Article:
Cicerone, G., McCann, P. and Venhorst, V.A. (2020) Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space. Journal of Economic Geography, 20 (1). pp. 293-316. ISSN 1468-2702

https://doi.org/10.1093/jeg/lbz001

Journal of Economic Geography (2019) pp. 1–24 doi:10.1093/jeg/lbz001

Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space

Gloria Cicerone, Philip McCann** and Viktor A. Venhorst***,†*

*Social Sciences Unit, Gran Sasso Science Institute, Via Michele Iacobucci, 2, 67100 L'Aquila, Italy
**University of Sheffield Management School, Conduit Road, Sheffield S10 1FL, UK
***Faculty of Spatial Sciences, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
†Correspondence to: *email*<v.a.venhorst@rug.nl>

Abstract

We adapt the product-space approach of Hausmann–Hidalgo et al. to the case of Italian provinces, examining the extent to which network connectedness and centrality of a province's exports is related to its economic performance. We construct a new Product Space Position (PSP) index which retains many of the Hausmann–Hidalgo et al. features but which is also much better suited to handling regional and provincial data. The PSP index is found to outperform other indices. Our comparison throws light on fundamental aspects of network-cognitive-distance-trade arguments. A better positioning in the export-network product space is indeed associated with better local economic outcomes.

Keywords: Relatedness, revealed comparative advantage, product space, regional development
JEL classifications: O11, R11, R12
Date submitted: 19 May 2017 **Editorial decision:** 30 December 2018 **Date accepted:** 5 January 2019

1. Introduction

The centrality, positioning and connectedness of a nation's tradeable sectors within global trade patterns are argued to be critical for a country's growth trajectories (Hausmann and Klinger, 2006; Hausmann et al., 2007; Hidalgo et al., 2007; Hidalgo and Hausmann, 2009), and similar arguments have also been put forward at the regional scale (Neffke et al., 2011). The underpinnings of this Hausmann–Hidalgo approach are based on widely held principles evident in fields such as economics, strategic management, international business and economic geography. Yet, while these approaches are useful for distinguishing between the development performance of rich, middle-income and poor countries, as we will demonstrate in this paper, the existing approaches not only have very limited powers to distinguish empirically between the development trajectories of different rich countries but they are even less well-adapted to examining the case of diversified regions within advanced economies. This would suggest that for such an approach to make a contribution to regional analysis in advanced economies, at the very least it would need to be adapted in a way which keeps the main underlying principles but does so in a more appropriate manner. Our research question is therefore, is the Hausmann–Hidalgo type of approach to trade centrality

and connectedness still useful for understanding the economic performance of advanced regions, and if so, can a significant adaptation of the existing frameworks better capture the economic performance of regions in advanced economies?

In order to answer this question, we investigated the development role played by the positioning and connectedness of a region's export patterns within the overall international trade system, over and above standard economic geography variables. Using province-level data from Italy, our analysis demonstrates that the existing Hausmann–Hidalgo types of approaches which are used to examine the performance of countries are less effective when discussing sub-national regional profiles in advanced economies. We therefore put forward a method for modifying the existing Hidalgo–Hausmann national-level indicators of trade network-relatedness and centrality (Hausmann and Klinger, 2006, 2007; Hidalgo et al., 2007; Hausmann et al., 2007; Hidalgo and Hausmann, 2009) in order to produce an index which is place-specific and much better suited to sub-national analyses. This new modified PSP index is shown to perform better than the existing Hausmann and Klinger (2006), Hausmann and Klinger (2007), Hidalgo et al. (2007), Hausmann et al. (2007) and Hidalgo and Hausmann (2009) indices, while still maintaining many of the features of the product-space method. Importantly, by using this new index we find that the original Hausmann–Hidalgo et al. type arguments do hold at the sub-national scale, even after controlling for more traditional regional growth factors.

This paper is structured as follows. Within the product space framework the next section discusses the interconnected ideas of relatedness, centrality and connectedness. By drawing on broader insights from other Hausmann–Hidalgo et al. papers we are then able to adapt and extend the methodological approach of Hausmann and Klinger (2006) to a wider context more suitable for addressing regional variations within advanced economies. We then apply our measure to an analysis of the economic and innovative performance of Italian provinces for the years 2007–12. Our analysis shows that in such a context this modified approach makes much more theoretical and empirical sense than the existing indices. Our findings demonstrate that a province's good positioning in the export network product space is indeed associated with enhanced regional development, over and above other more traditional regional economic variables such as variety, diversity, human capital and density.

2. Product and technological relatedness and network centrality

The product and network space arguments of Hausmann and Klinger (2006), Hidalgo et al. (2007) and Hidalgo and Hausmann (2009) suggest that within the overall global networks of trade countries which are represented relatively more in centrally located export activities are more likely to exhibit stronger growth and developments trajectories than countries which are more represented by the exporting of more peripheral products. This product-space approach is common to the arguments of Hausmann and Klinger (2006), Hidalgo et al. (2007) and Hidalgo and Hausmann (2009) and the conceptual foundations of the Hausmann–Hidalgo approach are 2-fold.

To begin with, their analysis posits that where two products or services share most of the same requisite production assets and capabilities, countries that export one will also tend to export the other. By the same token, goods or services that do not share many capabilities are less likely to be co-exported. As with the related variety literature

(Frenken et al., 2007; Boschma and Iammarino, 2009; Neffke et al., 2011) their fundamental conceptual ideas reflect the cognitive distance argument of Boschma (2005) in which it is assumed that greater cognitive proximity between products or services, defined in terms of the common production assets, competences and capabilities required, also offer greater possibilities for mutual technology transfer, learning and knowledge sharing. In turn all of these cognitive distance arguments originally derive from the various innovation-systems literatures (Iammarino and McCann, 2013). However, there are also fundamental differences in construction between the entropy-based related variety approach and the network-proximity approach of Hausmann–Hidalgo. The proximity indices measure the relatedness between two products by observing trade outcomes rather than the ex ante (sectoral classification) similarities between the products or inputs. Therefore, in contrast to the conventional related variety approach the new indicator is an ex post measure of relatedness, and should better capture all of the influences similarly affecting groups of industries. Indeed, the tentative evidence available suggests that the network proximity approach may actually perform better empirically than the conventional related variety approach (Boschma et al., 2012).

The product proximity index that Hausmann and Klinger (2006) propose is therefore a measure of the relatedness between pairs of products using cross-country export data. It is also a measure of the product-space distance between products, and one which avoids any priors as to the relevant dimensions of similarity. The similarity of requisite production assets and capabilities is revealed by the likelihood that where a country has a revealed comparative advantage (RCA) based on a Balassa Index (Balassa 1965) value of >1 in the exporting of one good, it will tend to have such an advantage in both goods.

Yet, these relatedness properties are themselves not sufficient to ensure strong development trajectories. Rather, the product-space framework also posits that countries with a revealed comparative advantage in groups of sectors which are centrally positioned within global trade networks will exhibit higher levels of economic development than those whose revealed comparative advantage is in sectors which are more peripherally positioned. The reason is that these products offer greater possibilities for technology transfer, learning and knowledge sharing. On average, core products are the most sophisticated and well-connected to the rest of the product space, and provide more opportunities to redeploy the capabilities that they embody, which facilitates the export of a large number of other products. The degree of centrality of a country's related exports in global trade networks is therefore critical in determining its long-term development trajectory, and the more centrally positioned are a country's exports the stronger will be its development trajectory.

Following Hausmann and Klinger (2006) and Hidalgo et al. (2007), it is possible to compute the proximity index between industry i and j by taking the minimum between the conditional probability of a region specializing in industry i given it specializes in industry j, and the conditional probability of a region specializing in industry j given it specializes in industry i, as follows (time subscript t suppressed for brevity throughout this introduction):

$$\phi_{i,j} = \min(P(x_i|x_j), P(x_j|x_i)), \tag{1}$$

where for any region or country c:

$$x_{i,c} = \begin{cases} 1 \ \textit{if} \ \text{RCA}_{i,c} \geq 1 \\ 0 \ \textit{otherwise} \end{cases} \tag{2}$$

and where the conditional probability is calculated using all regions (or countries). Since conditional probabilities are not symmetric we take the minimum of the probability of exporting product i given j and the reverse, to make the measure symmetric and more stringent.

One possible application of the proximity index can be found in the work of Hausmann and Klinger (2006). Firstly, they calculate a product i's centrality in the Product Space. A product that is more central in the Product Space will be connected to a greater proportion of the other products j, and therefore will have a higher value for centrality

$$C_i = \frac{\sum_j \phi_{i,j}}{J}. \tag{3}$$

This measure shows which goods are located in the dense part of the Product Space and which are located in the periphery by simply adding the row for that product in the matrix of proximities, and dividing by the maximum possible number of distance-weighted products J. Secondly, Hausmann and Klinger (2006) measure the density of the product space around the areas where different countries have specialized by calculating the average centrality of all products in which the country has comparative advantage. They also graph this variable against GDP per capita showing that in general, rich (poor) countries tend to be specialized in dense (sparse) parts of the product space. For convenience, we will call this index the 'Average Centrality' index

$$\text{AVG\_CENTR}_c = \frac{\sum_i (C_i{}^* x_{i,c})}{\sum_i (x_{i,c})}, \tag{4}$$

where for any region or country c:

$$x_{i,c} = \begin{cases} 1 \ \textit{if} \ \text{RCA}_{i,c} \ \geq 1 \\ 0 \ \textit{otherwise} \end{cases}. \tag{5}$$

The Hausmann–Hidalgo type of approach has been shown to be very effective in capturing the development performance across countries. However, when we apply this technique to regional data we get some very strange results. In order to demonstrate this in the case of Italy we use ISTAT international trade data (provided by the ISTAT Coeweb Section), disaggregated according to the Standardized International Trade Code at the three-digit level (SITC-3), providing the regional value share exported to the world for 118 product classes for each Italian province (NUTS 3) relative to the Italian national share. All of the export sectors in our regional trade dataset are manufacturing sectors, which in 2013 accounted for almost 82% of Italy's total exports (OECD, 2018a) and just under 29% of Italian GDP (OECD, 2018b). Applying

Equations (1) and (2) based on RCA ≥ cutoff[1] values, we calculate the proximity ϕ between product i and product j at year t, where the conditional probability is calculated using all Italian provinces P. We calculate these probabilities across 103 Italian provinces, for the period 2006–2013. As we have 118 industries in total in our dataset, we obtain a 118-by-118 matrix of proximities, which is common to all regions included in the analysis. Each row and column of this matrix represents a product and each off-diagonal element represents the proximity between a pair of products.

Applying the Hausmann–Klinger (AVERAGE CENTRALITY) methodology to the Italian provinces data for 2012 yields results which are rather curious.[2] Using the AVERAGE CENTRALITY index, we see that Italian provinces with higher values tend to be higher GDP regions ($\rho = 0.307$, $R^2 = 0.094$), but the relationship is very weak indeed. Moreover, many poorer southern Italian regions are ranked above rich areas such as Bolzano. A low income province such as Teramo is ranked above a high income province such as Padua, but this cannot be due to different specialization patterns because the same strange rankings are evident even between regions showing RCA in the same number of export sectors such as high income La Spezia and low income Sassari. The same picture is evident for other years of data. We have reported correlations for these other years, and for other indices which will be discussed further down, in Table 1.

The weak overall relationship between exports and provincial GDP per capita is not what we would expect from the Hausmann–Klinger types of arguments. Part of the problem is that the existing Hausmann–Klinger approach relies only on those products with a Balassa index of >1. The traditional Balassa index is asymmetrical and not homogeneous, in the sense that it varies between 0 and 1 for the cases of comparative disadvantage and between 1 and infinity, depending on the size the region, the country and the sector in question, for the cases of comparative advantage. Moreover, while considering only industry specializations (5) may be reasonable for low-income countries with few exporting sectors, in the case of an advanced economy with many intertwined sectors and agglomeration spillovers this misses much of the granularity of a region's economic fabric with multiple export sectors with Balassa values close to or below 1. Therefore, in a setting such as Italian provinces a more holistic approach is required which retains the basic AVERAGE CENTRALITY logic but which also takes account of the region's products which are both far and close to the well-connected core, as well as the products in which the local economy has both high and low RCA values. It is therefore necessary to move beyond a simple Balassa dummies weighting approach as in Equation (5).

1 Hausmann and Klinger (2006) and Hidalgo et al. (2007) built the dummies matrix applying RCA ≥ 1 in Equation (2). However, there is no widely accepted cut-off value that can explicitly delimit the specialization of an industry in a region (O'Donoghue and Gleave 2004, Cortinovis et al. 2017). Following Tian (2013), we compute a statistically significant cut-off value of the RCA for each industry in all Italian provinces. First, we calculate the Standardized Revealed Comparative Advantage (SRCA), as $\mathrm{SRCA}_{i,c} = \frac{\mathrm{RCA}_{i,c} - \overline{\mathrm{RCA}_{i,c}}}{\mathrm{std}(\mathrm{RCA}_{i,c})}$, where $\overline{\mathrm{RCA}_{i,c}}$ is the mean value of the RCA for industry i, and std(i, cRCA) is the standard deviation of the RCA for industry i. Second, we divide the SRCA into bootstrap samples for each industry. In particular, we re-sample with replacement 1000 times for each industry in order to obtain 1000 bootstrap samples, each having exactly the same length as the original sample of each industry. Third, for each bootstrap sample we use the sample mean of the 95th percentile as the estimate of the critical value at the 5% level of the true distribution. The advantage of this method is that it does not impose any assumptions in terms of the distribution of RCA.

2 Refer to Supplementary Data for scatterplots, Supplementary Figure S1.

Table 1. GDP correlations (2007–2013)

	2007		2008		2009		2010		2011		2012		2013	
	ρ	R^2	ρ	R^2	ρ	R^2	ρ	R^2	ρ	R^2	ρ	R^2	ρ	R^2
PSP EXP	0.66	0.43	0.66	0.44	0.65	0.42	0.67	0.44	0.66	0.44	0.67	0.44	0.65	0.42
PSP EMPL	0.25	0.06	0.33	0.11	0.44	0.19	0.42	0.18	0.36	0.13	0.43	0.18	0.37	0.14
Existing indices														
AVG CENTR	0.32	0.11	0.25	0.06	0.42	0.17	0.37	0.13	0.23	0.05	0.31	0.09	0.21	0.04
EXPY	0.70	0.50	0.66	0.43	0.62	0.39	0.63	0.40	0.63	0.40	0.63	0.40	0.67	0.45
LOCAL PRODY	0.78	0.62	0.74	0.55	0.74	0.55	0.81	0.66	0.77	0.59	0.80	0.64	0.78	0.62
LOCAL PRODYnoY	0.40	0.16	0.48	0.23	0.54	0.29	0.62	0.39	0.65	0.42	0.57	0.33	0.41	0.16
EXPYnoY	0.50	0.25	0.50	0.25	0.60	0.36	0.68	0.46	0.69	0.47	0.57	0.45	0.65	0.42
ECI	0.47	0.22	0.46	0.22	0.42	0.17	0.52	0.27	0.52	0.27	0.47	0.22	0.52	0.27
Standard controls														
VARIETY	0.65	0.42	0.65	0.42	0.65	0.43	0.67	0.45	0.64	0.42	0.67	0.45	0.67	0.45
DIVERSITY	0.37	0.14	0.31	0.10	0.33	0.11	0.31	0.09	0.29	0.08	0.30	0.09	0.30	0.09
GVApp	0.92	0.84	0.92	0.84	0.92	0.85	0.92	0.85	0.93	0.86	0.92	0.85	0.91	0.83
PAT	0.74	0.55	0.74	0.55	0.72	0.52	0.71	0.50	0.64	0.41	0.53	0.28		
POP	0.22	0.05	0.27	0.07	0.28	0.08	0.29	0.08	0.27	0.08	0.27	0.07	0.26	0.07
EDU	0.25	0.06	0.30	0.09	0.30	0.09	0.30	0.09	0.33	0.11	0.33	0.11	0.33	0.11
RD	0.05	0.00	0.11	0.01	0.18	0.03	0.18	0.03	0.17	0.03	0.25	0.06	−0.02	0.00
ADV SECT	0.58	0.34	0.62	0.39	0.61	0.37	0.62	0.38	0.54	0.29	0.59	0.35	0.55	0.30

Our method for doing this is first to calculate the measure of a product *i*'s centrality in the Product Space in time *t*, using Equation (3), and then we weight these values using a RCA definition which overcomes the limitation described above. Formally, we define the PSP of a local economy *p* as the sum of product *i*'s centralities in the Product Space weighted with the Revealed Symmetric Comparative Advantage RSCA values of province *p* for product *i*:

$$\mathrm{PSP}_p = \sum_i (C_i * \mathrm{RSCA}_{i,p}), \tag{6}$$

where RSCA values are constructed according to the approach of Iapadre (2001). The RSCA formula proposed by Iapadre (2001) is a variant of the one proposed by Dalum et al. (1998) and solves all statistical problems. The index used is the following:

$$\mathrm{RSCA}_{i,p} = \frac{(\mathrm{RCA}_{i,p} - \mathrm{RCA}_{j,p})}{(\mathrm{RCA}_{i,p} + \mathrm{RCA}_{j,p})}, \tag{7}$$

with

$$\mathrm{RCA}_{i,p} = \frac{\left(\frac{X_{i,p}}{X_p}\right)}{\left(\frac{X_{i,r}}{X_r}\right)}, \tag{8}$$

and

$$\mathrm{RCA}_{j,p} = \frac{\left(\frac{X_{j,p}}{X_p}\right)}{\left(\frac{X_{j,r}}{X_r}\right)}, \tag{9}$$

where p =province, i =product, r =total of other provinces net of p and j =total of the other products (net of i). This specialization of the value of exports (X) indicator varies between −1 and 1. Positive (negative) indicate advantages (disadvantages) compared with other Italian regions. Strictly speaking we use 1 + the Iapadre index to facilitate visualization within the network diagrams, and to simplify estimation. This framework implicitly assumes homogeneity within a product space but in a rich-country regional context we content that this is preferable to imposing an arbitrary discontinuity at a value of unity.

This new PSP index displays several more desirable properties than the AVERAGE CENTRALITY index when applied to the Italian provincial data. Firstly, the PSP index is seen to display a much stronger positive correlation ($\rho = 0.665$) than the AVERAGE CENTRALITY index and also produces sensible results in terms of the rank-ordering of regions, with the highest provincial PSP value is that of Milan at 4.759 while the minimum value is now that of Siracusa in Sicily at 0.369, with an overall mean for Italy of 2.328.[3]

Moreover, as an example, we are also able to draw here the export network-positioning of the provinces of Sassari and La Spezia using both indices. Figures 1 and 2 depict the network centrality and positioning of both Sassari and La Spezia, respectively, using the Hausmann–Klinger (AVERAGE CENTRALITY) approach, while Figures 3 and 4 depict their respective positioning using the PSP index. Because of the density of the networks possible in a 118–118 matrix, the complete network structure for each province looks like a hairball. Therefore, Hidalgo recommends that a good rule of thumb is to ensure that the average connectivity is not much more than four or five links per node (Hidalgo et al., 2007, 2009). In order to simplify the visual images, in each of these cases we therefore only depict those linkages with a cut-off value of at least 0.35. We also added 1 to the RSCA values just to improve the network visualization, with node size representing the RSCA value. Node gray shade represents the value for the centrality with darker shades being more centrally located sectors.

As already mentioned, although Sassari and La Spezia are very different provinces in terms of the levels of economic development, these two provinces have exactly the same number of sectors with RCA values greater than their respective cutoff values. Yet, what becomes clear from the visual network structure presented in Figures 3 and 4 is that it is very difficult using the AVERAGE CENTRALITY index to identify differences between these two provinces, even though they are very different economically. In contrast, the PSP approach clearly distinguishes between these two regions with La Spezia exhibiting far more sectors with a major presence in the center of the trade networks than Sassari. The PSP clearly captures these relationships very well and much better than AVERAGE CENTRALITY. Sassari has far fewer sectors with a major presence in the center of the global trade networks whereas La Spezia has a much greater presence in these central placings, as would be expected from a richer province. Our approach therefore moves beyond the existing approach because provinces with similar number of RCA sectors but with different network configurations will display different PSP values, and similarly provinces with similar network centrality values but with different RCA values will also display different PSP values. Our PSP findings as a whole therefore show that in general, richer (poorer) provinces tend to be specialized in

3 Refer to Supplementary Data, Supplementary Figure S2.

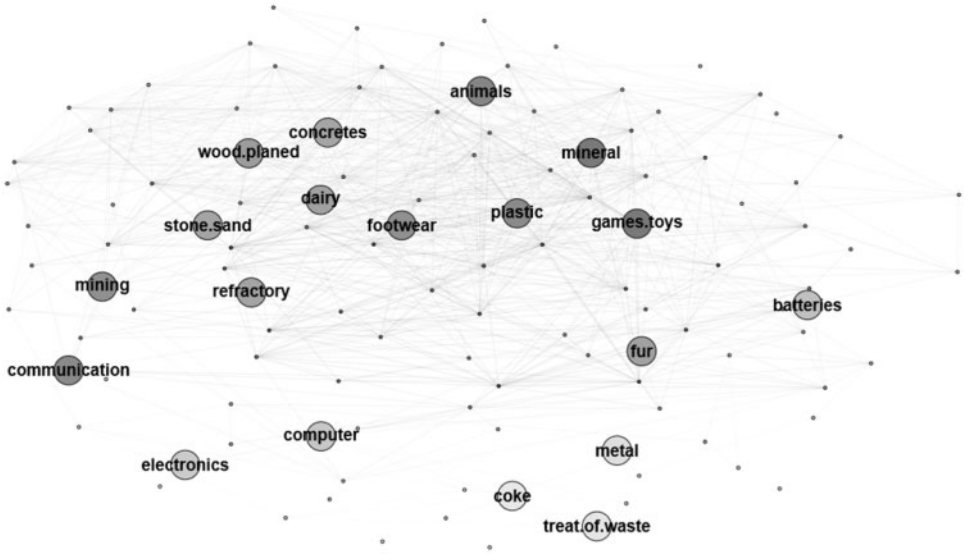

Figure 1. The network positioning for Sassari province using the AVERAGE CENTRALITY index.

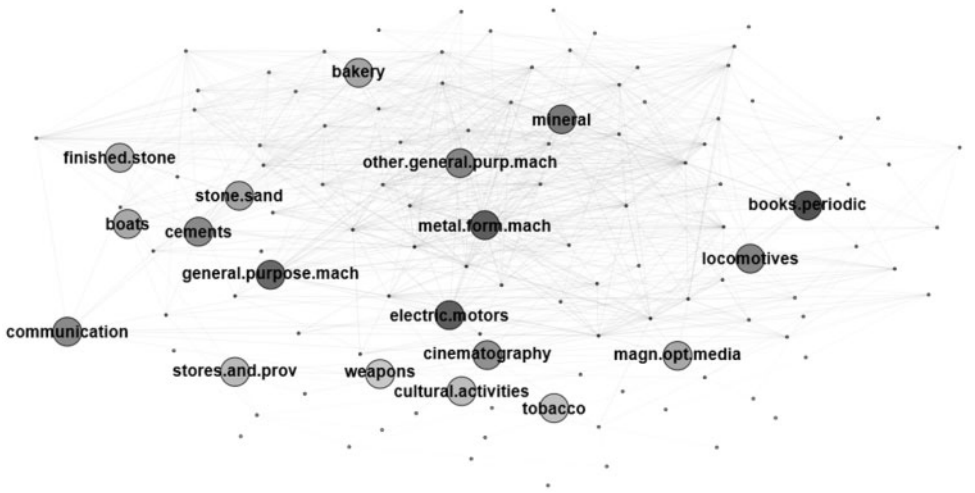

Figure 2. The network positioning for La Spezia province using the AVERAGE CENTRALITY index.

dense (sparse) parts of the product space, and therefore display a high (low) value of PSP.

However, this is not the end of the story, because the Hausmann–Hidalgo et al. tradition also produces other indices of trade and connectivity designed to analyze different contexts, which we will benchmark PSP against. When we consider the performance of these indices at the regional level it uncovers some other important analytical and conceptual issues which need to be addressed.

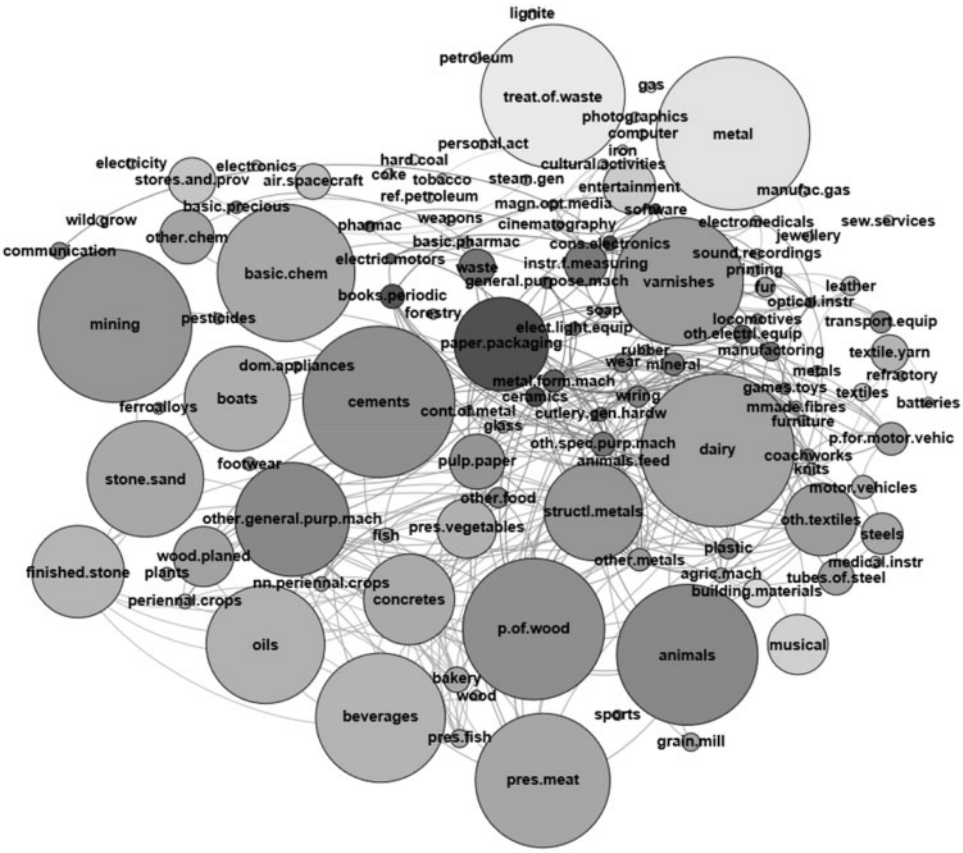

Figure 3. The network positioning for Sassari province using the PSP index.

Hausmann, Hidalgo and their co-researchers develop various different indices aimed at capturing other aspects of these development processes and relationships, and the most important of these are PRODY and EXPY. Hausmann et al. (2005, 2007) developed a measure, called EXPY, which aims to capture the productivity level associated with a country's exports. In order to calculate EXPY, first they construct an index called PRODY which is a measure of the sophistication of a product. Formally, this index is a weighted average of the per capita GDPs of countries exporting a given product, and thus represents the associated income/productivity level for each good:

$$\text{PRODY}_i = \sum_p (s_{i,p} * y_p), \tag{10}$$

where y_p stands for the real per capita GPD of the p-th ($p = 1, 2 \ldots N$) country (province in our case) exporting in sector i, while the weight:

$$s_{i,p} = \frac{\text{RCA}_{i,p}}{\sum_p \text{RCA}_{i,p}} \tag{11}$$

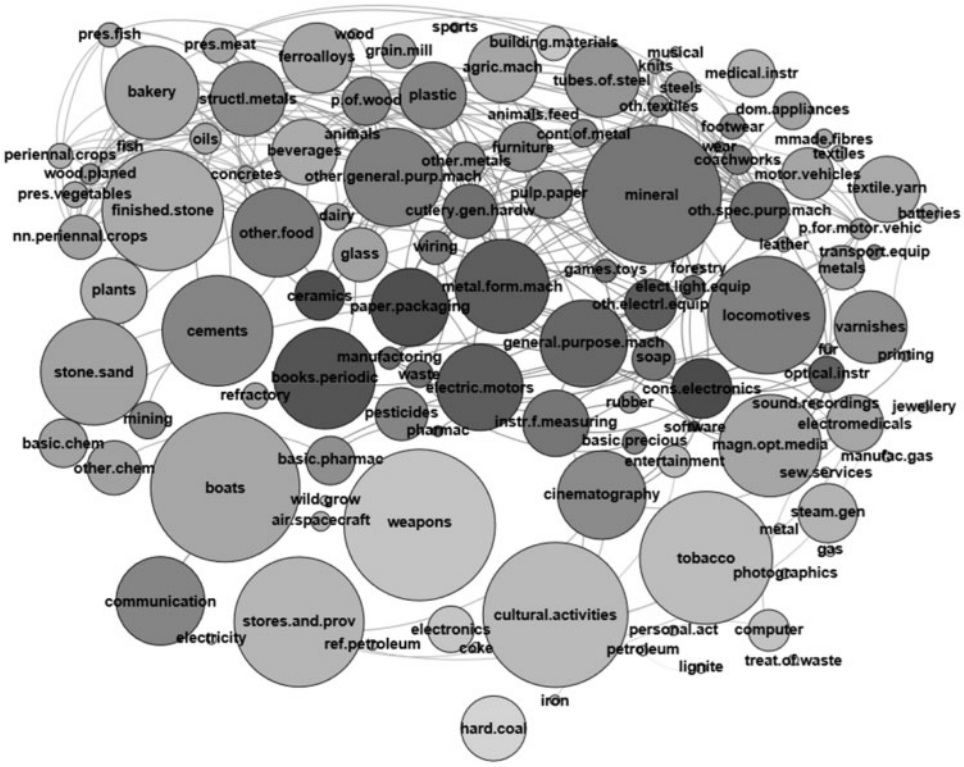

Figure 4. The network positioning for La Spezia province using the PSP index.

normalizes[4] country *p*'s Balassa index of RCA with respect to those of all the countries exporting in the same sector (Rodrik, 2006; Di Maio and Tamagni, 2008). The PRODY index is thus a sectoral measure returning a weighted average of the levels of development (proxied by per-capita income) of all the countries producing and exporting in a given sector. By construction, sectors with high values of PRODY are those where high income countries play a major role in world exports, displaying strong specializations where comparative advantages are determined by factors other than labor cost.

The EXPY index[5] is then in turn defined as the weighted sum of the PRODY indexes of all the sectors *i* wherein a country is exporting, with weights given by the share of

4 We computed two versions of EXPY. Following Hausmann et al. (2005, 2007) we compute PRODY and then EXPY without normalization and we find there is not a robust relationship between that EXPY and GDP per capita at regional level GDP. Following Rodrik (2006) and Di Maio and Tamagni (2008), and also after a discussion with Cesar Hidalgo, we compute PRODY and then EXPY with the normalization through the *s* weight. We find there is a robust relationship between that EXPY and GDP per capita GDP even at the regional level. In this study we use, for convenience of exposition, the words 'PRODY' and 'EXPY' to refer to their normalized versions.

5 A shortcoming of the EXPY indicator used by those authors is that it does not take into account the quality differences within exported products across countries (Minondo, 2010). In order to overcome this limitation, Minondo (2010) develops a new quality-adjusted EXPY indicator. His work shows that, once quality differences within products are taken into account, there is not a robust relationship between EXPY and subsequent growth even at the national level.

each sector in the export vector of the country p. It represents the productivity level associated with country p's export basket. Formally:

$$\mathrm{EXPY}_p = \sum_i \left(\frac{x_{p,i}}{X_p}\right) * \mathrm{PRODY}_i. \tag{12}$$

Hausmann et al. (2007) find a positive and robust relationship between EXPY, that is the productivity level associated to a country's exports, and subsequent economic growth. However, when we apply EXPY indices at the provincial level we see that, generally, the correlations (reported in Table 1) are slightly less strong than for the PSP index, and with still some strange observations such as low income Lecce having a higher value than high income Rome.[6]

However, Hidalgo et al. (2007) propose a further measure to summarize the position of a country in the product space. They average the PRODYs of the top N products of a country's export basket after M diffusion steps at ϕ_0 and denoted it by $<\mathrm{PRODY}>^{N}_{M.\ \phi}$. Following the Hidalgo et al. (2007) logic we average the PRODYs of the top $N = 6$ products[7] of a province's export basket and in our analysis we call this index LOCAL_PRODY. Applying the Hidalgo et al. (2007) LOCAL_PRODY methodology to the Italian provinces yields a strong correlation ($\rho = 0.800$ for 2012), which apparently makes it the best performing index.[8]

However, the use of GDP per capita income information in the creation of PRODY and EXPY is problematic in that given the definition, sectors with high values of PRODY are, by construction, those where high income countries play a major role in production, relative to the other participants in world exports in that sector. As a result, the observation that 'rich countries export rich country goods' is close to being a circular argument (Hidalgo, 2009) and this is also the case for regions. In order to try to answer this critique, it is therefore possible to separate the information on income Y from the information on network structure (RCA) in PRODY and EXPY. The contribution of income Y from PRODY and EXPY can be removed explicitly from their definitions in two steps (Hidalgo, 2009; Hidalgo and Hausmann, 2009). First, they set RCA = 1 if RCA is larger than a certain RCA* threshold. This is a simple way to build bipartite networks in which countries are connected to the products they export. Mathematically, they represent this network using the adjacency matrix Mi,c,[9] where $Mi,c = 1$ if country c is a significant exporter of product i and 0 otherwise. Finally, they make Y_c equal to the number of connections, or degree ($k_{c,0}$), that country has in this network. $K_{c,0}$ is therefore a measure that comes only from the structure of the network. Mathematically, these transformations are:

$$M_{i,c} = 1 \text{ if } \mathrm{RCA} \geq \mathrm{RCA}^*, \tag{13}$$

$$Y_c = k_{c,0}, \tag{14}$$

6 Refer to Supplementary Figure S3 in the Supplementary Data.
7 We also average the PRODYs of the top $N = 20$ products of a province's export basket and results are similar.
8 Refer to Supplementary Figure S4 in the Supplementary Data.
9 Following Hidalgo and Hausmann (2009), we avoid confusing the weighted and unweighted versions of the network by referring to the unweighted version as Mi,c while we continue to use RCA for the weighted version.

where $k_{c,0}$ is given by:

$$k_{c,0} = \sum_i M_{i,c}, \tag{15}$$

and represents the *diversification* of country c (the number of products that the country makes). Additionally, they define the degree, or *ubiquity*, of a product in this network as

$$k_{i,0} = \sum_c M_{i,c}. \tag{16}$$

They refer to k_i as the *ubiquity* of a product, as it is the number of countries that export that product. Hidalgo and Hausmann (2009) show that measures of knowledge complexity for both countries and products can be found by sequentially combining these measures of diversity and ubiquity in the following two equations over a series of n iterations.

When they apply these transformations to the definition of PRODY, they find that after removing the contribution of income, PRODY reduces to the average nearest neighbor degree of a product in the network, which they denote as $k_{i,1}$, where the 1 subscript is used to indicate that this is the average degree of the nodes that are at distance 1 from product i. Formally, this is

$$k_{i,n} = \frac{1}{k_i} \sum_c M_{i,c} k_{c,n-1}. \tag{17}$$

Similarly, these transformations take EXPY into a weighted average of the degree of nodes at distance two in the network of country c. Formally, this is

$$k_{c,n} = \frac{1}{k_c} \sum_i M_{i,c} k_{i,n-1}. \tag{18}$$

Hidalgo (2009) compares PRODY and EXPY with their pure network counterparts after having removed the income information, which we call here PRODYnoY and EXPYnoY. The R-squared values reported by Hidalgo (2009) are 0.51 and 0.75, respectively. Although there are significant differences in cross-country dispersion among sectors, these correlations suggest that in the case of PRODY half of the index information comes from just the network structure PRODYnoY while from EXPY three-quarters of the index information comes just from the network structure EXPYnoY connecting countries to the products they export. However, careful observation of Figure 1 from Hidalgo (2009) also shows that the relationship between EXPY and EXPYnoY for high income OECD countries is approximately zero, which casts doubt on the performance of these indices in high income contexts.

In terms of correlations, both EXPYnoY[10] and LOCAL_PRODYnoY perform less well than the PSP in 2012.[11] As such, these observations require us also to consider whether the network-only aspects of these indicators provide a useful test against which we can benchmark the performance of the PSP index.

10 EXPYnoY performs very slightly better than the PSP just in 2010 and 2011 (Table 1).
11 Refer to Table 1 and Supplementary Figures S5 and S6 in the Supplementary Data.

In order to do this we employ the so-called 'Method of Reflections' (Hidalgo and Hausmann, 2009) which allows us to extract relevant information about the availability of capabilities in a country. Here, we calculate two values for this with $n = 2$, which is the number of iterations (or reflections) Hidalgo (2009) uses, and $n = 12$, where n is the number of iterations used to calculate EXPYnoY (i.e., for $K_{c,n}$, where $n = 2$) and the *Economic Complexity Index* (*ECI*) (i.e., for $K_{c,n}$, where $n = 12$). The ECI is a measure of the knowledge intensity of economies and products that can be computed from trade data (Pinheiro et al., 2018). ECI is $K_{c,n}$ with n going to infinity (Hidalgo and Hausmann, 2009). Each additional iteration in $K_{c,n}$ provides a finer-grained estimate of the knowledge complexity of a region using information on the complexity of the product in which the region exhibits specialization. Although higher-order iterations in this technique become progressively more difficult to define, the method of reflections provides more and more precise measures of the ECI and PCI (Product Complexity Index)[12] as the noise and size effects are eliminated. The iterations are stopped when the ranking of regions and products is stable from one step to another (i.e., no further information can be extracted from the structure of the region-product network). For practical purposes, Hidalgo suggests to take $n \geq 12$ as being large enough. We stop our iteration process exactly at $n = 12$ as all of the ECI provincial values converge at this level. Following Hidalgo and Hausmann (2009) we extract information from the tiny deviations of these converging values. Plotting the provincial ECI index with respect to provincial GDP per capita yields a moderate positive correlation ($\rho = 0.469$ for 2012).[13]

In what follows, we benchmark our PSP index against a selection of these Hausman–Hidalgo indices. In our main econometric specifications we elect to use EXPYnoY, which correlated strongly with GDP, allowing us to test PSP against a measure with strong network-only characteristics. Furthermore, we report additional specifications using EXPY and ECI in the Supplementary Data document accompanying this article.[14] These indices were selected because of their strong correlations with GDP as well as their characteristics, thus providing the strongest benchmarks.

3. Econometric model, data and variables

In order to identify the extent to which the PSP of a region's tradeables network structure is related to the region's overall economic performance we also need to control for other local area characteristics. Using measures of a region's GDP per capita and also of its innovative performance we examine the extent to which PSP affects these outcomes over and above the standard urban and regional economic indicators on variety, diversity, human capital and agglomerative capacity.

In what follows we introduce our data and empirical specification, and we discuss our results followed by a robustness check where we consider alternative specifications of

12 PCI (Product Complexity Index) here is denoted as PRODYnoY when $n = 1$.

13 See Table 1 and Supplementary Figure S7 in the Supplementary Data.

14 Other specifications, using (LOCAL_)PRODY(noY) and AVERAGE_CENTRALITY, even though they gave very similar results, are not reported here. As noted earlier, AVERAGE_CENTRALITY gives some unexpected regional rankings, PRODY is essentially not a local indicator, and the modified LOCAL varieties either are very close to GDP by construction or show lower correlations than the envisaged end-product EXPYnoY, the latter therefore providing a stronger benchmark.

our agglomeration and density control variables as well as the alternative PSP formula PSP(EMPL) which is calculated using employment rather than trade data. In all of our model forms, we repeated the model with different independent trade network-related variables and sequentially included and removed each individual variable in order to check the change in goodness of fit associated with each variable. The specification using EXPYnoY is reported here. Specifications using EXPY and ECI can be found in the Supplementary Data.

We thus wish to ascertain the predictive power of PSP on GDP, and also on provincial productivity and innovative capacity, while benchmarking PSP against other indices, and, furthermore, checking its predictive power over and above standard explanatory variables. We therefore use three dependent variables, which in our various econometric specifications are also treated as independent variables. First, as a proxy for the economic prosperity of each province we use the per capita annual gross domestic product GDP per capita derived from the OECD regional database, available from 2001 to 2014, denoted here as GDP. Second, the model includes a labor productivity index defined as annual gross value added GVA per employed worker provided from the ISTAT local database, available for the years 2003–2014 and denoted as GVA. Third, as a measure of the innovation performance of each province we use patenting activity per capita PAT. In particular, we use the number of patent applications to the European Patent Office available from 2001 to 2012, classified by the inventors' residence.

We aim to identify whether PSP is also an important independent variable related to overall local economic prosperity GDP, local labor productivity GVA, local innovation PAT, over and above the other more conventional control variables used in urban and regional economic analysis.

We use the following standard control variables. As an indicator for the degree of the structural concentration of a local economy, we use the reciprocal of the Gini concentration coefficient VARIETY:

$$\mathrm{VARIETY} = \frac{1}{2\frac{\sum_{k=1}^{n} kE_k}{(n-1)\sum_{k=1}^{n} E_k} - \frac{n+1}{n-1}}, \tag{19}$$

where E_k is the sum of employees (E) for sector k, with sectors listed in increasing order. Given that the Gini coefficient is a measure of concentration, an increase of its reciprocal implies that the levels of provincial sectoral concentration are lower. Employment data are provided by the ISTAT statistical register ASIA, which is the Statistical Register of Active Enterprises, available for the period 2007–2014. In particular, we use employment data provided by the business register of local units. A local unit is defined by the Council Regulation on statistical units (N. 696/1993) as 'an enterprise or part thereof (e.g., a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place'. The ASIA-Local Units register provides information on location of the local unit, economic activity and the number of employees.

The measure of provincial specialization and diversity[15] at the local level that we use is given by the Duranton and Puga index DIVERSITY (available for the period 2007–2014). As with Duranton and Puga (2000), the degree of variety is measured by

15 There is currently no theoretical or empirical consensus on the role played by specialization versus diversity in economic development (De Groot et al., 2016) and indeed one of the advantages of the

summing for each province, over all sectors, the absolute value of the difference between each sector share on local employment and its share on national employment. Formally, it leads to:

$$\text{DIVERSITY} = \frac{1}{\sum_{p=1}^{118} |s_{k,p} - s_{k,c}|}, \tag{20}$$

where

$$s_{k,p} = \frac{E_{k,p}}{E_p}, \tag{21}$$

and

$$s_{k,c} = \frac{E_{k,c}}{E_c}, \tag{22}$$

and where $E_{k,p}$ is the employment in sector k in province p; E_p is the total employment in province p; $E_{k,c}$ the national sector employment in sector k and E_c is the total national employment.

In the typical regional production function approach, the innovative output of a region is also often argued to depend upon the level of research and development activities within the local economy. Therefore, we include a measure of the level of research and development activities RD defined as the level of provincial R&D employment divided by the total employment of each province.

The model also includes a variable ADV_SECT which reflects the provincial share of advanced tertiary sector employees relative to all employees of each province. The advanced tertiary sector of the economy includes organizations specialized in IT, marketing, research and development and legal, technical and financial consulting. We calculate this indicator ADV_SECT after excluding the share of employment on research and development sector.

The data for E, RD and ADV_SECT are all provided by ASIA-Local Units database, available for the period 2007–2014.

We also include a variable EDU, which is the share of the provincial population with a higher education (defined as a bachelor's degree or master's degree) as a proxy for the general quality of human capital. We use data provided by the Italian Ministry of Education, University and Research statistical section, collected with respect to the location of Universities.

Finally, we test whether urbanization economies matter by considering whether more densely populated provinces show higher levels of economic prosperity and innovation. To capture urbanization economies we take the population density of each province, that is, the number of inhabitants per squared kilometer POP, as derived from the OECD Regional Demographic Statistics, available for the period 2002–2014. The unstandardized sample statistics are reported in Table 2. We run our analysis

related variety literature is to chart a pathway through this blockage and to potentially reconcile often competing approaches. Our approach also offers further options for pushing these debates forward.

Table 2. Sample statistics of the variables

	Obs. period	Mean	Median	Std. dev.	Min	Max
GDP	2001–14	25,580	25,775	6577	13,741	51,837
PAT	1998–2012	6.74191e−05	4.94623e−05	6.76392e−05	0.0000	0.00072166
GVA	2004–13	55,493	55,632	6707	40,232	77,381
PSP	2006–13	2.335	2.364	0.850	0.369	4.854
EXPY	2006–13	24,056	24,307	1165	20,105	27,477
EXPYnoY	2006–13	26.608	26.144	2.595	20.559	33.473
LOCAL_PRODY	2006–13	1048	1049	79.747	814.753	1540
LOCAL_PRODYnoY	2006–13	27.120	26.358	3.904	19.715	46.063
AVG_CENTR	2006–13	0.002	0.002	0.001	4.06971e−05	0.009
VARIETY	2007–14	1.466	1.474	0.075	1.292	1.704
DIVERSITY	2007–14	0.027	0.026	0.006	0.014	0.048
EDU	2003–14	0.385	0.224	0.384	0.000	2.187
RD	2007–14	0.122	0.086	0.134	0.009	1.985
ADV_SECT	2007–14	15.175	14,719	2.940	10.234	30.658
POP	2002–14	245.712	172.078	327.762	36.588	2648

Note: Variables entered in standardized form in the models.

using the data for 2007–2012, which is the longest panel for which we have all desired variables available.

Clearly, industries which are successful exporters and which are central to the product space may relocate toward, or get started in, provinces with high levels of GDP and innovative capacity. In general, complex interactions may exist between our PSP and our three measures of provincial economic success. In our econometric approach, we therefore have to be mindful of reverse causality and simultaneity. We employ a Seemingly Unrelated Regression (SUR) analysis, which allows us to take potential correlations and interactions between the dependent variables into account. We include PSP as a fourth dependent variable, to this end. Furthermore, we take advantage of the panel structure of our data. We use lagged values for the independent variables and contemporaneous values for the dependent variables. All the independent variables are lagged by 1 year, except for our education and density control variables (EDU and POP) which are lagged 4 years, capturing the more structural characteristics of the provincial economy. The lag length of 4 years is a pragmatic choice, given data availability.[16] We acknowledge that a full causal interpretation of the PSP on provincial economic development would require, for example, a more fully fledged instrumental variables approach. However, we considered that having to find instruments for the very index whose properties we want to test would be counter to the scope of the present paper.

We adopt the following SUR[17] model with period-fixed effects, in which we jointly estimate the following equations with provincial GDP per capita GDP, local labor

16 We have also estimated a deep lag version, with all right-hand side variables lagged back 5 years. The results are very consistent with those presented here and are available upon request.

17 We preliminarily tested models adopting the pooled ordinary least squares (OLS). These models were rejected against the SUR models reported in this paper. The SUR correlation matrix shows that indeed there is correlation between the equations. The results are available upon request.

productivity—provincial gross value added per worker GVA, our innovation measure patents per capita PAT and the PSP all as dependent variables[18]:

$$\begin{aligned}\mathrm{GDP}_{p,t} = {} & \beta_0 + \beta_1 \mathrm{PSP}_{p,t-1} + \beta_2 \mathrm{PAT}_{p,t-1} + \beta_3 \mathrm{GVA}_{p,t-1} + \beta_4 \mathrm{EXPYnoY}_{p,t-1} \\ & + \beta_5 \mathrm{VARIETY}_{p,t-1} + \beta_6 \mathrm{DIVERSITY}_{p,t-1} + \beta_7 \mathrm{EDU}_{p,t-4} \\ & + \beta_8 \mathrm{POP}_{p,t-4} + \beta_9 \mathrm{RD}_{p,t-1} + \beta_{10} \mathrm{ADV\_SECT}_{p,t-1} + \lambda_1 \mathrm{d}t2 \\ & + \lambda_2 \mathrm{d}t3 + \lambda_3 \mathrm{d}t4 + \lambda_4 \mathrm{d}t5 + \lambda_5 \mathrm{d}t6 + \epsilon,\end{aligned}$$

(GDP equation in Reg 1.4)

where t denotes 1-year intervals, p denotes the province, ϵ denotes the error term and GDP, GVA, PAT, PSP, EXPYnoY,[19] VARIETY, DIVERSITY, EDU, POP, RD and ADV_SECT are the set of variables. We control for period-specific unobserved shocks by entering year-dummies, with 2007 being the reference year. In our analysis we consider 103 out of a possible 110 NUTS 3 provinces.[20]

At the local level there are conceptual problems linking employment and trade data (Boschma and Iammarino, 2009; Neffke et al., 2011; McCann, 2013), so as a robustness check we also use employment data instead of export data to compute an alternative measure of the PSP index. We use ISTAT employment data, providing the regional employment value for 80 sector classes for each Italian province (NUTS 3) relative to the Italian national share. Some 28 sectors are manufacturing and 52 sectors are services, accounting for 22% and 78% of total employment, respectively. This allows us to construct a network connecting all sectors of the economy, including non-tradeable service sectors. In order to compute the PSP, we calculate the proximity ϕ between all the 80 sectors at year t, across all the 110 Italian provinces, for the period 2007–2014. Plotting the provincial PSP index computed using employment data PSP(EMPL) against provincial GDP per capita displays a weaker positive correlation ($\rho = 0.428$) than the PSP index computed using export data.[21] Moreover, again this gives strange rank ordering with a higher value for low income Sassari than for higher income La Spezia and without the presence of Rome and Milan the correlation would be significantly lower. This suggests that the Product Space Index PSP constructed from export data is far superior to the PSP based on local employment data.

18 GVA, PAT and PSP equations likewise.

19 We checked EXPY and ECI as well as EXPYnoY. All the results are similar and are reported in the Supplementary Data file (Supplementary Tables S1–S3).

20 We excluded the new Italian provinces of Monza della Brianza, Fermo, Barletta-Andria-Trani, Carbonia Iglesias, Ogliastra, Medio Campidano, Olbia-Tempio. Monza della Brianza was officially created by splitting the north-eastern part from the province of Milan on 2004, and became executive in 2009. Fermo is a province in the Marche region of central Italy. It was established in 2004 and became operational in 2009. The Province of Barletta-Andria-Trani is a province of Italy in the Apulia region. The establishment of the province took effect in 2009, and Andria was appointed as its seat of government in 2010. Carbonia Iglesias, Ogliastra, Medio Campidano and Olbia-Tempio are provinces in the autonomous region of Sardinia. The formation of these province was announced in 2001 by the Autonomous Region of Sardinia and it officially became executive in 2005.

21 See Supplementary Figure S8 in the Supplementary Data.

Table 3. Seemingly unrelated regressions

GDP	Reg 1.1	Reg 1.2	Reg 1.3	Reg 1.4
PSP	0.122***	0.088***	0.139***	0.100***
PAT	0.218***	0.225***	0.196***	0.202***
GVA	0.756***	0.769***	0.720***	0.724***
EXPYnoY		0.008		0.036
VARIETY			−0.010	0.009
DIVERSITY			−0.066***	−0.056**
EDU			0.044**	0.043**
POP			−0.076***	−0.066***
RD			0.021	0.017
ADV_SECT			0.096***	0.099**
2008	−0.145***	−0.146***		
2009	−0.270***	−0.269***	−0.147***	−0.147***
2010	0.049	0.057	0.150***	0.161***
2011	−0.071	−0.052	0.030	0.100
2012	−0.189***	−0.171**	−0.084*	−0.020
cons	0.066*	0.057	−0.042	−0.075
R-sq	0.858	0.858	0.870	0.872
GVA	Reg 1.1	Reg 1.2	Reg 1.3	Reg 1.4
PSP	0.042**	0.041*	−0.003	0.009
PAT	−0.072***	−0.073***	−0.058**	−0.061**
GDP	0.951***	0.965***	0.955***	0.963***
EXPYnoY		−0.034		−0.046
VARIETY			0.014	0.019
DIVERSITY			−0.015	0.012
EDU			−0.068***	−0.068***
POP			0.044**	0.044**
RD			0.011	0.011
ADV_SECT			−0.010	−0.013
2008	−0.177***	−0.171***		
2009	−0.398***	−0.391***	−0.210***	−0.211***
2010	0.020	0.018	0.217***	0.204***
2011	−0.012	−0.071	0.188***	0.100
2012	−0.208***	−0.262***	−0.004	−0.084
cons	0.030	0.054	−0.167***	−0.126**
R-sq	0.842	0.841	0.847	0.847
PAT	Reg 1.1	Reg 1.2	Reg 1.3	Reg 1.4
PSP	0.170***	0.184***	0.209***	0.243***
GDP	0.981***	1.015***	0.912***	0.944***
GVA	−0.385***	−0.393***	−0.344***	−0.354***
EXPYnoY		−0.107		−0.150*
VARIETY			0.292***	0.306***
DIVERSITY			−0.196***	−0.205***
EDU			0.070	0.068
POP			−0.105**	−0.102**
RD			−0.043	−0.044
ADV_SECT			−0.211***	−0.217***
2008	−0.083	−0.063		
2009	−0.041	−0.023	0.063	0.061
2010	0.087	0.077	0.208*	0.164
2011	0.100	−0.086	0.230**	−0.059
2012	−0.083	−0.089	0.318***	0.056
cons	−0.048	0.030	−0.188***	−0.053
R-sq	0.447	0.445	0.476	0.475

(continued)

Table 3. continued

PSP	Reg 1.1	Reg 1.2	Reg 1.3	Reg 1.4
GDP	0.501***	0.305***	0.500***	0.313***
PAT	0.190***	0.174***	0.174***	0.178***
GVA	0.066	0.061	−0.087	−0.037
EXPYnoY		0.621***		0.498***
VARIETY			−0.014	−0.067
DIVERSITY			0.278***	0.280***
EDU			−0.057	−0.046
POP			0.338***	0.289***
RD			−0.098**	−0.083**
ADV_SECT			−0.013	−0.035
2008	−0.069	−0.184*		
2009	0.009	−0.101	0.055	0.058
2010	0.172*	0.213**	0.143	0.278***
2011	0.106	1.174***	0.102	1.050***
2012	0.109	1.095***	0.115	0.977***
cons	−0.080	−0.524***	−0.101	−0.544***
R-sq	0.449	0.544	0.602	0.656

Notes: Reg 1.1 and 1.2 Obs. = 618; Reg 1.3 and 1.4 Obs. = 515. *$p < 0.10$, **$p < 0.05$; ***$p < 0.01$.

4. Estimation results

The estimation results are shown in Table 3 and for the four dependent variables. All specifications show high R^2 values.

In the first section of Table 3 we present the main results concerning per capita GDP as the dependent variable. In Column 1, PSP is the sole variable in the model, along with the other dependent variables. The PSP coefficient is significant and positive, indicating that a standard deviation rise in PSP is associated with a 0.122 standard deviation increase in GDP. In order to investigate the relevance of PSP over and above other measures of economic complexity, in Columns 2, 3 and 4 we include step by step the EXPYnoY, VARIETY and DIVERSITY indices. PSP maintains a positive and significant impact on economic prosperity with an effect size of similar magnitude in all regressions, whereas EXPYnoY is not significant. In Columns 3 and 4 we include all our variables, finding an insignificant value for the VARIETY coefficient, while the DIVERSITY coefficient is significant and presents a negative effect.

In the second section of Table 3 we present the main results concerning the dependent variable capturing local labor productivity, namely provincial gross value added per worker (GVA). Just in the first regression, Model 1.1, PSP shows a significant and positive coefficient. In the next steps when we include the EXPYnoY index in the model, PSP no longer shows a significant impact on labor productivity GVA, while the coefficient of the EXPYnoY variable is also negative and not significant. In Models 1.3 and 1.4, we add DIVERSITY and VARIETY, but their estimated coefficients are not significant.

In the third section of Table 3 we turn to the equation regarding innovative behavior. Once more, in the first regression, Model 1.1, we only include PSP and we find a significant and positive coefficient. In the next step we include the EXPYnoY index in

the model. PSP maintains a positive and significant impact on innovation, whereas the coefficient of the EXPYnoY variable is negative and not significant. In Models 1.3 and 1.4, we add DIVERSITY and VARIETY. PSP maintains again a positive and significant impact on innovation, while VARIETY also displays a high and significant value, whereas we find a negative and significant value for the DIVERSITY coefficient in both regressions.

In the last part, Table 3 includes the main results concerning the PSP index as a dependent variable. In the first regression, Model 1.1, we only include the other dependent variables and we find EXPYnoY displaying positive and statistically significant effects on PSP, whereas GVA is not statistically significant for all regressions. In the last regressions, we include both VARIETY and DIVERSITY. While DIVERSITY displays a high and significant value, VARIETY does not have any statistically significant effects on PSP.

Our results suggest that PSP always matters strongly for our overall indicator of economic prosperity GDP and also for innovation PAT at the provincial level. It matters to a lesser extent for our labor productivity measure GVA, losing its significance as other variables are added. Moreover, the PSP index shows stable correlations and positive significance with respect to both GDP and PAT over and above all other control variables and is generally a better performer than EXPYnoY. All of the regression results are very similar no matter which type of EXPY we use. We also experimented with equations including the ECI and EXPY, and again these performed even less well than EXPYnoY and had no real, stable or consistent explanatory power.[22]

These results support our original assumption that the more related is the productive structure and the knowledge base of the province, the wider is the contribution of cognitive proximity to local economic prosperity and innovation behavior. Moreover, these effects are positive and significant over and above the standard controls emanating from the urban and regional economics literature. However, only the PSP index consistently demonstrates this.

Regarding the other control variables, the econometric results show the crucial role of innovation PAT in the GDP equation, and also the crucial role of economic prosperity GDP in the PAT equation. Not surprisingly the results also show a strong positive effect of labor productivity GVA on GDP and also a strong impact of economic prosperity GDP on labor productivity GVA. More surprisingly, we find a negative relation between GVA and PAT. The correlations between these variables are positive but decline from 0.668 to 0.449 in the years we have available. After adding our control variables and applying SUR, we find a negative effect. This may reflect the fact that in the immediate post-crisis era lower- (or non)-patenting firms have tended to contract employment more sharply than higher-patenting firms, although without firm-specific data this can only be a tentative suggestion. At the same time, the effect of ADV_SECT is strongly positive and significant on GDP, whereas it is insignificant with respect to labor productivity GVA and PSP and negatively related to innovation PAT. Similarly, the effect of the RD variable is insignificant with respect to GDP, GVA and PAT, although this might have been expected to be positively related. Instead of an employment-based R&D variable, a more suitable measure for R&D inputs may be the

22 Refer to Supplementary Data, Supplementary Tables S1–S3.

total R&D expenditure per capita for each area. Unfortunately, however, R&D expenditure data disaggregated at the level of the Italian provinces do not exist, and are only reported at the much larger spatial units of the broader Italian regions.

In order to control for regional human capital endowments, we also included the variable EDU in the model. The impact of EDU on economic prosperity GDP is positive and significant, on PAT it is positive but just outside of the 10% significance range, and with respect to GVA it is negative. As already said, these education data are collected considering the location of universities, so in our models we also tried with data collected according to the residence of students, and the results did not change markedly. The period under examination involves very high levels of unemployment and especially among younger people and in southern regions, so EDU may not as closely reflect actual worker participation as in other situations. Moreover, in times of severe unemployment labor productivity GVA and GDP per capita GDP tend to move in opposite directions to each other. These opposing effects may also partially account for the fact that when we control for population density POP displays a significant and negative coefficient with respect to GDP, but a significant and positive coefficient with respect to both GVA and PAT, suggesting that density is associated with both positive externalities and increasing costs. Finally, our SUR results using a PSP index calculated on the basis of employment data PSP(EMPL) generate results with respect to both GDP and PAT which are very different, if not almost the opposite, of what would be expected from a priori theoretical arguments, and in particular those posited by the Hausmann–Hidalgo et al. tradition. The results are reported in Table 4. Part of the problem here is likely to be related to the fact that 78% of the data used in constructing the PSP(EMPL) is related to services which only account for 22% of Italy's exports, such that a majority of these jobs actually reflect non-exporting activities. In contrast, all of the data used to calculate PSP index reflect not only actual exporting, but the overwhelming majority of Italian exports.

5. Conclusion

Returning to our original question as to whether the Hausmann–Hidalgo type of trade centrality and connectedness argument is useful for understanding the economic development of regions in advanced economies, and if so, whether a significant adaptation of their framework would empirically provide a better reflection of these realities, on both counts our paper demonstrates that the answer is yes. In this paper, we have applied the Hausmann–Hidalgo type logic in order to examine the extent to which the network-positioning, simultaneously measured in terms of centrality and comparative advantage, of a sub-national region's exports accounts for its level of economic development. At the regional level of an advanced economy, we explained using Italian provincial export data why it is necessary to construct a new PSP index, which better captures the features of these types of economies than the other proposed indices employed in the international trade and development literatures. Part of the reason is related to the specialization discontinuities evident in the current indices which are inappropriate for capturing regional systems of innovation and part of the reason is because the pure proximity-network dimensions of the current indices also insufficiently capture regional characteristics. As such, our analysis has also demonstrated that both the specialization features and also the pure network dimensions of the current

Table 4. Robustness checks PSP(EMPL): SURs

GDP	Reg 1.5
PSP(EMPL)	−0.091***
PAT	0.157***
GVA	0.757***
EXPYnoY	0.082**
VARIETY	0.064*
DIVERSITY	−0.037
EDU	0.052**
POP	−0.034*
RD	0.006
ADV_SECT	0.149***
2008	
2009	−0.189***
2010	0.103*
2011	0.069
2012	−0.035
Cons	−0.027
R-sq	0.872
GVA	Reg 1.5
PSP(EMPL)	0.209***
PAT	0.018
GDP	0.979***
EXPYnoY	−0.006
VARIETY	−0.091**
DIVERSITY	0.016
EDU	−0.085***
POP	0.029
RD	0.011
ADV_SECT	−0.127***
2008	
2009	−0.110**
2010	0.410***
2011	0.469***
2012	0.230**
cons	−0.381***
R-sq	0.848
PAT	Reg 1.5
PSP(EMPL)	−0.700***
GDP	0.620***
GVA	−0.117
EXPYnoY	−0.129*
VARIETY	0.568***
DIVERSITY	−0.116**
EDU	0.113
POP	0.031
RD	−0.042
ADV_SECT	0.206***
2008	
2009	−0.278***
2010	−0.463***
2011	−1.019***
2012	−0.780***
Cons	0.679***
R-sq	0.507

(continued)

Table 4. continued

PSP(EMPL)	Reg 1.5
GDP	−0.271***
PAT	−0.415***
GVA	0.318***
EXPYnoY	−0.148***
VARIETY	0.522***
DIVERSITY	−0.026
EDU	0.132***
POP	−0.075***
RD	−0.025
ADV_SECT	−0.461***
2008	
2009	−0.430***
2010	−0.977***
2011	−1.273***
2012	−1.052***
cons	0.799***
R-sq	0.733

Notes: Reg 1.5 Obs. = 515. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

frameworks are inappropriate for regional analyses and need to be adapted accordingly. Once these adaptations are made, our analysis suggests that in an advanced economy the basic insights of Hausmann, Hidalgo et al. still continue to hold even at the regional level over and above the more traditional assumed drivers of local economic performance. Nor is the distinction between tradeables and non-tradeables critical for local economic development, but rather which particular combinations of tradeables are produced (Hausmann et al., 2007).

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

Acknowledgments

We are very grateful to Cesar Hidalgo, Frank Neffke, Lelio Iapadre, Ron Boschma and Nicola Cortinovis who provided us with many ideas and helpful insights and whose comments greatly helped us to improve the paper.

References

Balassa, B. (1965) Trade liberalization and revealed comparative advantage. *The Manchester School of Economic and Social Studies*, 33: 99–123.

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

Boschma, R., Iammarino, S. (2009) Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, 85: 289–311.

Boschma, R., Minondo, A., Navarro, M. (2012) Related variety and regional growth in Spain. *Papers in Regional Science*, 91: 241–256.
Cortinovis, N., Xiao, J., Boschma R., Van Oort F. G. (2017) Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography*, 17: 1179–1208.
Dalum, B., Laursen, K., Villumsen, G. (1998) Structural change in the OECD export specialisation patterns: de-specialization and stickiness. *International Review of Applied Economics*, 12: 423–443.
De Groot, H. L., Poot, J., Smit, M. J. (2016) Which agglomeration externalities matter most and why? *Journal of Economic Surveys*, 30: 756–782.
Di Maio, M., Tamagni F. (2008) The evolution of world export sophistication and the Italian trade anomaly. *Rivista di Politica Economica*, 98: 135–174.
Duranton, G., Puga, D. (2000) Diversity and specialisation in cities: why, where and when does it matter? *Urban Studies*, 37: 533–555.
Frenken, K., Oort, F. G. V., Verburg, T. (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41:685–697.
Hausmann, R. (2010) Ricardo Hausmann on 'product space' and development. Available online at: http://www.hks.harvard.edu/news-events/publications/insight/markets/ricardo-hausmann.
Hausmann, R., Hwang, J., Rodrik, D. (2007) What you export matters. *Journal of Economic Growth*, 12: 1–25.
Hausmann, R., Klinger, B. (2006) Structural transformation and patterns of comparative advantage. CID Working Paper No. 128, Center for International Development, Harvard University.
Hausmann, R., Klinger, B. (2007) The structure of the product space and the evolution of comparative advantage. *CID Working Paper No. 146*, Center for International Development, Harvard University, Cambridge, MA.
Hidalgo, C. (2009) The dynamics of economic complexity and the product space over a 42 year period. *CID Working Paper No. 146*, Center for International Development, Harvard University.
Hidalgo, C., Hausmann, R. (2009) The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106: 10570–10575.
Hidalgo, C., Klinger, B., Barabasi, A., Hausmann, R. (2007) The product space conditions the development of nations. *Science*, 317: 482–487.
Hidalgo, C., Blumm, M., Barabasi, A., Christakis, N. A. (2009) A dynamic network approach for the study of human phenotypes. *PLos One: Computational Biology*, 5: e1000353.
Iammarino, S., McCann, P. (2013) *Multinationals and Economic Geography: Location, Technology and Innovation*. Cheltenham: Edward Elgar.
Iapadre, L. (2001) Measuring international specialisation. *International Advances in Economic Research*, 7: 173–183.
McCann, P. (2013) *Modern Urban and Regional Economics*. Oxford: Oxford University Press.
Minondo, A. (2010) Exports' quality-adjusted productivity and economic growth. *The Journal of International Trade & Economic Development*, 19: 257–287.
Neffke, F., Henning, M., Boschma, R. (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87: 237–265.
O'Donoghue, D., Gleave, B. (2004) A note on methods for measuring industrial agglomeration. *Regional Studies*, 38: 419–427.
OECD. (2018a) Trade in goods (indicator) (doi: 10.1787/1ea6b5ed-en) (accessed 16 August 2018).
OECD. (2018b) Trade in goods and services (indicator) (doi: 10.1787/0fe445d9-en) (accessed 16 August 2018).
Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., Hidalgo, C. (2018) Shooting high or low: do countries benefit from entering unrelated activities? *Working Paper, MIT Media Lab*, Cambridge, MA. Available online at: https://arxiv.org/ftp/arxiv/papers/1801/1801.05352.pdf.
Rodrik, D. (2006) What's so special about China's exports? *Chinese Academy of Social Sciences*, 14: 1–19.
Tian, Z. (2013) Measuring agglomeration using the standardized location quotient with a bootstrap method. *The Journal of Regional Analysis & Policy*, 43: 186–197.