



Structure of the Region-Technology Network as a Driver for Technological Innovation

Dion R. J. O'Neale^{1,2*†}, Shaun C. Hendy^{1,2} and Demival Vasques Filho^{3*†}

¹Department of Physics, University of Auckland, Auckland, New Zealand, ²Te Pūnaha Matatini—The Centre for Complex Systems and Networks, Auckland, New Zealand, ³Leibniz Institute of European History, Mainz, Germany

Agglomeration and spillovers are key phenomena of technological innovation, driving regional economic growth. Here, we investigate these phenomena through technological outputs of over 4,000 regions spanning 42 countries, by analyzing more than 30 years of patent data (approximately 2.7 million patents) from the European Patent Office. We construct a bipartite network—based on revealed comparative advantage—linking geographic regions with areas of technology and compare its properties to those of artificial networks using a series of randomization strategies, to uncover the patterns of regional diversity and technological ubiquity. Our results show that the technological outputs of regions create nested patterns similar to those of ecological networks. These patterns suggest that regions need to dominate various technologies first (those allegedly less sophisticated), creating a diverse knowledge base, before subsequently developing less ubiquitous (and perhaps more sophisticated) technologies as a consequence of complementary knowledge that facilitates innovation. Finally, we create a map—the Patent Space Network—showing the interactions between technologies according to their regional presence. This network reveals how technology across industries co-appear to form several explicit clusters, which may aid future works on predicting technological innovation due to agglomeration and spillovers.

Keywords: innovation networks, patents, knowledge spillover, agglomeration advantage, bipartite networks, patent space, evolutionary economic geography (EEG), technological innovation

OPEN ACCESS

Edited by:

Michele Coscia,
IT University of Copenhagen, Denmark

Reviewed by:

Jake Carr,
Moody's Analytics, United States
Justyna Majewska,
Poznań University of Economics and
Business, Poland

*Correspondence:

Dion R. J. O'Neale
d.oneale@auckland.ac.nz
Demival Vasques Filho
vasquesfilho@ieg-mainz.de

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Big Data Networks,
a section of the journal
Frontiers in Big Data

Received: 31 March 2021

Accepted: 18 June 2021

Published: 14 July 2021

Citation:

O'Neale DRJ, Hendy SC and
Vasques Filho D (2021) Structure of the
Region-Technology Network as a
Driver for Technological Innovation.
Front. Big Data 4:689310.
doi: 10.3389/fdata.2021.689310

1 INTRODUCTION

Innovation is facilitated by the combination of diverse yet complementary knowledge inputs (Jacobs, 1969). Indeed, some of the most influential conceptualisations of the innovation process regard technological change as originating from the combination of new and existing technological capabilities (Weitzman, 1998). A recent study of more than 2 centuries of patents granted in the United States suggests that more than half of all patented inventions in this period arose through novel recombination of pre-existing technologies (Youn et al., 2015).

While the potential for new combinations of the world's current portfolio of technologies is vast, the economic geography of innovation may constrain the ability of inventors to explore all technological combinations. Furthermore, the value of adding an additional technology to a set of pre-existing capabilities will vary depending on both the new technology and the existing combination. This would suggest that those regions with a diverse knowledge base are at an advantage when it comes to regional technological progress (Feldman and Audretsch, 1999).

The economic geography of innovation is known to be dominated by agglomeration effects and spillovers, where innovation and economic growth are facilitated by geographical proximity (David and Rosenbloom, 1990; Krugman, 1991; Saxenian, 1996) and localized learning processes (Glaeser, 1999). According to Feldman (1999, 5), “knowledge is not easily contained and geography provides one means to define knowledge spillovers,” such that firms, including competitors, can benefit from being locally proximate (Baldwin and Okubo, 2006). In short, agglomeration effects occur when firms or people accrue benefit from being located near to one another, while knowledge spillovers are one process by which firms and individuals can derive such benefits, by taking advantage of new knowledge that has been created by others.

While developments in communication technology make the dissemination of codified knowledge ever faster and cheaper, the transmission of tacit knowledge may still be difficult in the absence of face-to-face interactions. Consistent with this, the effects of knowledge spillovers on innovation have been found to be most evident when people or firms are geographically proximate (Jaffe et al., 1993; Glaeser and Ponzetto, 2010; Baicker and Chandra, 2010; Buzard et al., 2017; Buzard et al., 2020). As a result, inventors who have access to a more highly connected and diverse local innovation system—and the firms that employ them—face lower costs in exploring a larger set of potential technologies.

Knowledge spillovers have been considered to act in two ways. First, concerning firms or inventors within particular industries, spillovers favor *localization*: because of the decay in spillover benefits over distance, similar firms will co-locate so that they can learn from each other (Porter, 1990; Jaffe et al., 1993; Ellison and Glaeser, 1997; Duranton and Overman, 2005; Chatterji et al., 2013). Second, concerning firms or inventors in different industries, spillovers favor *urbanization*: firms from a variety of industries will choose locations where they can benefit from a diverse range of knowledge spillovers (Glaeser et al., 1992).

Regarding the latter, the effect of spillovers weakens across different industrial or technological domains (Jaffe et al., 1993). Due to this, regional innovative performance becomes path dependent: regions are only able to explore new combinations of technologies if the relevant technological and organizational capabilities are already present (Hidalgo and Hausmann, 2009). Thus, our first question is: do regions with a broader knowledge base tend to create high-tech—more unique—innovations than regions with a small set of capabilities? To address that, we have constructed a bipartite *revealed* technological network by observing how specialties appear across geographical regions. Using over 3 decades of data from the European Patent Office we have computed the revealed comparative advantage in specific technological domains for more than 4,000 regions spanning 42 countries. We compare the observed bipartite network with several null models to assess the technology co-occurrence—and possibly knowledge spillover—effects on regional diversity and technological ubiquity. We find that regions with a low diversity of technological capabilities tend to have a more ubiquitous set of technologies than regions with a

higher diversity. In other words, technologies that are less ubiquitous (allegedly more sophisticated) tend to occur in regions with high technological diversity. This is consistent with the idea that regional innovation is constrained by access to diverse technological inputs that are available locally.

However, not all co-occurrence and combinations of technologies will have utility [e.g. the “espresso-making toothbrush” (Youn et al., 2015)], while some technologies may lend themselves to many more useful combinations than others (e.g. general purpose technologies, such as the integrated circuits within both smartphones and automobile engines). Then, we pose a second question: how are baskets of technologies organized, thanks to regional output, such that they might favor spillovers and facilitate the prediction of new technologies? From the bipartite network, we have constructed a (projected) network map of technological proximity—based on the co-occurrence of technologies. This approach is similar to that of Hidalgo and Hausmann (2009), who used exported products rather than patents to examine the links between the type of goods exported and economic success at a national level.

The advantage of using patents rather than exported products is that patents can be tied to a particular region, whereas export data is typically aggregated at a national scale. The disadvantage of using patents lies in the greater difficulty in assessing their value: many, if not most patents will have little market value, while a few may be of considerable worth (Hall et al., 2005). Moreover, our approach is complementary to that of Youn et al. (2015), who categorize individual patents by the combination of technological classification codes assigned to them during the application process. However, by using regional co-location of patents as a measure of technological proximity, our method has the potential to capture the incorporation of tacit knowledge in new technologies that would not necessarily be evident in a codified classification scheme. In what follows, we demonstrate that our measure of proximity does differ from a codified classification scheme.

We begin with a brief discussion about the scholarship that underpins this study. Next, we explain the data set used and the methodological approach taken to create the region-technology (bipartite) network, including a description of a variety of null models used to test hypotheses about the spillover effect, and the Patent Space (projected) network of technological co-occurrence in regions. Then, we discuss our results focused on the relationship between diversity and ubiquity found in regional patent portfolios. We conclude with comments on the importance of this relationship.

2 DIVERSIFICATION, PATH-DEPENDENCE AND PROXIMITY

Diversity has a long history in theoretical and empirical debates about its effects on several fronts of regional economic development as, for instance, growth, resilience against cycles and unemployment, stability, per capita income, and firms performance (Attaran, 1986; Sherwood-Call, 1990; Malizia and Ke, 1993; Wagner and Deller, 1998; Qian et al., 2008). Under the

lenses of evolutionary economics, diversity plays a leading role also in technological progress: due to cumulative and path-dependent processes, regions with a diverse knowledge base are more likely to produce new technologies (Feldman, 1999).

How regions evolve following particular paths is a key question in the recent body of studies dubbed Evolutionary Economic Geography (Boschma and Martin, 2010). On the one hand, the latter draws from evolutionary thinking where the “current state of affairs cannot be derived from current conditions only” (Boschma and Frenken, 2017, 280), characteristic of a path-dependent process: regional capabilities determine which industries—or technologies—are most likely to develop in the future (Hidalgo et al., 2007; Hausmann and Klinger, 2007). On the other hand, evolutionary economic geography builds upon New Economic Geography models concerning agglomeration mechanisms and the formation of dense clusters of industries (Krugman, 1991).

These clusters happen thanks to the advantages of reducing transportation costs, generating economies of scale and easing factor mobility that result in manufacturing belts with a high concentration of people (Krugman, 1991). Ultimately, agglomeration leads to concentrate diversity—through knowledge spillover mechanisms of localization and urbanization (Porter, 1990; Glaeser et al., 1992; Jaffe et al., 1993; Chatterji et al., 2013)—which, in turn, leads to more agglomeration, in an iterative process. Cumulative knowledge (i.e. technological output) follows a path of evolution, such that regions that acquire a vast technological base are more likely to produce ubiquitous technologies.

Regional technological innovation, therefore, depends on diversity and spillovers effects. Even though technology has facilitated learning exchange at distance, spillover effects on innovation are most evident with spatial proximity, especially thanks to tacit knowledge (Howells, 2002). However, Boschma (2005) argues that learning and innovation are facilitated by, besides spatial proximity, other four dimensions of proximity—cognitive, organisational, social and institutional proximity. In fact, he says that “geographical proximity per se is neither a necessary nor a sufficient condition for learning to take place: at most, it facilitates interactive learning, most likely by strengthening the other dimensions of proximity” Boschma (2005, 62). Among these other four dimensions, cognitive proximity is more relevant to this study, as it relates to the co-occurrence of technological codes in regions.

Cognitive proximity accounts for how different the knowledge base and the capabilities between two actors is. “The ability to evaluate and utilize outside knowledge is largely a function of the level of prior related knowledge” (Cohen and Levinthal, 1990, 128). Thus, again, we have a path-dependent process of innovation *via* combinations of technologies, the so-called recombinant innovation (Weitzman, 1998; Neffke et al., 2011; Castaldi et al., 2015). Although recombinant innovation is more common within sectors (related to the localization mechanism), it also happens in extra-sector (urbanization) contexts. In the latter case, innovations are more likely to fail but also to be disruptive (Frenken et al., 2007; Castaldi et al., 2015). The Patent Space

network we propose attempts to provide insights on these combinations based on regional technological co-occurrence.

3 DATA AND METHODS

3.1 Data

We use patent records from the February 2016 edition of the Organization for Economic Co-operation and Development (OECD) REGPAT database (PATSTAT), which is itself derived from two complementary sources: the European Patent Office (EPO) Worldwide Statistical Patent Database (PATSTAT, 2018) and the EPO Bibliographic Database and Abstracts (November 2015). The data covers 2,892,607 patent applications filed to the EPO from 1977 to 2012, with partial data from 2012 to 2015.

Patent records in REGPAT are “regionalized” by matching applicant addresses with one of 4,106 micro-statistical regions (TL3) covering the 46 countries in the data set (Maraut et al., 2008). During the filing process, patents are assigned one or more International Patent Classification (IPC) code which attempts to categorize the type of technology that relates to the essential novel component of the invention described in the patent.

The classification system is hierarchical: technologies described by lower levels of the code are subdivisions of the technologies at higher levels. A complete IPC code consists of four levels with additional classification at an additional fifth sub-level in some cases. This divides the technologies described by the patents into roughly 70,000 subdivisions. In the analysis presented in this paper, we use the third level of the IPC codes, consisting of 635 categories.

We provide results with aggregated micro-regions (TL2—639 regions) and the fourth level (IPC4—7,823 codes) of the IPC codes in the **Supplementary Material**.

3.2 Methods

We constructed a matrix of regions and IPC codes where each entry in the matrix is the number of times that a particular IPC code was used on patents from that region. To determine when the count for a particular region-code pair is significant, we used the method of revealed comparative advantage (RCA) (Balassa, 1965):

$$RCA(r, c) = \frac{x(r, c)}{\sum_c x(r, c)} \bigg/ \frac{\sum_r x(r, c)}{\sum_{r,c} x(r, c)} \quad (1)$$

where $x(r, c)$ is the number of times code c appears on patents filed by an applicant from region r .

The RCA method takes account of both the total amount of patenting activity within a region, and the global prevalence of each IPC code as a fraction of all those used in the data. For a given region-technology pair, a RCA value greater than one indicates that the region produces more patents using that technology than would be expected, given the total number of patents produced by that region and the fraction of the world's patents that also use that technology.

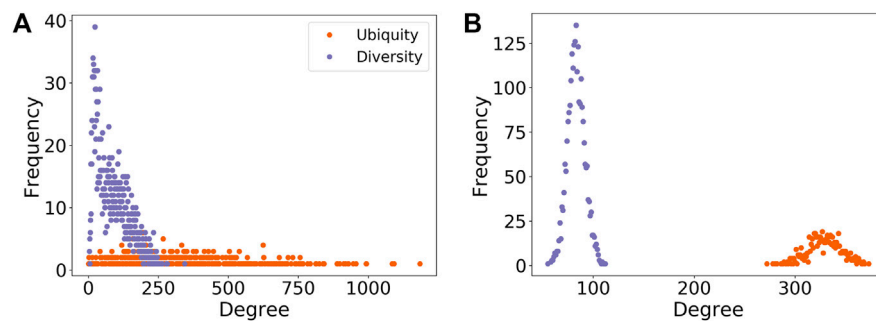


FIGURE 1 | Degree distributions of regions (diversity) and technologies (ubiquity) for the **(A)** empirical and **(B)** random (null model 1) networks. The latter is a bipartite variant of the Erdős and Rényi model (Erdős and Rényi, 1959; Vasques Filho and O'Neale, 2020a). The randomization of connections between regions and IPC codes makes the diversity distribution less right-skewed and changes completely the shape of the ubiquity distribution from a uniform-like to a normal-like distribution. The randomization tests the effects of localization and urbanization (for regions), and sophistication (for technologies), by assuming that there are no underlying attributes of regions and technologies that might be responsible for the observed degree distributions.

A bipartite network is defined by connecting regions and technology codes when $RCA(r, c) \geq 1$. The network can be represented by an unweighted adjacency matrix

$$M_{r,c} = \begin{cases} 1, & \text{if } RCA(r, c) \geq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The diversity of a region (the number of technology codes for which it has a revealed comparative advantage) is given by $d_r = \sum_c M_{r,c}$, while the ubiquity of a technology code (the number of regions which have a revealed comparative advantage with respect to that code) is given by $u_c = \sum_r M_{r,c}$. These correspond to the degrees of the region and technology nodes, respectively, in the bipartite network (**Figure 1A**). For each region, we also calculate the average ubiquity of the technologies for which that region has $RCA \geq 1$: $\langle u \rangle_r = \frac{1}{d_r} \sum_c M_{r,c} u_c$.

3.2.1 Bipartite Network Null Models

In order to determine whether the structure we observe in the region-technology network diverges from that where regional and/or technological attributes are ignored, we compared its properties with those from random networks produced by five different null models. We associate each null model with a particular hypothesis about what effects may be causing the observed structure. Four null models rewire the empirical region-technology network in different ways, based on Hidalgo and Hausmann (2009). The fifth null model generates alternative region-technology networks by randomly reallocating to regions the patents, as recorded in the PATSTAT database. This accounts for the co-occurrence of multiple technology codes on the same patent. For each of the null models we run 100 realisations. Below we describe these five models and the hypothesis we test with them.

- **Model 1:** Randomly reassigns the edges in the network conserving only the number of nodes and number of edges.

This is a bipartite variant of the Erdős and Rényi model (Erdős and Rényi, 1959; Vasques Filho and O'Neale, 2020a). The model preserves the mean degree for the two node types, but not the degree of individual nodes neither the shape of the degree distribution (**Figure 1B**). It has the highest level of randomization, meaning that regions and technologies have similar levels of diversity and ubiquity, respectively. Here, we test the effects of localization and urbanization in regions and sophistication of technologies due to recombination at the same time, by ignoring the attributes of both region (e.g. population, number of research institutions) and IPC codes (e.g. different amounts of capability or resources required for their development) related to these effects.

- **Model 2:** Randomly reassigns edges while preserving the degree sequence of regions (i.e. preserving d_r for each region node and the diversity distribution of **Figure 1A**), in addition to preserving the total numbers of nodes and edges. This model can be interpreted as treating all technologies as being identical, while each region possesses some property (and the localization and urbanization effects) that affects its ability to develop a RCA in any technology.
- **Model 3:** Randomly reassign edges while preserving the degree sequence of technology codes (i.e. preserving u_c for each code and the ubiquity distribution of **Figure 1A**). Such a model represents the situation where all regions are identical, but technologies each possess some attribute that affects the likelihood of them being employed in any invention. This would be consistent with the hypothesis that some technologies are easier to acquire and hence are more prolific than others.
- **Model 4:** Randomly swap pairs of existing edges such that for a pair of edges (a, b) and (u, v) they are replaced by (a, v) and (u, b) if neither of the two new edges were already part of the network. This is the most stringent of the rewiring models, preserving the degree sequences of both

the region and technology nodes (as in **Figure 1A**), but removing correlations that might exist between specific technologies and their associations with specific regions.

- **Model 5:** Randomly reassign individual patents to regions, such that each region retains the same total number of patents, before calculating a new bipartite network. Individual patents can, and often do, list more than one technology code. This model preserves the co-occurrence of these codes on patents, but not co-occurrence of technology codes due to regional clustering.

3.2.2 The Patent Space (Projected) Network

The five null models each test the relationship between the diversity of technologies used by a region for its inventions and the ubiquity of those technologies. However, while they do test the observed structure of the region-technology network as a result of agglomeration or spillover effects, they do not explain how these effects occur.

To investigate that, we look at the co-occurrence of different technologies within regions by projecting the region-technology network to get a technology-technology co-occurrence network. For this projection, we follow the example of Hidalgo et al. (2007) who constructed a similar network for products exported by different countries. The proximity (weight of the edge) between any pair of IPC technology codes is given by the pairwise conditional probability that a region with $RCA \geq 1$ for code c_i also has $RCA \geq 1$ for code c_j , that is

$$\phi(c_i, c_j) = \min\left\{P(RCA(c_i) \geq 1 | RCA(c_j) \geq 1), P(RCA(c_j) \geq 1 | RCA(c_i) \geq 1)\right\} \quad (3)$$

4 RESULTS AND DISCUSSION

4.1 Empirical Data

We can begin to understand the distribution of technical capabilities among regions by looking at the structure of the adjacency matrix $M_{r,c}$. Ordering the rows and columns of the matrix according to the values of d_r and u_c shows that the matrix is approximately triangular (**Figure 2**).

The triangular structure implies that more technically advanced regions—those with high diversity—file patents involving both low and high ubiquity technologies, while low diversity regions are only able to make use of the most ubiquitous—and presumably less sophisticated—technologies. Thus, the triangular structure indicates that the technologies associated with inventions from low diversity regions tend to be subsets of those used by high diversity regions. Also, this structure indicates that the Ricardian model of producing goods for trade does not apply here, if we were to extrapolate the model for the production of technology. Such a model for technological production would lead to an adjacency matrix with a block-diagonal structure as regions would specialize in only those technologies that they have a comparative advantage for and which will have the lowest ubiquity, at the expense of having high diversity through also producing some more ubiquitous technologies.

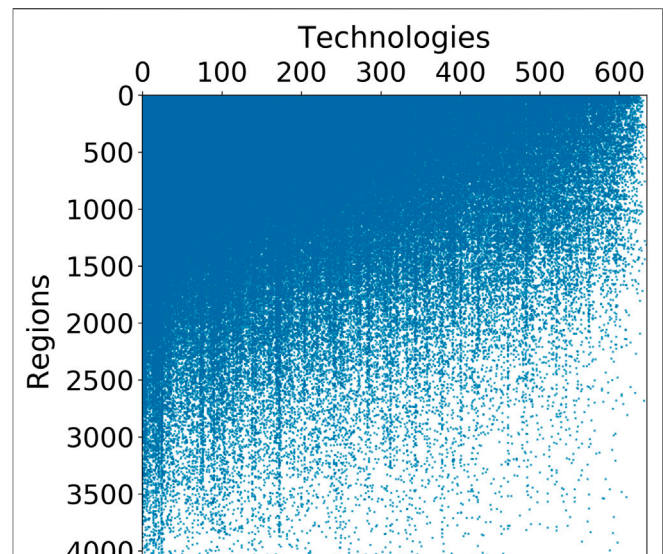


FIGURE 2 | Adjacency matrix of the region-technology network ordered by degree value, from highest to lowest diversity (**top-bottom**) and ubiquity (**left-right**). The triangular-like structure tells that high diversity regions file patents with both low and high ubiquity IPC codes, while low diversity regions are only able to make use of the most ubiquitous—and presumably less sophisticated—technologies.

A consequence of the triangular structure for the adjacency matrix is that those regions with higher diversity d_r tend to have a lower mean ubiquity $\langle u \rangle_r$. In the case of a perfectly triangular $m \times n$ adjacency matrix

$$M_{r,c} = \begin{cases} 1, & \text{if } r/c \leq m/n \\ 0, & \text{otherwise,} \end{cases}$$

it can be shown analytically that mean ubiquity $\langle u \rangle_r$ is a linear function of diversity d_r , given by

$$\langle u \rangle_r \approx -\frac{m}{2n}d_r + m.$$

The empirical correlation between diversity d_r and regional mean ubiquity $\langle u \rangle_r$, with a linear least squares fit, gives a slope of -0.61 and an intercept of 537.0 . The Pearson correlation coefficient is $r = -0.52$ ($R^2 = 0.27$) (**Figure 3**). While the R^2 value is relatively low, this is to be expected since, particularly for regions with only a small number of patents, there will be a range of factors that influence the likelihood of a technology being recorded in a region. (To avoid more of this noise, we make this analysis considering only the regions with more than ten patents in the whole period. This is a rather weak constraint that filters out only those regions which have close to zero patenting activity over the 30 year period of the study data.) The nature of this correlation makes it possible to approximately partition the regions into those with high diversity and relatively unique technologies, and those with low diversity and common technologies (**Figure 3**).

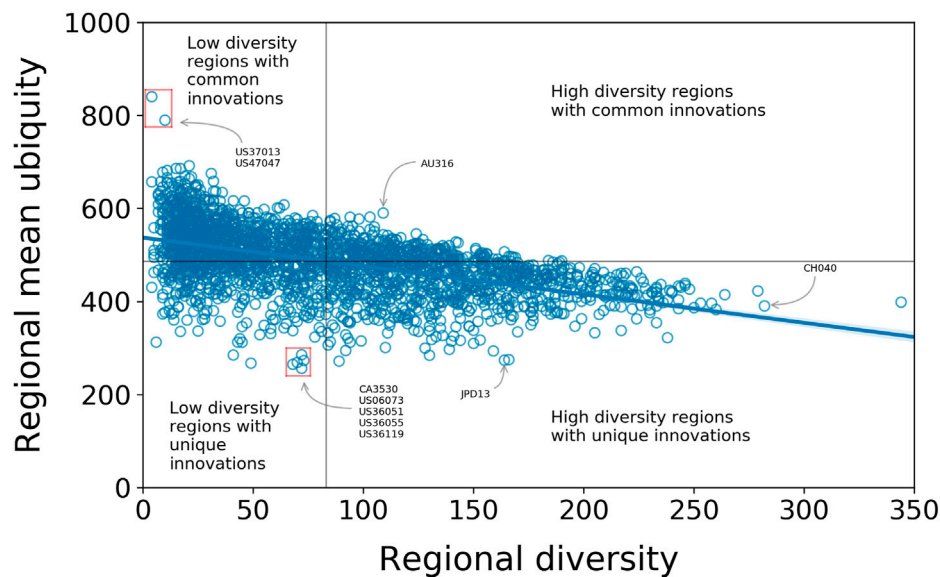


FIGURE 3 | Correlation between regional mean ubiquity and diversity of regions with Pearson correlation coefficient $r = -0.52$ ($R^2 = 0.27$). We partition the regions into those with high diversity and relatively unique technologies, and those with low diversity and common technologies and give some examples of regions in these quadrants.

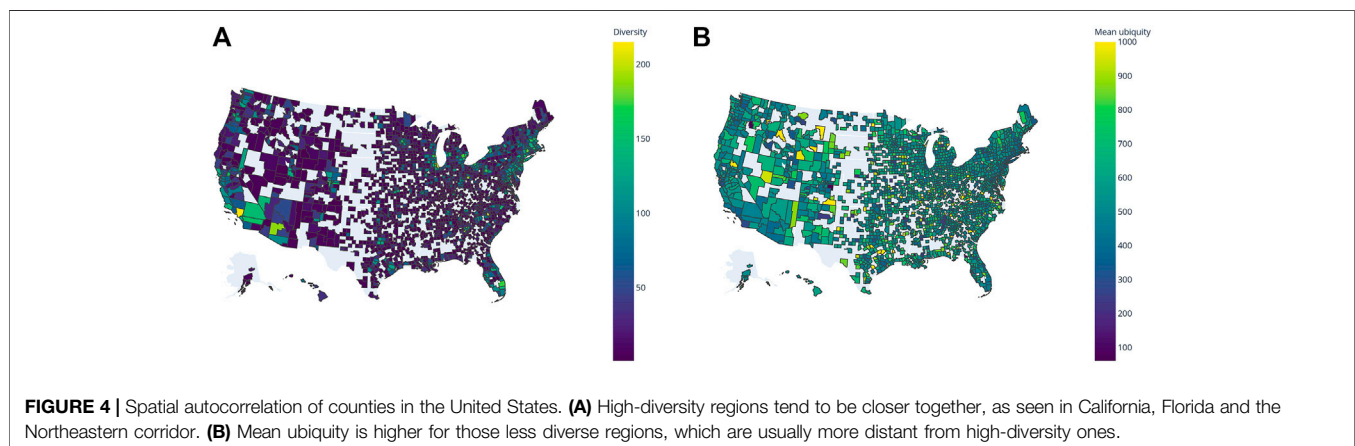


FIGURE 4 | Spatial autocorrelation of counties in the United States. **(A)** High-diversity regions tend to be closer together, as seen in California, Florida and the Northeastern corridor. **(B)** Mean ubiquity is higher for those less diverse regions, which are usually more distant from high-diversity ones.

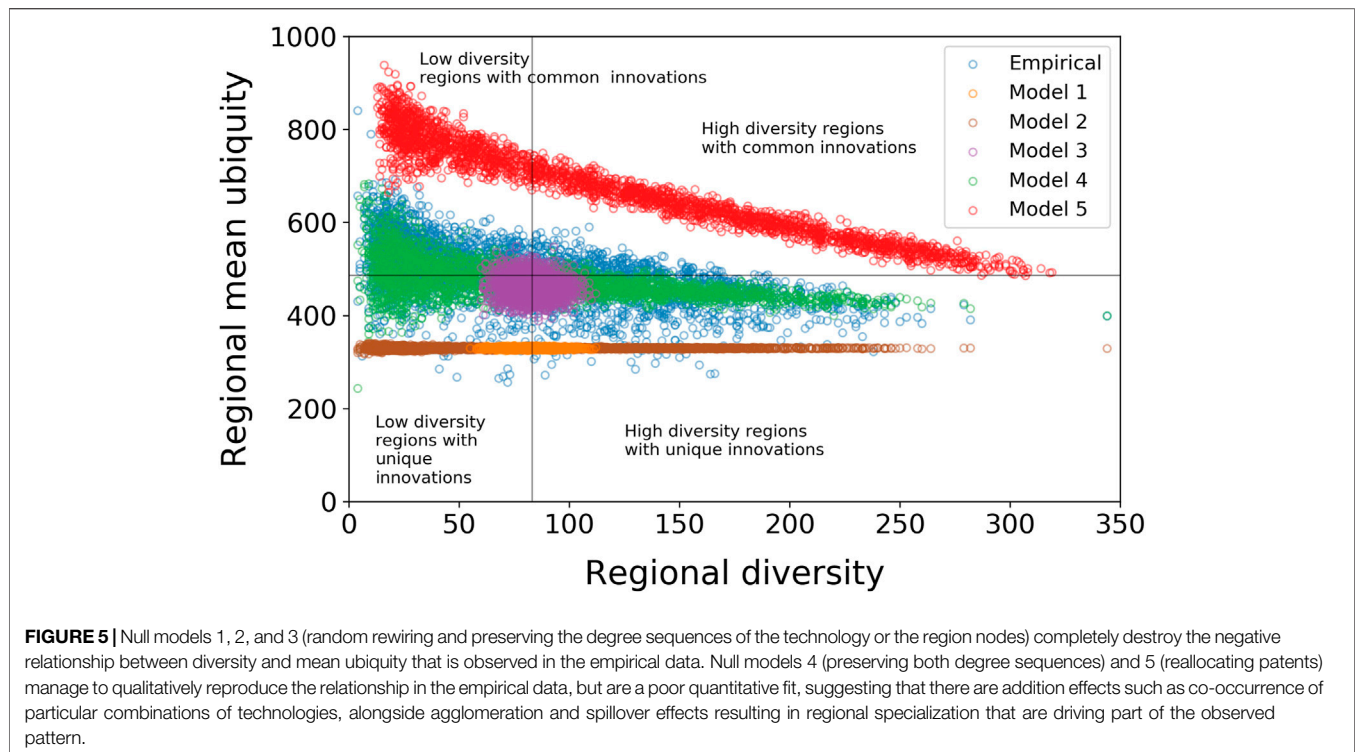
While we do not carry out a detailed analysis of the spatial distributions of technologies and the correlation between them, we would like to highlight some high-level trends using some examples, especially in the United States. From the about 3,200 United States counties, almost 2,100 have a RCA in at least one IPC code, and 90% of those are in either one of the two low-diversity quadrants of **Figure 3**. Some (as Beaufort County, South Carolina and Fayette County, Tennessee) have high mean ubiquity, implying these regions produce a small set of low-tech innovations; others have low mean ubiquity (as San Diego, in California, and Livingston, Monroe and Westchester counties in New York), suggesting they present more sophisticated innovations as part of their technological output. The former are usually far from high diversity regions in contrast to the latter, which are closer (**Figure 4**). This pattern also indicates a

clustering of regions with similar profiles. That is, high-diversity regions tend to be geographically proximate to one another, and next to them regions with both low diversity and mean ubiquity tend to appear. This idea is supported in a global scale as the tendency of high-diversity regions having a RCA in less ubiquitous technology holds when regions at the TL3 level are aggregated at the TL2 level (**Supplementary Material**).

On the quadrants with high diversity in **Figure 3**, there is a small fraction of regions with high mean ubiquity (e.g., Sunshine Coast, in Australia), and several with low mean ubiquity (e.g., Tokyo and Zurich).

4.2 Null Models

As expected, the correlations of the null models behave differently from that of the empirical data, with the differences increasing



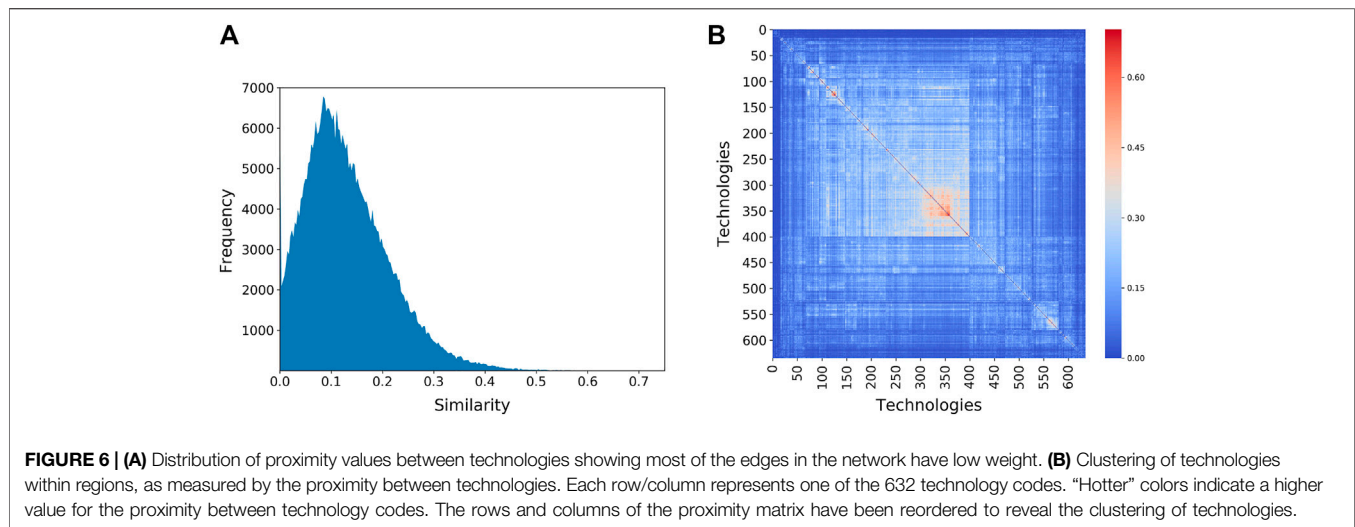
with the level of randomization (**Figure 5**). In null model 1 (orange rings in **Figure 5**), where regions and technologies are connected completely at random, just keeping the total number of connections, all regions have similar diversity (as shown in **Figure 1B**) and virtually the same regional mean ubiquity. The latter is significantly lower than the average regional mean ubiquity in the empirical data thanks to all technologies having similar ubiquity after rewiring (**Figure 1B**). The combination of low-diversity regions with high-ubiquity IPC codes that brings mean ubiquity up do not exist when all codes have similar ubiquity. These observations are indications that knowledge spillover might be affecting regional and technological attributes through localization and urbanization of regions, and recombination of technological codes, respectively.

With null models 2 and 3, we decrease the level of randomization and test separately IPC codes co-occurrence effects on technological and regional characteristics, respectively. On one hand, null model 2 (brown rings in **Figure 5**) preserves only the diversity of regions. Again, we see that regions have similar regional mean ubiquity thanks to all technologies being treated as equals, in terms of their acquisition by regions. That is, regional attributes alone (represented by regional diversification) cannot explain technological ubiquity and regional mean ubiquity as seen in the empirical data. On the other hand, null model 3 (purple rings in **Figure 5**) preserves the ubiquity of technologies and ignores regional attributes. The model results in what resembles a two-dimensional normal distribution for regional diversity and mean ubiquity.

Thus, technological ubiquity cannot explain these two variables by itself.

Then, in null model 4, we preserve both regional diversity and technological ubiquity but rewire the links between regions and technologies. In this case, the correlation between regional diversity and mean ubiquity is much closer to the empirical data than the previous null model (green rings in **Figure 5**). It becomes meaningful again look at the statistics of the correlation. In this case, the linear least squares fit gives a mean slope (95% CI) of -0.35 ($-0.33 < m < -0.37$) and an intercept of 505.2 ($502.5 < c < 507.8$) over 100 realisations. The Pearson correlation coefficient is $r = -0.51$ ($R^2 = 0.26$). The intercept and the correlation coefficient are very similar to those of the empirical data, however the slope is significantly lower than the -0.61 for the empirical data. We credit this difference in the slope to the possible correlations between specific technologies and their associations with specific regions not captured when rewiring the links. That is, even when we preserve regional diversity and technological ubiquity in the network, the model still does not capture the path-dependence of specific regions specializing in certain technologies and the clustering of technologies that can result from localized knowledge transfer and agglomeration.

Finally, each patent can be associated with more than one IPC technology code. It is therefore necessary to also test whether the structure we observe can be explained by the co-occurrence of technologies on patents, rather than co-occurrences within regions. Null model 5 (red rings in **Figure 5**) differs from the four rewiring null models in that it randomly reassigns the existing stock of patents to regions. It is the most stringent of



the null models we apply; it preserves the co-occurrence of IPC codes on patents and allocates each region the same number of patents that it originally held. As with null model 4, the results of this model are qualitatively similar to the empirical data. The linear least squares fit gives a mean slope (95% CI) of -1.12 ($-1.10 < m < 1.14$) and an intercept of 816.4 ($811.7 < c < 821.0$). The Pearson correlation coefficient is $r = -0.96$ ($R^2 = 0.92$). We attribute these and the increase in the regional mean ubiquity to the model not preserving any regional specialization in technologies. Hence, all technologies are more ubiquitous than they would be if certain regions were more likely to be allocated patents with particular IPC codes, according to regional specialization.

4.3 Co-occurrence of International Patent Classification Codes

In order to better explain the features of the regional structure of technical innovation that are unaccounted for by null models four and five, we now look at the co-occurrence of technologies within regions and the associated network of technological dependency that this implies.

The co-occurrences of technologies that result from regional specialization were used to define a network of the IPC technology codes, with connections between pairs of codes when they tended to co-occur within regions more than would be expected by chance. This network of technologies can be thought of as a projection of the region-technology network, where the weight of the edges in the projection is defined by Eq. 3. In analogy with the network in (Hidalgo et al., 2007), we refer to the resulting network of technology codes as the Patent Space.

The Patent Space network is highly connected—around 98% of all possible proximity links are present. This is an expected structure thanks to the high diversity of several regions that create large clique of codes (Vasques Filho and O’Neale, 2018; Vasques Filho and O’Neale, 2020b). Most links, however, are rather weak due to the projection method: the mean proximity (mean edge weight) between technologies is only 0.125 (the distribution of these values is shown in Figure 6A). The block diagonal structure

of the proximity matrix reveals clusters of technologies that tend to co-occur within the same sets of regions. These are indicated in Figure 6B where the rows and columns of the proximity matrix have been reordered so as to maximize the clustering.

The large number of weak connections in the proximity matrix means it is practical to visualize the Patent Space network. We do so by first extracting the maximal spanning tree of the full network and then adding in all those proximity links above a certain threshold. In Figure 7 we use a threshold of $\phi \geq 0.4$, which gives a network with an average degree of around four. This threshold was chosen primarily for visual clarity; smaller or larger values will result in a similar network with more or less detail.

We applied a modularity maximizing community detection method (Rosvall and Bergstrom, 2008) to the network in Figure 7. This partitions the 632 technology codes into around 80 distinct communities. The largest of these contains around 75 (12%) of the technology codes and is located at the center of the network. It presents all sort of technologies, including agriculture, animal husbandry, dentistry, furniture, ceramics, treatment of water, and so on. The remaining communities are roughly an order of magnitude smaller and typically are associated with a particular industry, application or area of technology, even when the IPC codes within that community come from different branches of the IPC hierarchy. Furthermore, the clusters of technologies that relate to a common industry tend to be proximately located on the network; for example the large branch on the upper left of the network contains a number of technology clusters that are all associated with fiber and textile processing, or derivative products. However, some communities, especially of chemicals, can present a surprising mix due to their wide range of applications as, for instance, a community with technologies related to preservation of bodies (human, animals or plants) and adhesive materials.

Although we cannot infer causation, the heterogeneous network structure implies that there is a non-trivial clustering of specific technologies within geographic regions, consistent with the expected effects of both technological spillovers and agglomeration.

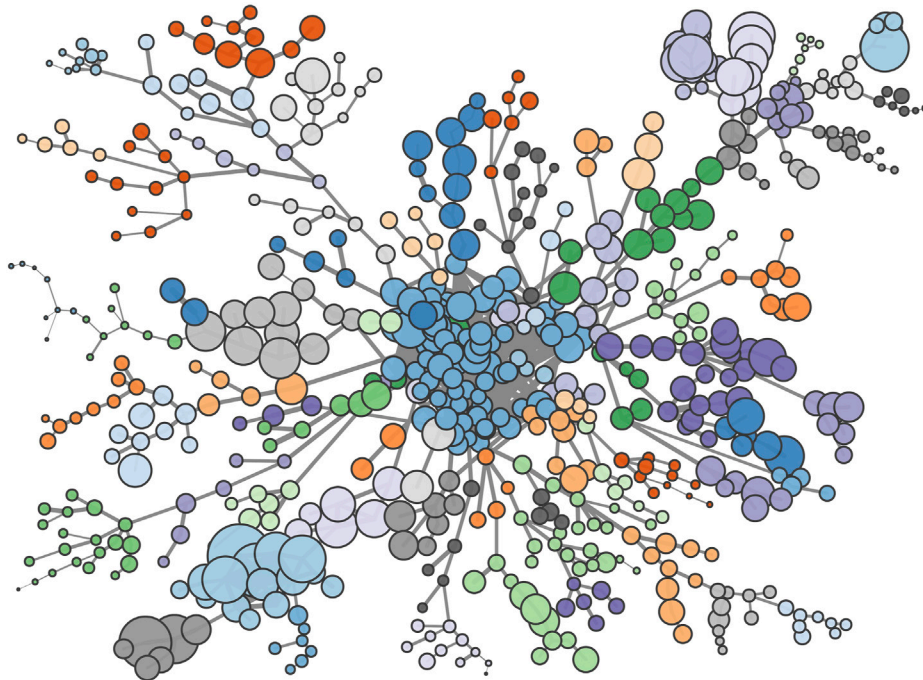


FIGURE 7 | Patent-space: nodes indicate the 632 IPC technology codes, linked by their likelihood of co-occurring within a geographic region. Node sizes are proportional to the number of times each code appears while colors indicate communities as determined by a modularity maximizing community detection method.

5 CONCLUSION

In this paper we used meta-data on the geographic location of patent applicants, and the technologies specific to the patents they filed, to investigate the regional structure of technological innovation. Using revealed comparative advantage as a metric to identify when regions produced a greater than expected number of patents related to a particular technology, we constructed a network linking regions to technologies.

The adjacency matrix associated with this network can be re-arranged to give a roughly triangular structure. This indicates that regions continue to produce patents relating to ubiquitous technologies even when they are producing patents that use technologies accessed by few other regions; a behavior that is inconsistent with a Ricardian model of technological specialization. We found that those regions which have a high diversity of technologies present in their patent portfolio tended to have a lower mean ubiquity of those technologies, relative to less technologically diverse regions. That is likely related to cumulative and path-dependent processes, characteristic of the evolutionary economic geography framework.

Examples of regions with high diversity are Zurich; Nord-Pas de Calais, a port/industrial area in north France, bordering the English Channel and Belgium; Gauteng, the province where Johannesburg is situated; Bern; Cologne; and Torino. On one hand, ubiquitous technologies include containers for storage (bags, barrels, bottles, etc.); treatment of water, waste water, and sewage; general building construction, such as walls, roofs,

insulation and others; shaping of plastic materials; and transporting or storage devices. On the other, some low ubiquity technologies are related to energy production, as fusion and nuclear reactors; computing (e.g. neural networks for image processing, cryptography, sensors); and aerospace activity.

The negative correlation between the diversity and mean ubiquity of the technologies in a region's patent portfolio is not simply a consequence of properties of the individual regions (e.g. population) of technologies (e.g. difficulty to access). Re-wiring null models that take account of such effects are insufficient to explain the observed structure. Nor is the structure simply due to the co-occurrence of multiple technology codes on individual patents. Randomizing which patents are assigned to which regions causes a significant increase in the mean ubiquity of the patent portfolio of all regions.

The null models suggest that the observed structure is due to co-occurrence of different technologies within geographic regions. The structure of the co-occurrences defines a network of technologies that tend to be located within the same sets of regions, related to the cognitive proximity of actors present in these regions. This network has a heterogeneous structure with a core of ubiquitous technologies surrounded by branches with communities of related technologies, associated with specific industries and product types. This non-trivial structure suggests that spillover and agglomeration effects are involved in the distribution of technological innovation. The Patent Space network gives a powerful tool for understanding the role that combinations of knowledge play in determining the success of innovation activities, and hence the economic prosperity, of regions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: <https://data.epo.org>.

AUTHOR CONTRIBUTIONS

DO'N, DV, and SH contributed to the design of the study. DO'N and DV performed the data analysis and wrote the manuscript. All authors approved the submitted version.

REFERENCES

- Attaran, M. (1986). Industrial Diversity and Economic Performance in U.S. Areas. *Ann. Reg. Sci.* 20, 44–54. doi:10.1007/bf01287240
- Baicker, K., and Chandra, A. (2010). "Understanding Agglomerations in Health Care," in *Agglomeration Economics*. Editor E. L. Glaeser (Chicago: Chicago University Press), 211–236.
- Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage. *Manchester Sch.* 33, 99–123. doi:10.1111/j.1467-9957.1965.tb00050.x
- Baldwin, R. E., and Okubo, T. (2006). Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting. *J. Econ. Geogr.* 6, 323–346. doi:10.1093/jeg/lbi020
- Boschma, R. A., and Frenken, K. (2017). Why Is Economic Geography Not an Evolutionary Science? towards an Evolutionary Economic Geography. *Economy* 6, 127–156. doi:10.4324/9781351159203-6
- Boschma, R., and Martin, R. (2010). The Aims and Scope of Evolutionary Economic Geography," in *The Handbook of Evolutionary Economic Geography*. Editors R. Boschma and R. Martin (Cheltenham: Edward Elgar), 3–42.
- Boschma, R. (2005). Proximity and Innovation: a Critical Assessment. *Reg. Stud.* 39, 61–74. doi:10.1080/0034340052000320887
- Buzard, K., Carlino, G. A., Hunt, R. M., Carr, J. K., and Smith, T. E. (2020). Localized Knowledge Spillovers: Evidence from the Spatial Clustering of R&D Labs and Patent Citations. *Reg. Sci. Urban Econ.* 81, 103490. doi:10.1016/j.regsciurbeco.2019.103490
- Buzard, K., Carlino, G. A., Hunt, R. M., Carr, J. K., and Smith, T. E. (2017). The Agglomeration of American R&D Labs. *J. Urban Econ.* 101, 14–26. doi:10.1016/j.jue.2017.05.007
- Castaldi, C., Frenken, K., and Los, B. (2015). Related Variety, Unrelated Variety and Technological Breakthroughs: an Analysis of US State-Level Patenting. *Reg. Stud.* 49, 767–781. doi:10.1080/00343404.2014.940305
- Chatterji, A., Glaeser, E. L., and Kerr, W. R. (2013). "Clusters of Entrepreneurship and Innovation," in NBER Working Paper Series, (Cambridge, MA: National Bureau of Economic Research (NBER)), 19013.
- Cohen, W. M., and Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Sci. Q.* 35, 128–152. doi:10.2307/2393553
- David, P. A., and Rosenbloom, J. L. (1990). Marshallian Factor Market Externalities and the Dynamics of Industrial Localization. *J. Urban Econ.* 28, 349–370. doi:10.1016/0094-1190(90)90033-j
- Duranton, G., and Overman, H. G. (2005). Testing for Localization Using Micro-geographic Data. *Rev. Econ. Stud.* 72, 1077–1106. doi:10.1111/0034-6527.00362
- Ellison, G., and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *J. Polit. Economy* 105, 889–927. doi:10.1086/262098
- Erdős, P., and Rényi, A. (1959). On Random Graphs I. *Publ. Math. Deb.* 6, 290–297.
- Feldman, M. P., and Audretsch, D. B. (1999). Innovation in Cities. *Eur. Econ. Rev.* 43, 409–429. doi:10.1016/s0014-2921(98)00047-6
- Feldman, M. P. (1999). The New Economics of Innovation, Spillovers and Agglomeration: A Review of Empirical Studies. *Econ. Innovation New Technol.* 8, 5–25. doi:10.1080/10438599900000002
- Frenken, K., Van Oort, F., and Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Reg. Stud.* 41, 685–697. doi:10.1080/00343400601120296
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *J. Polit. Economy* 100, 1126–1152. doi:10.1086/261856
- Glaeser, E. L. (1999). Learning in Cities. *J. Urban Econ.* 46, 254–277. doi:10.1006/juec.1998.2121
- Glaeser, E. L., and Ponzetto, G. A. M. (2010). "Did the Death of Distance Hurt Detroit and Help New York?," in *Agglomeration Economics*. Editor E. L. Glaeser (Chicago: Chicago University Press), 303–337.
- Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market Value and Patent Citations. *RAND J. Econ.* 36, 16.
- Hausmann, R., and Klinger, B. (2007). "The Structure of the Product Space and the Evolution of Comparative Advantage," in CID Working Paper Series, (Cambridge, MA: Center for International Development (CID), Harvard University), 146.
- Hidalgo, C. A., and Hausmann, R. (2009). The Building Blocks of Economic Complexity. *Proc. Natl. Acad. Sci.* 106, 10570–10575. doi:10.1073/pnas.0900943106
- Hidalgo, C. A., Klinger, B., Barabasi, A.-L., and Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science* 317, 482–487. doi:10.1126/science.1144581
- Howells, J. R. L. (2002). Tacit Knowledge, Innovation and Economic Geography. *Urban Stud.* 39, 871–884. doi:10.1080/00420980220128354
- Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.
- Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Q. J. Econ.* 108, 577–598. doi:10.2307/2118401
- Krugman, P. R. (1991). *Geography and Trade*. Cambridge: MIT press.
- Malizia, E. E., and Ke, S. (1993). The Influence of Economic Diversity on Unemployment and Stability*. *J. Reg. Sci.* 33, 221–235. doi:10.1111/j.1467-9787.1993.tb00222.x
- Maraut, S., Dernis, H., Webb, C., Spiezia, V., and Guellec, D. (2008). *The OECD REGPAT Database: A Presentation, STI working papers 2008/2*. Paris: OECD.
- Neffke, F., Henning, M., and Boschma, R. (2011). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Econ. Geogr.* 87, 237–265. doi:10.1111/j.1944-8287.2011.01121.x
- PATSTAT (2018). *Searching for Patents*. Available at: <https://data.epo.org>.
- Porter, M. E. (1990). *The Comparative Advantage of Nations*. New York: Free Press.
- Qian, G., Li, L., Li, J., and Qian, Z. (2008). Regional Diversification and Firm Performance. *J. Int. Bus. Stud.* 39, 197–214. doi:10.1057/palgrave.jibs.8400346
- Rosvall, M., and Bergstrom, C. T. (2008). Maps of Random Walks on Complex Networks Reveal Community Structure. *Proc. Natl. Acad. Sci.* 105, 1118–1123. doi:10.1073/pnas.0706851105
- Saxenian, A. (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, with a New Preface by the Author*. Cambridge: Harvard University Press.
- Sherwood-Call, C. (1990). *Assessing Regional Economic Stability: A Portfolio Approach*. San Francisco: Economic Review-Federal Reserve Bank of San Francisco, 17.
- Vasques Filho, D., and O'Neale, D. R. J. (2018). Degree Distributions of Bipartite Networks and Their Projections. *Phys. Rev. E* 98, 022307. doi:10.1103/PhysRevE.98.022307

ACKNOWLEDGMENTS

The authors thank two reviewers for their constructive comments and suggestions that helped us to significantly improve the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdata.2021.689310/full#supplementary-material>

- Vasques Filho, D., and O'Neale, D. R. J. (2020a). Transitivity and Degree Assortativity Explained: The Bipartite Structure of Social Networks. *Phys. Rev. E* 101, 052305. doi:10.1103/PhysRevE.101.052305
- Vasques Filho, D., and O'Neale, D. R. (2020b). The Role of Bipartite Structure in R&D Collaboration Networks. *J. Complex Networks* 8, cnaa016. doi:10.1093/comnet/cnaa016
- Wagner, J. E., and Deller, S. C. (1998). Measuring the Effects of Economic Diversity on Growth and Stability. *Land Econ.* 74, 541–556. doi:10.2307/3146884
- Weitzman, M. L. (1998). Recombinant Growth. *Q. J. Econ.* 113, 331–360. doi:10.1162/003355398555595
- Youn, H., Strumsky, D., Bettencourt, L. M. A., and Lobo, J. (2015). Invention as a Combinatorial Process: Evidence from Us Patents. *J. R. Soc. Interf.* 12, 20150272. doi:10.1098/rsif.2015.0272

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 O'Neale, Hendy and Vasques Filho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.