I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12223

Relatedness, Complexity and Local Growth

Benjamin Davies
David C. Maré

MARCH 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12223

Relatedness, Complexity and Local Growth

Benjamin Davies
Motu Economic and Public Policy Research

David C. Maré
Motu Economic and Public Policy Research, University of Waikato and IZA

MARCH 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Relatedness, Complexity and Local Growth*

We derive a measure of the relatedness between economic activities based on weighted correlations of local employment shares, and use this measure to estimate city and activity complexity. Our approach extends discrete measures used in previous studies by recognising the extent of activities' local over-representation and by adjusting for differences in signal quality between geographic areas with different sizes. We examine the contribution of relatedness and complexity to urban employment growth, using 1981–2013 census data from New Zealand. Complex activities experienced faster employment growth during our period of study, especially in complex cities. However, this growth was not significantly stronger in cities more dense with related activities. Relatedness and complexity appear to be most relevant for analysing how large, complex cities grow, and are less informative for understanding employment dynamics in small, less complex cities.

JEL Classification: R11, R12
Keywords: relatedness, complexity, smart specialisation

Corresponding author:
David C. Maré
Motu Economic and Public Policy Research
Level 1, 97 Cuba Street
PO Box 24390
Wellington 6142
New Zealand
E-mail: dave.mare@motu.org.nz

* Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in the study are the work of the authors, not Statistics New Zealand. The research was supported by Te Pūnaha Matatini, and by the Building Better Homes, Towns and Cities National Science Challenge. The authors thank Andrew Coleman and Guanyu Zheng for early discussions, Dean Hyslop and Edward Glaeser for helpful suggestions, and seminar participants at Motu for their comments. Any remaining errors are the authors' alone.

1 Introduction

The spatial concentration of economic activities in cities generates agglomeration economies arising from labour market pooling, input sharing, and knowledge spillovers (Marshall, 1920). The principle of relatedness (Hidalgo et al., 2018; Vicente et al., 2018) suggests that the advantages of proximity may also accrue to interacting activities that are similar in ways other than spatially. Such interactions support the growth of complex activities, which rely on specialised combinations of complementary knowledge and skills (Hidalgo and Hausmann, 2009; Balland et al., 2018b). Relatedness and complexity capture "the risks and rewards of competing diversification strategies" (Balland et al., 2018a), making their connection with employment growth important to understand for regional development and innovation policy-making.

In this paper, we derive a measure of the pairwise relatedness between economic activities based on weighted correlations of local employment shares, and use this measure to estimate activity and city complexity. Our approach extends discrete measures used in previous studies (Hidalgo et al., 2007; Hidalgo and Hausmann, 2009; Balland and Rigby, 2017; Farinha Fernandes et al., 2018; Balland et al., 2018a) by recognising the extent of activities' local over-representation and by adjusting for differences in signal quality between geographic areas with different sizes. These attributes make our measure more suitable than previous measures for studying relatedness and complexity in small geographic areas, in which measurement errors and random fluctuations are proportionally large.

We examine the contribution of relatedness and complexity to urban employment growth, using 1981–2013 census data from New Zealand. These data cover a range of urban areas that are smaller than, but contain similar activities to, previously studied regions. This property allows us to investigate whether the mechanisms through which relatedness drives employment growth operate only in sufficiently large cities.

Complex activities experienced faster employment growth during our period of study, especially in complex cities. However, this growth was not significantly stronger in cities more dense with related activities. Relatedness and complexity appear to be relevant for analysing how large, complex cities grow, but offer little information about local growth trajectories in small areas. This result is consistent with the idea propagated throughout the urban economics literature that cities are dense networks of interacting activities; in our data, the benefits of such interaction are more apparent in larger cities, in which activity networks are more dense and in which people with complementary skills interact more frequently.

This paper is structured as follows. Section 2 discusses the links between related activities' interaction and employment growth. Section 3 presents our approach to estimating relatedness and complexity, and compares it with the approach used in previous studies. Sections 4 and 5 summarise our empirical findings from applying our methods to employment data from New Zealand. Section 6 concludes. Appendix A offers a brief primer on networks and graph theory, ideas from which we use throughout the paper.

2 Relatedness-driven growth

The principle of relatedness is applied extensively in studies of regional and urban growth and innovation (Hidalgo et al., 2018; Vicente et al., 2018). A dominant focus within the existing literature is on the relevance of relatedness for processes of innovation (Feldman and Audretsch, 1999; Boschma, 2005), entrepreneurship (Neffke et al., 2018) and industrial diversification (Neffke and Henning, 2013). This focus reflects the microfoundations of the relatedness literature, which emphasise knowledge spillovers and the consequent knowledge creation occurring between related knowledge bases (Asheim and Gertler, 2005).

The principle of relatedness influences current European regional policy. It provides a rationale for spatially differentiated policy approaches. Local policies are designed to be context-specific, in light of the relatedness patterns among local economic activities as well as the local institutional context (Barca et al., 2012; Boschma, 2014). The policy emphasis, as with the relatedness literature, is on innovation, entrepreneurial, and research and development processes, and the support of innovation-led growth. Such processes are the focus of smart specialisation policies (Foray et al., 2009, 2011), which encourage regions to upgrade their economic structure "by building on their existing capabilities" (Balland et al., 2018a) and which are a core component of the reformed EU Cohesion policy (Barca, 2009; McCann and Ortega-Argilés, 2015).

Balland et al. (2018a) appeal to the principle of relatedness in their framework for analysing smart specialisation. Cities more dense with related activities are more able to sustain employment growth through labour market shocks because workers can reallocate into activities that require similar knowledge and skills (Neffke and Henning, 2013; Morkutė et al., 2016). This capacity for reallocation reduces the growth risk of investing in activities related to cities' existing knowledge base vis-à-vis activities that are less locally related. The expected return on such investment is greatest when cities expand into complex activities, which "form the basis for long-run competitive advantage" (Balland et al., 2018a). Based on these arguments, Balland et al.'s (2018a) framework suggests that cities' optimal diversification strategy is to encourage employment growth within complex activities that are related to existing local competencies.

Relatedness-driven innovation processes also encourage employment growth. Innovation involves combining existing knowledge bases to produce new ideas (Schumpeter, 1934; Weitzman, 1998). Clusters of related activities promote innovation (Delgado et al., 2014) by bringing together complementary ideas (Jacobs, 1969; Feldman and Audretsch, 1999). To the extent that such innovation produces long-term economic growth, competitive forces drive employment growth in local clusters of related activities in order to capitalise on their potential to facilitate knowledge creation and spillovers (Asheim and Gertler, 2005).

Hendy and Callaghan (2013) discuss the potential benefits of relatedness-driven innovation for New Zealand. New Zealand's small size and geographic isolation mean that its labour market has less access to agglomeration economies than such markets in other developed countries (McCann, 2009). Hendy and Callaghan (2013) argue that these features of New Zealand's economic geography can be overcome by "establishing collaborative networks of researchers in areas where New Zealand could be internationally competitive," and by "encouraging the mobility of researchers between organisations and firms in the innovation ecosystem." Such efforts promote innovation by connecting researchers with complementary ideas and by facilitating knowledge flows between firms, raising New Zealand's potential to produce complex products in the "densely connected core" of Hidalgo et al.'s (2007) product space.

Hendy and Callaghan (2013) highlight the tension between building on existing strengths and expanding into more complex activities. This tension permeates current regional development and innovation policy debates in New Zealand, which discuss the merits of expanding the primary sector relative to diversifying into more knowledge-intense products and services. Our analysis informs these debates by evaluating whether historical employment dynamics in New Zealand provide ex-post evidence of relatedness-driven growth, thereby indicating the capacity for such growth in the future.

3 Measuring relatedness and complexity

3.1 Activity relatedness

Two activities are related if they require similar knowledge or inputs (Hidalgo et al., 2018). We infer such similarities from employee colocation patterns, which reveal firms' shared preferences for using knowledge and other spatially heterogeneous resources. Previous studies infer activity relatedness from worker flows (Neffke and Henning, 2013; Jara-Figueroa et al., 2018), input-output linkages and shared labour pools (Delgado et al., 2016), and patent applications (Boschma et al., 2015; Balland et al., 2018a).

We estimate activity relatedness as follows. Consider an economy that comprises a set $\mathcal{C}$ of cities and a set $\mathcal{A}$ of activities. Let E_c^a denote the number of people employed in city $c \in \mathcal{C}$ and activity $a \in \mathcal{A}$. Then total city c employment is given by the sum

$$E_c = \sum_{a\in\mathcal{A}} E_c^a,$$

while national activity a employment is equal to

$$E^a = \sum_{c\in\mathcal{C}} E_c^a.$$

Summing over all cities and activities yields national employment:

$$E = \sum_{c\in\mathcal{C}} \sum_{a\in\mathcal{A}} E_c^a.$$

Comparing the local share

$$LS_c^a = \frac{E_c^a}{E_c}$$

of activity a in city c with its share E^a/E of national employment reveals whether the activity is relatively over-represented in city c. Such over-representation reflects local specialisation in activity a relative to the national economy.

We estimate the relatedness of activities a_1 and a_2 using the correlation between the corresponding vectors $(LS_1^{a_1}, LS_2^{a_1}, \ldots, LS_{|\mathcal{C}|}^{a_1})$ and $(LS_1^{a_2}, LS_2^{a_2}, \ldots, LS_{|\mathcal{C}|}^{a_2})$ of local employment shares. This correlation is high when a_1 and a_2 are relatively over-represented in similar cities, revealing firms' tendency to colocate in pursuit of agglomeration economies (Marshall, 1920). First, we compute the weighted covariance

$$\begin{aligned}\Omega_{a_1 a_2} &= \sum_{c\in\mathcal{C}} \frac{E_c}{E}\left(LS_c^{a_1} - \sum_{c\in\mathcal{C}} \frac{E_c}{E} LS_c^{a_1}\right)\left(LS_c^{a_2} - \sum_{c\in\mathcal{C}} \frac{E_c}{E} LS_c^{a_2}\right) \\ &= \sum_{c\in\mathcal{C}} \frac{E_c}{E}\left(\frac{E_c^{a_1}}{E_c} - \frac{E^{a_1}}{E}\right)\left(\frac{E_c^{a_2}}{E_c} - \frac{E^{a_2}}{E}\right) \end{aligned} \tag{3.1}$$

in local shares for activities a_1 and a_2, where the weighting factor E_c/E is equal to city c's share of national employment. Second, we normalise $\Omega_{a_1 a_2}$ by the city-share weighted standard deviations of $(LS_1^{a_1}, LS_2^{a_1}, \ldots, LS_{|\mathcal{C}|}^{a_1})$ and $(LS_1^{a_2}, LS_2^{a_2}, \ldots, LS_{|\mathcal{C}|}^{a_2})$, yielding the weighted correlation between the local shares of activities a_1 and a_2. Finally, we map this correlation to the closed unit interval $[0, 1]$ using the linear transformation $x \mapsto (x + 1)/2$. Hence, our estimate of the relatedness between activities a_1 and a_2 is given by

$$R_{a_1 a_2} = \frac{1}{2}\left(\frac{\Omega_{a_1 a_2}}{\sqrt{\Omega_{a_1 a_1}\Omega_{a_2 a_2}}} + 1\right). \tag{3.2}$$

This estimate is largest when activities a_1 and a_2 have equal local shares in each city $c \in \mathcal{C}$, and is smallest when the percentage point difference between activity a_1's local and national shares has equal magnitude but opposite sign to that difference for a_2 in all cities. We assume that there is variation among the components of each local share vector $(LS_1^a, LS_2^a, \ldots, LS_{|\mathcal{C}|}^a)$, so that $\Omega_{aa} > 0$ for each $a \in \mathcal{A}$ and hence that (3.2) is well-defined.

3.1.1 Comparison with previously used relatedness measures

The bracketed terms in the summand of (3.1) are equal to the percentage point difference between activities' local and national employment shares, and thus measure the extent to which activities are locally over-represented. An alternative measure of local over-representation is the location quotient

$$LQ_c^a = \frac{E_c^a/E_c}{E^a/E}, \tag{3.3}$$

which exceeds unity if and only if activity a comprises a larger share of city c employment than of national employment. Following Balassa (1965), Hidalgo et al. (2007) use an analogous metric to identify the commodities in which different countries exhibit revealed comparative advantage (RCA) and infer commodities' similarity from RCA co-occurrence patterns. Boschma et al. (2015) and Balland et al. (2018a) use this RCA approach to estimate the similarity between different technologies using patent data from US cities and European regions, respectively.

Inferring activity relatedness from RCA co-occurrence patterns is problematic for at least three reasons. First, if activity a represents a small share of national employment then measurement errors in the numerator of (3.3) are exacerbated by the denominator being close to zero.

Our measure (3.2) of activity relatedness prevents this exacerbation by comparing percentage point differences in local and national shares rather than ratios of such shares.

Second, RCA co-occurrence patterns ignore the extent to which activities are locally over-represented and are sensitive to small pertubations in employment in cities with location quotients near unity. To see why, consider the indicator variable

$$\mathrm{RCA}_c^a = \begin{cases} 1 & \text{if } LQ_c^a \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

for the event in which city c has revealed comparative advantage in activity a. This variable is constant with respect to LQ_c^a on the intervals $[0, 1)$ and $(1, \infty)$, and is discontinuous at $LQ_c^a = 1$ and, therefore, in E_c^a. In contrast, our relatedness measure (3.2) recognises different extents of local over-representation and varies continuously with local activity employment.

Third, the RCA approach is sensitive to external influence and noise in employment counts within small cities. If an activity exits a small city then all other activities in that city are likely to become over-represented relative to the national average because the increase in their local shares will be proportionally larger than any change in national shares. Thus, the identification of activities in which small cities appear to be specialised is sensitive to internal migration and to fluctuations in local employment in other activities. Activity specialisations in large cities are less noisy because local shares are less sensitive to absolute fluctuations in local employment. The RCA approach does not recongise differences in signal quality between cities of different size. In contrast, our relatedness measure (3.2) is more robust to noise induced by small cities because it gives such cities less weight than large cities, where the relationship between local specialisations and local activity employment is more stable.

3.1.2 Mean local relatedness and relatedness density

Observe from (3.2) that $R_{aa} = 1$ for all activities $a \in \mathcal{A}$. Therefore, the local share-weighted mean relatedness of activity a with the activities in city c can be written as

$$\sum_{a' \in \mathcal{A}} \frac{E_c^{a'}}{E_c} R_{aa'} = LS_c^a + RD_c^a, \tag{3.5}$$

where we define

$$RD_c^a = \sum_{a' \in \mathcal{A} \setminus \{a\}} \frac{E_c^{a'}}{E_c} R_{aa'}$$

as the *relatedness density* of activity a in city c. Boschma et al. (2015) and Balland et al. (2018a) suggest an alternative measure

$$\frac{\sum_{a' \in \mathcal{A} \setminus \{a\}} \mathrm{RCA}_c^{a'} R_{aa'}}{\sum_{a' \in \mathcal{A} \setminus \{a\}} R_{aa'}}$$

of relatedness density, which estimates the share of activity a's relatedness with all other activities that is contributed by locally over-represented activities. However, this measure suffers the same problems as measuring relatedness using RCA co-occurrences: amplified measurement errors, disregard for the extent of local over-representation, discontinuity at unit location quotients, and fragility in small cities.

3.2 Activity complexity

Complexity captures the knowledge intensity of economic activities (Balland et al., 2018b) by encoding the extent to which they rely on specialised combinations of knowledge. We define activity complexity using the second eigenvector of the row-standardised activity relatedness matrix. Our approach extends Caldarelli et al.'s (2012) Markov chain interpretation of Hidalgo and Hausmann's (2009) Method of Reflections.

It is instructive to first apply Hidalgo and Hausmann's approach to estimating complexity before decribing how we extend that approach. Recall the indicator variable (3.4) for the event in which city c has revealed comparative advantage in activity a. The *diversity* of city c,

$$\phi_c^{(0)} = \sum_{a\in\mathcal{A}} \mathrm{RCA}_c^a,$$

counts the activities in $\mathcal{A}$ in which city c has RCA. Similarly, the *ubiquity* of activity a,

$$\psi_a^{(0)} = \sum_{c\in\mathcal{C}} \mathrm{RCA}_c^a,$$

counts the cities in $\mathcal{C}$ that have RCA in activity a. Following Hidalgo and Hausmann (2009), we define the sequences $(\phi_c^{(k)})_{k\geq 0}$ and $(\psi_a^{(k)})_{k\geq 0}$ recursively by the system

$$\phi_c^{(k)} = \frac{1}{\phi_c^{(0)}} \sum_{a\in\mathcal{A}} \mathrm{RCA}_c^a\, \psi_a^{(k-1)} \tag{3.6}$$

$$\psi_a^{(k)} = \frac{1}{\psi_a^{(0)}} \sum_{c\in\mathcal{C}} \mathrm{RCA}_c^a\, \phi_c^{(k-1)} \tag{3.7}$$

of difference equations to obtain generalised measures of diversity and ubiquity. Table 1 interprets $\phi_c^{(k)}$ and $\psi_a^{(k)}$ for small k. Hidalgo and Hausmann argue that the limit point of $(\psi_a^{(k)})_{k\geq 0}$ measures the complexity of activity a because it captures "the complexity that emerges from the interactions between the increasing number of individual activities that conform an economy."

Let $\psi^{(k)} = (\psi_1^{(k)}, \psi_2^{(k)}, \ldots, \psi_{|\mathcal{A}|}^{(k)})$ be the vector of generalised ubiquities obtained on the k^{th} iteration of the Method of Reflections. Substituting (3.6) into (3.7) and letting $k \to \infty$ yields the linear system

$$P\psi^{(\infty)} = \psi^{(\infty)},$$

where

$$\psi^{(\infty)} = \lim_{k\to\infty} \psi^{(k)}$$

is the limiting vector of generalised ubiquities and where $P = (p_{a_i a_j})$ is the matrix with

$$p_{a_i a_j} = \sum_{c\in\mathcal{C}} \frac{\mathrm{RCA}_c^{a_i}}{\psi_{a_i}^{(0)}} \frac{\mathrm{RCA}_c^{a_j}}{\phi_c^{(0)}} \tag{3.8}$$

as the entry in row a_i and column a_j. Then P is the transition matrix for a Markov chain on the set $\mathcal{A}$ of activities. We explain this interpretation in Appendix B.

Activity complexity can be captured from the spectral properties of P as follows. Consider the standardised vector

$$\hat{\psi}^{(k)} = \frac{\psi^{(k)} - \overline{\psi^{(k)}}\mathbf{1}}{\mathrm{sd}(\psi^{(k)})}, \tag{3.9}$$

where $\overline{\psi^{(k)}}$ denotes the arithmetic mean of the components of $\psi^{(k)}$ and $\mathrm{sd}(\psi^{(k)})$ denotes the standard deviation of those components, and where $\mathbf{1}$ is the $|\mathcal{A}| \times 1$ vector of ones. According to Hidalgo and Hausmann (2009), the vector of activity complexities is given by the limit

$$\hat{\psi}^{(\infty)} = \lim_{k\to\infty} \hat{\psi}^{(k)}.$$

Caldarelli et al. (2012) show that if the matrix P has eigenvectors $e_1, e_2, \ldots, e_n$ and corresponding eigenvalues of decreasing absolute value then $\psi^{(k)}$ becomes proportional to e_2 as $k \to \infty$.[1] As shown in Appendix C, it follows from (3.9) that

$$\hat{\psi}^{(\infty)} = \frac{e_2 - \overline{e_2}\mathbf{1}}{\mathrm{sd}(e_2)}. \tag{3.10}$$

The Markov transition matrix P is derived from RCA co-occurrence patterns that can produce unreliable relatedness estimates for reasons identified in Section 3.1.1. Consequently, the spectral properties of P are not robust to, for example, small city and activity sizes, or pertubations in local employment near unit location quotients. We overcome this weakness by replacing P with a row-standardised copy of the activity relatedness matrix $R = (R_{a_i a_j})$, where $R_{a_i a_j}$ is the relatedness between activities a_i and a_j defined in (3.2). In the resulting Markov chain on the activity set $\mathcal{A}$, transitions from node a_i to node a_j occur with probability

$$\frac{R_{a_i a_j}}{\sum_{a\in\mathcal{A}} R_{a_i a}}. \tag{3.11}$$

Using our relatedness measure, rather than RCA co-occurrences, to define the stochastic structure of the inter-activity Markov chain retains potentially important information about activities' spatial distribution. Our approach thus improves upon the Method of Reflections, which discards such information by discretising the extent of activities' local over-representation.

We define the complexity C^a of activity a as the a^{th} component of the standardised second eigenvector of the row-standardised relatedness matrix R. Thus, our procedure is consistent with the eigenvector approximation suggested by Caldarelli et al. (2012) except that we replace the transition probability $p_{a_1 a_2}$ with (3.11). The resulting vector $(C^1, C^2, \ldots, C^{|\mathcal{A}|})$ partitions the set $\mathcal{A}$ into two subsets according to the sign of each component. Such a partition is invariant to reversing the sign of each component C^a. In order to overcome this ambiguity, we choose

[1] Mealy et al. (2019) show that Caldarelli et al.'s (2012) approach is equivalent to a spectral clustering solution to the problem of partitioning a weighted graph into two components of similar order (Shi and Malik, 2000). Each activity a is assigned to one of two clusters according to the sign of e_2's a^{th} component; the absolute value of that component encodes the "distance" between activity a and the two clusters' shared boundary.

the sign of C^1 such that C^a is positively correlated with the weighted mean size

$$\sum_{c \in \mathcal{C}} \frac{E_c^a}{E^a} E_c$$

of cities that contain activity a. This choice recongises that large cities facilitate a deeper division of labour than do small areas (Jacobs, 1969), and that such division is needed for complex knowledge to develop (Hidalgo and Hausmann, 2009; Balland et al., 2018b).

3.3 City complexity

We estimate city complexity symmetrically to activity complexity. For each pair $c_1, c_2 \in \mathcal{C}$, we compute the activity size-weighted covariance

$$\sum_{a \in \mathcal{A}} \frac{E^a}{E} \left(\frac{E_{c_1}^a}{E^a} - \sum_{a \in \mathcal{A}} \frac{E^a}{E} \frac{E_{c_1}^a}{E^a} \right) \left(\frac{E_{c_2}^a}{E^a} - \sum_{a \in \mathcal{A}} \frac{E^a}{E} \frac{E_{c_2}^a}{E^a} \right) = \sum_{a \in \mathcal{A}} \frac{E^a}{E} \left(\frac{E_{c_1}^a}{E^a} - \frac{E_{c_1}}{E} \right) \left(\frac{E_{c_2}^a}{E^a} - \frac{E_{c_2}}{E} \right)$$

in city shares of activity employment, from which we derive the relatedness between city c_1 and c_2 by converting to a weighted correlation and mapping the result linearly to $[0, 1]$. Hence, our city relatedness index measures the extent to which cities have more similar local activity portfolios than would be expected if employees were spatially distributed in proportion to city size. We define the complexity C_c of city c as the c^{th} component of the standardised second eigenvector of the row-standardised city relatedness matrix, consistent with our definition of activity complexity. We choose the sign of C_1 such that C_c is positively correlated with the local share-weighted mean complexity

$$\sum_{a \in \mathcal{A}} \frac{E_c^a}{E_c} C^a$$

of activities in city c. Thus, by construction, complex activities tend to concentrate in complex cities.

4 Data

We apply our methods for estimating relatedness and complexity to historical New Zealand census data aligned to current industry, occupation and urban area codes. These data provide usual resident employment counts in each census from 1981 to 2013. We capture cities by 2013 urban area code and identify activities using industry-occupation pairs. We capture industries by a manual grouping of New Zealand Standard Industry Output Category codes and identify occupations by one-digit 1999 New Zealand Standard Classification of Occupations code.

We study urban areas and industry-occupation pairs with persistently high employment. We identify 50 urban areas with at least 1,400 employed usual residents in census years 1981 through 2013. Tables 8–10 present local employment counts in these areas. We identify 199 industry-occupation pairs with at least 800 usually resident employees in each census year.[2]

[2] We filter industry-occupation pairs with minimum employment below 800 and then filter urban areas with minimum employment below 1400. Our data contains four activities with minimum employment counts below 800 due to some filtered urban areas containing employees within unfiltered industry-occupation pairs.

These pairs span 61 industries and nine occupations. We pool all remaining pairs into a single residual activity that represents about 18% of national employment across census years 1981–2013. Tables 11 and 12 show national employment by industry class and occupation in each census year, while Table 13 identifies the industry-occupation pairs included in our selection.

We restrict our analysis to persistently large urban areas and activities in order to mitigate the impact of two confidentiality requirements imposed by Statistics New Zealand, the agency that provides our data. First, the employment count in each cell—that is, each combination of urban area, industry, occupation, and census year—is randomly rounded to base three. Second, cells with unrounded employment counts below six are suppressed.[3] Table 14 provides suppression rates by census year in our data, both for our selected 50 urban areas and for the New Zealand economy as a whole. Employment in our selected areas is suppressed at a rate at least 50% lower than national employment in each census year. Our data represent 91.6% of unsuppressed national employment across all available census years.

We use our relatedness measure (3.2) to estimate local shares, relatedness densities, and activity and city complexities for each census year. By definition, both activity and city complexity have zero mean and unit variance in each year.[4] We exclude all observations corresponding to the residual activity after generating our estimates. We also exclude all observations corresponding to census years 1986, 1996 and 2006. Thus, our data comprises a panel of the selected 50 urban areas and the 199 non-residual activities in census years 1981, 1991, 2001 and 2013.[5] Observations correspond to city-activity pairs in a given census year.

5 Empirical analysis

5.1 Activity space

We first define an "activity space" that captures the network structure of economic activities based on our activity relatedness estimates. Our construction echoes the "product space" of traded commodities defined by Hidalgo et al. (2007). We describe activity space by a weighted network $N = (\mathcal{A}, \mathcal{E})$, where $\mathcal{A}$ is the set of 199 non-residual activities in our data and where each inter-activity edge $\{a_i, a_j\} \in \mathcal{E}$ has weight equal to the pairwise relatedness $R_{a_i a_j}$ between activities a_i and a_j.

Figure 1 presents two network maps of activity space based on 2013 census employment data; one coloured by sector and one by occupation.[6] Each node in Figure 1 has size propor-

[3] See http://archive.stats.govt.nz/Census/2013-census/methodology/confidentiality-how-applied (retrieved August 29, 2018) for further information about these two confidentiality rules.

[4] We use Bessel's correction to normalise our complexity estimates within each year.

[5] We analyse decade-separated censuses, rather than consecutive censuses, in order to balance the anticipated trade-off between predicting local employment growth using out-of-date relatedness and complexity estimates, and allowing too little time for local conditions to affect employment dynamics.

[6] We assign industry classes to sectors as follows. The primary sector includes agriculture, forestry and fishing (AA), and mining (BB). The goods-producing sector includes manufacturing (CC), electricity, gas, water and waste services (DD), and construction (EE). The distributive services sector includes wholesale trade (FF), retail trade and accomodation (GH), and transport, postal and warehousing (II). The person-centred services sector includes rental, hiring and real estate services (LL), professional, scientific, technical, administrative and support services

tional to the corresponding activity's share of national employment. We use Fruchterman and Reingold's (1991) force-directed algorithm to position nodes in the plane based on the edge weights of N. In order to reveal the strongest inter-activity connections, we show only those edges and nodes contained within the subnetwork of N induced by the 500 edges of largest weight.[7] This subnetwork contains 155 nodes with mean degree 6.45.

At the centre of our maps is a tightly connected, nest-shaped cluster of low-skill occupations in the manufacturing, road transport, and postal, courier and warehousing industries. To the right of this cluster is a group of medium- to low-skill occupations in the construction sector, which provide links to service industries such as healthcare, supermarkets and grocery stores, and motor vehicle and parts retailing. Further to the right is a branch of low-skill occupations within the forestry and logging, wood product manufacturing, and meat and seafood processing industries. The close proximity of these industries in Figure 1 reflects their mutual tendency to be relatively over-represented in small urban areas. Below the nest of manufacturing activities is a collection of high-skill occupations within various wholesaling industries. These occupations provide links to activities within the professional, support and financial service industries, which are connected to activities in the public sector through civil, professional and other interest groups.

Traversing activity space counter-clockwise, from its branch of primary industries through to its public sector tail, is analogous to traversing the urban-rural continuum from small towns to large cities. Activities in professional services and public administration rely on more diverse skills and a deeper division of labour than do activities in primary industries. Therefore, the former industries tend to cluster together in large cities, where thick labour markets facilitate knowledge specialisation and skill complementarities, while the latter are comparatively concentrated in small, rural areas where such markets are relatively thin and natural resources are more abundant. Our network map reveals the concentration among high-skill employees in the person-centred and information services sectors, and among low-skill employees in the goods-producing and distributed services sectors.

On average, activity complexity rises as we traverse activity space counter-clockwise from low-skill occupations in primary and retail industries to high-skill occupations in professional services and public administration. Figure 2 plots activity complexity against the annualised percentage point growth rate

$$G^a = 100\left(\left(\frac{E^a}{\text{L}.E^a}\right)^{1/n} - 1\right) \tag{5.1}$$

(MN), public administration and safety (OO), health care and social assistance (QQ), and arts, recreation and other services (RS). Finally, the information services sector includes information media and telecommunications (JJ), financial and insurance services (KK), and education and training (PP). These assignments reflect those used by the New Zealand Productivity Commission (2014).

[7] Our approach differs from that adopted by Hidalgo et al. (2007), who construct a maximum-weight spanning tree (MST) of product space and reintroduce the edges of largest weight until a desirable mean node degree (about four) is achieved. However, the resulting network map may fail to encode the strongest connections if, for example, the underlying network comprises a small cluster of nodes with heavy internal edges that are ignored in order to connect relatively unrelated nodes to the MST.

in national activity employment, where n is the number of years between consecutive observations in our data and where "L." is the lag operator. On average, high-skill occupations are more complex and experienced faster growth during our period of study than low-skill occupations. Standardising lagged activity complexity $\mathrm{L}.C^a$ to have zero pooled mean and unit pooled variance across census years 1991, 2001 and 2013, and estimating the linear model

$$G^a = \beta_0 + \beta_1 \, \mathrm{L}.C^a + \mathrm{error}$$

using ordinary least squares (OLS) provides the coefficient estimate $\hat{\beta}_1 = 1.038$ and associated standard error $\mathrm{se}(\hat{\beta}_1) = 0.158$. Thus, on average, a one standard deviation increase in activity complexity is associated with a one percentage point increase in overall activity employment growth in our data. This relationship reflects the concentration of complex activities in large cities (Balland et al., 2018b), which experienced faster employment growth than small areas during our period of study (New Zealand Productivity Commission, 2017).

5.2 Smart specialisation opportunities

We embed our relatedness and complexity estimates within Balland et al.'s (2018a) framework for analysing smart specialisation.[8] They characterise smart specialisation as a way to "leverage existing strengths" and "generate novel platforms on which regions can build competitive advantage in high value-added activities," arguing that such activities are precisely those with high complexity. Balland et al.'s (2018a) framework aims to highlight "the potential risks and rewards of adopting alternative diversification strategies." On the one hand, expanding into locally related activities carries low risk because local workers already possess the knowledge and skills necessary to carry out those activities. On the other hand, the highest expected returns are obtained through expanding into complex activities because they "form the basis for long-run competitive advantage." Balland et al.'s (2018a) framework identifies low-risk, high-return development opportunities as those locally under-represented activities that have both high mean local relatedness and high complexity.

For example, Figure 3 plots mean local relatedness against complexity for activities that are locally under-represented in the Central Auckland Zone; the urban area with the greatest usual resident working population as at the 2013 New Zealand census. The zone is relatively specialised in complex activities, such as professionals in the telecommunications, finance and insurance, and administration and support service industries. Such complex activities are facilitated by the large and diverse local labour market (Blackie and Lynch, 2012). Figure 3 confirms that there is limited scope for the Central Auckland Zone to develop new low-risk, high-return specialisations because it is already specialised in such activities.[9]

[8] The "feasibility charts" at `http://atlas.cid.harvard.edu/explore/feasibility/` (retrieved September 11, 2018) use a similar framework to identify export diversification opportunities.

[9] Two of Central Auckland's highest-Sharpe-ratio development options are legislators and professionals in central government administration and justice. However, expanding into these activities is unnecessary because New Zealand's central government and justice system are already established in Wellington, the nation's capital and second most populous city.

We also observe a positive relationship between mean local relatedness and activity complexity in Queenstown, a lakeside town in New Zealand's South Island with about 6,700 employed usual residents according to the 2013 census. Queenstown is a hub for New Zealand's largest export industry: tourism. As a result, Queenstown boasts many activities often found in areas with large populations. However, because its population comprises transient tourist flows, Queenstown is relatively specialised in few of the complex activities derived through deep divisions of labour. Thus, as shown in Figure 3, Queenstown appears poised for employment growth in activities typically reserved for large cities because such activities are locally under-represented but also highly related to Queenstown's existing activity portfolio.

Smaller areas in our data tend to have worse opportunities for smart specialisation. As an example, Figure 3 plots the mean local relatedness and complexity of locally under-represented activities in Huntly, a small coal mining town in New Zealand's North Island with about 1,600 employed usual residents as at the 2013 census. The negative relationship between mean local relatedness and complexity suggests that Huntly's local activity portfolio does not contain the knowledge and skills necessary for sustainable expansion into complex activities.

Huntly's small size makes it appear relatively specialised in all but the nationally largest activities because few local employees are needed to obtain location quotients that exceed unity. For example, only 876 of the 1,503,018 usual residents employed nationally at the date of the 2013 census were employed as trades workers in the specialised food retailing industry. Thus, a single employee in that activity provides Huntly with a location quotient of about 1.07. This exemplifies the instability of RCA-based relatedness measures for small cities and activities.

5.3 Do relatedness and complexity predict employment growth?

Finally, we use Balland et al.'s (2018a) characterisation of the smart specialisation framework to estimate whether relatedness and complexity promote employment growth. We define the growth rate in city c, activity a employment as the annualised percentage point change

$$G_c^a = 100\left(\left(\frac{E_c^a}{\text{L}.E_c^a}\right)^{1/n} - 1\right),$$

where n is the number of years between consecutive observations in our data and where "L." is the lag operator. We omit year subscripts throughout our analysis for symbolic clarity.

We test for predictive power by regressing G_c^a on lagged values of local share, relatedness density, activity complexity, and city complexity. We perform our analysis using the 21,352 observations for which G_c^a is computable[10] and weight observations by the corresponding lagged share $\text{L}.E_c^a$ / $\text{L}.E$ of total employment.

[10]We lose observations for two suppression-induced reasons. First, there are 9,143 observations for which the unrounded value of E_c^a is below the suppression threshold, preventing us from computing the local share LS_c^a, relatedness density RD_c^a, and growth rate G_c^a. Second, there are another 9,305 observations for which the unrounded value of $\text{L}.E_c^a$ falls below the suppression threshold, preventing us from computing G_c^a. We account for suppressed values of LS_c^a by computing relatedness indices based on pairwise complete observations.

We transform our local share estimates by subtracting their weighted mean and multiplying the result by 100 to obtain percentage point demeaned shares. We also standardise our estimates of relatedness density, activity complexity and city complexity to have zero weighted mean and unit weighted variance.

Table 2 reports descriptive statistics for our transformed data before and after weighting by lagged employment shares. Comparing the weighted and unweighted means reveals that, on average, observations with larger city-activity employment counts are associated with greater local shares, lower relatedness density, higher activity complexity and higher city complexity. Annualised city-activity growth ranges from -32.9% to 34.7% per year.[11] The largest local share of 23.6% is attained by service and sales workers in Queenstown's accommodation and food services industry as at the 1981 census.

Table 3 presents our regression results. Consistent with the positive relationship between activity complexity and employment growth displayed in Figure 2, columns (1) and (2) show that more complex activities grew faster during our period of study. On average, and holding both local share and relatedness density constant at their weighted mean value, a one standard deviation increase in activity complexity is associated with a 0.89 percentage point increase in local employment growth per year. This effect increases to 0.98 percentage points when we control for city complexity, and its interaction with local share and relatedness density. More locally related activities experienced slower growth, especially in complex cities.

The coefficient estimates in columns (1) and (2) of Table 3 may be biased by unobservable, time-varying activity and city factors that are correlated with our covariates of interest. We control for these factors in column (3) by introducing activity-year and city-year fixed effects. This allows us to identify the effects of cross-sectional variation in local growth rates, controlling both for the period-specific growth experienced by the activity across New Zealand and for the period-specific growth experienced by the city as a whole. However, we lose the ability to separately identify coefficients on activity and city complexity. Our estimates in column (3) suggest that cities diversified their local activity portfolios during our period of study, and that this diversification was faster into more complex activities and within more complex cities.

Balland et al.'s (2018a) framework characterises high relatedness density as an indicator of low-risk local investment options, and activity complexity as an indicator of high reward options. Their framework thus suggests that complex activities with high local relatedness offer the strongest prospects for future growth. If this were true then we would expect a strong positive coefficient on the interaction of relatedness density and activity complexity. Our estimates in columns (1)–(3) of Table 3 show only a weak and insignificant interaction.

5.3.1 Are the effects of relatedness and complexity context dependent?

Estimating regression coefficients by averaging across all observations in our data may mask effects that are only relevant for particular activities or local contexts. We analyse subsamples

[11] These limit rates correspond to city-activity employment counts that fell from 489 to 9 and rose from 9 to 177 between consecutive observations. Our regression results do not significantly change when we restrict our analysis to activities $a \in \mathcal{A}$ with $\min\{E_c^a, \mathrm{L}.E_c^a\} > 100$ across all cities $c \in \mathcal{C}$.

of our data in order to investigate the variation in attributes of the activities and cities to which the growth benefits of relatedness and complexity accrue. Our subsamples isolate groups of locally over- and under-represented activities, groups of complex activities and cities, and levels of city urbanisation. Table 4 reports weighted means and standard deviations within each subsample. We transform our data so that, within each subsample, local share has zero within-subsample weighted mean, and relatedness density, activity complexity and city complexity have zero within-subsample weighted mean and unit within-subsample weighted variance. We use the model specification in column (3) of Table 3 throughout our analyses.

Balland et al. (2018a) argue that the benefits of relatedness and complexity are most important for locally under-represented activities. Such activities have $\text{RCA}_c^a = 0$, where RCA_c^a is the indicator variable for the event in which city c has revealed comparative advantage in activity a. Table 5 reports results from estimating our preferred model specification after partitioning our data into two subsamples: activities with $\text{L.RCA}_c^a = 1$ and activities with $\text{L.RCA}_c^a = 0$. This partitioning allows us to identify the potential benefits of relatedness and complexity for new specialisations, which are the focus of the smart specialisation framework. The interaction of relatedness density and activity complexity was negative and significant in our subsample of locally under-represented activities. This suggests that smart specialisation, as captured by Balland et al. (2018a), does not explain employment dynamics within our data.

We next partition our data by levels of complexity in order to identify differences in the effect of relatedness and complexity for different types of activities and within different types of cities. We define subsamples of complex activities—comprising all activities with above-mean complexity in each of the census years 1981, 1991, 2001 and 2013—and "simple" activities—comprising activities with below-mean complexity in each of those years. We pool all activities that meet neither of these criteria into an "other activities" subsample. We define subsamples of complex, "simple" and other cities using analogous criteria for city complexity. Figures 5 and 6 present the activity and city complexity series, respectively, within each subsample.

Table 6 reports results from our activity and city complexity subsample analyses. Activities with large local shares experienced significantly slower growth in all of our subsamples. Relatedness density had an insignificant positive direct effect overall but a significant negative effect in each activity subsample, suggesting that the positive overall effect is due to between-group variation. The interaction of relatedness density and activity complexity was small and insignificant in all six subsamples. Relatedness density and city complexity appear to be complementary in complex cities, suggesting that being relatedness dense promotes growth only for activities in the most complex cities.

The mechanisms underlying relatedness and complexity may operate only at certain city scales. For example, knowledge spillovers may be relevant only in cities with sufficiently thick labour markets, which facilitate such spillovers. We examine the scale-dependence of relatedness and complexity by partitioning our data into four subsamples according to cities' urban area type as classified in Tables 8–10. Main urban areas represent the most urbanised areas in New Zealand, while secondary and minor urban areas tend to be smaller and less urbanised. We distinguish the four Auckland zones from other main urban areas due to Auckland's relatively large size and unique labour market conditions. Table 4 shows that, on average, larger urban areas experienced greater employment growth during our period of study. Such areas also had more diverse local labour markets, as shown by the weighted means of local share and relatedness density, and tended to contain more complex activities.

Table 7 summarises our analysis of our urban area type subsamples. Auckland diversified faster than other areas—as shown by the large, negative coefficient on local share—although this diversification was slower for more complex activities. Comparing the interaction of city complexity and relatedness density across subsamples supports our result from Table 6 that relatedness density appears to promote growth only in the most complex cities. A one standard deviation rise in relatedness density has an insignificant effect on employment growth overall, but a significant effect in the most complex part of Auckland—the Western Auckland Zone—in which the rise is associated with a 0.544 percentage point increase in annualised employment growth based on our 2013 complexity estimates. Finally, the continued lack of joint effect that relatedness density and activity complexity have on local employment growth suggests that relatedness and complexity provide information about the way that large, complex cities grow, rather than information about how activity complementarities foster local growth. This result reflects cities being dense networks of interacting activities: the benefits of activities' interaction are more apparent in larger cities, within which networks are more dense and where people engaged in related activities interact more frequently.

6 Conclusion

This paper analyses the relatedness and complexity of economic activities using employment data on geographic areas with small populations. We develop a measure of activity relatedness based on weighted correlations of local employment shares. Our approach extends previously used measures by allowing for continuous variation in activities' local over-representation and by accounting for differences in signal quality across geographies of different size. We use our measure to estimate city and activity complexity, and to define an "activity space" that encodes the network structure of related activities. Our network map reveals clusters of activities associated with high-skill occupations in the person-centred and information services sectors, and of activities associated with low-skill occupations in the goods-producing and distributed services sectors. These clusters respectively reflect strong colocation patterns among activities with high and low levels of complexity.

We apply our estimates within Balland et al.'s (2018a) framework for analysing smart specialisation, and investigate the mechanisms underlying that framework by evaluating the ex-post ability of relatedness and complexity to predict subsequent growth in local activity employment within a selection of New Zealand urban areas. Complex activities in our selection experienced faster growth between 1981 and 2013, especially in complex cities. However, this growth was not significantly stronger in cities more dense with related activities. Relatedness and complexity appear to be most relevant for analysing how large, complex cities grow, and provide less information about growth trajectories for small cities.

Overall, we do not identify strong effects of relatedness and complexity on growth in local activity employment. It is an open question whether this absence means that these effects do not operate or that New Zealand cities lack the scale for such operation. Our results may reflect the limited capacity for knowledge specialisation within New Zealand's local labour markets, which are smaller than those in previously studied geographic areas. Alternatively, our failure to identify strong effects may reflect how, within New Zealand and during our period of study, policies were not explicitly designed to encourage or support relatedness and complexity. The absence of such targeted policies may have prevented any potential employment growth benefits of smart specialisation policies from being realised.

References

Asheim, B. T. and Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. C., and Nelson, R. R., editors, *The Oxford Handbook of Innovation*, pages 291–317, Oxford. Oxford University Press.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, 33(2):99–123.

Balland, P.-A., Boschma, R., Crespo, J., and Rigby, D. L. (2018a). Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*.

Balland, P.-A., Jara-Figueroa, C., Petralia, S., Steijn, M., Rigby, D., and Hidalgo, C. A. (2018b). Complex economic activities concentrate in large cities. Papers in Evolutionary Economic Geography 18.29, Utrecht University.

Balland, P.-A. and Rigby, D. L. (2017). The geography of complex knowledge. *Economic Geography*, 93(1):1–23.

Barca, F. (2009). An agenda for a reformed cohesion policy: A place-based approach to meeting European Union challenges and expectations.

Barca, F., McCann, P., and Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. *Journal of Regional Science*, 52(1):134–152.

Blackie, E. and Lynch, C.-A. (2012). The Auckland labour market in 2011: An overview.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1):61–74.

Boschma, R. (2014). Constructing regional advantage and smart specialisation: Comparison of two European policy concepts. *Italian Journal of Regional Science*, 13(1):51–68.

Boschma, R., Balland, P.-A., and Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24(1):223–250.

Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., and Tacchella, A. (2012). A network analysis of countries' export flows: Firm grounds for the building blocks of the economy. *PLoS ONE*, 7(10).

Delgado, M., Porter, M. E., and Stern, S. (2014). Clusters, convergence and economic performance. *Research Policy*, 43(10):1785–1799.

Delgado, M., Porter, M. E., and Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, 16(1):1–38.

Farinha Fernandes, T., Balland, P.-A., Morrison, A., and Boschma, R. (2018). What drives the geography of jobs in the US? Unpacking relatedness. Papers in Evolutionary Economic Geography 18.13, Utrecht University.

Feldman, M. P. and Audretsch, D. B. (1999). Innovation in cities: Science-based diversity, specialisation and localized competition. *European Economic Review*, 43(2):409–429.

Foray, D., David, P. A., and Hall, B. (2009). Smart specialisation – the concept. Knowledge Economists Policy Brief No. 9, European Commission, Brussels.

Foray, D., David, P. A., and Hall, B. (2011). Smart specialisation: From academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation. MTEI working paper, École Polytechnique Fédérale de Lausanne.

Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(1):1129–1164.

Hendy, S. and Callaghan, P. (2013). *Get off the Grass: Kickstarting New Zealand's Innovation Economy*. Auckland University Press, Auckland.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., and Zhu, S. (2018). The principle of relatedness. In Morales, A. J., Gershenson, C., Braha, D., Manai, A. A., and Bar-Yam, Y., editors, *Unifying Themes in Complex Systems IX*, pages 451–457, Cham. Springer International Publishing.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487.

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

Jara-Figueroa, C., Jun, B., Glaeser, E. L., and Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the survival of new firms. *Proceedings of the National Academy of Sciences*, 115(50):12646–12653.

Marshall, A. (1920). *Principles of Economics*. Macmillan, London, 8 edition.

McCann, P. (2009). Economic geography, globalisation and New Zealand's productivity paradox. *New Zealand Economic Papers*, 43(3):279–314.

McCann, P. and Ortega-Argilés, R. (2015). Smart specialisation, regional growth and applications to European Union cohesion policy. *Regional Studies*, 49(8):1291–1302.

Mealy, P., Doyne Farmer, J., and Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, 5(1).

Morkutė, G., Koster, S., and Van Dijk, J. (2016). Employment growth and inter-industry job reallocation: Spatial patterns and relatedness. *Regional Studies*, 51(6):958–971.

Neffke, F., Hartog, M., Boschma, R., and Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, 94(1):23–48.

Neffke, F. and Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297–316.

New Zealand Productivity Commission (2014). Boosting productivity in the services sector.

New Zealand Productivity Commission (2017). Better urban planning: Final report.

Schumpeter, J. (1934). *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.

Shi, J. and Malik, J. (2000). Normalised cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Vicente, J., Balland, P.-A., and Crespo, J. (2018). Les fondements micro du changement structurel régional: Que nous enseignent 25 ans de proximités? *Revue d'Économie Régionale & Urbaine*, 5(6):1013–1041.

Weitzman, M. L. (1998). Recombitant growth. *Quarterly Journal of Economics*, 113(2):331–360.

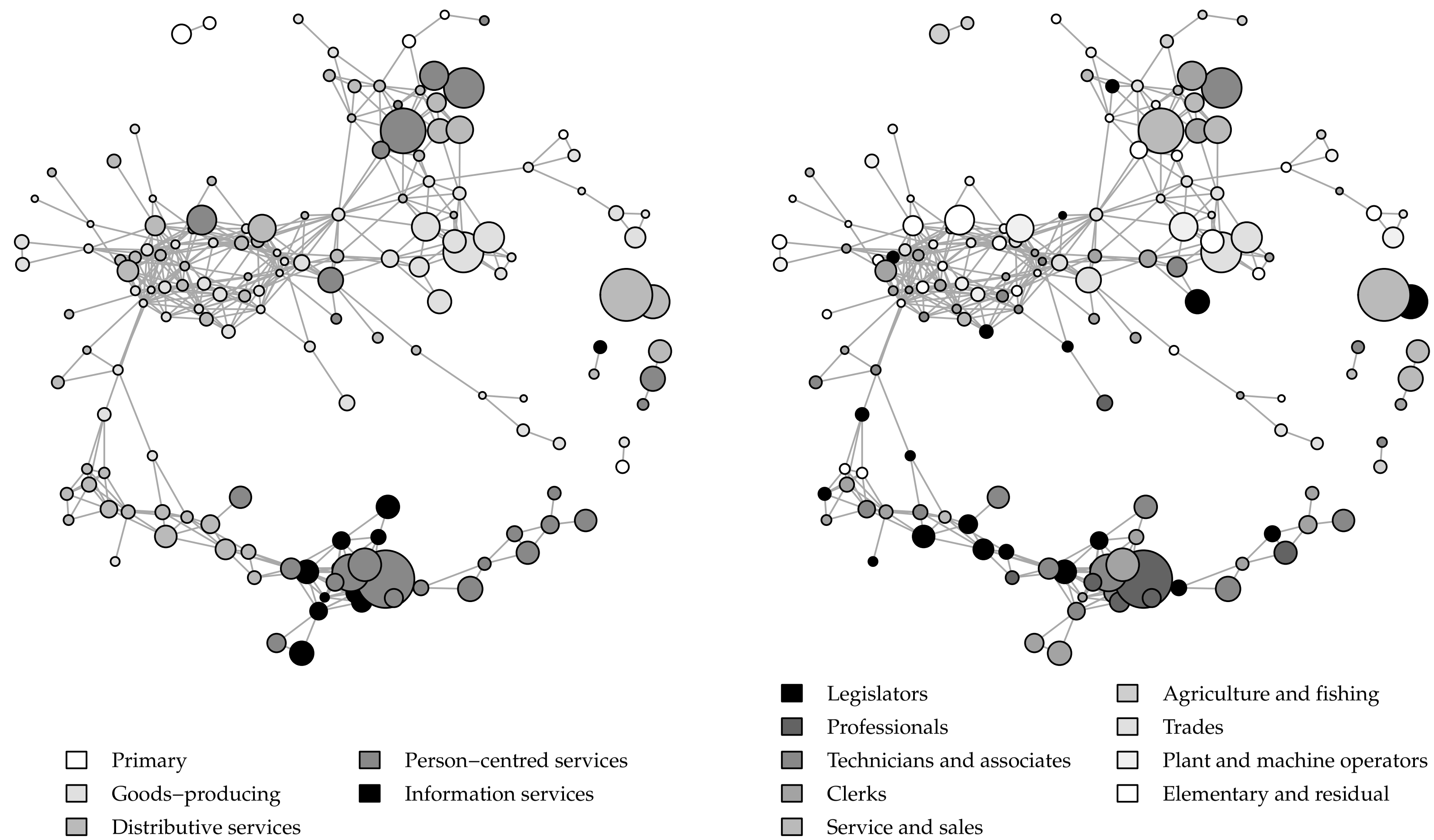

Figure 1: Network maps of activity space with nodes coloured by sector (left) and occupation (right)

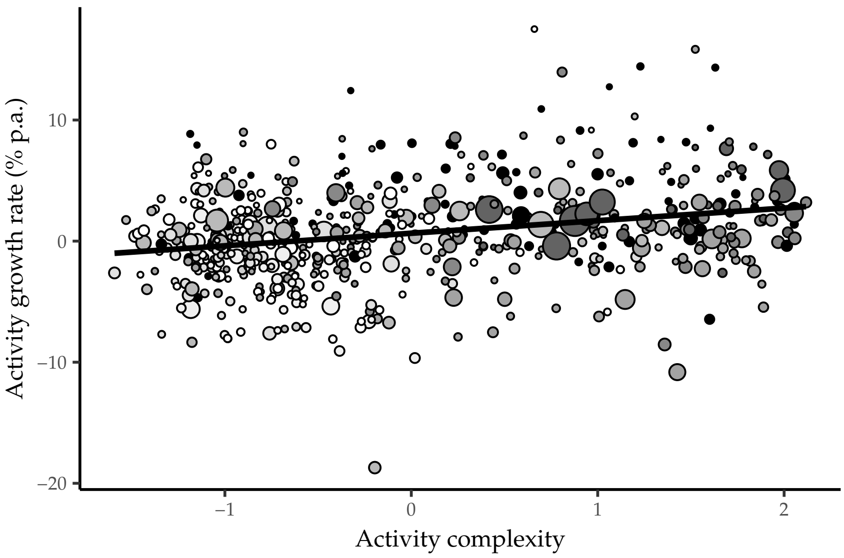

Figure 2: Activity complexity and activity employment growth

Notes: The plotted data are derived from 1981, 1991, 2001 and 2013 census usual resident employment counts, with random rounding to base three and with all cells with true counts below six suppressed. Plotted points correspond to activities in census years 1991, 2001 and 2013, and are scaled by lagged total activity employment in our data and coloured by occupation as in Figure 1. The fitted line is given by $G^a = \hat{\beta}_0 + \hat{\beta}_1 \, \mathrm{L.}C^a$, where G^a is the growth rate in total activity employment defined in (5.1), where $\mathrm{L.}C^a$ is lagged activity complexity, and where $\hat{\beta}_0$ and $\hat{\beta}_1$ are OLS estimates.

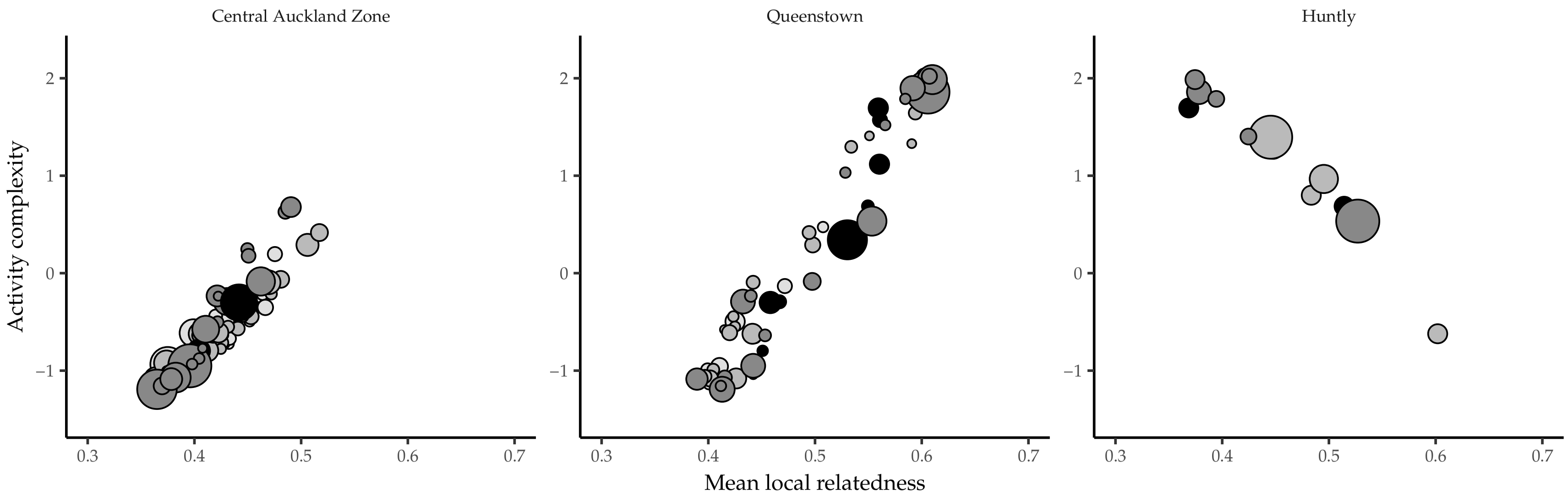

Figure 3: Smart specialisation opportunities the Central Auckland Zone (left), Queenstown (middle) and Huntly (right)

Notes: The plotted data are derived from 2013 census usual resident employment counts with random rounding to base three and with all cells with true counts below six suppressed. Only those activities with computable location quotients below unity are shown. Points are scaled by local share and coloured by sector as in Figure 1.

Table 1: Interpretation of $\phi_c^{(k)}$ and $\psi_a^{(k)}$ for $k \in \{0, 1, 2\}$

Variable	Description
$\phi_c^{(0)}$	Diversity of city *c* (i.e., number of activities in which city *c* has RCA)
$\phi_c^{(1)}$	Mean ubiquity of activities in which city *c* has RCA
$\phi_c^{(2)}$	Mean diversity of cities that have RCA in similar activities to city *c*
$\psi_a^{(0)}$	Ubiquity of activity *a* (i.e., number of cities that have RCA in activity *a*)
$\psi_a^{(1)}$	Mean diversity of cities that have RCA in activity *a*
$\psi_a^{(2)}$	Mean ubiquity of activities in which cities with RCA in activity *a* also have RCA

Table 2: Descriptive statistics

	Unweighted		Weighted			
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Min	Max
City-activity growth rate (G_c^a)	0.627	5.012	0.190	4.420	-32.935	34.702
Local share (L.LS_c^a)	-0.818	0.921	0.000	1.709	-1.391	22.244
Relatedness density (L.RD_c^a)	0.211	1.066	0.000	1.000	-4.301	4.809
Activity complexity (L.C^a)	-0.197	0.975	0.000	1.000	-1.699	1.890
City complexity (L.C_c)	-0.729	0.955	0.000	1.000	-2.136	1.384

Notes: Our data includes 21,352 observations of each variable. These observations correspond to city-activity pairs in a given census year. We exclude observations corresponding to the residual activity. "L." is the lag operator.

Table 3: Main regression results

	City-activity growth rate (G_c^a)		
	(1)	(2)	(3)
Local share (L.LS_c^a)	-0.060	-0.079	-0.266***
	(0.038)	(0.048)	(0.044)
Relatedness density (L.RD_c^a)	-0.209***	-0.131*	0.059
	(0.063)	(0.059)	(0.032)
Activity complexity (L.C^a)	0.887***	0.975***	
	(0.062)	(0.068)	
Activity complexity × local share	0.189***	0.204***	-0.137*
	(0.043)	(0.043)	(0.060)
Activity complexity × relatedness density	0.076	0.035	0.030
	(0.053)	(0.057)	(0.040)
City complexity (L.C_c)		0.288***	
		(0.058)	
City complexity × local share		-0.085*	-0.084***
		(0.041)	(0.023)
City complexity × relatedness density		-0.362***	-0.142***
		(0.053)	(0.026)
Activity-year and city-year fixed effects			Yes
Observations	21,352	21,352	21,352
R^2	0.047	0.057	0.719

Notes: This table reports OLS estimates with analytic weights equal to lagged shares of total employment in our data and with heteroskedasticity-robust standard errors provided in parentheses. One, two and three asterisks indicate significance at the 5%, 1% and 0.1% levels, respectively. Observations correspond to city-activity pairs in a given census year. We omit year subscripts for symbolic clarity. Each model specification includes the vector of ones as a covariate, the estimated coefficient and standard error of which we omit from the table.

Table 4: Descriptive statistics by regression subsample

	City-activity growth rate	Local share	Relatedness density	Activity complexity	City complexity
All data	0.190 (4.420)	0.000 (1.709)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
Activities for which $\text{L.RCA}_c^a = 1$	-0.207 (4.341)	0.273 (1.986)	0.433 (0.874)	-0.008 (0.993)	-0.083 (1.014)
Activities for which $\text{L.RCA}_c^a = 0$	0.774 (4.469)	-0.403 (1.067)	-0.638 (0.815)	0.011 (1.010)	0.123 (0.967)
Complex activities	1.123 (3.986)	0.479 (1.771)	-0.188 (1.073)	0.985 (0.519)	0.072 (0.953)
Simple activities	-0.844 (4.311)	-0.238 (1.826)	0.140 (0.953)	-0.963 (0.342)	-0.093 (1.041)
Other activities	0.356 (5.068)	-0.574 (0.704)	0.121 (0.846)	-0.114 (0.579)	0.043 (0.996)
Complex cities	0.441 (4.251)	-0.194 (1.250)	-0.046 (1.076)	0.114 (1.006)	0.700 (0.488)
Simple cities	-0.432 (4.749)	0.421 (2.268)	0.206 (1.012)	-0.211 (0.958)	-1.429 (0.288)
Other cities	-0.051 (4.568)	0.219 (2.145)	-0.005 (0.776)	-0.147 (0.970)	-0.835 (0.568)
Auckland	0.844 (4.162)	-0.261 (1.195)	-0.040 (1.170)	0.102 (1.008)	0.999 (0.248)
Main urban areas (excluding Auckland)	-0.084 (4.464)	0.008 (1.551)	-0.040 (0.822)	-0.010 (0.995)	-0.383 (0.805)
Secondary urban areas	-0.417 (4.861)	0.630 (2.669)	0.324 (1.108)	-0.288 (0.934)	-1.249 (0.609)
Minor urban areas	-0.958 (4.422)	2.123 (4.793)	0.738 (1.556)	-0.447 (0.892)	-1.609 (0.354)

Notes: This table reports weighted means and weighted standard deviations—the latter shown in parentheses—by regression subsample.

Table 5: Local over-representation subsample regression results

	City-activity growth rate (G_c^a)		
		Local over-representation	
	All data	L.RCA$_c^a = 1$	L.RCA$_c^a = 0$
Local share (L.LS_c^a)	-0.266***	-0.326***	-1.851***
	(0.044)	(0.050)	(0.524)
Relatedness density (L.RD_c^a)	0.059	0.236***	0.258***
	(0.032)	(0.057)	(0.066)
Activity complexity × local share	-0.137*	-0.144*	0.289
	(0.060)	(0.060)	(0.565)
Activity complexity × relatedness density	0.030	0.052	-0.268***
	(0.040)	(0.057)	(0.072)
City complexity × local share	-0.084***	-0.149***	-0.068
	(0.023)	(0.026)	(0.041)
City complexity × relatedness density	-0.142***	-0.208***	-0.242***
	(0.026)	(0.047)	(0.035)
Observations	21,352	10,963	10,389
R^2	0.719	0.755	0.774

Notes: This table reports OLS estimates with analytic weights equal to lagged shares of total within-subsample employment and with heteroskedasticity-robust standard errors shown in parentheses. One, two and three asterisks indicate significance at the 5%, 1% and 0.1% levels, respectively. Observations correspond to city-activity pairs in a given census year. We omit year subscripts for symbolic clarity. Each model specification includes the vector of ones as a covariate, the estimated coefficient and standard error of which we omit from the table. We include activity-year and city-year fixed effects in all models.

Table 6: Activity and city complexity subsample regression results

	City-activity growth rate (G_c^a)						
		Activity subsample			City subsample		
	All data	Complex	Simple	Other	Complex	Simple	Other
Local share (L.LS_c^a)	-0.266***	-0.291***	-0.260***	-1.586***	-0.585**	-0.458***	-0.344***
	(0.044)	(0.079)	(0.063)	(0.180)	(0.179)	(0.068)	(0.060)
Relatedness density (L.RD_c^a)	0.059	-0.178*	-0.664***	-0.271***	0.033	0.211*	-0.094
	(0.032)	(0.072)	(0.076)	(0.065)	(0.061)	(0.095)	(0.066)
Activity complexity × local share	-0.137*	0.062	-0.423**	-1.086***	0.234	-0.397***	-0.060
	(0.060)	(0.056)	(0.143)	(0.317)	(0.140)	(0.108)	(0.085)
Activity complexity × relatedness density	0.030	-0.112	-0.047	0.050	-0.047	-0.163	0.065
	(0.040)	(0.064)	(0.044)	(0.060)	(0.072)	(0.089)	(0.072)
City complexity × local share	-0.084***	-0.152***	0.008	-0.309**	-0.115***	-0.035	0.017
	(0.023)	(0.026)	(0.053)	(0.095)	(0.032)	(0.038)	(0.027)
City complexity × relatedness density	-0.142***	-0.063	0.272**	0.116*	0.097**	-0.067	0.098***
	(0.026)	(0.045)	(0.089)	(0.051)	(0.031)	(0.051)	(0.030)
Observations	21,352	6,399	10,750	4,203	5,275	8,216	7,861
R^2	0.719	0.783	0.623	0.830	0.836	0.603	0.727

Notes: This table reports OLS estimates with analytic weights equal to lagged shares of total within-subsample employment and with heteroskedasticity-robust standard errors shown in parentheses. One, two and three asterisks indicate significance at the 5%, 1% and 0.1% levels, respectively. Observations correspond to city-activity pairs in a given census year. We omit year subscripts for symbolic clarity. Each model includes the vector of ones as a covariate, the estimated coefficient and standard error of which we omit from this table. We include activity-year and city-year fixed effects in all models.

Table 7: Urban area type subsample regression results

	City-activity growth rate (G_c^a)				
			Urban area type		
	All data	Auckland	Main (exc. Auckland)	Secondary	Minor
Local share (L.LS_c^a)	-0.266***	-0.816**	-0.343***	-0.208*	-0.411***
	(0.044)	(0.256)	(0.084)	(0.094)	(0.105)
Relatedness density (L.RD_c^a)	0.059	0.127	0.074	0.154	0.405
	(0.032)	(0.081)	(0.043)	(0.128)	(0.325)
Activity complexity × local share	-0.137*	0.377*	-0.068	0.012	-0.074
	(0.060)	(0.147)	(0.087)	(0.139)	(0.160)
Activity complexity × relatedness density	0.030	-0.127	0.082	0.009	-0.314
	(0.040)	(0.095)	(0.054)	(0.128)	(0.310)
City complexity × local share	-0.084***	0.008	-0.026	-0.022	-0.023
	(0.023)	(0.053)	(0.031)	(0.045)	(0.039)
City complexity × relatedness density	-0.142***	0.178***	0.011	-0.036	-0.232*
	(0.026)	(0.036)	(0.044)	(0.056)	(0.107)
Observations	21,352	2,388	11,901	5,128	1,935
R^2	0.719	0.921	0.732	0.592	0.618

Notes: This table reports OLS estimates with analytic weights equal to lagged shares of total within-subsample employment and with heteroskedasticity-robust standard errors shown in parentheses. One, two and three asterisks indicate significance at the 5%, 1% and 0.1% levels, respectively. Observations correspond to city-activity pairs in a given census year. We omit year subscripts for symbolic clarity. Each model includes the vector of ones as a covariate, the estimated coefficient and standard error of which we omit from this table. We include activity-year and city-year fixed effects in all models.

A Primer on networks and graph theory

A *graph* is an order pair $G = (\mathcal{V}, \mathcal{E})$ comprising a *vertex set* $\mathcal{V}$ and an *edge set* $\mathcal{E}$ of two-element subsets of $\mathcal{V}$. The elements of $\mathcal{V}$ are called *vertices* while the elements of $\mathcal{E}$ are called *edges*. A *network* is a graph in which vertices represent named entities and are called *nodes*. We introduce several properties of networks below using the language of graph theory.

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph. Two vertices $u \in \mathcal{V}$ and $v \in \mathcal{V}$ are *adjacent* if $\{u, v\} \in \mathcal{E}$, while the edge $\{u, v\}$ is *incident* with u and v. The *degree* of v, denoted $\deg(v)$, is equal to the number of vertices with which v is adjacent. Since every edge is incident with two vertices, we have

$$\sum_{v \in \mathcal{V}} \deg(v) = 2|\mathcal{E}|.$$

Thus, the vertices in $\mathcal{V}$ have mean degree $2|\mathcal{E}|/|\mathcal{V}|$.

A *walk* in a graph G is an alternating sequence $v_1 e_1 v_2 e_2 \cdots$ of vertices $v_1, v_2, \ldots \in \mathcal{V}$ and edges $e_1, e_2, \ldots \in \mathcal{E}$. A *path* is a walk in which each edge (and therefore each vertex) is unique. Two vertices are *connected* if there is a path between them. We say that G is *connected* if every pair of vertices is connected; otherwise, we say that G is *disconnected*.

It is often useful to denote the vertex set of a graph G by $\mathcal{V}(G)$ and its edge set by $\mathcal{E}(G)$. We call $|\mathcal{V}(G)|$ the *order* of G and $|\mathcal{E}(G)|$ the *size* of G. A graph G' is called a *subgraph* of G if both $\mathcal{V}(G') \subseteq \mathcal{V}(G)$ and $\mathcal{E}(G') \subseteq \mathcal{E}(G)$. A *component* of G is a connected subgraph of maximal size. A subgraph G' of G is *induced* by a set of edges $\mathcal{X} \subseteq \mathcal{E}(G)$ if $\mathcal{E}(G') = \mathcal{X}$ and

$$\mathcal{V}(G') = \{v \in \mathcal{V}(G) : v \text{ is incident with } e \text{ for some edge } e \in \mathcal{X}\}.$$

A *cycle* in a graph is a path with equal first and last vertices. A *tree* is a connected graph with no cycles, while a *forest* is a graph in which all components are trees. A *spanning tree (forest)* of a connected graph G is a subgraph G' of G that is a tree (forest) and has vertex set $\mathcal{V}(G') = \mathcal{V}(G)$.

Finally, a *weighted graph* is a graph G within which a numerical *weight* is assigned to each edge by a function $w : \mathcal{E}(G) \to \mathbb{R}$. Thus, an unweighted graph can be intepreted as a weighted graph for which w has range $w(\mathcal{E}(G)) = \{1\}$. A *maximum-weight spanning tree (forest)* of G is a spanning tree (forest) G_* that obtains the maximum value of the sum

$$\sum_{e \in \mathcal{E}(G_*)} w(e).$$

B Interpreting P as a Markov transition matrix

Consider the city-activity network shown in Figure 4, in which city $c \in \{c_1, c_2, c_3, c_4\}$ is adjacent to activity $a \in \{a_1, a_2, a_3, a_4\}$ if $\mathrm{RCA}_c^a = 1$. Suppose that a random walker traverses this network by transitioning across edges with uniform probability. For example, if the walker starts at activity node a_1 then it transitions to city nodes c_1, c_2 and c_3 with probability $1/3$ each, whereas if the walker starts at city node c_1 then it transitions to activity nodes a_1 and a_2 with probability $1/2$ each. Then the probability that a walker initially positioned at activity node a_i

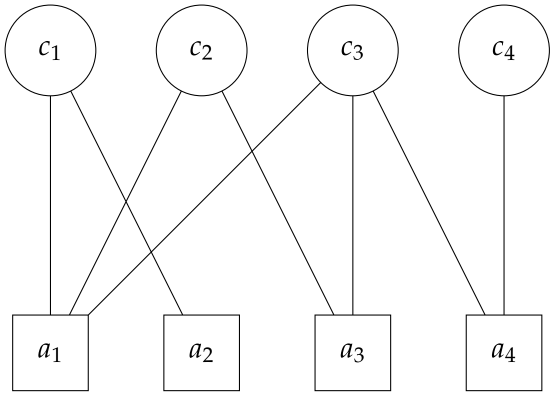

Figure 4: Example RCA network with four cities and four activities

transitions to activity node a_j after two steps is equal to (3.8). For example, the network shown in Figure 4 is associated with the Markov transition matrix

$$P_4 = \begin{bmatrix} 4/9 & 1/6 & 5/18 & 1/9 \\ 1/2 & 1/2 & 0 & 0 \\ 5/12 & 0 & 5/12 & 1/6 \\ 1/6 & 0 & 1/6 & 2/3 \end{bmatrix} \tag{B.1}$$

between the activities a_1, a_2, a_3 and a_4. The only city with RCA in activity a_2 is c_1, which has RCA in activities a_1 and a_2. Thus, if a random walker starts at activity node a_2 then its position after two steps is either a_1 or a_2 with equal probability as confirmed by the second row of P_4.

The intuitive meaning of $p_{a_i a_j}$ is as follows. Suppose that a specialist in activity a_i relocates to another city specialising in that activity and that, on arrival, they change jobs to one of the local specialisations in the new city. If all possible outcomes of the relocation and job-change decision are equally likely, then $p_{a_i a_j}$ is the probability that the specialist shifts to activity a_j.

C Proof that $\hat{\psi}^{(\infty)} = (e_2 - \overline{e_2}\,\mathbf{1})/\mathrm{sd}(e_2)$

First let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of P and define $\psi^{(0)} = \alpha_1\lambda_1 e_1 + \alpha_2\lambda_2 e_2 + \cdots + \alpha_n\lambda_n e_n$, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the coordinates of $\psi^{(0)}$ with respect to the eigenbasis $e_1, e_2, \ldots, e_n$. Then

$$\psi^{(k)} = \alpha_1 e_1 + \alpha_2\lambda_2^k e_2 + \mathcal{O}(\lambda_3^k)$$

by linearity and the Perron-Frobenius theorem. It follows that

$$\psi^{(k)} - \overline{\psi^{(k)}}\mathbf{1} = \alpha_2\lambda_2^k(e_2 - \overline{e_2}\mathbf{1}) + \mathcal{O}(\lambda_3^k)$$

and that

$$\mathrm{sd}(\psi^{(k)}) = |\alpha_2\lambda_2^k|\mathrm{sd}(e_2) + \mathcal{O}(\lambda_3^k)$$

as $k \to \infty$. These limits yield (3.10) because the sign of e_2 is arbitrary.

D Additional figures and tables

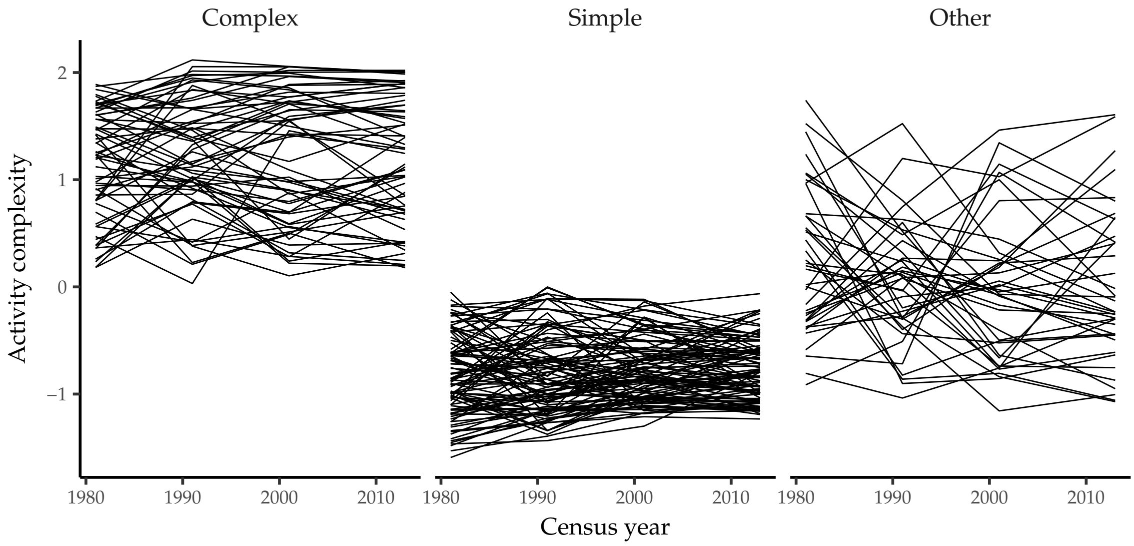

Figure 5: Activity complexity series by subsample membership

Notes: The plotted data are derived from 1981, 1991, 2001 and 2013 census usual resident employment counts, with random rounding to base three and with all cells with true counts below six suppressed. Each activity is assigned to the "complex," "simple" or "other" subsample according to whether its complexity is greater than zero, is smaller than zero, or has variable sign across the four census years included in our data. We omit the complexity series for the residual activity.

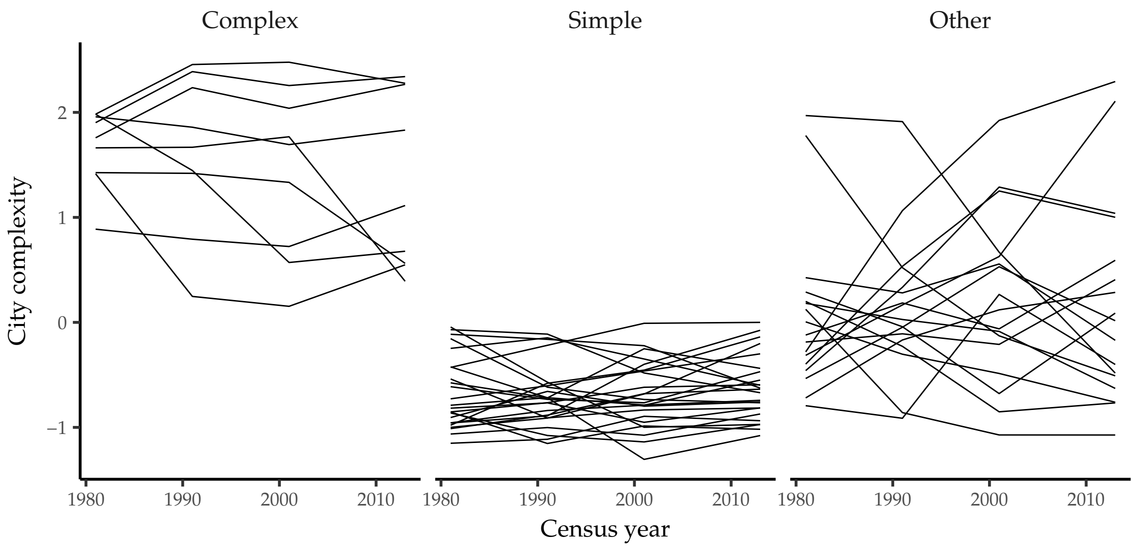

Figure 6: City complexity series by subsample membership

Notes: The plotted data are derived from 1981, 1991, 2001 and 2013 census usual resident employment counts, with random rounding to base three and with all cells with true counts below six suppressed. Each city is assigned to the "complex," "simple" or "other" subsample according to whether its complexity is greater than zero, is smaller than zero, or has variable sign across the four census years included in our data.

Table 8: Employment in main urban areas

Urban area	Local employment by census year						
	1981	1986	1991	1996	2001	2006	2013
Central Auckland Zone	129,429	136,896	124,854	144,315	160,935	187,641	198,603
Christchurch	126,714	133,062	124,662	143,397	152,742	176,325	173,631
Southern Auckland Zone	91,974	108,609	99,825	113,184	125,328	149,748	154,890
Northern Auckland Zone	68,439	80,313	81,636	94,956	104,640	123,930	131,112
Wellington Zone	77,145	79,944	74,526	81,063	86,853	98,268	102,024
Western Auckland Zone	52,749	62,295	61,254	69,789	75,237	85,851	88,545
Hamilton Zone	48,621	52,968	49,734	56,967	61,671	72,207	76,125
Tauranga	21,426	26,118	25,488	31,905	38,481	48,666	51,288
Dunedin	46,683	46,164	40,833	45,099	46,125	51,267	49,746
Lower Hutt Zone	43,899	45,264	40,356	41,421	43,080	46,386	44,664
Palmerston North	29,886	30,207	28,413	31,773	32,271	36,375	34,977
Nelson	17,481	19,536	18,933	22,200	23,934	27,393	28,137
Hastings Zone	23,091	24,453	21,912	22,911	23,622	26,802	26,259
Napier Zone	21,138	22,764	19,809	22,143	23,004	26,013	25,197
New Plymouth	18,339	20,892	18,408	19,962	19,440	22,812	24,234
Invercargill	24,030	23,475	20,283	21,054	19,986	22,140	22,323
Porirua Zone	17,244	19,551	17,259	16,935	19,224	20,748	21,834
Rotorua	19,581	22,263	18,429	20,670	21,072	22,722	20,943
Whangarei	15,573	19,218	15,195	16,266	17,067	19,398	18,012
Upper Hutt Zone	16,539	17,010	15,411	15,342	15,510	17,217	17,322
Kapiti	7,071	8,601	9,681	10,821	12,855	15,174	15,966
Wanganui	15,675	16,215	13,677	14,160	14,487	15,438	14,406
Blenheim	8,736	9,696	9,249	11,037	12,069	14,010	13,644
Gisborne	12,810	13,275	10,449	11,313	11,595	12,765	12,477
Cambridge Zone	3,639	4,584	4,653	5,220	5,835	6,933	8,028
Te Awamutu Zone	4,467	4,803	4,266	4,914	5,277	6,186	6,123

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are suppressed and do not contribute to local employment.

Table 9: Employment in secondary urban areas

Urban area	Local employment by census year						
	1981	1986	1991	1996	2001	2006	2013
Timaru	11,676	11,991	10,101	10,770	10,884	12,096	12,267
Pukekohe	5,034	5,820	5,700	6,294	7,476	9,687	10,953
Taupo	5,982	7,221	6,789	8,046	8,616	9,939	9,783
Ashburton	5,844	6,210	5,784	6,432	6,951	7,863	8,652
Masterton	7,881	7,914	6,672	7,197	7,470	8,208	8,262
Whakatane	5,466	6,042	5,475	6,087	6,771	7,248	6,954
Queenstown	1,434	2,178	2,589	4,188	4,848	6,159	6,705
Rangiora	2,544	3,018	2,997	3,906	4,413	5,022	6,555
Levin	6,585	7,146	6,489	6,360	6,441	6,732	6,363
Feilding	4,584	4,926	4,791	5,355	5,160	5,790	6,042
Oamaru	5,319	5,226	4,575	4,749	4,644	5,094	5,040
Hawera	4,395	4,389	4,062	4,467	4,446	4,428	4,647
Greymouth	4,524	4,650	3,672	3,876	3,834	4,431	4,071
Tokoroa	7,680	7,551	5,559	4,758	4,644	4,218	3,696

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are suppressed and do not contribute to local employment.

Table 10: Employment in minor urban areas

Urban area	Local employment by census year						
	1981	1986	1991	1996	2001	2006	2013
Gore	4,878	4,566	3,909	4,248	4,140	4,182	3,933
Motueka	1,953	2,169	2,034	2,223	2,583	2,778	2,694
Morrinsville	1,740	1,989	1,833	1,965	2,145	2,421	2,571
Matamata	1,674	1,971	1,818	1,953	2,145	2,202	2,436
Te Puke Community	1,773	2,106	1,623	1,686	1,968	2,409	2,334
Thames	2,316	2,289	2,175	2,346	2,253	2,316	2,094
Stratford	1,884	1,905	1,443	1,653	1,569	1,677	1,857
Huntly	2,253	2,541	1,551	1,419	1,644	1,746	1,611
Dannevirke	1,878	2,025	1,770	1,656	1,689	1,845	1,515
Balclutha	1,830	1,716	1,467	1,521	1,563	1,542	1,473

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are suppressed and do not contribute to local employment.

Table 11: National employment by industry class

Code	Industry class	National employment by census year						
		1981	1986	1991	1996	2001	2006	2013
AA	Agriculture, forestry and fishing	41,517	45,036	36,444	46,545	45,750	41,034	38,598
BB	Mining	2,406	3,306	2,244	2,016	1,614	2,199	3,321
CC	Manufacturing	270,903	274,176	191,283	192,240	182,391	181,374	153,354
DD	Electricity, gas, water and waste services	14,454	15,909	11,016	9,024	6,936	7,578	10,974
EE	Construction	75,588	89,688	73,377	81,159	89,007	126,357	127,605
FF	Wholesale trade	56,766	69,822	62,052	79,521	83,823	85,848	83,460
GH	Retail trade and accomodation	173,439	198,093	186,585	215,604	230,034	267,210	261,939
II	Transport, postal and warehousing	60,165	51,906	55,251	58,200	60,972	69,459	68,535
JJ	Information media and telecommunications	54,870	59,190	36,321	32,031	32,826	33,852	31,758
KK	Financial and insurance services	42,303	48,825	51,699	48,501	47,472	58,227	62,277
LL	Rental, hiring and real estate services	9,126	17,469	19,860	31,740	33,453	44,904	39,402
MN	Professional, scientific, technical, administrative and support services	55,464	69,315	95,154	123,477	151,947	198,768	209,079
OO	Public administration and safety	68,787	70,656	76,329	73,014	67,842	71,430	85,317
PP	Education and training	73,011	70,554	80,673	90,663	109,866	123,603	139,716
QQ	Health care and social assistance	83,919	89,034	90,339	94,725	122,967	142,905	170,604
RS	Arts, recreation and other services	46,830	52,776	51,354	72,810	77,316	93,351	95,595

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are suppressed and do not contribute to national employment.

Table 12: National employment by occupation

Code	Occupation	National employment by census year						
		1981	1986	1991	1996	2001	2006	2013
1	Legislators	65,736	105,420	140,802	161,556	181,782	240,213	259,122
2	Professionals	126,648	140,160	147,585	165,987	209,499	255,684	296,211
3	Technicians and associate professionals	107,418	116,247	132,402	153,267	163,563	205,914	224,346
4	Clerks	196,047	225,609	177,921	189,942	188,982	188,151	170,556
5	Service and sales workers	159,456	168,930	156,690	194,220	208,896	230,505	234,744
6	Agriculture and fishery workers	49,515	52,143	39,819	47,241	46,425	42,033	39,510
7	Trades workers	156,744	163,020	128,622	122,862	118,548	136,482	118,692
8	Plant and machine operators and assemblers	146,769	147,243	108,375	107,502	113,634	117,273	102,225
9	Elementary occupations (inc. residuals)	121,215	106,983	87,765	108,693	112,887	131,844	136,128

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are suppressed and do not contribute to national employment.

Table 13: Included industry-occupation pairs

Industry		Occupation code								
Code	Name	1	2	3	4	5	6	7	8	9
AA111	Horticulture and fruit growing						X			
AA121	Sheep, beef cattle and grain farming						X			
AA131	Dairy cattle farming						X			
AA141	Poultry, deer and other livestock farming						X			
AA211	Forestry and logging						X			
AA310	Fishing and aquaculture									
AA320	Agriculture, forestry and fishing support services						X			
BB11_	Mining								X	
CC1_	Meat product manufacturing and seafood processing				X			X	X	X
CC131	Dairy product manufacturing								X	
CC141	Fruit, oil, cereal and other food product manufacturing	X			X	X		X	X	X
CC151	Beverage and tobacco product manufacturing									
CC211	Textile and leather manufacturing								X	
CC212	Clothing, knitted products and footwear manufacturing								X	
CC311	Wood product manufacturing							X	X	X
CC321	Pulp, paper and converted paper product manufacturing								X	
CC411	Printing				X			X		
CC5_	Petroleum, chemical, polymer and rubber product manufacturing	X	X	X	X			X	X	X
CC611	Non-metallic mineral product manufacturing								X	
CC711	Primary metal and metal product manufacturing								X	
CC721	Fabricated metal product manufacturing	X		X	X			X	X	X
CC811	Transport equipment manufacturing							X	X	
CC821	Electronic and electrical equipment manufacturing	X		X	X				X	X
CC822	Machinery manufacturing	X		X	X			X	X	
CC91_	Furniture and other manufacturing	X						X	X	X
DD1_	Electricity, gas, water and waste services			X	X				X	X
EE1_	Other construction	X	X	X	X			X	X	X
EE110	Residential building construction				X			X		X
EE113	Non-residential building construction							X		
FF11_	Machinery, equipment and motor vehicle parts wholesaling	X		X	X	X		X		X
FF110	Other goods and commission based wholesaling	X		X	X	X				X
FF111	Basic material wholesaling	X		X	X	X			X	X
FF114	Grocery, liquor and tobacco product wholesaling	X		X	X	X			X	X
GH111	Motor vehicle and parts retailing	X			X	X		X		X
GH112	Fuel retailing	X				X				

Table 13 (continued): Included industry-occupation pairs

Industry		Occupation code								
Code	Name	1	2	3	4	5	6	7	8	9
GH121	Supermarket and grocery stores	X			X	X				X
GH122	Specialised food retailing	X				X		X		
GH130	Other store based retailing; non store and commission based wholesaling	X	X	X	X	X				X
GH131	Furniture, electrical and hardware retailing	X		X	X	X		X		X
GH132	Recreational, clothing, footwear and personal accessory retailing	X			X	X				
GH133	Department stores				X	X				
GH21_	Accomodation and food services	X			X	X				X
II111	Road transport	X			X				X	X
II120	Other transport									
II121	Rail transport									
II123	Air and space transport			X		X				
II13_	Postal, courier transport support and warehousing services				X				X	X
JJ11_	Information media services			X	X					X
JJ120	Telecommunications services (inc. internet service providers)	X	X	X	X					
JJ123	Library and other information services				X					
KK1_	Other financial and insurance services	X		X	X					
KK110	Banking and financing; financial asset investing	X	X	X	X					
LL110	Rental and hiring services (exc. real estate); non-financial asset leasing									
LL12_	Property operators and real estate services			X	X					
MN11_	Professional, scientific and technical services		X	X	X					
MN21_	Administrative and support services		X	X	X	X				X
OO111	Local government administration		X	X	X					
OO21_	Public safety and defence				X	X				
OO211	Central government administration and justice	X	X	X	X	X				
PP11_	Education and training	X	X	X	X	X				X
QQ11_	Health care and social assistance	X	X	X	X	X			X	X
RS11_	Arts and recreational services			X	X	X	X			
RS210	Personal services; domestic household staff					X			X	
RS211	Repair and maintenance				X			X		
RS219	Religious services; civil, professional and other interest groups	X	X	X	X					

Table 14: Suppression rates by census year

Year	Total employment in our selection			National employment		
	Suppressed data	Unsuppressed data	Suppression rate (%)	Suppressed data	Unsuppressed data	Suppression rate (%)
1981	1,063,506	1,079,712	1.501	1,129,548	1,169,256	3.396
1986	1,155,735	1,171,155	1.317	1,225,755	1,266,735	3.235
1991	1,064,073	1,083,381	1.782	1,119,981	1,166,229	3.966
1996	1,191,972	1,209,462	1.446	1,251,270	1,299,777	3.732
2001	1,280,637	1,298,052	1.342	1,344,216	1,393,902	3.565
2006	1,472,448	1,489,365	1.136	1,548,099	1,599,579	3.218
2013	1,503,018	1,519,932	1.113	1,581,534	1,632,342	3.113

Notes: The data shown are based on 1981–2013 census usual resident employment counts in urban area-industry-occupation cells. Such counts are randomly rounded to base three. Cells with true counts below six are declared as missing in the suppressed data and so do not contribute to total employment.