Technology Analysis & Strategic Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ctas20

A toolbox for smart specialization: relatedness, complexity and KETs based on technology flow matrices

Haixiao Wu

To cite this article: Haixiao Wu (13 Oct 2022): A toolbox for smart specialization: relatedness, complexity and KETs based on technology flow matrices, Technology Analysis & Strategic Management, DOI: 10.1080/09537325.2022.2131516

To link to this article: https://doi.org/10.1080/09537325.2022.2131516

TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT
https://doi.org/10.1080/09537325.2022.2131516

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

A toolbox for smart specialization: relatedness, complexity and KETs based on technology flow matrices

Haixiao Wu

Department of Economics, Nanjing Audit University, Nanjing, People's Republic of China

ABSTRACT

This paper proposes a toolbox for smart specialisation strategy (S3) that can provide region-specific suggestions about the prioritisation of technology domains based on technology flow matrices estimated with patent citation data. Upstream–downstream relations across technologies are found using citations as proxies, and the key enabling technologies and knowledge complexity are defined accordingly, integrating three principles of S3 that were formerly analyzed separately in the literature. The average propagation lengths model in production theory, which captures long-term and indirect linkages, is then utilised to improve the measure of relatedness across technologies in the current literature. The S3 domains of Lombardy, Italy, are re-examined, and suggestions are made regarding previously 'undiscovered gems'.

ARTICLE HISTORY
Received 23 February 2022
Revised 28 July 2022
Accepted 26 September 2022

KEYWORDS
Smart specialisation strategy; innovation network; relatedness; complexity

1. Introduction

The Smart Specialization Strategy (S3) proposed by Foray, David, and Hall (2009) is the central strategy of the EU Cohesion Policy. Three aspects of the S3 set it apart from traditional industrial policies: Specialisation is driven by entrepreneurial innovation activities rather than government guidance (Foray, David, and Hall 2009); regions specialise in technology fields (IPC, USPC, and CPC)[1] rather than industrial categories (e.g. HS, ISIC, and SITC)[2] (D'Adda, Iacobucci, and Palloni 2020); and knowledge linkages are emphasised rather than input–output relations, which promotes innovation 'at the intersection of different industries' (D'Adda, Iacobucci, and Palloni 2020).

Although the S3 has been widely studied, it is more a concept than a practical tool for policy design and implementation (Balland et al. 2019). Its drawbacks can be summarised as follows, based on Capello and Kroll (2016): Firstly, local authorities lack the analytical capacity to identify new technological domains related to existing local strengths. Consequently, local governments, especially those in less developed regions, simply replicate the high-tech activities adopted elsewhere. Secondly, while advanced regions usually show advantages in multiple fields, less innovative regions may not exhibit advantages in any high-tech field. Finally, the positive externality in knowledge spillover raises the concern that innovations in many technology fields are insufficiently provided.

To overcome these drawbacks, this paper presents a new analytical toolbox for policy making in relation to regional S3 domains. It uses the patent citation data and the average propagation lengths (APL) model to identify the extent of direct/indirect knowledge spillovers across technology fields. The toolbox for S3 policies is represented by a few result matrices encapsulating the complex

CONTACT Haixiao Wu wuhx@nau.edu.cn Department of Economics, Nanjing Audit University, Nanjing, JiangSu, 211815, People's Republic of China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09537325.2022.2131516 .

calculation of innovation networks. Based on the upstream–downstream technological linkages found in the matrices, three key concepts of S3 – knowledge relatedness, complexity, and key enabling technologies (KETs) – can be defined collectively in the same model, whereas they have thus far been treated individually in the literature. As depicted in Figure 1, for any researcher performing regional analysis who inputs a vector of existing advantages in a region of interest, the matrices generate indicators of the potential for branching into new related tech fields. This represents an improvement over the revealed technological advantages (RTAs) simply based on preexisting patent counts.

The externalities in knowledge spillovers relate this study to public economics debates about whether development policies like the S3 should be decentralised. The pros stem from the 'decentralization theorem' by Oates (1972). The cons argue that, when there are interjurisdictional externalities, factors like the structure of political institutions (Ponce-Rodríguez et al. 2020) and the conditionality attached to intergovernmental grants (Garzarelli and Keeton 2018) affect whether decentralisation can fulfil its welfare improving promises. The toolbox developed in this paper can be generally applied in both scenarios.

2. Literature review

The conventional wisdom in traditional industrial policies is for regions to specialise in the industries with revealed comparative advantage, in which economic activities are classified into industries by output (HS, ISIC, and SITC). In the context of high-tech activities, the outputs are often not physical (i.e. they consist of knowledge products such as patents and computer programmes) and the linkages among them are more difficult to track than transactions of physical goods or the movement of workers. As a result, researchers use patent data, with classification codes provided by patent offices (IPC, USPC, and CPC), to measure the revealed comparative advantage of technologies (Montresor and Quatraro 2017; Vlčková, Kaspříková, and Vlčková 2018).

2.1. Relatedness

Traditional industrial policies emphasise industrial chain development, exploring the potential for regions to branch into related industries. Relatedness in traditional industrial organisational theories is usually identified in input– output relations in production or inter-industry labour force flows (Capasso et al. 2019). Discussions about the S3 mainly focus on knowledge relatedness. Researchers widely acknowledge that the development of new high-tech activities should build on existing specialisations (Grillitsch and Asheim 2018). However, there is no agreement on a measure of relatedness. For instance, Santoalha (2019) and D'Adda, Iacobucci, and Palloni (2020) measure relatedness using 'the probability that a region is specialized in a specific pair of IPC classes,' while Kogler, Essletzbichler, and Rigby (2017), Vlčková, Kaspříková, and Vlčková (2018) and Balland et al. (2019) use the probability of 'two IPC classes appearing on the same patent.' The co-occurrence

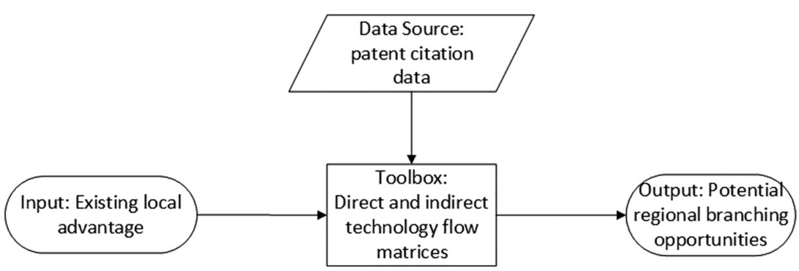

Figure 1. The toolbox identifies regional branching opportunities.

indicators of relatedness in the literature are not directional, which is a major limitation considering knowledge flows may be asymmetric. The knowledge sender is unlikely to have many innovations built on the existence of an advantage for the knowledge taker. In this case, the co-occurrence indicators that suggest regional branching into the knowledgesending field may fail. The knowledge spillover literature, represented by Jaffe, Trajtenberg, and Henderson (1993), Maurseth and Verspagen (2002), Malerba, Mancusi, and Montobbio (2013), Murata et al. (2014), and Kwon et al. (2022), supports the argument that knowledge spillover is directional and heterogeneous across technology fields.

However, very little of the S3 research focuses on exploiting the rich information about directional knowledge flows in patent citation relations. Industrial policy discussions are usually more concerned with technology spillovers across industrial sectors (Verspagen and de Loo 1999), which does not match the proposition of S3. The social/knowledge networks literature (Dosso and Lebert 2020; Lee et al. 2016) discusses the importance of regions, industries, firms and technologies in the innovation network. However, to build new advantage, regions need to know not only their current role in the network, but also their potential to rise in the hierarchy or branch out into new territory within the network. This cannot be achieved without an understanding of upstream-downstream relations. Rigby (2015) and Acemoglu, Akcigit, and Kerr (2016) represent two of the few studies to estimate the innovation network, regarding new patents as knowledge outputs of the technologies they cite. The current paper follows a similar path in the context of the EU and further investigates both direct and indirect relatedness in the long-term innovation network. Such knowledge flows are particularly conducive to industrial diversification in that a region can build advantage in a downstream technological territory (regional branching) by taking advantage of existing strengths.

2.2. Complexity

In the literature about related variety, Castaldi, Frenken, and Los (2015), Miguelez and Moreno (2018) and Grillitsch, Asheim, and Trippl (2018) argue that unrelated variety promotes more significant technological breakthroughs, which hints that the complexity of economic activities must be considered in addition to relatedness. Complexity has long been associated with development and efficiency (Hidalgo and Hausmann 2009). However, knowledge complexity is an understudied concept pertaining to the S3, argued by Balland and Rigby (2017) and Balland et al. (2019), possibly due to the lack of a common approach to measuring the complexity of technologies (Deegan, Broekel, and Fitjar 2021). Balland and Rigby (2017) apply Hidalgo and Hausmmann's economic complexity index based on the spatial co-occurrence probability. This method does not apply to the aspatial data used in this paper. Fleming and Sorenson (2001) take another approach to define complexity, using the USPC subclasses assigned as the number of components (complexity) of a patent. The present paper uses upstream classes (classes cited) instead of sub-classes assigned by patent examiners as the components of a technology, a case in which the Fleming and Sorenson (2001) approach reduces to something very similar to the Herfindahl-Hirschman Index.

2.3. Key enabling technologies

In addition to related and/or complex fields, fields that complement and enable innovations in many downstream technologies may be of interest. The KETs are six general-purpose technologies (GPTs) recommended by the European Commission (Sörvik, Rakhmatullin, and Palazuelos Martínez 2013): industrial biotechnology, nanotechnology, micro- and nanoelectronics, photonics, advanced materials, and advanced manufacturing technologies. Although not explicitly distinguished from GPTs, KETs emphasise enabling a region to acquire new technologies related to preexisting ones (Montresor and Quatraro 2017). In general-purpose technology theories (which are defined based on six characteristics, see the review by Bekar, Carlaw, and Lipsey 2018), this concept is referred to as the characteristic of generality or pervasiveness. Because other characteristics of GPTs overlap with complexity and relatedness, this paper uses a narrow concept of key enabling technology (KET) rather than GPT when referring only to the pervasive impact of technologies. The method by Moser and Nicholas (2004) is used to define KETs. After all, the GPT is implicitly considered if a S3 domain satisfies the requirements for relatedness, complexity and KET.

In sum, discussions about S3 face challenges with insufficient integration of three principles: relatedness, complexity and KETs. Two main streams of literature, innovation networks and knowledge spillover, can be referenced to find a solution. The former is concerned with the inter-regional or inter-field features of innovation networks (Érdi et al. 2013; Lee et al. 2016); the latter studies the factors affecting knowledge spillover (Jaffe, Trajtenberg, and Henderson 1993; Murata et al. 2014; Singh and Marx 2013). The specific question in S3 is the idiosyncratic prioritisation of innovation activities, which usually requires indices to measure the regional relative advantages or unrealised potentials across all technology domains. In contrast to the revealed comparative advantage based simply on patent counts and the relatedness measures based on the co-occurrence, this paper contributes to the literature by estimating directional relations in the innovation network, using the average propagation lengths (APL) model by Dietzenbacher, Romero, and Bosma (2005), Dietzenbacher and Romero (2007) and Jiang et al. (2020) to identify the direct and indirect relatedness of potential S3 domains to local strengths. Unlike many current studies, such as Vlčková, Kaspříková, and Vlčková (2018) and Santoalha (2019), which focus solely on relatedness, the estimated innovation network in this paper also helps to define knowledge complexity and KETs. This method is conducive to the development of 'undiscovered gems', the presently disadvantaged fields that are linked to strong local upstream conditions.

3. Data

Patent citation data is the most accessible high-quality micro-data source for studies of knowledge flows. This paper uses patent citations as proxies for upstream–downstream relations. The data is from the OECD citations database and is extracted primarily from the EPO's Worldwide Statistical Patent Database (PATSTAT, Spring 2020). It includes citation-pair information for citing EPO patents with priority years from 1977 to 2019. Because there is an average lag of 18 months between application date and publication date (Malerba, Mancusi, and Montobbio 2013), 2017 was selected as the ending priority year for the analysis. The OECD REGPAT database (January 2020) provides supplemental information about the locations of patent inventors, covering 283 Nomenclature of Territorial Units for Statistics 2 (NUTS2) regions.

Kuhn, Younge, and Marco (2020) summarised the implicit assumptions underlying the use of patent citation datasets in the analyses of knowledge spillover. Compared with their statements about the United States Patent and Trademark Office (USPTO) data, there are some challenges associated with using EPO data that should be noted or addressed in this study:

(1) Patents filed to the EPO are a selected sample.

Not all of the patents filed in the national patent offices of EPO member states are also filed with the EPO. Therefore it is likely that EPO patents represent a select group of high-quality patents that applicants are willing to pay extra to file with the organisation. Selecting high-quality patents is potentially beneficial for the study because it filters out low-quality patents that do not represent technological advancements.

This paper also limits the sample to patent citation pairs filed at the EPO and located in EPO member countries, because the locations of patents outside of EPO member countries are of low precision.

(1) In EPO data, a patent's IPC codes[3] and inventors are not ranked, so it can be challenging to determine the main IPC classes and locations of patents.

In data processing, a patent is assigned the IPC class (the first three digits of the IPC code) that appears most frequently in its list of IPC codes in documentation. To designate the location of an inter-regional coinvention, the patent is coded for multiple locations corresponding to the NUTS2 categorizations of the inventors. Each co-inventor is assumed to hold an equal proportion of the patent so that patent counts can be aggregated by region accordingly.

(1) The 'duty of disclosure' imposed by the USPTO does not exist at the EPO (Bacchiocchi and Montobbio 2010). Consequently, most citations are added by examiners at the EPO.

In USPTO data, researchers generally treat applicant-added citations as indicators of knowledge spillover. In EPO data, the lack of incentive for applicants to disclose all prior art means that the examiner-added citations combined with applicant-added citations (20% of total citations) are the best available proxy of knowledge flow. There is evidence that examiner-added citations are on average more relevant (Alcácer, Gittelman, and Sampat 2009; Hegde and Sampat 2009; Kuhn, Younge, and Marco 2020). Without the 'duty of disclosure', applicant-added citations at the EPO may be a less effective indicator of knowledge-relatedness than examiner-added citations.

4. Methods

4.1. The technology flow matrices

A technology flow matrix records the average propensity of a patent in the knowledge-taking field (column names of the matrix) to cite a patent in the knowledge-sending field (row names). Because the matrix can change with technological advancement, researchers usually perform their analysis over a 10-year time span (Acemoglu, Akcigit, and Kerr 2016; Singh and Marx 2013). The cohort of patent pairs chosen as the sample is limited to cited patents with priority years ranging from 1998 to 2007. The forward citations received by the sample cohort are restricted to citations that occur within a 10-year window after the cited patent's priority year, following Singh and Marx (2013) and Acemoglu, Akcigit, and Kerr (2016). The backward citations are limited to a 10-year window prior to the sample patents' priority year to avoid bias due to the truncation in forward citations, referencing Kuhn, Younge, and Marco (2020).

The three-digit IPC class codes were used as the technology fields for patents, allowing for a manageable number of 123 technology domains (Balland et al. 2019; D'Adda, Iacobucci, and Palloni 2020). An element of matrix K, k_{mn} represents the average increase in the number of new citing patents in field n due to spillover from a cited patent in field m:

$$k_{mn} = \frac{\sum_{i \in C_{mn}} \theta_i}{N_m} \tag{1}$$

where C_{mn} denotes the set of all forward citations field m received from field n. And i is an index for individual citations. θ_i is the average contribution of a cited patent in a citation and is equivalent to giving a weight to each citation pair equal to the reciprocal of the number of citations made by the citing patent.[4] Compared to prior work (Jiang et al. 2020), this weighting method corrects for the bias introduced by the patents that have an unrealistic number of citations made by the citing patent, which is discussed in Kuhn, Younge, and Marco (2020). N_m represents the total number of patents in technology field m. The pairwise relatedness measures, k_{mn}'s, form a technology flow matrix K:

$$\begin{bmatrix} k_{1,1} & \cdots & k_{1,123} \\ \vdots & \ddots & \vdots \\ k_{123,1} & \cdots & k_{123,123} \end{bmatrix}. \tag{2}$$

Similarly to the production theory, there is a Gosh inverse matrix, G:

$$G = (I - K)^{-1} = I + K + K^2 + \ldots \tag{3}$$

Matrix G gives the long-run multiplier effects of a new patent, considering indirect effects due to multi-generation diffusion. Given the number of local patents in each IPC class, one can predict the short-run and long-run output of new patents based on the inter-class relatedness shown in Matrices K and G. Imagine that a new patent has been granted in field m. Then Row m of Matrix K represents the first-generation spillover to other fields due to this marginal change, or the short-run effect. Consider Row_m as the vector of new patents as a result of first-generation spillover. These new patents will induce a vector of second-generation spillovers given by $Row_m \times K$. This propagation process goes on, with the vector of new patents multiplied by Matrix K one more time for each new generation, eventually leading to a long-run effect on all technological fields represented by Row m of Matrix G.

Similarly, an element of the absorption coefficient matrix A represents the average quantity of patents from field m 'absorbed' by a new EPO patent in field n:

$$a_{mn} = \frac{\sum_{i \in C_{mn}} \eta_i}{N_n} \tag{4}$$

where η_i is the reciprocal of the number of patent citations that the cited patent received.[5] N_n represents the total number of patents in citing field n.

4.2. The APL model

Although matrices K and G give a prediction of the short-run and long-run output in new patents, it remains uncertain how long it will take for the longrun effect to occur, which matters for policy making. The APL model by Dietzenbacher, Romero, and Bosma (2005), Dietzenbacher and Romero (2007) and Jiang et al. (2020) solves this problem by estimating the average propagation lengths in knowledge spillover:

$$H = G(G - I)$$
$$APL_{mn} = \begin{cases} \dfrac{h_{mn}}{g_{mn}};\ \text{if } m \neq n \\ \dfrac{h_{mn}}{g_{mn} - 1};\ \text{if } m = n. \end{cases} \tag{5}$$

h_{mn} stands for an element of the H matrix in row m and column n. APL_{mn} is a measure of the average 'steps' needed for field m to spill over to field n. *A region can more easily generate new patents in technology field n if its current advantaged fields have small APLs towards field n*. If patents in all IPC classes are regarded as equally 'valuable', the downstream classes with smaller APLs from the existing local strengths should be chosen by a region as the prioritised S3 fields.

4.2. Measure of KETs and knowledge complexity

Policy makers and economists often believe that patents in different IPC classes are not equal in that certain fields yield more value-added or are central for economic development. This paper considers the KETs from the perspective of the extent of knowledge diffusion. For each row (cited IPC class) of Matrix K, a Herfindahl-Hirschman Index (HHI-K) can be calculated following Moser and Nicholas (2004). The HHI-K of a technology decreases with its pervasive spillovers to other classes. An IPC class is called a KET if its HHI-K is less than the mean, 0.460.

Balland et al. (2019) assert that more complex innovation activities generate more value-added, proposing the use of knowledge complexity in addition to relatedness in smart specialisation. An HHI-A can be calculated for each column (IPC class) of Matrix A. A small HHI-A means a dispersed distribution of backward citations (knowledge input) across upstream technology fields. An HHI-A that is less than the mean of 0.452 is a sign of complexity. If there is a concern about the arbitrariness in setting the HHI-K and HHI-A thresholds, one can always turn back to the HHI values for more quantitative perceptions of complexity and KET.

5. Results: APL and S3 domains, using Lombardy, Italy, as an example

There are 123 IPC classes in the OECD citations database, so the resulting matrices are 123×123.[6] Matrices K, G, A, H, and APL and the associated indices are the main results of this paper. As briefly shown in Figure 1, these matrices can be used to analyze the relatedness across, and the complexity of, technology classes and help to select which technology domains should be prioritised in smart specialisation. However, the matrices are too large to be shown in the paper and are therefore provided online as supplementary materials (see Tables A.1 – A.6 in Online Appendices).

The basic features of the innovation network can be better understood through a brief review of all result Tables. In Table A.1, without considering multi-generation diffusion, the short-run technology flows are shown as Matrix K, reflecting how each IPC class affects other classes. Elements along the diagonal indicate average contributions of intra-class spillovers. Table A.1 also shows that interclass-spillovers are asymmetric. For example, the average contribution of a patent in A23 to a patent in A22 is 0.0034, while A22 has an average contribution of 0.0088 per patent to A23. After considering multi-generational diffusion, the long-run average contributions of knowledge spillovers are shown in Table A.2 as Matrix G. Table A.3 exhibits Matrix A, the absorption coefficient matrix, representing the technological ‘composition’ of IPC classes. Matrix A is also a short-run technology flow matrix. For example, the element in Row 4 and Column 1 of Matrix A denotes that the existence of a patent in A01 on average requires 0.0024 pre-existing patents in A23. Matrix H in Table A.4 is an intermediate step towards Matrix APL in A.5. An APL of 2.85 (e.g. Row 5 and Column 1 of Matrix APL) means an average of 2.85 propagation steps are required for the knowledge in A24 to reach A01. Table A.6 reports resulting indices based on the matrices.

To show how the resulting matrices and indices contribute to policy analysis, this section implements the approach in the context of Lombardy, Italy. Lombardy has declared a few regional domains in its S3 document. Table 1 from D’Adda et al. (2019) lists the IPC classes (Column 3) associated with Lombardy’s S3 domains. A debate exists concerning whether smart specialisation should be an outcome of a bottom-up or topdown approach, as discussed in McCann and Ortega-Argiles

Table 1. IPCs associated with the S3 domains of Lombardy.

Region	Regional domains	Corresponding IPC classes
Lombardy	Aerospace	A23 A61 A62 B01
	Agrifood	B44 B60 B63 B64
	Green manufacturing	B81 B82 C22 D01
	Health	G02 G06 G09 G21
	Artistic and cultural industries	H01 H02 H04
	Advanced manufacturing	
	Sustainable mobility	

Source: D’Adda et al. (2019).

(2013, page 409) and Santoalha (2019). The method presented in this paper can be applied to either a centralised or decentralised policy.

5.1. Application to a centralized policy

When sub-national specialisation is part of a national development project, the regional S3 domains are predetermined. A regional policy can only seek to strengthen upstream technology fields of the selected domains. To use the toolbox, first create a 123×1 vector of technology domains, M_d. There are 19 S3 domains selected by Lombardy. In vector M_d, an element = 1 if the corresponding domain is chosen by Lombardy, and 0 otherwise. Matrix APL contains all APL_{mn}'s calculated according to Equation 5. Then elements of vector $P = \frac{1}{19}\mathrm{APL} * M_d$ represent the average steps from each technology domain to the 19 domains chosen by Lombardy. The IPC classes with the smallest average APLs are linked to the chosen S3 domains most closely.

Table 2 lists the seven upstream fields most closely linked to the IPC classes chosen by Lombardy, three of which are already its S3 domains. The RTA based on the number of patents since 2008 is calculated using a standard Balassa method for each class.[7] Column 3 shows whether a class is presently advantaged (RTA>1). Clearly, Lombardy is mainly prioritising its advantaged classes as its S3 domains. If Lombardy can build additional advantages in the fields in Table 2, benefiting from the spillovers from these upstream fields, its S3 domains may have a better foundation for generating new patents. In some intuitive instances, S3 domain G06 (Computing; Calculating or Counting) benefits from spillovers from G01 (Measuring; Testing); the fact that S3 domains B63 and B64 are related to ships and aircraft can utilise innovations in upstream class C09 (Dyes; Paints; Adhesives; etc.); upstream classes of B29 (Working of Plastics) and C08 (Organic Macromolecular Compounds) can extensively contribute to S3 domains like B82 Nanotechnology, B81 Microstructural Technology, and A61 Medical and Veterinary Science.

The central government can also benefit from this method of analysis. Suppose a country plans to establish new advantage in electronic circuitry and needs to choose a region as the innovation centre. The APL indices show the relative relatedness of the 123 IPC classes to electronic circuitry. And the region that is advantageous in the IPC classes with the smallest APL is the region with best opportunity of developing circuitry technology.

5.2. Application to a decentralized policy

A decentralised policy allows regions to determine their S3 domains at their discretion. Here applying the toolbox helps to determine the best opportunities. Given a 1×123 vector M_s of binary indicators of whether a field is presently advantaged (RTA>1), the elements of vector $Q = M_s * \mathrm{APL}$ represent the steps from presently advantaged fields to each field of interest. *The fields with the smallest APLs, as shown in Table 3 for Lombardy, should be chosen as the regional S3 domains to be developed.* The average APL<2 in Table 3.

Table 2. Top 7 upstream fields by APL for Lombardy.

IPC class	S3 domain	RTA	Description
A61	1	>1	MEDICAL OR VETERINARY SCIENCE; HYGIENE
G01	0	<1	MEASURING; TESTING
C09	0	<1	DYES; PAINTS; POLISHES; NATURAL RESINS; ADHESIVES; COMPOSITIONS NOT OTHERWISE PROVIDED FOR; APPLICATIONS OF MATERIALS NOT OTHERWISE PROVIDED FOR
H01	1	>1	BASIC ELECTRIC ELEMENTS
B01	1	>1	PHYSICAL OR CHEMICAL PROCESSES OR APPARATUS IN GENERAL
B29	0	>1	WORKING OF PLASTICS; WORKING OF SUBSTANCES IN A PLASTIC STATE IN GENERAL
C08	0	<1	ORGANIC MACROMOLECULAR COMPOUNDS; THEIR PREPARA-TION OR CHEMICAL WORKING-UP; COMPOSITIONS BASED THEREON

Refer to Online Appendices for complete list of 123 domains.

Eight of the top ten IPC classes suggested by the toolbox also have RTA>1. Lombardy has already prioritised three out of these eight fields. Overall, only 8 of the 36 presently advantaged fields are selected as the S3 domains, indicating that RTA is not the only factor considered. Some advantaged classes like A47 and B65 are not prioritised with good reason: these are domains in which patents are not good indicators of local strengths. This matter relates to the discussions of the KETs and the complexity of technologies in the next part.

More importantly, a region can also branch into disadvantaged classes (G01 and C08 in Table 3) if the region is advantageous in their corresponding upstream classes. This has been largely ignored in the analysis of S3 and has also been ignored by Lombardy. For example, the underlying assumption of D'Adda, Iacobucci, and Palloni (2020) is that S3 domains should be advantaged fields with high degrees of relatedness. In another example, although Vlčková, Kaspříková, and Vlčková (2018) assess the current specialisation of German regions based on measures of RTA and average relatedness, they admit that the methodology to identify prospective industries is insufficiently developed. A policy focusing on the already-advantaged fields misses a key principle of S3: diversifying or branching into new technologies which are initially disadvantaged. The toolbox in this paper provides a data-driven approach to discovering such opportunities.

5.2.1. KETs and knowledge complexity

When multiple domains are suggested according to the APL standard, KETs and knowledge complexity become the tie-breakers.

Column 4 of Table 4 shows the KETs of the top 10 related IPC classes. If HHI-K < 0.469, the class is defined as a KET (= 1). The S3 domains do not match well with KETs among IPC classes with small APLs. However, 12 of all the 19 S3 domains are KETs. This indicates that Lombardy paid some attention to the concept of KET but ignored the most promising KETs in terms of APLs that are presently disadvantaged (e.g. C08). This may be due to the cognitive limitation to catch up with technological transitions that have 'generalized the purpose' of technologies that were formerly not KETs. For instance, cleaning technologies in class B08 do not seem to have wide applications and are reasonably omitted by researchers and government authorities who take a static view of the scope of KETs. Possibly because of the prevailing emission controls and the rise of green manufacturing, B08 now has sizable spillovers to traditional manufacturing classes like F16 and green energy classes like F24, according to either Matrix K or Matrix APL.

A domain is complex if it absorbs knowledge from a dispersed range of technology fields. Let Complex = 1 in Column 7 if HHI-A<0.452. In Table 4, the S3 domains do not match well with complex IPC classes. However, 10 of the 19 S3 domains are complex, suggesting that Lombardy

Table 3. Top 10 related fields by APL in Lombardy.

IPC class	S3 domain	RTA	Description
B29	0	>1	WORKING OF PLASTICS; WORKING OF SUBSTANCES IN A PLASTIC STATE IN GENERAL
B01	1	>1	PHYSICAL OR CHEMICAL PROCESSES OR APPARATUS IN GENERAL
A47	0	>1	FURNITURE; DOMESTIC ARTICLES OR APPLIANCES; COFFEE MILLS; SPICE MILLS; SUCTION CLEANERS IN GENERAL
G01	0	<1	MEASURING; TESTING
A61	1	>1	MEDICAL OR VETERINARY SCIENCE; HYGIENE
F16	0	>1	ENGINEERING ELEMENTS OR UNITS; GENERAL MEASURES FOR PRODUCING AND MAINTAINING EFFECTIVE FUNCTIONING OF MACHINES OR INSTALLATIONS; THERMAL INSULATION IN GENERAL
B65	0	>1	CONVEYING; PACKING; STORING; HANDLING THIN OR FILAMENTARY MATERIAL
H01	1	>1	BASIC ELECTRIC ELEMENTS
B08	0	>1	CLEANING
C08	0	<1	ORGANIC MACROMOLECULAR COMPOUNDS; THEIR PREPARATION OR CHEMICAL WORKING-UP; COMPOSITIONS BASED THEREON

Refer to Online Appendices for complete list of 123 domains.

Table 4. Top 10 fields by APL in Lombardy, with KET and complexity.

IPC classes	S3 domain	RTA	KET	Complex
B29	0	>1	1	1
B01	1	>1	1	1
A47	0	>1	0	0
G01	0	<1	0	0
A61	1	>1	0	0
F16	0	>1	1	0
B65	0	>1	0	0
H01	1	>1	0	0
B08	0	>1	1	1
C08	0	<1	1	0

Refer to Online Appendices for complete list of 123 domains.

have considered complexity to a partial extent, but ignored some other complex fields for similar reasons as in the case of KET.

5.3. Discussions of the application to policy making

The role of local government in S3 implementation is to avoid market and government failures in the implementation of S3. The first type of market failure that can be experienced in S3 policy stems from positive externality in knowledge spillover, resulting in an insufficient supply of innovations, especially for KETs. The second type of market failure originates from imperfect information. Economic agents may not have all the necessary information about technologies in fields other than their own specialisations, which makes it difficult to develop innovations in complex technologies. Finally, government failure arises from the tendency to dissipate public resources in activities that do not need much intervention and a lack of capability to discover opportunities and support new domains.

The approach taken in this paper helps to avoid market and government failures by standardising the procedures to discover advantaged, related, complex and/or key enabling technologies of regions, alleviating the problem of imperfect inofrmation. The KETs insufficiently provided by the private sector can now be better supported by well-informed local governments. In the application to Lombardy in Table 4, Classes A61 and H01, with RTA>1, are already prioritised as S3 domains. Conversely, A47, F16 and B65, with RTA>1, are not prioritised. The five advantaged classes mentioned above are neither KETs nor complex, meaning that the problems of externality and imperfect information may not seriously impede their development. Therefore, extensive supports for classes A61 and H01 are not necessary. The classes of B01 (already an S3 domain), B08, and B29, all being presently advantaged, complex and KETs, are more obvious candidates for S3 domains. When Lombardy is looking for regional branching into presently disadvantaged fields (RTA<1), G01 and C08 are candidates. They closely relate to Lombardy's local strengths in terms of APL and are KETs that spill over to other fields.

In summary, the toolbox determines the S3 domains following the standard steps:

(1) Look up the RTAs calculated based on new local patents in the last decade.
(2) Combine the RTAs with the APL matrix to generate a list of most-related candidate domains in terms of APL.
(3) Prioritise KET and complex domains with RTA>1 in the list.
(4) When branching into disadvantaged fields, prioritise domains with small APL. KETs and complexity act as tie-breakers.

Conclusion

Many implementations of S3 rely on RTAs based on simple patent counts, and co-occurrence measures may ignore 'undiscovered gems' in presently disadvantaged fields. Efforts to measure relatedness, KETs and complexity have thus far not been integrated into the same model. This paper differs from other S3 research in using directional upstream-downstream relations rather than co-occurrences to measure relatedness, which is more carefully defined by accounting for asymmetric inter-field knowledge flows. Use of the APL model in this approach is more advantageous than using simple short-run technology flow matrices in that it takes into account both the multi-generation diffusion and the 'velocity' of diffusion. Furthermore, application of the APL model helps in determining the fields that possess the potential advantages to generate new patents. Properties generated from the directional technology flow matrices are also used to define the KETs and knowledge complexity of IPC classes. The methodology in this paper provides a data-driven analytical tool that properly defines and encompasses the principles of relatedness, KETs, knowledge complexity, and local RTAs, which can be easily operationalised with the matrices given. For more detailed results, this method can also be applied at a finer IPC classification level or geographical level.

This method can be applied to determine or re-examine the S3 domains of regions and to make suggestions regarding formerly undiscovered opportunities, either in a centralised or decentralised paradigm. Reference to this paper can complement the current policy-making process of 'debate and engagement among local parties with different interests and preferences and also between local and central government actors' (McCann and Ortega-Argiles 2013, 409). The inter-regional externality in knowledge spillover violates the condition for the 'decentralization theorem' (Oates 1972) to hold, which possibly necessitates some degree of centralisation in the form of political institutional changes (Ponce-Rodríguez et al. 2020) or conditional grants (Garzarelli and Keeton 2018). It has been shown that the S3 toolbox can apply generally, but the optimal degree of centralisation and the instruments for inter-governmental coordination with interregional knowledge spillovers require further investigation. Future research can also explore the impact of interregional spillovers on local advantages.

Notes

1. Three major patent classification systems: International Patent Classification (IPC), United States Patent Classification (USPC) and Cooperative Patent Classification (CPC).
2. Classifications of products or economic activities: Harmonized Commodity Description and Coding Systems (HS), International Standard Industrial Classification (ISIC) and Standard International Trade Classification (SITC).
3. A patent may be assigned multiple IPC codes by the EPO examiner.
4. This number includes non-patent literature and cited patents outside of EPO.
5. Unlike in the calculation of k_{mn}, non-patent literature and forward citations outside of the EPO are not observable, so only EPO-citing patents are included in the calculation. This means that the absorption coefficient matrix is estimated with less precision.
6. There are three newly added classes that are extremely inactive: B33, G12 and G16. Not many applications occur in the chosen cohort of patents from 1998 to 2007.
7. A similar application of RTA to patents can be found in Montresor and Quatraro (2017) and Vlčková, Kašpříková, and Vlčková (2018). In applications to regions with very few patents, users of the RTA and other indicators developed in this paper should be cautious. These indicators are not good indicators of local strength for the regions without a foundation for any kind of smart specialisation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper was partly supported by the National Natural Science Foundation of China (Grant number 72103099).

Notes on contributor

Haixiao Wu received the Ph.D. degree from the George Washington University, USA. His research interests includegeography of innovation, regional and urban economics. He is currently an Assistant Professor with Nanjing Audit University.

ORCID

Haixiao Wu http://orcid.org/0000-0002-2078-9665

References

Acemoglu, D., U. Akcigit, and W. R. Kerr. 2016. "Innovation Network." *Proceedings of the National Academy of Sciences of the United States of America* 113 (41): 11483–11488.

Alcácer, J., M. Gittelman, and B. Sampat. 2009. "Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis." *Research Policy* 38 (2): 415–427.

Bacchiocchi, E., and F. Montobbio. 2010. "International Knowledge Diffusion and Home-Bias Effect: Do USPTO and EPO Patent Citations Tell the Same Story?" *Scandinavian Journal of Economics* 22: 441–470.

Balland, P.-A., R. Boschma, J. Crespo, and D. L. Rigby. 2019. "Smart Specialization Policy in the European Union: Relatedness, Knowledge Complexity and Regional Diversification." *Regional Studies* 53 (9): 1252–1268.

Balland, P.-A., and D. Rigby. 2017. "The Geography of Complex Knowledge." *Economic Geography* 93 (1): 1–23.

Bekar, C., K. Carlaw, and R. Lipsey. 2018. "General Purpose Technologies in Theory, Application and Controversy: A Review." *Journal of Evolutionary Economics* 28 (5): 1005–1033.

Capasso, M., E. J. Iversen, A. Klitkou, and T. Sandven. 2019. "Which Region to Choose for an Industrial Policy? a Research Path to Highlight Restructuring Opportunities." *European Planning Studies* 27 (8): 1461–1482.

Capello, R., and H. Kroll. 2016. "From Theory to Practice in Smart Specialization Strategy: Emerging Limits and Possible Future Trajectories." *European Planning Studies* 24 (8): 1393–1406.

Castaldi, C., K. Frenken, and B. Los. 2015. "Related Variety, Unrelated Variety and Technological Breakthroughs: An Analysis of US State-Level Patenting." *Regional Studies* 49 (5): 767–781.

D'Adda, D., E. Guzzini, D. Iacobucci, and R. Palloni. 2019. "Is Smart Specialisation Strategy Coherent with Regional Innovative Capabilities?" *Regional Studies* 53 (7): 1004–1016.

D'Adda, D., D. Iacobucci, and R. Palloni. 2020. "Relatedness in the Implementation of Smart Specialisation Strategy: A First Empirical Assessment." *Papers in Regional Science* 99 (3): 405–425.

Deegan, J., T. Broekel, and R. D. Fitjar. 2021. "Searching Through the Haystack: The Relatedness and Complexity of Priorities in Smart Specialization Strategies." *Economic Geography* 97 (5): 497–520.

Dietzenbacher, E., and I. Romero. 2007. "Production Chains in an Interregional Framework: Identification by Means of Average Propagation Lengths." *International Regional Science Review* 30 (4): 362–383.

Dietzenbacher, E., I. Romero, and N. Bosma. 2005. "Using Average Propagation Lengths to Identify Production Chains in the Andalusian Economy." *Estudios de economía aplicada, ISSN 1697-5731* 23 (2): 405–422.

Dosso, M., and D. Lebert. 2020. "The Centrality of Regions in Corporate Knowledge Flows and the Implications for Smart Specialisation Strategies." *Regional Studies* 54 (10): 1366–1376.

Érdi, P., K. Makovi, Z. Somogyvári, K. Strandburg, J. Tobochnik, P. Volf, and L. Zalányi. 2013. "Prediction of Emerging Technologies Based on Analysis of the US Patent Citation Network." *Scientometrics* 95 (1): 225–242.

Fleming, L., and O. Sorenson. 2001. "Technology as a Complex Adaptive System: Evidence from Patent Data." *Research Policy* 30 (7): 1019–1039.

Foray, D., P. A. David, and B. Hall. 2009. "Smart Specialisation: The Concept." *Knowledge Economists Policy Brief* 9 (85): 100.

Garzarelli, G., and L. Keeton. 2018. "Laboratory Federalism and Intergovernmental Grants." *Journal of Institutional Economics* 14 (5): 949–974. doi:10.1017/S1744137417000595.

Grillitsch, M., and B. Asheim. 2018. "Place-Based Innovation Policy for Industrial Diversification in Regions." *European Planning Studies* 26 (8): 1638–1662.

Grillitsch, M., B. Asheim, and M. Trippl. 2018. "Unrelated Knowledge Combinations: The Unexplored Potential for Regional Industrial Path Development." *Cambridge Journal of Regions, Economy and Society* 11 (2): 257–274.

Hegde, D., and B. Sampat. 2009. "Examiner Citations, Applicant Citations, and the Private Value of Patents." *Economics Letters* 105 (3): 287–289.

Hidalgo, C. A., and R. Hausmann. 2009. "The Building Blocks of Economic Complexity." *Proceedings of the National Academy of Sciences* 106 (26): 10570–10575.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3): 577–598.

Jiang, Z., H. Song, J. Wang, M. Hou, and D. Xie. 2020. "APL Identification Method for Breakthrough Opportunities of Emerging Technologies in Technology Networks (in Chinese)." *Science of Science and Management of S.&.T* 41 (12): 71–82.

Kogler, D. F., J. Essletzbichler, and D. L. Rigby. 2017. "The Evolution of Specialization in the EU15 Knowledge Space." *Journal of Economic Geography*, 17 (2): 345–373.

Kuhn, J., K. Younge, and A. Marco. 2020. “Patent Citations Reexamined.” *The RAND Journal of Economics* 51 (1): 109–132.
Kwon, H.-S., J. Lee, S. Lee, and R. Oh. 2022. “Knowledge Spillovers and Patent Citations: Trends in Geographic Localization, 1976–2015.” *Economics of Innovation and New Technology* 31 (3): 123–147.
Lee, S., W. Kim, H. Lee, and J. Jeon. 2016. “Identifying the Structure of Knowledge Networks in the US Mobile Ecosystems: Patent Citation Analysis.” *Technology Analysis & Strategic Management* 28 (4): 411–434.
Malerba, F., M. L. Mancusi, and F. Montobbio. 2013. “Innovation, International R&D Spillovers and the Sectoral Heterogeneity of Knowledge Flows.” *Review of World Economics* 149 (4): 697–722.
Maurseth, P. B., and B. Verspagen. 2002. “Knowledge Spillovers in Europe: A Patent Citations Analysis.” *Scandinavian Journal of Economics* 104 (4): 531–545.
McCann, P., and R. Ortega-Argiles. 2013. “Transforming European Regional Policy: A Results-Driven Agenda and Smart Specialization.” *Oxford Review of Economic Policy* 29 (2): 405–431.
Miguelez, E., and R. Moreno. 2018. “Relatedness, External Linkages and Regional Innovation in Europe.” *Regional Studies* 52 (5): 688–701.
Montresor, S., and F. Quatraro. 2017. “Regional Branching and Key Enabling Technologies: Evidence from European Patent Data.” *Economic Geography* 93 (4): 367–396.
Moser, P., and T. Nicholas. 2004. “Was Electricity a General Purpose Technology? Evidence from Historical Patent Citations.” *American Economic Review* 94 (2): 388–394.
Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura. 2014. “Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach.” *Review of Economics and Statistics* 96 (5): 967–985.
Oates, W. E. 1972. *Fiscal Federalism*. New York [etc.]: Harcourt Brace Jovanovich.
Ponce-Rodríguez, R. A., C. R. Hankla, J. Martinez-Vazquez, and E. Heredia-Ortiz. 2020. “The Politics of Fiscal Federalism: Building a Stronger Decentralization Theorem.” *Journal of Theoretical Politics* 32 (4): 605–639.
Rigby, D. L. 2015. “Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes.” *Regional Studies* 49 (11): 1922–1937.
Santoalha, A. 2019. “New Indicators of Related Diversification Applied to Smart Specialization in European Regions.” *Spatial Economic Analysis* 14 (4): 404–424.
Singh, J., and M. Marx. 2013. “Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity.” *Management Science* 59 (9): 2056–2078.
Sörvik, J., R. Rakhmatullin, and M. Palazuelos Martínez. 2013. “Preliminary Report on KETs Priorities Declared by Regions in the Context of Their Work on Research and Innovation Strategies for Smart Specialisation (RIS3).” *JRC Technical Report. Seville, Spain: European Commission*.
Verspagen, B., and I. de Loo. 1999. “Technology Spillovers Between Sectors.” *Technological Forecasting and Social Change* 60 (3): 215–235.
Vlčková, J., N. Kaspříková, and M. Vlčková. 2018. “Technological Relatedness, Knowledge Space and Smart Specialisation: The Case of Germany.” *Moravian Geographical Reports* 26 (2): 95–108.