coagglomeration index for nonprofits defined at the county level.

Table A4. Poisson quasi-maximum likelihood (PQML) estimation results: variations by sector, facility expansions

	Mining & utilities		Construction		Manufacturing		Trade & transportation		FIRE & business services		Other services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Social\ proximity_{rt}^{C-type}$	-0.012	0.022	0.018	-0.005	-0.067**	-0.062*	-0.003	0.006	-0.033	-0.033	-0.051***	-0.027
	(0.060)	(0.055)	(0.041)	(0.041)	(0.034)	(0.034)	(0.013)	(0.012)	(0.021)	(0.021)	(0.017)	(0.017)
$Social\ proximity_{rt}^{K-type}$	-0.018	-0.013	0.018	-0.008	0.059	0.044	0.025**	0.019	-0.004	-0.005	-0.004	-0.036**
	(0.065)	(0.060)	(0.051)	(0.050)	(0.040)	(0.041)	(0.013)	(0.012)	(0.014)	(0.015)	(0.016)	(0.018)
LQ_{irt}	0.301***	0.250***	0.214***	0.193***	0.166***	0.148***	0.125***	0.093***	0.142***	0.135***	0.140***	0.118***
	(0.042)	(0.032)	(0.035)	(0.033)	(0.012)	(0.012)	(0.009)	(0.009)	(0.014)	(0.013)	(0.008)	(0.008)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$	0.052	0.025	-0.048***	-0.045**	0.035***	0.027***	0.030***	0.037***	0.055***	0.049***	0.036***	0.002
	(0.034)	(0.033)	(0.017)	(0.018)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)	(0.015)	(0.009)	(0.009)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$	-0.056**	0.010	-0.059	-0.067	-0.064***	-0.062***	-0.027***	-0.042***	-0.088***	-0.079***	-0.087***	-0.063***
	(0.028)	(0.023)	(0.051)	(0.060)	(0.008)	(0.010)	(0.007)	(0.010)	(0.010)	(0.012)	(0.011)	(0.010)
$IO_{proximity\ irt}$	0.005		-0.061		0.180***		0.167***		0.054***		0.164***	
	(0.035)		(0.053)		(0.023)		(0.014)		(0.011)		(0.018)	
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$	-0.035		-0.004		0.008		0.004		-0.005		-0.035***	
	(0.024)		(0.031)		(0.013)		(0.006)		(0.006)		(0.011)	
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$	0.042**		-0.093*		-0.006**		-0.005		-0.007		0.008	
	(0.020)		(0.053)		(0.003)		(0.006)		(0.008)		(0.010)	
$Labor\ proximity_{irt}$	0.022		0.455***		0.487***		0.319***		0.048		0.093*	
	(0.194)		(0.083)		(0.082)		(0.024)		(0.033)		(0.049)	
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$	0.090*		-0.018		0.051*		0.013		-0.022*		0.045***	
	(0.047)		(0.024)		(0.027)		(0.011)		(0.013)		(0.011)	
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$	-0.099*		-0.039		-0.037		-0.034***		0.012		-0.032***	
	(0.060)		(0.033)		(0.026)		(0.012)		(0.009)		(0.010)	
$Knowledge\ proximity_{irt}$	-0.028		0.052***		0.102***		0.005		0.063***		0.020	
	(0.050)		(0.020)		(0.019)		(0.007)		(0.015)		(0.015)	
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$	0.022		-0.012		-0.003		-0.029***		-0.044***		-0.011	
	(0.039)		(0.021)		(0.015)		(0.006)		(0.014)		(0.012)	
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$	-0.019		0.067**		0.028***		0.008		-0.004		0.017	
	(0.047)		(0.026)		(0.008)		(0.007)		(0.011)		(0.011)	
$Composite\ proximity_{irt}$		0.344***		0.065***		0.464***		0.260***		0.096***		0.375***
		(0.038)		(0.021)		(0.032)		(0.011)		(0.014)		(0.035)
$Social\ proximity_{rt}^{C-type} \times Composite_{proximity\ irt}$		0.014		-0.018		0.009		-0.028***		-0.008		-0.006
		(0.018)		(0.018)		(0.025)		(0.011)		(0.014)		(0.023)
$Social\ proximity_{rt}^{K-type} \times Composite_{proximity\ irt}$		-0.083***		0.033		-0.014		-0.046***		0.023		0.069***
		(0.021)		(0.022)		(0.023)		(0.010)		(0.015)		(0.024)
Log pseudo-likelihood	-13,085	-12,765	-15,553	-15,581	-40,798	-40,460	-221,138	-219,964	-222,255	-222,418	-247,341	-244,837
Observations	26,136	26,136	32,670	32,670	277,695	277,695	238,491	238,491	124,146	124,146	215,622	215,622

Notes: See Table 5. Estimations utilize the count of new establishments that are part of an existing firm as the dependent variable.

Table A5. Zero-inflated Poisson estimation results: variations by sector

	Mining & utilities		Construction		Manufacturing		Trade & transportation		FIRE & business services		Other services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Social\ proximity_{rt}^{C-type}$	-0.078	-0.115**	-0.033**	-0.009	-0.059***	-0.020	-0.035*	-0.007	-0.004	-0.017*	-0.046***	-0.018**
	(0.059)	(0.052)	(0.015)	(0.011)	(0.022)	(0.021)	(0.018)	(0.014)	(0.010)	(0.010)	(0.009)	(0.009)
$Social\ proximity_{rt}^{K-type}$	0.004	0.187**	0.001	-0.013	0.052***	0.028	0.024**	-0.001	-0.013*	0.002	0.001	-0.012*
	(0.078)	(0.086)	(0.009)	(0.008)	(0.020)	(0.018)	(0.012)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
LQ_{irt}	0.222***	0.201***	0.378***	0.273***	0.166***	0.132***	0.097***	0.086***	0.181***	0.167***	0.183***	0.134***
	(0.051)	(0.043)	(0.043)	(0.040)	(0.018)	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)	(0.019)	(0.016)
$Social\ proximity_{rt}^{C-type} \times LQ_{irt}$	0.045	0.039	-0.080***	-0.058***	0.054**	0.034**	0.078***	0.029**	0.017**	0.020***	0.026**	0.019
	(0.047)	(0.040)	(0.010)	(0.010)	(0.021)	(0.015)	(0.022)	(0.011)	(0.007)	(0.008)	(0.011)	(0.013)
$Social\ proximity_{rt}^{K-type} \times LQ_{irt}$	-0.002	0.035*	-0.036***	-0.026***	-0.072***	-0.054***	-0.037***	-0.017**	-0.076***	-0.069***	-0.047**	-0.018
	(0.020)	(0.020)	(0.008)	(0.008)	(0.015)	(0.009)	(0.010)	(0.007)	(0.007)	(0.008)	(0.020)	(0.014)
$IO_{proximity\ irt}$	0.009		0.024**		0.166***		-0.014		0.066***		0.112***	
	(0.015)		(0.012)		(0.015)		(0.040)		(0.012)		(0.016)	
$Social\ proximity_{rt}^{C-type} \times IO\ proximity_{irt}$	-0.009		-0.025***		-0.018		0.046***		-0.012***		0.002	
	(0.021)		(0.006)		(0.016)		(0.014)		(0.004)		(0.012)	
$Social\ proximity_{rt}^{K-type} \times IO\ proximity_{irt}$	0.004		-0.006		-0.040***		-0.012		0.018**		0.004	
	(0.014)		(0.010)		(0.007)		(0.009)		(0.007)		(0.008)	
$Labor\ proximity_{irt}$	0.349**		0.110***		0.318***		0.292***		0.086***		-0.096***	
	(0.171)		(0.035)		(0.061)		(0.055)		(0.027)		(0.031)	
$Social\ proximity_{rt}^{C-type} \times Labor\ proximity_{irt}$	0.106**		0.048***		0.082***		-0.029		0.027***		0.008	
	(0.053)		(0.008)		(0.027)		(0.019)		(0.006)		(0.010)	
$Social\ proximity_{rt}^{K-type} \times Labor\ proximity_{irt}$	-0.086		-0.039***		-0.079***		-0.040**		0.001		0.006	
	(0.065)		(0.009)		(0.019)		(0.016)		(0.007)		(0.008)	
$Knowledge\ proximity_{irt}$	0.002		0.024***		0.040***		0.032***		0.035**		-0.033*	
	(0.061)		(0.009)		(0.015)		(0.012)		(0.015)		(0.019)	
$Social\ proximity_{rt}^{C-type} \times Knowledge\ proximity_{irt}$	-0.003		-0.007		-0.005		-0.039***		-0.027***		0.004	
	(0.031)		(0.009)		(0.016)		(0.013)		(0.007)		(0.009)	
$Social\ proximity_{rt}^{K-type} \times Knowledge\ proximity_{irt}$	0.051		-0.002		0.026***		0.002		0.029***		0.003	
	(0.050)		(0.009)		(0.009)		(0.010)		(0.009)		(0.009)	
$Composite\ proximity_{irt}$		0.245***		0.167***		0.275***		0.307***		0.164***		0.207***
		(0.040)		(0.024)		(0.010)		(0.017)		(0.032)		(0.013)
$Social\ proximity_{rt}^{C-type} \times Composite_{proximity\ irt}$		0.024		0.038***		0.040***		-0.056***		-0.028		-0.052***
		(0.018)		(0.011)		(0.012)		(0.012)		(0.018)		(0.018)
$Social\ proximity_{rt}^{K-type} \times Composite_{proximity\ irt}$		-0.117***		-0.016		-0.002		-0.031**		0.056***		0.029**
		(0.027)		(0.017)		(0.014)		(0.013)		(0.020)		(0.014)
Log pseudo-likelihood	-12,927	-12,676	-82,165	-81,234	-141,920	-136,344	-338,046	-328,520	-250,324	-251,842	-368,017	-360,676
Observations	26,136	26,136	32,670	32,670	277,695	277,695	238,491	238,491	124,146	124,146	215,622	215,622

Notes: See Table 5. Estimations utilize a zero inflated Poisson model with a logit link for the inflation model. Only estimates for the count model are reported for brevity.

References

Acs, Z.J., Armington, C. (2006) *Entrepreneurship, geography, and American economic growth*. Cambridge University Press Cambridge.

Agrawal, A., Cockburn, I., McHale, J. (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*. 6: 571–591.

Alesina, A., La Ferrara, E. (2000) Participation in heterogeneous communities. *The Quarterly Journal of Economics*. 115: 847–904.

Andersson, M., Larsson, J.P. (2016) Local entrepreneurship clusters in cities. *Journal of Economic Geography*. 16: 39–66.

Arrow, K.J. (2000) Observations on social capital. In P. Dasgupta & I. Serageldin, eds. *Social capital: a multifaceted perspective*. Washington, DC, DC: The World Bank, pp. 3–5.

Audia, P.G., Freeman, J.H., Reynolds, P.D. (2006) Organizational foundings in community context: Instruments manufacturers and their interrelationship with other organizations. *Administrative Science Quarterly*. 51: 381–419.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A. (2018) Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*. 32: 259–280.

Bathelt, H., Malmberg, A., Maskell, P. (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*. 28: 31–56.

Boschma, R., Minondo, A., Navarro, M. (2012) Related variety and regional growth in Spain. *Papers in Regional Science*. 91: 241–256.

Bourdieu, P. (1986) The forms of capital. In J. G. Richardson, ed. *Handbook of theory and research for the sociology of education*. New York, NY: Greenwood, pp. 241–258.

Broekel, T., Boschma, R. (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*. 12: 409–433.

Bürker, M., Minerva, G.A. (2014) Civic capital and the size distribution of plants: Short-run dynamics and long-run equilibrium. *Journal of Economic Geography*. 14: 797–847.

Burt, R.S. (2005) Brokerage and closure : an introduction to social capital.

Cameron, A.C., Trivedi, P.K. (2013) *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press.

Capello, R. (2009) Spatial spillovers and regional growth: a cognitive approach. *European Planning Studies*. 17: 639–658.

Caragliu, A., Nijkamp, P. (2016) Space and knowledge spillovers in European regions: the impact of different forms of proximity on spatial knowledge diffusion. *Journal of Economic Geography*. 16: 749–774.

Choi, E.J., Spletzer, J.R. (2012) The declining average size of establishments: evidence and explanations. *Monthly Labor Review*. 135: 50.

Coleman, J.S. (1988) Social capital in the creation of human capital. *American Journal of Sociology*. 94: S95–S120.

Combes, P.-P., Duranton, G. (2006) Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*. 36: 1–28.

Currid, E., Williams, S. (2010) The geography of buzz: art, culture and the social milieu in Los Angeles and New York. *Journal of Economic Geography*. 10: 423–451.

Dahl, M.S., Sorenson, O. (2009) The embedded entrepreneur. *European Management Review*. 6: 172–181.

Davidsson, P., Honig, B. (2003) The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*. 18: 301–331.

Delgado, M., Porter, M.E., Stern, S. (2014) Clusters, convergence, and economic performance. *Research Policy*. 43: 1785–1799.

Delgado, M., Porter, M.E., Stern, S. (2010) Clusters and entrepreneurship. *Journal of Economic Geography*. 10: 495–518.

Durlauf, S.N. (2002) On The Empirics Of Social Capital. *The Economic Journal*. 112: F459–F479.

Ellison, G., Glaeser, E.L. (1997) Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of political economy*. 105: 889–927.

Ellison, G., Glaeser, E.L., Kerr, W.R. (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. *The American Economic Review*. 100: 1195–1213.

Figueiredo, O., Guimarães, P., Woodward, D. (2015) Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail. *Journal of Urban Economics*. 89: 21–31.

Fruchterman, T.M.J., Reingold, E.M. (1991) Graph drawing by force-directed placement. *Software: Practice and Experience*. 21: 1129–1164.

Fukuyama, F. (1995) *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Ghani, E., Kerr, W.R., O'Connell, S. (2014) Spatial determinants of entrepreneurship in india. *Regional Studies*. 48: 1071–1089.

Glaeser, E.L., Gottlieb, J.D. (2009) The wealth of cities: agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*. 47: 983–1028.

Glaeser, E.L., Kerr, S.P., Kerr, W.R. (2014) Entrepreneurship and urban growth: an empirical assessment with historical mines. *The Review of Economics and Statistics*. 97: 498–520.

Glaeser, E.L., Kerr, W.R. (2009) Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*. 18: 623–663.

Glaeser, E.L., Kerr, W.R., Ponzetto, G.A.M. (2010) Clusters of entrepreneurship. *Journal of Urban Economics*. 67: 150–168.

Glückler, J. (2007) Economic geography and the evolution of networks. *Journal of Economic Geography*. 7: 619–634.

Goldschlag, N., Lybbert, T.J., Zolas, N.J. (2016) *An 'Algorithmic Links with Probabilities' Crosswalk for USPC and CPC Patent Classifications with an Application Towards Industrial Technology Composition*. US Census Bureau CES Working Paper No. 16-15.

Granovetter, M. (2005) The impact of social structure on economic outcomes. *The Journal of Economic Perspectives*. 19: 33–50.

Granovetter, M. (1973) The strength of weak ties. *American Journal of Sociology*. 78: 1360–1380.

Hall, B.H., Jaffe, A.B., Trajtenberg, M. (2001) *The NBER patent citation data file: lessons, insights and methodological tools*. NBER Working Paper No. 8498.

Helsley, R.W., Strange, W.C. (1990) Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*. 20: 189–212.

Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R. (2007) The product space conditions the development of nations. *Science*. 317: 482–487.

Ioannides, Y.M. (2013) *From Neighborhoods to Nations: The Economics of Social Interactions*. Princeton University Press.

Isserman, A.M., Westervelt, J. (2006) 1.5 million missing numbers: Overcoming employment suppression in County Business Patterns data. *International Regional Science Review*. 29: 311–335.

Iyer, S., Kitson, M., Toh, B. (2005) Social capital, economic growth and regional development. *Regional Studies*. 39: 1015–1040.

Jacobs, J. (1969) *The Economy of Cities*. New York: Vintage.

Jacomy, M., Venturini, T., Heymann, S., Bastian, M. (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS One*. 9: e98679.

Jaffe, A.B., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 577–598.

Jofre-Monseny, J., Marín-López, R., Viladecans-Marsal, E. (2011) The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*. 70: 61–74.

Knack, S., Keefer, P. (1995) Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*. 7: 207–227.

Krugman, P. (1991) *Geography and Trade*. Cambridge, MA: MIT Press.

Lin, N. (2001) *Social capital: A theory of social structure and action*. Cambridge, UK: Cambridge Univ. Press.

Malecki, E.J. (2012) Regional Social Capital: Why it Matters. *Regional Studies*. 46: 1023–1039.

Manski, C.F. (1993) Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*. 60: 531–542.

Marshall, A. (1919) *Industry and Trade*. London: Macmillan.

Marshall, A. (1890) *Principles of Economics*. London: MacMillan.

McPherson, M., Smith-Lovin, L., Cook, J.M. (2001) Birds of a feather: homophily in social networks. *Annual Review of Sociology* 415–444.

Nyström, K. (2007) An industry disaggregated analysis of the determinants of regional entry and exit. *The Annals of Regional Science*. 41: 877–896.

Porter, M.E. (1998a) Clusters and the new economics of competition. *Harvard Business Review*. 76: 77–90.

Porter, M.E. (1998b) Location, Clusters, and the New Microeconomics of Competition. *Business Economics* 7–13.

Porter, M.E. (2003) The economic performance of regions. *Regional studies*. 37: 549–578.

Portes, A. (2014) Downsides of social capital. *Proceedings of the National Academy of Sciences*. 111: 18407–18408.

Portes, A. (1998) Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*. 24: 25.

Putnam, R.D. (2001) *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Putnam, R.D., Leonardi, R., Nanetti, R.Y. (1993) *Making democracy work: Civic traditions in modern Italy*. Princeton university press.

Rosenthal, S.S., Strange, W.C. (2003) Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*. 85: 377–393.

Rosenthal, S.S., Strange, W.C. (2001) The Determinants of Agglomeration. *Journal of Urban Economics*. 50: 191–229.

Rupasingha, A., Goetz, S.J., Freshwater, D. (2006) The production of social capital in US counties. *The Journal of Socio-Economics*. 35: 83–101.

Safford, S. (2009) *Why the garden club couldn't save Youngstown: The transformation of the rust belt*. Harvard university press.

Samila, S., Sorenson, O. (2017) Community and capital in entrepreneurship and economic growth. *American Sociological Review*. 82: 770–795.

Saxenian, A. (1996) *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Sobel, J. (2002) Can We Trust Social Capital? *Journal of Economic Literature*. 40: 139–154.

Sorenson, O., Audia, P.G. (2000) The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940–1989. *American Journal of Sociology*. 106: 424–462.

de Souza Briggs, X. (2004) Social capital: easy beauty or meaningful resource? *Journal of the American planning Association*. 70: 151–158.

Storper, M. (1995) Competitiveness policy options: the technology-regions connection. *Growth and Change*. 26: 285–308.

Storper, M. (2005) Society, community, and economic development. *Studies in Comparative International Development*. 39: 30–57.

Storper, M. (1997) *The regional world: territorial development in a global economy*. New York, London: Guilford Press.

Storper, M., Venables, A.J. (2004) Buzz: face-to-face contact and the urban economy. *Journal of economic geography*. 4: 351–370.

Stuart, T.E., Sorenson, O. (2005) Social networks and entrepreneurship. In *Handbook of entrepreneurship research*. Springer, pp. 233–252.

Stuart, T.E., Sorenson, O. (2003) The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. *Research Policy*. 32: 229–253.

Tura, T., Harmaakorpi, V. (2005) Social capital in building regional innovative capability. *Regional Studies*. 39: 1111–1125.

Walton-Roberts, M., Hiebert, D. (1997) Immigration, entrepreneurship, and the family: Indo-Canadian enterprise in the construction industry of Greater Vancouver. *Canadian Journal of Regional Science*. 20: 119–140.

Westlund, H., Larsson, J.P., Olsson, A.R. (2014) Start-ups and local entrepreneurial social capital in the municipalities of Sweden. *Regional studies*. 48: 974–994.

Wilson, K.L., Portes, A. (1980) Immigrant enclaves: an analysis of the labor market experiences of Cubans in Miami. *American Journal of Sociology* 295–319.

Woolcock, M. (2001) The place of social capital in understanding social and economic outcomes. *Canadian Journal of Policy Research*. 2: 11–17.

Woolcock, M., Narayan, D. (2000) Social capital: Implications for development theory, research, and policy. *The world bank research observer*. 15: 225–249.

Zipf, G.K. (1949) *Human behavior and the principle of least effort*. New York: Hafner.

Acknowledgment

This publication was prepared by Jae Beum Cho at Indiana University using Federal funds awarded to the Trustees of Indiana University and as a subcomponent under award number ED17HDQ3120040 from the U.S. Economic Development Administration, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the Economic Development Administration or the U.S. Department of Commerce.