

The Role of Network Topology and the Spatial Distribution and Structure of Knowledge in Regional Innovation Policy: A Calibrated Agent-Based Model Study

Ben Vermeulen^{1,2} · Andreas Pyka¹

Accepted: 2 December 2017 / Published online: 13 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract Recently, new regional innovation policy paradigms emerged that transcend a long-lived dispute on whether either regional specialization, diversification, or rather related variety is most conducive to regional innovativeness. This includes ‘smart specialization’ in which regions are deliberately specialized and connected following technological relatedness, ‘gatekeepers’ in which there are pipelines between dense regional networks, and a ‘regional mix’ of knowledge that allows sustained ‘branching’. We develop and use an agent-based model to study the conjecture that, to stimulate (supra)regional innovativeness, social planners need to consider both, in conjunction, the mix of technological knowledge in regions and the (inter)regional innovation network topology. We use this agent-based model to evaluate the performance of and study the internal mechanisms of these new policy paradigms in a variety of scenarios. To increase the external validity of our findings, we calibrate the knowledge graph searched by the agents in our model to the OECD patent database. We

This work was realized with financial support of the DACH research program *Innovation Networks for Regional Development* funded by the German science foundation DFG, Grant PY 70/8-1, extending work done within the *Netwerken van netwerken* program funded by the Dutch science foundation NWO, Grant 458-03-112. Moreover, part of the writing was finished during an extended stay at the AI-ECON Research Center of the National Chengchi University (NCCU) funded by the MOST-DAAD Germany-Taiwan PPP, Project 57219050. We are grateful for helpful and inspiring feedback from Bin-Tzong Chie, Shu-Heng Chen, Giuseppe Vizzari, Holger Graf, Ingrid Ott, Matthias Müller, Tobias Buchmann, several anonymous reviewers, and participants at EMAEE 2017, ERSa 2014, WEHIA 2014, and IEEE-CIFer 2014.

✉ Ben Vermeulen
b.vermeulen@uni-hohenheim.de

¹ Institute of Economics, Chair of Innovation Economics, Hohenheim University, Wollgrasweg 23, 70599 Stuttgart, Germany

² Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Den Dolech 2, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

confirm that access to related variety is important, yet that, on top of that, access to incidentally related knowledge is crucial to prevent high-level lock-in and ensure long term technological progress. Moreover, we find that networks with regional gatekeepers are particularly innovative, because these gatekeepers form ‘knowledge hubs’, create short paths to potential partners, and enlarge the total pool of knowledge. In case agents have few relationships, we find exceptionally high performance for the gatekeeper network in combination with regional diversification. The smart specialization network is a solid second option, although it lacks access to incidentally related knowledge and thus will ultimately fall behind whenever agents have relatively few relationships. The study elaborates on specific scenarios to reveal intricacies in the relationship between knowledge distribution, network topology, and the structure of interrelationships between knowledge fields.

Keywords Regional innovativeness · Innovation network · Specialization · Diversification · Related variety · Gatekeeper · Smart specialization · Branching · Agent-based model · Patent · Social planner

1 Introduction

In the knowledge-based perspective on innovation economics literature, there is an ongoing debate on whether technological *specialization* or rather *diversification* of research and development activities is conducive to regional innovativeness (Glaeser et al. 1992; Panne 2004; Paci and Usai 2000; Feldman and Audretsch 1999), or, as an additional alternative, whether regions should contain a mix of technologically related research activities (cf. Balland et al. 2015; Bristow 2010; Frenken et al. 2007). Taking technological discoveries as due to finding a suitable combination of existing knowledge (cf. Arthur 2009), firms in diversified regions face many unrelated knowledge flows and ‘infeasible’ combinations, thus lowering dynamic efficiency, while, in contrast, firms in specialized regions may be dynamically efficient but may not have access to technological path-breaking combinations and thus face a lock-in (Menzel and Fornahl 2010; Saxenian 1996; Hassink 2005; Martin and Sunley 2011). Arguably, this body of literature may overemphasize the regional character of knowledge flows. While actual co-location (and tighter governance) adds to the efficiency of product definition, knowledge recombination, and diffusion (Maskell and Malmberg 1999; Pinch et al. 2003), organizational proximity of innovation activities compensates for geographical distance (Capaldo and Petruzzelli 2014). As such, arguably, to enhance regional innovativeness, a social planner¹ should not only look at the composition of the regional knowledge, but also at the (possibly interregional) topology of innovation networks and notably whether there is collaboration with agents active in different technological fields (cf. Grant and Baden-Fuller 1995, 2004; Pyka and Küppers 2002; Pyka 2002). In this paper, the contention is that the regional specialization versus diversifi-

¹ We deliberately use the term ‘social planner’ to underline the power required to impose the knowledge mix and network topology onto regions in the various scenarios in our agent-based model study. We are aware that, in real-world cases, policy makers rarely have such power.

cation versus related variety debate (and the related purely regional innovation policy paradigms) cannot be settled without taking the (inter)regional innovation network into the consideration. Regional innovation systems feature a mix of public and private, intra- and inter-industry, and local and global knowledge flows (Autant-Bernard et al. 2013; Camagni and Capello 2013). Following these insights, social planners were confronted with new regional innovation policy paradigms. In this paper, we focus on three prominent ones. Firstly, the ‘smart innovation’ policy paradigm (Foray 2014; Foray et al. 2009, 2011) argues that a social planner, such as the European Commission, best prioritizes technological development in regions in particular technological fields, building upon existing strengths, and assigns other regions ‘co-innovator’ roles with specializations in other fields. In following the smart specialization paradigm, the social planner thus deliberately creates a patchwork of complementary, specialized regions connected through an interregional network. This network is crucial as, under fully fledged regional specialization, realizing path-breaking inventions requires access to alien technological knowledge that may be located in other regions (Bathelt et al. 2004; Rosenkopf and Almeida 2003). Secondly, the ‘branching’ policy paradigm (Asheim et al. 2011; Boschma 2011) [and related ‘regional resilience’ paradigm (cf. Balland et al. 2015; Bristow 2010)] argues that regions may pick particular mixes of technological fields that can continue to branch in a self-sustained manner, independent of knowledge from outside the region. Thirdly, in the ‘regional gatekeeper’ literature, a more specific network form is prescribed. Hereby innovation networks are regionally dense, and there are regionally central gatekeepers which monitor and filter technological knowledge in other regions to then diffuse that knowledge within their own region (Graf 2011; Graf and Krüger 2011; Spencer 2003; Breschi and Lenzi 2015; Bathelt et al. 2004).²

The research question now is: *given* a particular regional distribution and interrelationship of technological knowledge, notably of specialized fields, which network topology should a social planner pick to enhance technological progress in these regions, on average? In fact, we are able to prescribe the combination of knowledge distribution and network topology which a social planner should pick. To transcend the ongoing qualitative arguments and conflicting empirical findings, we define and use an agent-based model to experimentally investigate the relationship of specialization/related variety/diversification and intra-/inter-regional innovation network topologies (and notably the smart and gatekeeper network) in conjunction. Given that such an agent-based model provides full control over the topology of the network, we study combinations of spatial knowledge distribution and (inter)regional innova-

² Despite the use of network topology terminology, analysis of economic innovation networks is different from social network analysis. The latter emanates from the concept of structural holes (Burt 1992) and social capital (Bourdieu 1980; Coleman 1990) and focuses on the role of *homogeneous* agents (e.g. in *diffusion* of knowledge) purely based on the *position* within a network. When it comes to a knowledge-based perspective on *creation* of new knowledge in economic innovation networks, agents are *heterogeneous* in that they have particular technological specializations and related knowledge endowments. Apart from purely the adjacency of agents or the local structure of the network, the focus in the economic network literature is on *complementarities* of technological knowledge of agents collaborating and the innovation potential of combinations of their knowledge. Readers interested in a social network perspective on knowledge exchange and diffusion (rather than the recombination for the endogenous creation of new knowledge) are referred to Cowan and Jonard (e.g. 2007).

tion networks which are not discussed in literature. Moreover, the agent-based model also allows us to study the number of relationships of each of the agents on technological progress. Ultimately, the agent-based model enables us to get insights in mechanisms at work on the complex interplay of spatial distribution, technological interrelationships, and network topology, and insights in the performance of policy paradigms followed by social planners.

In addition to addressing the topical debate on which regional network innovation policy is to be preferred, we extend the debate. Our claim is that the significance of collaboration with firms in other fields and in other regions depends on *structural features of the underlying technological knowledge* and the distribution of knowledge over regions. In preceding simulation studies, we revealed how collaboration of agents in different specialized regions is predominantly significant if technological knowledge accumulation is ‘progressive’ (i.e. not building upon own ancestors, which would be ‘conservative’), which is the case for some technological fields more than for others (see Vermeulen and Pyka 2014b, a). In these studies, however, the technological structure is highly simplified such that conclusions have limited external validity. In the study at hand, the structural features of the technological knowledge searched by the agents in the model is empirically calibrated to citation and classification statistics of patents in the OECD patent database. In the spatial agent-based model presented here, the world consists of sea and land regions where a (controlled and fixed) number of firm agents resides in each land region. Each agent is specialized in one technological field, is engaged in knowledge discovery, and has a number of relationships with agents with whom it can (but need not) collaborate in knowledge discovery. The specializations of the agents in each region as well as the topology of the innovation network are now independent variables and the absolute ‘amount’ and ‘advancedness’ of knowledge discovered are studied. We see this paper and its model as the first step in a research line in which various regional innovation network policies are studied for realistic knowledge structures while controlling innovation system properties.

The structure of this paper is as follows. In the following section, we provide an overview of the literature on (inter)regional innovation policy paradigms and notably their preoccupation with knowledge specialization/diversity and with the role of (inter)regional innovation network relationships. As our principal claim in previous work already was that the structure of the knowledge determines which (inter)regional innovation network structure is most efficient in technology development, we subsequently discuss existing ‘technology discovery models’.³ We then specify our own ‘technological knowledge graph’ model, how it is generated and calibrated to the OECD patent database, followed by an operational definition of the spatial agent-based simulation model and the search heuristics followed by agents in searching and unlocking new parts of that knowledge graph. This is followed by a section with an overview and discussion of simulation results. The last section provides conclusions.

³ These models generally talk about ‘technology’ rather than ‘knowledge’. Instead of changing the vocabulary in our discussion of this literature to concur with our knowledge-based perspective, we ask the reader to ascertain that these models have in fact a level of abstraction at which they may just as well deal with *technological knowledge*.

2 Economics of Innovation: The Spatial and Network Dimensions

In the knowledge-based perspective of the firm, it is dynamically more efficient to conduct research in collaborative rather than in integrated or market governance forms (Grant and Baden-Fuller 1995, 2004). This is corroborated by the fact that strategic alliances (Hagedoorn 2002) and innovation networks (cf. Pyka and Küppers 2002; Pyka 2002) have emerged as persistent forms of organizing research and development of new technologies. Given that technological knowledge with a substantial tacit component (cf. Polanyi 1966; Johnson et al. 2002) is conveyed and combined more efficiently when done face-to-face (Gertler 2003), one expects and also observes geographical clustering of vertically specialized but technologically related firms active in the same industry (Feldman 1994). Moreover, firms active in the same industry tend to co-locate to find specialized component suppliers, to share a pool of skilled labor, capture (unintended) knowledge flows, and reap localized scale economies: the so-called Marshall–Arrow–Romer externalities. Not surprisingly, scholars of economics of innovation focus on the regional nature of innovation collaboration (e.g. Cooke 2001; Asheim and Coenen 2005; Bathelt et al. 2004; Neffke et al. 2011; Karlsson and Stough 2005; MacKinnon et al. 2002; Vermeulen and Paier 2016). Consequently, popular regional innovation policy paradigms (including also new paradigms such as smart specialization and branching, discussed later) have in common that innovation-based economic growth is stimulated by formation of technologically specialized clusters (e.g. Cumbers and MacKinnon 2004; Boschma 2014). It is acknowledged, though, that technologically specialized regions (and, notably, ‘industrial districts’) tend to focus on merely incremental extension of the locally supported product designs. Consequently, a region (and the local networks contained in it) may thus fall behind others if they fail to absorb, imitate, or leapfrog the technology developed elsewhere (Menzel and Fornahl 2010; Saxenian 1996; Hassink 2005; Martin and Sunley 2011). Other authors argue that regions should, rather than exploit the Marshall–Arrow–Romer externalities, stimulate technological diversity to stimulate innovation. The argument is that (path-breaking) innovation comes about by cross-fertilization of different technological fields. For an in-depth discussion of the conflict between these two bodies of literature see Glaeser et al. (1992), Panne (2004), Paci and Usai (2000), Feldman and Audretsch (1999), and the possibly methodological causes of this conflict, see Beaudry and Schiffauerova (2009).

In recent literature, this perspective is refined in two directions. Firstly, the related variety literature (Frenken et al. 2007) argues that diversity is not beneficial per se. It is only the case if two or more knowledge bases can be fruitfully combined. Problematic is that it is not known a priori whether or not particular knowledge bases can be fruitfully combined into path-breaking innovations (cf. Castaldi et al. 2015). Policy interventions may of course ensure sustained presence of fruitful multi-industry knowledge diversity (cf. Balland et al. 2015; Bristow 2010). Note that, purely from the perspective of regional growth, a diversified region is more resilient to shocks to or falling behind of one of the industrial activities (see Christopherson et al. 2010, for an extensive discussion of resilience). The ‘regional resilience’ paradigm calls for a balanced, diverse portfolio of sectors to prevent a region from getting into a technological lock-in and regional economic stagnation. Regions with particular combinations of sectors

may allow for sustained creation of opportunities for cross-fertilization, by ‘branching’ into new but technologically related directions (Asheim et al. 2011; Boschma 2011). Regional innovation policy aiming for regional resilience may stimulate a ‘smart mix’ of related fields providing options for sustained branching.

Given that it may not be known *a priori* what other technological knowledge base gives rise to fruitful cross-fertilization, the, in retrospect, unrelated variety in a diverse region may hence be a source of dynamic *inefficiency*, while the positive agglomeration externalities of regional specialization are not reaped. That said, being co-located in a region is not sufficient nor necessary for innovation to take place (Boschma 2005): a channel of communication has to be present and all parties have to be willing to exchange and cross-fertilize knowledge, which can then take place regardless of the location. Moreover, so we add, diversity in the region need not impair innovation efficiency as long as the agents active in unrelated fields are not connected in the innovation network. While actual co-location (and tighter governance) would add to the efficiency of product definition, knowledge recombination, and diffusion (Maskell and Malmberg 1999; Pinch et al. 2003), organizational proximity of innovation activities compensates for geographical distance (Capaldo and Petruzzelli 2014).

Innovation policy literature stresses that there is no one-size-fits-all regional innovation policy, but needs to be tailored to circumstances specific for the region (cf. Tödtling and Trippl 2005). Each regional innovation system calls for its particular mix of public and private, intra- and inter-industry, and local and global knowledge flows (Autant-Bernard et al. 2013; Camagni and Capello 2013). Moreover, given regional disparities, it makes sense for a social planner such as the European Commission to differentiate the extent and direction in which innovation is stimulated in the various regions. The ‘smart innovation’ concept (Foray 2014; Foray et al. 2009, 2011) elaborates on this idea in that a social planner, such as the European Commission, best prioritizes technological development in particular sectors, thus deliberately creating a patchwork of complementary, specialized regions. Based on existing competitive advantages and knowledge strengths, regions are stimulated to specialize in particular fields, whereby regions specialized in other fields are designated to support innovative activities in the focal regions. A necessary complement to this patchwork of specialized regions is a network of research collaborations between organizations in different regions following technological relationships. Extending Camagni and Capello (2013), we define smart innovation network policies as those policies that seek to “enhance local knowledge production by exploiting existing local knowledge specificities through intra- and interregional research and development relationships”.

Literature with an innovation network perspective on regional development argues that regional lock-in may be overcome by ensuring that local agents have relationships with agents in other regions to acquire and absorb (only) relevant knowledge otherwise unavailable to them. In literature, there is a particular interest for so-called gatekeepers (Graf 2011; Graf and Krüger 2011; Spencer 2003; Breschi and Lenzi 2015) or pipelines (Bathelt et al. 2004). These gatekeepers function as filter for unrelated knowledge available outside the region and translate and diffuse related knowledge in the regional network. Generally, we take the purpose of innovation networks to, on the one hand, connect pools of technological knowledge, regardless of the geographical distances, and, on the other hand, increase efficiency by limiting the relationships to

Table 1 Claims in literature on the efficiency and mechanisms at work in regions with a particular knowledge mix (specialized, related variety, diversified) versus the (inter)regional innovation network topology (regional, gatekeeper, or smart interregional network)

	Regional network	Gatekeeper network	Smart network
Specialized	‘Cluster/industrial district’: efficient technology development (Marshall), yet risk of technological lock-in		‘Smart specialization’: efficient development within region and filtered, efficient collaboration with co-inventor regions
Diversified	Innovative potential, limited risk of lock-in. Many combinations not fruitful, so inefficiency		
Related variety	‘Branching’: potential for ‘indefinite’ branching in limited number of fields		
<i>Unspecified</i>		Efficient intraregional recombination. Filtered knowledge with ‘branching’ and ‘path breaking’ potential	

those technological relevant. In this perspective, the topology of innovation networks may overcome ‘pollution’ with technologically unrelated knowledge in diversified regions as well as overcoming the lock-in in specialized regions devoid of technological related knowledge. Stimulating innovation and regional economic development combines interventions on either side of the continuum: (i) enhancing innovation networks by facilitating access of (supposedly) relevant technological knowledge given its location, and (ii) altering the mix of industrial knowledge present in the region, e.g. by ensuring presence of agents active in other sectors, given the network topology.

Table 1 contains the claims discussed above and actually reveals how these claims are unspecific or rather specific with regard to the network (implicitly) presumed. Using the agent-based model presented below, we study all combinations.

3 Knowledge-Based Model of Technology Discovery

3.1 Existing Models of Technology Discovery

In order to study our claims on technology discovery using a computational agent-based model, we need an operational model of both technological knowledge and how this is obtained. Literature does not provide a standard one, but rather many different ones, which are often tailored to particular research questions. Conceptually, though, there is consensus that technology discovery is an evolutionary process of recombination of, experimentation with, and selection of technologies, followed by diffusion, (imperfect) imitation, and adaptation of technologies, ultimately culminating in localized, evolving repositories of artifacts, tools, and associated technological

knowledge (cf. Arthur 2009; Childe 1936; Basalla 1988). Given our explicit focus on the knowledge-based perspective in the geography of innovation literature, we follow Arthur (2009, p. 21): “novel technologies arise by combination of existing technologies. [...] The overall collection of technologies bootstraps itself upward from the few to the many and from the simple to the complex.”. In Arthur’s perspective, technologies harness phenomena (e.g. optical, electrical) and put particular effects to use. Metaphorically, Arthur perceives technologies as hidden underground in ‘seams’ of related effects and mankind gradually ‘mines’ deeper from the top down. At the top, effects are primitive and may be serendipitously discovered. The discovery of ‘deep’ effects (e.g. chemical, quantum) is preceded by discovery of simpler effects, and requires advanced knowledge and modern technologies. Technologies based on quantum effects could not have been discovered without using tools harnessing electrical effects, for instance. “Phenomena cumulate by bootstrapping their way forward: effects are captured and devices using these effects are built and uncover further effects.”. While some clusters of phenomenologically related technologies may require research or tool technologies to be discovered, other technologies simply combine ‘modules’ as developed in different fields, e.g. a microscope combines optical lens technologies with mechanical positioning technologies.

Conclusively, a model of technology search ideally features that (i) technology gradually accumulates, (albeit possibly in a bursty nature, upon discovery of particular ‘seams’), (ii) there is an inherent order in which technologies are (to be) discovered, and (iii) new technology comes about by combining existing technologies or its discovery requires using existing technologies. For accumulation and expansion, the search space should expand non-convexly or be so large that search does not reach the boundaries during the modeled time horizon. Combining or extending existing technologies should (be able to) yield technologies outside the readily acquired set. Secondly, to model the inherent order in which technologies are to be discovered, one can think of, e.g. distance in search space that has to be traversed. Thirdly, there is a precedence relationship between technologies. Note that ancestors can be certain tools or research instruments required to ‘uncover’ certain phenomena or ‘seams’ of effects (to stick to Arthur’s metaphors) as well as that particular technology indeed simply is a modular combination of existing technologies. If there are multiple ancestors required to discover a more advanced technology, this meets the requirement of ‘non-convex accumulation’, ‘inherent order’, as well as ‘combinatorial search’.

Few combinatorial technology search models found in literature model meet all three requirements. We briefly go over the most prominent ones and then present our own model. In the models discussed here, technology search is modeled by having agents combine (discovered) elements, thereby seeking to meet or maximize a particular objective (e.g. Gilbert et al. 2001; Arthur and Polak 2006; Chie and Chen 2013; Chen and Chie 2006; Korhonen and Kasmire 2013; Padgett et al. 2003; Vermeulen et al. 2016a). Often the set of elements is fixed and given, and technology search merely is finding the ‘fittest’ or most ‘attractive’ combination. In the NK -landscape approach, technology search is modeled as the process of combining elements (from a fixed set) which jointly and interactively determine the combination’s fitness (cf. Frenken 2006; Woodard and Clemons 2012; Auerswald et al. 2000). In Padgett et al. (2003), a technology is successfully constructed when a combination of ‘processes’

forms a ‘hyper-cycle’. Technology search concerns randomly combining processes from a given set of processes. In Korhonen and Kasmire (2013), technical elements (numbers) are interconnected by transformations (arithmetic operators) into a system which is considered technically feasible if it has a preset outcome. Technology search concerns finding a combination of elements and operators that has that particular outcome. In Gilbert et al. (2001) and Vermeulen et al. (2016a), units of technological knowledge are combined into a product that generates a payoff. Technology search concerns finding a set of units with the highest payoff, both by altering the firm’s portfolio of units and trying different combinations of units from that portfolio (or from the portfolios of firms collaborating in innovation). Korber and Paier (2014) extend the Gilbert–Pyka–Ahrweiler model and calibrate knowledge discovery opportunities using empirical co-occurrence rates. In the current setup, these models feature accumulation of technology, but in a narrow sense: there is no bootstrapping into new sets of technologies, so ultimately, the landscape is confined. In Morone and Taylor (2010), technologies are ordered in a predefined, fixed network structure. Agents explore this network to discover new technologies, however these technologies only feasible when all of its ancestors are discovered and feasible. So, also here there is accumulation and a natural order, plus new technologies that can be generated on the fly. Below we will present a model in the vein of Morone and Taylor (2010).

3.2 Our Directed Graph Search Approach

In line with notably Morone and Taylor (2010) and following our earlier experimental work contained in Vermeulen and Pyka (2014b) and Vermeulen and Pyka (2014a), technological knowledge is perceived as ‘units’ that are ordered in *directed graphs*, from primitive (upstream) to advanced (downstream). Moreover, like Morone and Taylor (2010), one agent or a collective of agents tries to unlock units in this directed graph by combining units that it already possesses. A single agent or a collective of agents discovers a knowledge unit downstream in the graph if it has discovered and puts together the exact set of primitive ancestors found upstream in the graph. The idea is that primitive knowledge has to be acquired before more advanced knowledge can be understood. An example of this is that man needs to know ‘how to control fire’ and ‘how to contain liquid’ before it can learn how to ‘boil liquid’. The discovery of technological knowledge thus expands the technology search space non-convexly. Note that our model thus meets the three requirements mentioned above: (i) knowledge *accumulates* (yet it is stored in the agents’ knowledge repositories), (ii) there is an *inherent order* in which knowledge is (to be) discovered (namely from primitive to advanced), and (iii) new knowledge is discovered *by combining* readily discovered knowledge.

3.3 Empirical Calibration with Patents?

In Vermeulen and Pyka (2014b) and Vermeulen and Pyka (2014a), we used a parameterized algorithm to generate a *stylized* knowledge graph. In order to study the controversies in the innovation policy paradigms and the empirical findings on which they are based, it is paramount to use a knowledge graph with a real-world structure

and real-world complexity. Inspired by Korber and Paier (2014), we seek to empirically calibrate the search process. In our case, we calibrate the directed graph of knowledge units using statistics on the classifications and citations of patents in the European Patent Office's patent database PATSTAT. By doing so, the simulation results are arguably externally more valid than when using a highly stylized and simplified structure.

A patent database such as the one of the European Patent Office forms an extensive, valuable source of combinatorial precedence relationships of knowledge. After all, each patent provides backward citations to older patents and thereby related technological knowledge. Moreover, as patents receive a (highly detailed) classification code, actual interrelationships of different 'fields' of technological knowledge can be derived from the patent database as well. We can thus determine how often knowledge from one field is used in the discovery of knowledge of the same or another field. In seeking to stimulate innovativeness, this 'co-occurrence' of fields has ramifications for the desirable topology of innovation networks, hereby mediated by the geographical distribution of the (agents with) technological knowledge in these fields.

For now, we limit ourselves to the top-level of the International Patent Classification system, i.e. to the level of 'sections' purely because of scale issues in our computer implementation. We do acknowledge that also the classification of patents into a 'class', 'subclass', 'group' and 'subgroup' is valuable, and even that the knowledge structure may change by adding these lower-level classifications. The 'section' level of the International Patent Classification consists of the eight 'sections': (A) Human Necessities, (B) Performing operations; Transporting, (C) Chemistry; Metallurgy, (D) Textiles; Paper, (E) Fixed constructions, (F) Mechanical engineering; Lighting; Heating; Weapons; Blasting; (G) Physics, and (H) Electricity. In our directed knowledge graph, we will classify each knowledge unit as to be part of one of eight 'technological knowledge fields', associated with these eight sections.

Apart from having to limit the depth of classification in our calibration, there are several other caveats in using patents to study technological knowledge. If anything, the patent database is not all-encompassing: not all inventions are patentable and not all are patented. Moreover, the inventions that are patented differ in 'quality' (for these last points, see Griliches 1957). This may in part relate to strategic patenting of firms as they may create 'thickets' of patents around the same technology to (i) discourage competitors from moving into a particular technology field and to secure a de-facto monopoly (cf. Hall et al. 2012, 2015) or (ii) defend against encroachment by competitors pending particular regulation (see e.g. Vermeulen et al. 2016b). Despite the occurrence of e.g. low quality patents or patent thickets, the technological relationships of the various fields and hence the importance of bringing knowledge from these fields together remain valid.

3.4 Algorithm for Graph Generation

We use a straightforward method to 'statistically calibrate' the cumulative, combinatorial structure of the technological knowledge graph to the European Patent Office's database PATSTAT. Each knowledge unit in the knowledge graph is assigned to one of

Table 2 Notation

Variable	Dimension/range	Meaning
m	$1/\mathbb{N}^+$	Number of fields (equals the number of sections in the patent classification). Note that for the International Patent Classification (IPC) $m = 8$
\mathcal{F}	$m/[0, 1]^m$	Relative occurrence rates of the first field in a patent classification
\mathcal{B}	$m/[0, 1]^m$	Empirical probability distribution of the number of backward citations per field
\mathcal{C}	$m \times m/[0, 1]^{m \times m}$	Co-occurrence rates of the first fields of a citing and cited patent
\mathcal{L}	$u \in \mathbb{N}^+/[0, 1]^u$	Distribution of the lag in number of periods between citing and cited patent. Dimension u depends on timespan of patent database
N_t	$1/\mathbb{N}$	Number of knowledge units on tier $t \geq 0$
ρ	$1/\mathbb{R}$	Period-on-period growth rate in the total number of knowledge units
r	$1/\mathbb{N}^+$	Number of regions in the spatial agent-based model, here we set $r = m$
n	$1/\mathbb{N}^+$	Number of agents per region, here we set $n = m$
k	$1/\mathbb{N}^+$	Mediating variable. Number of connections established by an agent
d	$1/\mathbb{N}^+$	Mediating variable. Relational depth of access to knowledge units for recombination. If $d = 1$ ('direct partnering'), agents just collaborate with their own partners. If $d = 2$ ('triadic closure'), agents also collaborate with partners of partners. Etc

$m = 8$ fields, associated with the eight patent sections in the International Patent Classification code. The knowledge graph is arranged in 'tiers', the equivalent of years in the patent database. We initialize Tier 0 with N_0 unique initial knowledge units evenly distributed over the $m = 8$ fields. We then use the algorithm described below to expand the knowledge graph recursively, hereby using the citation and classification statistics of patents in the PATSTAT database. See Table 2 for the specification of the notation used in our description.

After studying the distribution of the number of backward and forward citations, we decided to use only the patents granted during the years 1988–1994 for calibration. Starting with an earlier year would underestimate the number of *cited* patents ('input knowledge') while ending with a later year would underestimate the number of *citing* patents ('output knowledge').

The first step is to generate the number N_t of knowledge units for tier t . For this we use a simple period-on-period growth model $N_t = (1 + \rho)N_{t-1}$, where $\rho = 0.15$ is the period-on-period growth rate as computed from our extensive patent dataset.

The second step is to determine the field for each of these N_t knowledge units. For each unit, determine its field x by drawing it from the discrete, empirical distribution \mathcal{F} of the occurrence rates of fields. The distribution \mathcal{F} is obtained by counting the number

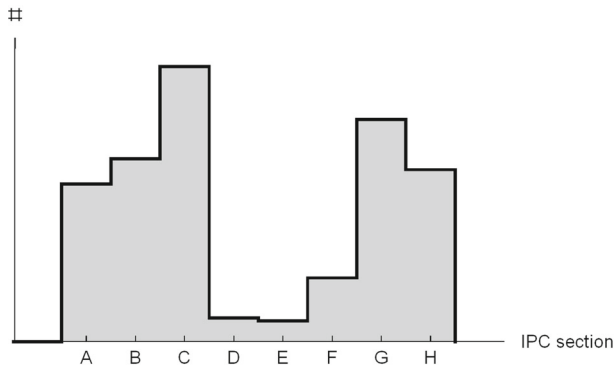
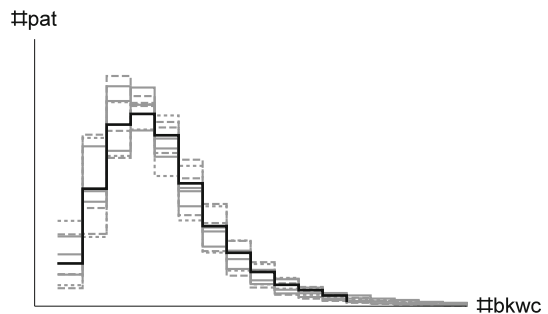


Fig. 1 Number of patents by first classification in the eight patent sections

Fig. 2 Histogram of normalized number of backward citations by the first patent section (thin lines) and the empirical approximation used (thick)



of times a patent has a (first) classification in a particular field and then normalizing by the total number of patents in our set. The empirical distribution \mathcal{F} is plotted in Fig. 1.

The third step is to determine the number of backward citations B for each of these N_i units. For each unit, draw the number B from the discrete, empirical distribution \mathcal{B} . The empirical distribution \mathcal{B} of the number of backward citations by patents' first field is plotted in Fig. 2. While the number of backward citations may differ from field to field, we find that this difference is actually limited. Given this similarity in the distributions for the various fields, the citation data \mathcal{B} is aggregated and approximated with one and the same empirical distribution (plotted in thick lines).

The fourth step is to determine the field for each of these B citations. For this we first construct the 'co-occurrence' matrix \mathcal{C} of the first field of a citing and the first field of the cited patent.⁴ For each of these B citations we have to determine the field of the cited unit. Say, from a unit in field x , draw the field y from the empirical distribution in the x -th row $\mathcal{C}_{x\cdot}$ of the matrix \mathcal{C} . For clarity of exposition, the matrix \mathcal{C} as extracted from the patent database is depicted in the chord diagram in Fig. 3. Hereby, the width of the circle arc represents the fraction of patents with a particular first field (see also the

⁴ Own analysis revealed that this matrix is highly similar if one aggregates over *all* fields in cited and citing patents.

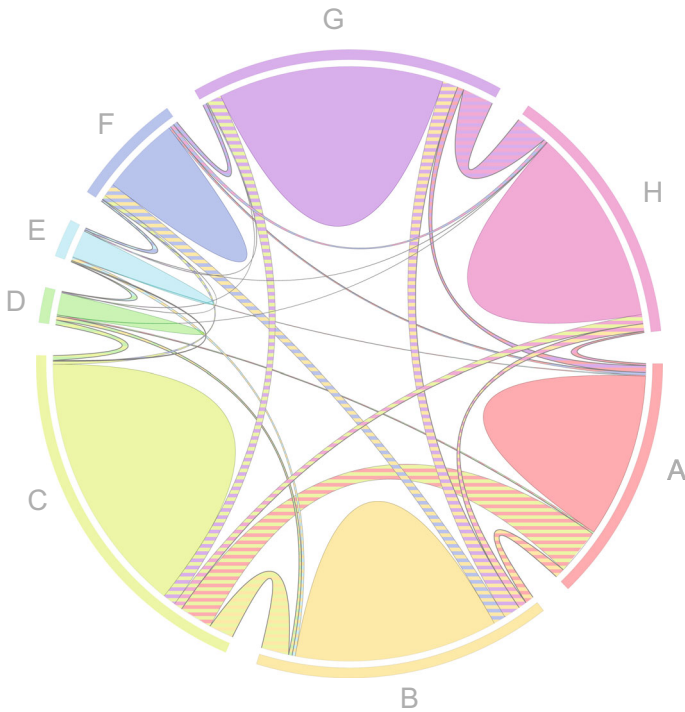


Fig. 3 Chord diagram of the first patent sections of citing and cited patents

distribution \mathcal{F}). Each of the chords connect two circle arcs, indicating a relationship of two fields. The width of a chord at a circle arc x connected to a circle arc y specifies the relative number of patents in field x citing a patent in field y (i.e. C_{xy}).

The fifth step is to determine the period-lag for each of these B citations. This lag is drawn from the discrete, empirical distribution \mathcal{L} of the lag (in number of periods) from cited and citing patent. This period-lag is obtained by looking at the number of years between the year of publication of the citing and cited patents. This distribution is plotted in the dark-shaded histogram in Fig. 4. Ignoring a short period between patent application and grant during which there are no (or very few) backward citations, the number of backward citations can be approximated within a few percentages with $\hat{\mathcal{L}}(\Delta) = g\Delta^p$ with Δ the lag in number of periods, $g = 13$ and $p = 3.1$. While generating Tier $i = 1, \dots$ of the knowledge graph, the lag Δ is drawn based on the truncated discrete distribution, i.e. from $\hat{\mathcal{L}}(i - 1), \hat{\mathcal{L}}(i - 2), \dots, \hat{\mathcal{L}}(0)$.

As such, we have created a number of knowledge units, which have been assigned a field according to the empirical occurrence rate and a number of ancestors with a field and time-lag which are also drawn from empirical distributions. We thus generate a next tier in a directed graph of knowledge units that is calibrated to rate of (co-)occurrences of patent classifications as well as the lag and backward citations figures of a real-world patent database. A flowchart of this algorithm is given in Fig. 5.

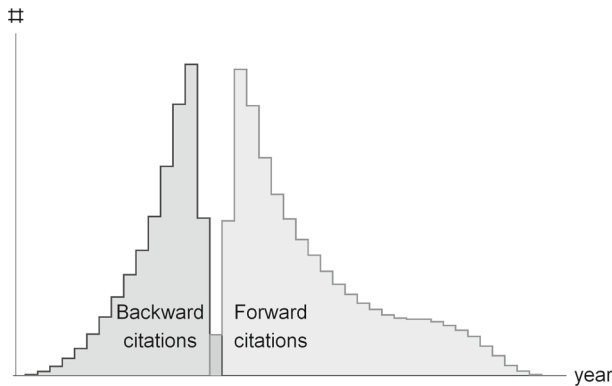


Fig. 4 Histogram of year-lag of citing to cited patent

At the start of a simulation run, fifteen tiers are generated recursively, such that, given the distribution of the citation time-lag \mathcal{L} , new tiers are unlikely to add additional descendants ('forward citations') to existing knowledge units. Moreover, as soon as search (described later) unlocks knowledge units closer than ten tiers from the most advanced tier, new tiers are generated, again with the reason that all descendants of primitive knowledge should be known. We do this because of the search heuristic followed by the agents, as will be described in the next section. Figure 6 contains an illustration of a generated knowledge graph, here with only four tiers.

4 Spatial Agent-Based Model

To determine what network topology or spatial distribution of knowledge a social planner should implement to stimulate the discovery of technological knowledge, we use an agent-based model in which we can study the topology and the spatial distribution in conjunction. In the agent-based model, there are $r = 8$ regions ('land' cells as opposed to 'sea' cells), $m = 8$ fields (each associated with a patent classification section in the patent database), $n = 8$ agents per region, and a bipartite matrix assigning knowledge sections to agents (as their technological specialization).

Each agent starts with all of the primitive knowledge units within its specialization (i.e. the 'nodes' at the bottom in Fig. 6). In each period of the simulation, each agent investigates a single combination of knowledge units in its own repository and the repositories of its partners. There is no economic principle (e.g. a demand market) that rewards knowledge discovery or paces search. Moreover, there is no entry or exit of agents. Finally, the circular layout of regions is fixed for all simulations. This simple setup ensures undistorted observations of the effects the role of the network topology and spatial distribution of knowledge the advancement of technological knowledge.

As explained above, in this paper, certain combinations of technological knowledge units 'unlock' a unique, more advanced knowledge unit. Agents are hence engaged in searching for successful combinations of knowledge units found in their own repos-

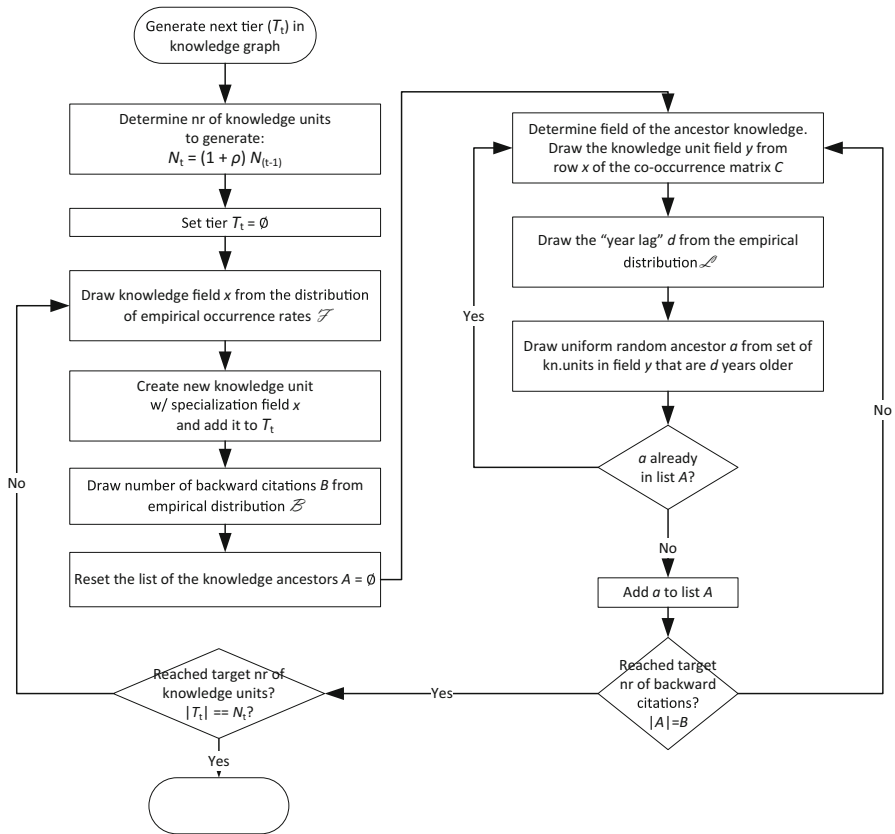


Fig. 5 Flowchart of the algorithm to generate the next tier in the knowledge graph using the empirical statistics of the PATSTAT patent database

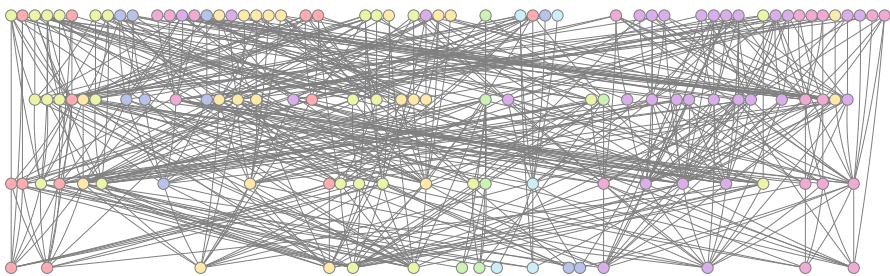


Fig. 6 Illustration of a few tiers of directed graph of technological knowledge units as generated by calibrated using the patent database

itory complemented with knowledge units found in the repositories of their network partners. As described in Sect. 3, many technology discovery models found in literature have agents that are engaged in *trial-and-error* recombination of technical elements, whereby the number of elements available is constant (and generally low)

and thus form a concave set. In the model presented here, combinations of primitive knowledge units produce more advanced knowledge units that may very well not yet be in the pool of available units. As the knowledge units that are discovered can in turn be combined to try to discover even more advanced knowledge, the number of units is endogenously and non-convexly growing. Consequently, the number of possible combinations explodes with the number of simulation steps. In case the number of feasible combinations (i.e. those ‘unlocking’ a new unit) is a non-trivial percentage, there is a rapid plateauing in the number of discoveries if such a trial-and-error search heuristic is followed. As such, it is disputable that research is, in the real world, naive trial-and-error.

In reality, inventors have visions of the technology that they seek to create, then isolate operating principles to be exploited and technological solutions to try (cf. Arthur 2009). Our operationalization of Arthur’s observation is that the agents, firstly, serendipitously ‘see’ a potential extension of a randomly drawn knowledge unit that they already possess (i.e. a descendant in the knowledge graph). In the current simulation results, the agent only sees extensions on the next tier of the knowledge graph. Secondly, agents know which technical *specializations* are required for this extension, but without knowing which knowledge *units* exactly. The searching agent approaches partners with one of the required specializations in uniform random order with the request to test whether any of their knowledge units would be adequate. Once all of the required knowledge units are present, the new knowledge unit is ‘unlocked’. This newly discovered unit is then also added to the repository of all agents that have contributed and are specialized in the field of the output knowledge. This search heuristic is depicted in the flowchart in Fig. 7. In new rounds, this knowledge can be used for further recombination.

Under these assumptions, the knowledge discovery search is then not a purely random ‘trial-and-error’ of recombining knowledge units, but rather inquiries with a deliberately chosen, technologically specific selection of partners. Note that this does not alter the pool of possible discoveries, but merely compresses the time scale compared to simulation runs with a random selection of knowledge units and/or collaboration partners.

5 Simulation Setup and Results

5.1 Simulation Scenarios

Simulations are run for 1000 period, with $r = 8$ regions placed in a circle (such that each region has exactly two immediate neighboring regions for a purpose explained below), and $n = 8$ agents per region, whereby each agent is specialized in only one of eight fields. With regard to the spatial distribution of technological fields over regions, we study three scenarios. Firstly, the ‘specialized’ scenario in which the $n = 8$ agents in a particular region are active in the same field and start out with the same primitive knowledge units. Each region has a different specialization (Field A, B, ..., H). Secondly, the ‘diversified’ scenario in which each agent in a particular region has a different field of specialization, i.e. the first agent is specialized in (and starts with

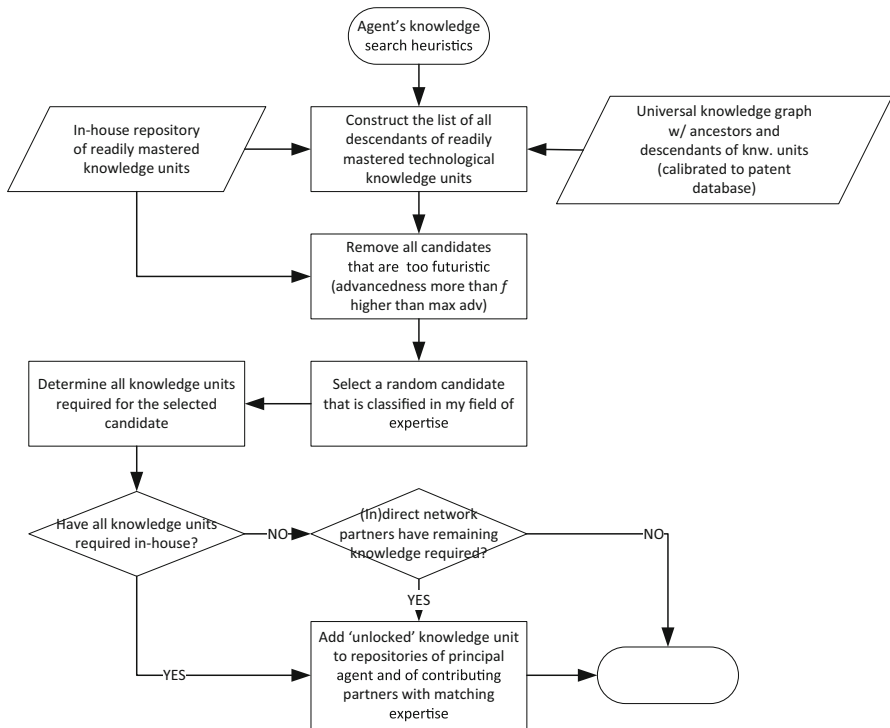


Fig. 7 Search heuristic

primitive knowledge units in) field A, the second in field B, and so forth, and the eighth agent is specialized in field H. The setup is the same in each of the $r = 8$ regions. Thirdly, the ‘related variety’ scenario in which each region has agents active in a mix of fields. Hereby, each region has its unique, main field of activity (like in the ‘specialized’ scenario), yet the region possibly contains agents engaged in probabilistically most related fields. This is initialized as follows. Each region gets an agent specialized in a unique field, say field x , while the other $n - 1$ agents are randomly assigned a field based on the empirical co-occurrence distribution \mathcal{C}_x , i.e. the x -th row of the co-occurrence matrix \mathcal{C} . To be able to have perfectly specialized regions while covering all fields (in certain simulation scenarios), we set r equal to m , and to be able to have perfectly diversified regions while covering all fields, we set n equal to m .

With regard to the innovation network imposed by the social planner, we study three network topologies. Firstly, the ‘regional’ network topology, in which each agent is recursively connected to k randomly drawn other agents *in the same region*. As such, the work consists of r disconnected subnetworks. The total number of ties is rnk and the average degree is $2k$.

Secondly, the ‘gatekeeper’ topology in which each region has exactly one ‘gatekeeper’. Only this gatekeeper is connected to agents outside the region, and then in particular exclusively to other gatekeepers. Within the region, each non-gatekeeper agent is connected to the gatekeeper. Moreover, each non-gatekeeper is connected to

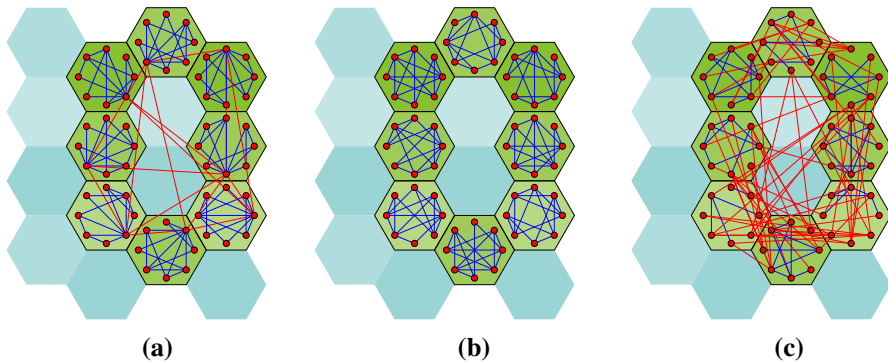


Fig. 8 Illustrations of each of the network structures for $k = 2$, here with specialized regions. **a** Gatekeeper, **b** regional, **c** smart

$k - 1$ other agents in the region, whereby these connections are recursively picked randomly from all possible connections not yet existing. To meet the correlation of geographical and organizational proximity, we have chosen to lay out the regions in a circle and thus make sure that each gatekeeper is in contact with the two gatekeepers in adjacent regions. Any other configuration with 8 regions would yield asymmetries in the number of contacts. As such, there are r connections between gatekeepers in total. To make sure that the total number of ties (and thereby the average degree and average ‘transaction costs’) are equal across the three network topologies, $r(k - 1)$ random, unique ties *between gatekeepers* are created.

Thirdly, the ‘smart’ topology connects each agent with k other agents based on field of specialization. Hereby, an agent with specialization x is connected to an agent with specialization y with probability C_{xy} (see the description of \mathcal{C} given before). Empirically, the higher C_{xy} , the more likely a patent in field x is citing a patent in field y . So, the higher C_{xy} , the more often collaboration yields an invention. In case there are more agents with specialization y , one is picked randomly from those with whom there is not yet a connection. The number of ties is rnk and the average degree is $2k$. Figure 8 contains illustrations of each of the three network topologies. Note that the ‘regionality’ of the ties in the ‘smart’ topology reflects the spatial distribution of knowledge. In Fig. 9, examples of the ‘smart’ network topology for both the case with specialized and the case with diversified regions are shown. We see that the smart network with diversified regions is in fact *more* interregional. This is caused by the fact that many inventions are actually combinations of pieces of knowledge *within* one and the same field, i.e. the on-diagonal elements on the co-occurrence matrix \mathcal{C} are relatively high. Alternatively, the chord diagram in Fig. 3 reveals that *cited* patents often have a classification in the same field as the *citing* patent. In case of diversified regions, agents active in the same field are in fact found in other regions.

For $k = 1, 3$, the total number of ties are 64, 192 respectively in all network topologies. Consequently, any differences in knowledge discovery (rates) are then due to structural differences in the topology (that is, with the spatial distribution of technological knowledge held equal), not because of differences in the number of relationships. In reality, firms incur transaction costs for maintaining (innovation network)

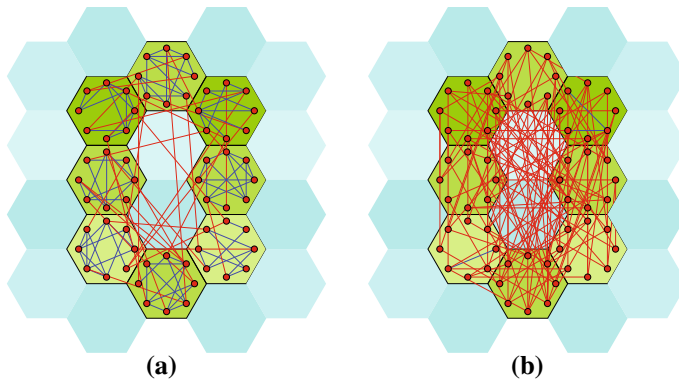


Fig. 9 Examples of smart topologies for specialized and diversified regions, with $k = 2$ ties initiated by each agent. With diversified regions, more ties are interregional so as to collaborate with others active in the same technological field. **a** Example of a smart network topology for specialized regions, **b** example of a smart network topology for diversified regions

relationships, so by ensuring that the topologies have equal numbers of relationships, we can ignore these costs in our analysis and conclusions.

We thus get a 3×3 contingency table of possible scenarios for the spatial distribution of technological fields versus the innovation network topology. For each of the 3×3 scenarios, we also varied the number of links $k = 1, 3$ each agent itself creates (explained above as to why we pick this ‘ego-centered’ operationalization) and the ‘depth’ $d = 1, 2$ of the collaboration, whereby $d = 2$ means there is triadic closure. For each possible scenario, we simulated 100 cases for 1000 periods. To provide more insight in the potential innovativeness, we provide the $m \times m$ matrices of the average minimum shortest path lengths between agents by technological specialization in “Appendix A”. Despite the appearance and naming of the network topologies, the traditional metrics of social network structure provide limited insight in the performance in terms of rate and advancedness of inventions.⁵ As such, we refrain from reporting these metrics.

5.2 Scenario Simulation Results

We study the technological progress using two measures. Firstly, we measure the ‘average maximum advancedness’ of knowledge. Over time, agents discover knowledge at increasingly higher tiers in the knowledge graph. By taking the average of the indices of the most advanced tier reached by each agent, we obtain the ‘average maximum advancedness’. We plot this advancedness over time. However, while this

⁵ After all, agents have technological specializations in different fields and, very often, a combination of knowledge units from different fields is required to unlock new knowledge. For instance, whether an agent specialized in Field A is connected to an agent in Field C or in Field E matters strongly in the probability to unlock new knowledge units. As such, the fact that or how agents are connected (e.g. in a cluster) says very little of ultimate innovativeness of the collective, for this one needs to know the distribution of the knowledge over these agents and their technological focus. Social network analysis does not (yet) provide metrics for ‘centrality’, ‘density’, etc. under node heterogeneity.

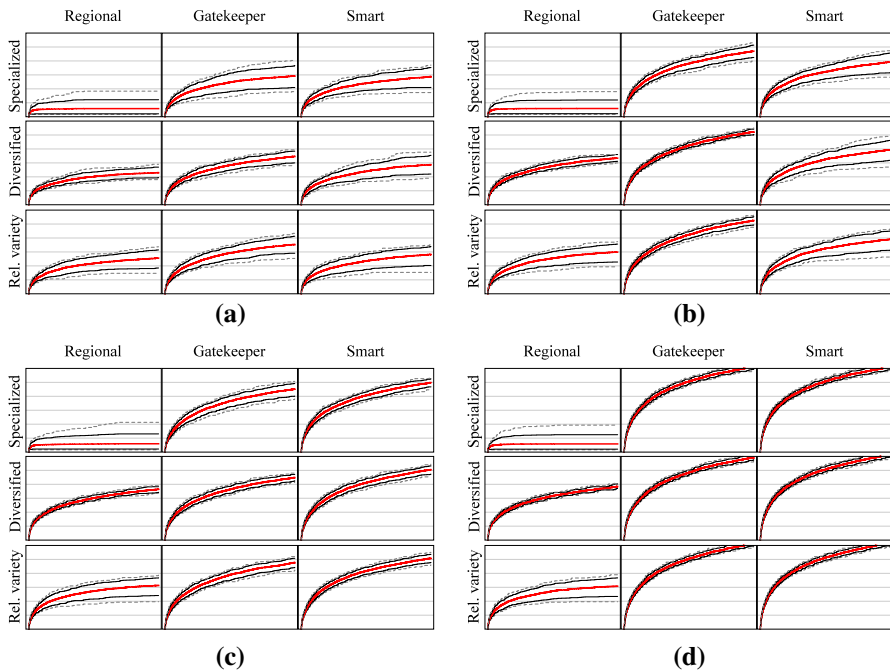


Fig. 10 Plots of the average maximum advancedness of the knowledge discovered for different parameter settings. Plotted are the average (thick), the top and bottom 20% (thin), and the top and bottom 5% over all simulation runs. **a** $k = 1, d = 1$, **b** $k = 1, d = 2$, **c** $k = 3, d = 1$, **d** $k = 3, d = 2$

measures how advanced technology becomes, i.e. the ‘depth’ of the technological knowledge, it does not measure the ‘breadth’ of discoveries, i.e. how many of the possible technologies at each tier have actually been discovered. So, secondly, we measure the total size of the ‘technological knowledge pool’ by counting the total number of *unique* knowledge units across all agents. We plot this cumulative number over time as well. We also study the percentage of knowledge units discovered by tier and by field using density plots. As explained in “Appendix B”, we use uniform occurrence rates ($\mathcal{F} = (1/m, \dots, 1/m)$) rather than the original rates plotted in Fig. 1 to overcome distortion in the simulation results caused by very low frequencies of knowledge units in fields D, E, and F.

In Fig. 10, we plot the average maximum advancedness for the 3×3 scenarios (specialized/related variety/diversified regions \times regional/gatekeeper/smart network topologies) for the four combinations of the number of ties $k = 1, 3$ (rows) and the partnering depth $d = 1, 2$ (columns) for collaboration with own partners only ($d = 1$) or also with partners of partners ($d = 2$, triadic closure). In general, we see that there is an initial phase with fast technological progress (caused by the low number of tiers yet generated and the fact that the lion share of knowledge units on these tiers are in the possession of agents for further recombination, see Sect. 3.4), the increase in average maximum advancedness settles at slowly decreasing rate (see Fig. 10), while the total number of unique inventions keeps on increasing steadily (see Fig. 11). Comparative analysis reveals that technological progress in regional networks is almost

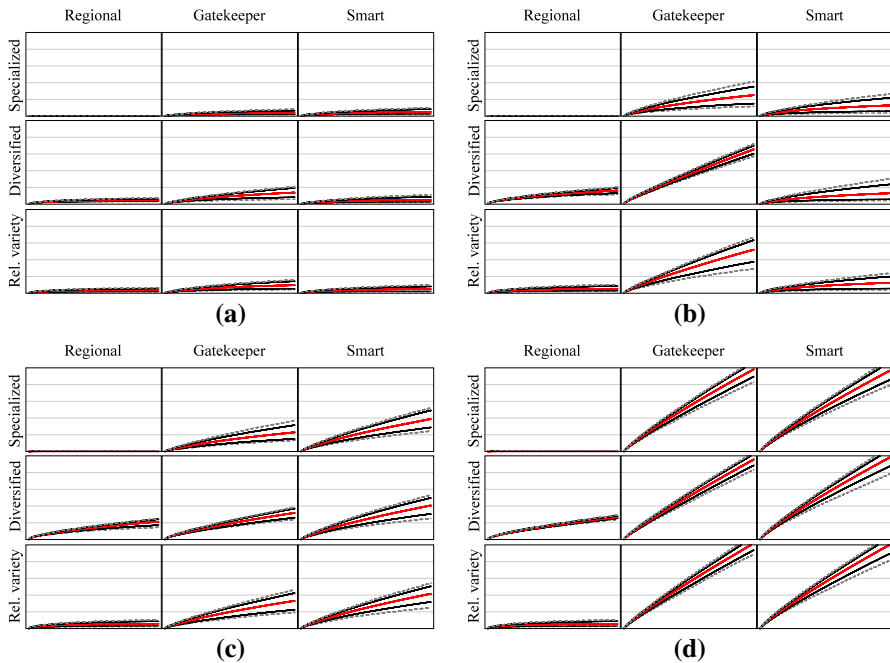


Fig. 11 Plots of the cumulative number of unique discoveries over time for different parameter settings. Plotted are the average (thick), the top and bottom 20% (thin), and the top and bottom 5% over all simulation runs. **a** $k = 1, d = 1$, **b** $k = 1, d = 2$, **c** $k = 3, d = 1$, **d** $k = 3, d = 2$

always outperformed by the progress in the other two topologies, regardless of spatial knowledge distribution, while smart networks are generally slightly outperformed by the gatekeeper networks particularly if the number of partners is limited (k and d low). In regional networks, the size of the pool of knowledge in various fields is simply more limited, either in terms of diversity or magnitude. Fundamental results on the average minimum path lengths between agents active in different fields contained in “Appendix A” show that regional networks (i) lack access to knowledge in different fields in case of a specialized region, or (ii) lack access to knowledge in the same field in case of a diversified region, so has a limited pool of knowledge. Given the high number of backward citations and the frequent cross-references between knowledge fields (as captured by the co-occurrence matrix \mathcal{C} , see the chord diagram in Fig. 3), access to knowledge in other fields is crucial. In fact, comparing the advancedness for the regional networks, we see that the advancedness—although not high in general—is higher for the case with diversified knowledge than for the case with related variety in the region. So, while related variety is to be preferred over specialization (in which the region gets into a lock-in soon), the access to *incidentally related* knowledge provided by diversification is crucial for long term ‘branching’. Across all network topologies, the technological advancedness is as high as or higher in the diversified region as it is in the other two. The fundamental results in “Appendix A” also reveal that the average minimum shortest path length is—on average—also slightly lower in case of

diversified regions than in case of regions with related variety, with a notable exception for the shortest path to agents in the same knowledge field.

We also see that both the advancedness and total number of discoveries are high in case of the gatekeeper topology for most connection settings (k and d), but saliently for low k ($k = 1$) with triadic closure ($d = 2$). Analysis of the knowledge repository of the gatekeepers reveals that they accumulate an extensive set of knowledge, provide access to the aforementioned incidentally related knowledge and thus are the motor for discoveries in their principal region, in connected regions, but indirectly also in other regions. For high connectivity cases ($k = 3$), the performance of the smart network is somewhat higher than that of the gatekeeper network because agents in the smart network now also have partners active in not only the most but also less related knowledge fields, while gatekeeper networks have a geographical bias as they connect in the first place to other gatekeepers the two adjacent regions.

In Fig. 11, we plot the total number of unique knowledge units for each of 3×3 scenarios and the four parameter combinations. In general, the number of discoveries is low if the connectivity is low ($k = 1$ and $d = 1$), plus, as we already saw, also the advancedness of the discoveries is low. The number of unique discoveries increases substantially with more connections ($k = 3$) particularly in conjunction with triadic closure ($d = 2$). Looking at the average minimum shortest path length statistics in “Appendix A”, the root cause is the fact that, apart from the exceptional regional case, agents in all fields are often directly connected to agents in any other field and with relatively short path lengths. While this holds in general, there are more subtle insights gained from detailed analysis. Firstly, in case of a regional topology without interregional ties, we see that the effect of more ties or triadic closure is relatively limited and, if there is any salient effect, this is only the case when the region is diversified. All in all, agents in other regions make discoveries in other branches of the knowledge graph and, for further progress in each of these branches, agents active in these branches should be able to access knowledge in the other branches. From this we conclude, that there is a strong effect of the ‘size of the knowledge pool’ and that access to the knowledge bases of other agents is significant in both the absolute number and advancedness of inventions. Secondly, if the region is diversified, the regional topology generates more discoveries than the smart topology. This is indicative of a phenomenon we call ‘high level lock-in’: in the smart topology, agents may lack access to knowledge in incidentally related fields, which effectively stifles discoveries in certain subgraphs of the knowledge graph. We discuss this ‘high level lock-in’ phenomenon in detail, later. In short, the cause of the ‘high level lock-in’ is that smart topologies lack connections to merely incidentally required knowledge, yet have (excessive) access to knowledge in the same and technologically strongly related fields. This is corroborated by the high average minimum shortest path lengths reported in the $k = 1$ and $k = 3$ tables in “Appendix A” for technologically weakly connected fields. Interestingly, the *reverse* is true if the region contains related variety or if the region is specialized: smart topologies generally outperform the merely regionally connected networks in case the region hosts specialized or a mix of related knowledge. Thirdly, there is an exceptionally high relative number of discoveries for the gatekeeper topology with triadic closure for diversified regions, which is essentially the culmination of three effects: (i) diversification provides access to ‘incidentally related’ knowledge, (ii) gatekeepers function

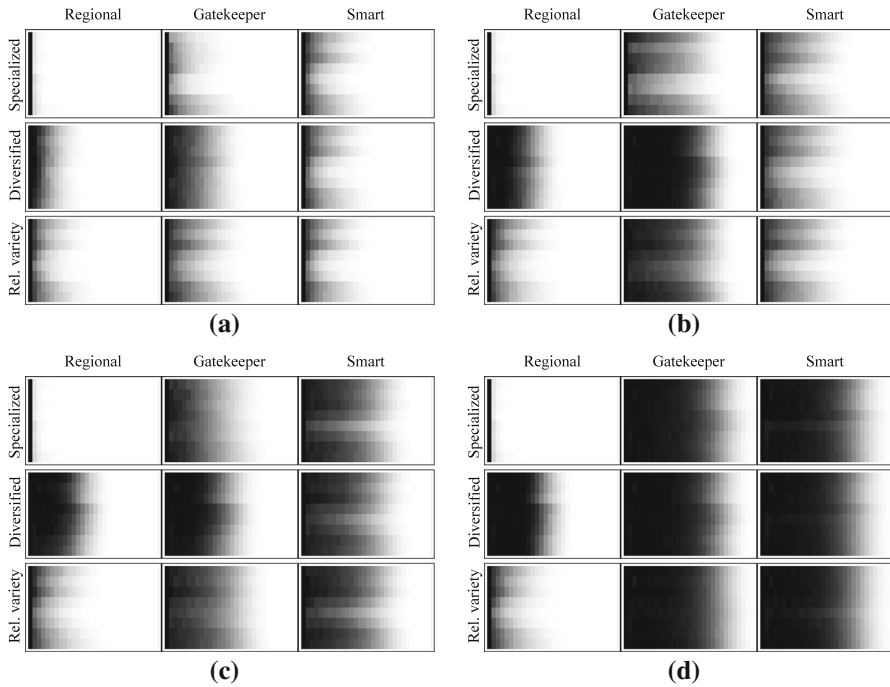


Fig. 12 Density plots of the percentage of discoveries possible per tier (X-axis) per each of the m fields (Y-axis) after T periods for different parameter settings. The darker, the higher the percentage. Completely black (white) means 100% (0%) of knowledge units available in the knowledge graph is actually discovered at $T = 1000$. **a** $k = 1, d = 1$, **b** $k = 1, d = 2$, **c** $k = 3, d = 1$, **d** $k = 3, d = 2$

as knowledge hubs, and (iii) triadic closure provides access to a bigger pool of knowledge. Triadic closure through gatekeepers provides agents with short paths to many other agents, even already for $k = 1$. Fourthly, the increase in the curves seem to taper off, whereby the increase in advancedness decreases rather quickly and in the number of inventions only very gradually. A primary cause of the decreasing rate of progress and discovery is the increasing ‘breadth’ of undiscovered potential ancestors from tier to tier.⁶ Note that all combinatorial technology search models reviewed in this paper feature this leveling off. However, unlike these models, the cause is not combinatorial trial-and-error search. In our model, each agent asks *all* of its partners with a particular specialization (in random order) for suitable knowledge units to unlock targeted knowledge *in a single period*, so *all* combinations of knowledge units present among partners are considered. As such, the tapering off of progress is due to the fact that more and more advanced branches in the knowledge graph remain unlocked and hence knowledge therein cannot be exploited to unlock even more and more advanced knowledge.

In Fig. 12, we plot the percentage of discovered knowledge units for each tier (with Tier 0 on the left and Tier 30 on the right) for each of the fields (with Field A at the top

⁶ Initially also the ‘depth’ increases, but this is related to the simulation onset in the number of tiers from which backward cited units can be drawn. The depth is at some point in time completely governed by the lag distribution \mathcal{L} and no longer truncated.

and Field H at the bottom). The differences in discovery rates from scenario to scenario are striking. In line with the results presented above, the rate of discovery increases strongly with connectivity (i.e. increasing k and d) and whether or not agents connect their partners (i.e. for $d = 2$ there is triadic closure, while for $d = 1$ there is not). When agents have more (potential) connections, they have access to more and more varied sources of (related) knowledge. Insight is obtained through more detailed analysis. We see that, for the gatekeeper and smart topologies, the frontier is about equally advanced, yet the percentage of knowledge discovered across all tiers is superior in diversified regions, generally. For regional topologies, both the advancedness and percentage of discoveries is particularly high if the region is diversified. Conclusively, although the rate of discovery within regions may be enhanced by diversification, there may still be high level lock-in due to limited access to knowledge in branches developed in other regions. Moreover, a salient feature of knowledge discovery not showing up in the plots of the advancedness and the number of inventions is the ‘profile’ in the percentage of discoveries across different knowledge fields. We see that, for many simulation scenarios, knowledge discovery in Field E and, to lesser extent, Field D and Field F is falling behind (has a lighter shade) the rest. Surprisingly, though, for several scenarios with a diversified knowledge mix and many local relationships (regional and gatekeeper network topologies), the percentage of discoveries is rather higher (has a darker shade). The root cause is that particularly for these knowledge fields, only a few other fields matter mostly (e.g. Field D relies strongly on Field B and C, Field F relies strongly on B) and the network topologies exploit profound access to these fields in the region *and* access to fields with incidentally related knowledge. In fact, smart network topologies have *less* access to the incidentally related knowledge. Looking at the path length statistics contained in “Appendix A”, we see that, for instance, agents in Field D are relatively well-connected to agents in Field B and C (high number of connections, which is reported between square brackets and low average path length), but the opposite holds for agents in other fields (so, low number of connections and high average path length). Moreover, where the local density of the gatekeeper and regional networks enhances access to incidentally related knowledge in case of diversified regions, this incidentally related knowledge is not or rarely present in case of regions with specialized knowledge or related variety. The density plots thus highlight the intricate relationship between knowledge distribution, network topology, and the technological interrelationships between the various knowledge fields.

5.3 High Level Lock-In

While the term ‘lock-in’ is generally associated with the inability to innovate due to technological specialization, there is a form of ‘lock-in’ also in the case of recombining exclusively within a set of (strongly) related variety knowledge. In the simulation results for the different scenarios, we observed that the ‘smart network’ may ultimately be outperformed by the ‘gatekeeper network’ particularly if the collaboration depth is $d = 2$ and $k = 3$. Inspection of simulation logs revealed that agents in smart networks sooner or later do not have access to knowledge in ‘incidentally related’ fields. In this $d = 2$ and $k = 3$ case, the few random connections of the gatekeeper

to the gatekeepers in other regions gives access to path-breaking (or rather ‘subgraph unlocking’) knowledge. This is corroborated by the path length statistics contained in “Appendix A” in the light of the co-occurrence matrix plotted in Fig. 3: in smart network topologies, agents have many connections and short path lengths to agents active in related fields, but few connections and long path lengths to agents active in less-related fields. Simulation results not included in this paper revealed that also pure random network topologies outperform smart networks for the same reason. The key implication is that while smart networks and access to related variety is performance enhancing, access to incidentally related variety remains crucial for discoveries in the long run. We consider this a counterpart of the ‘weakness of strong ties’ and how ‘weak ties’ may be a long-run competitive advantage.

A similar principle of ‘high level lock-in’ applies for purely regional networks and has ramifications for ‘branching’ policies. Looking at Fig. 11, we see that the cumulative number of discoveries in regional network is increasing faster if the region is diversified than if the region contains related variety. Analysis of simulation logs revealed that quite early on knowledge that is only incidentally related is not found, effectively blocking discovery of large subgraphs in the knowledge graph. From this, we see that a branching strategy in a region with merely related variety (see Boschma 2011; Asheim et al. 2011, advocating this regional innovation policy) is risky if there is no pipeline to such merely incidentally related knowledge.

5.4 Mediating Effect of the Knowledge Structure

The findings in this papers are particular to the knowledge graph as calibrated to the patent database. As discussed, there are several caveats to the patent classification code level we used for calibration and also to using patents in the first place. Our motivation to do this paper is to analyze the role of innovation network topology in knowledge discovery given a particular spatial distribution of knowledge. Yet, we also claim that the underlying technological knowledge structure is a (strong) mediator. Here we study more in depth, *how* the knowledge structure affects the innovativeness of the network topology given a spatial knowledge distribution. From this, a social planner can derive which network topology to pick. However, we restrict our analysis to changing the co-occurrence matrix \mathcal{C} and leave the distribution of the number of backward citations \mathcal{B} and the distribution of the citation lag \mathcal{L} unaltered. We construct the knowledge co-occurrence matrix as follows:

$$\mathcal{C}(\zeta) \leftarrow (1 - m \zeta) I + \zeta U,$$

whereby I is the identity matrix and U is the unit matrix with all entries equal to 1. In our simulations, we pick $\zeta = 0.005, 0.010, \dots, 0.120$ to vary the co-occurrence matrix between the case where each field uses nearly only knowledge from its own field (i.e. mostly self-referential) when $\zeta \downarrow 0$ to the case where fields are all nearly equally related to one another when $\zeta \uparrow 1/m$.

After inspection, we dismiss less interesting cases in which many agents are connected to many others already (e.g. $k = 3$ and $d = 2$). Rather, we select the case

with $k = 2$ and $d = 1$ with a few, yet non-trivial number of connections. Figure 13 contains plots of the cumulative number of unique discoveries after 1000 periods for the range of ζ values for each of the studied network topologies (regional, gatekeeper, smart), for each of the three studied geographical knowledge distributions (specialization, diversification, related variety). We see that, in general, in the gatekeeper network cases, high numbers of unique discoveries are attained. This is caused by the fact that the gatekeepers function as knowledge hub either between different regions specialized in different fields or between diversified pools of knowledge and thus stimulate innovativeness in all regions connected to. Gatekeeper networks even significantly outperform the smart and regional networks for near-identity co-occurrence matrices, i.e. for more self-referential knowledge. Interestingly, we see in Fig. 13a that for specialized regions and regions with related variety, the performance of smart and gatekeeper network topologies is (statistically) equal for near-uniform co-occurrence matrices. This is caused by the fact that, for near-uniform co-occurrence matrices, smart network topologies are nearly random, thus have relatively short paths to virtually all types of knowledge (particularly because the number of links $k > 1$). In contrast, gatekeepers still first connect to gatekeepers that are immediate neighbors and connect with uniform randomly drawn other gatekeepers if the number of ties k permits. The focus on regional ties for non-gatekeepers and the focus on immediate neighbors for gatekeepers limits the access to other types of knowledge, hence the relatively weak performance.

Conclusively, if knowledge is highly cross-referential, relationships in the (inter) regional innovation networks should be formed on the basis of knowledge relationships and rather *ignore* geographical factors. This is particularly the case for (strongly) uniform knowledge structures with specialized regions; any geographical bias in collaboration relationships hampers innovativeness. This is corroborated by looking at the performance of the smart network topology. Although not the best performing network, the number of discoveries in the smart network cases is consistent across the three scenarios for the spatial knowledge distribution. This is because of the focus on technological rather than geographical relatedness. The merely sub-optimal performance is due to absence of access to *incidentally* related technological knowledge.

Another phenomenon observed is that the number of discoveries in the gatekeeper (dashed) and regional (dotted) network changes from down- to upward sloping if the regions contain diverse rather than specialized knowledge or related variety. Whenever regions contain specialized knowledge or related variety while discoveries often require knowledge from other fields (high ζ), the local network density and geographical bias towards immediate neighbors in gatekeeper networks hampers access to that knowledge from other fields. Whenever regions are diversified, though, the local network density is in fact conducive to a high discovery rate. This observation validates the role of the geographical bias in innovativeness in the previous paragraphs.

5.5 Technological Focus Stifles Progress

One of the first findings that actually led to a change in our operational search algorithm is a form of lock-in which may very well exist in the real world. Analysis of logs of

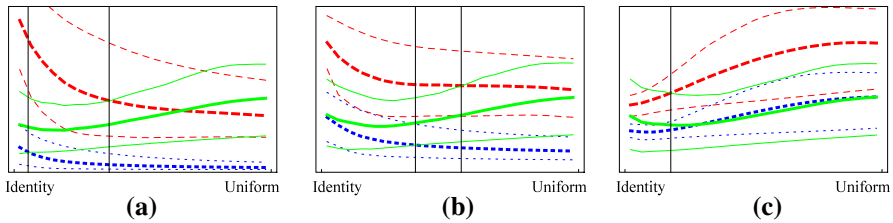


Fig. 13 Number of discoveries new to the world at time 1000 for different stylized knowledge matrices, ranging from a near-identity matrix ($\zeta = 0.005$) to a near-uniform matrix ($\zeta = 0.12$), for a regional (dotted), smart (continuous), and gatekeeper (dashes) network topology. Plotted are the averages and 90% confidence interval. **a** Regions contain specialized knowledge, **b** regions contain related variety, **c** regions contain diverse knowledge

single runs revealed that if agents initiate R&D only if the discovered knowledge falls within their own field of specialization, the discovery of knowledge in fields different than the fields sought to extend is inhibited. Suppose that a knowledge unit, say X in field x can be combined with a knowledge unit Y in field y (which may or may not be equal to field x) and that the outcome is knowledge unit Z in field z with $z \neq x$. If a technologically focused agent active in field x focuses exclusively on discoveries in its own field, it will not conduct R&D to discover Z unless approached by an agent active in field z . However, if $x \neq z$ and $y \neq z$, no agent will seek to discover technological knowledge Z . The effects on further development may be substantial, after all, given the high number of backward citations, see \mathcal{B} , this knowledge Z is itself likely to be required to discover more advanced knowledge. Consequently, the technological focus of agents in their R&D activities in fact inhibits discovery of whole subgraphs of the knowledge graph.

Figure 14 plots the descriptive statistics for both the case in which agent do (bottom row) and do not (top row) have an explicit technological focus, i.e. whether agents only seek to discover knowledge in its own field of expertise (bottom row) or not (top row). We see that a technological focus of agents causes the total number of discoveries to drop significantly (compare Fig. 14c with Fig. 14f), but that this is not only due to lower advancedness of the discoveries (compare Fig. 14b with Fig. 14e), but rather that the percentage of discoveries across all tiers drops (compare Fig. 14a with Fig. 14d). Since (partnerships of) agents do not unlock technologies that are not in their field of expertise, whole subgraphs remain undiscovered, thereby limiting future options for recombination as well, with—as we see here—significant consequences. In Sect. 6, we provide recommendations on how policy makers may stimulate agents to broaden their research scope.

6 Conclusions and Policy Implications

In this paper, we studied how a social planner seeking to enhance innovativeness of one or multiple regions should take into account the (inter)regional innovation network and the spatial distribution and structure of technological knowledge. Our findings confirm our conjecture that the regional specialization versus diversification versus

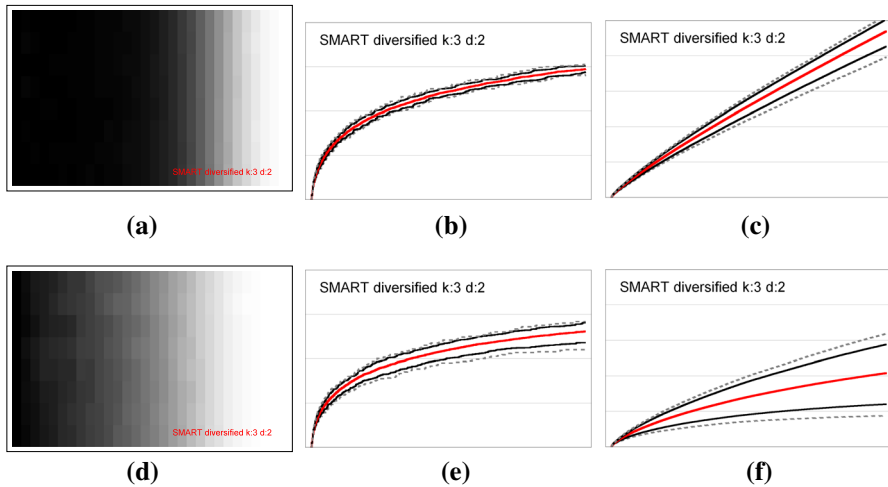


Fig. 14 Plots of descriptive statistics for $k = 3$ and $d = 2$ in case without (top row) or with (bottom row) explicit technological focus. Here we use the knowledge structure $\mathcal{C}(\zeta)$ with $\zeta = 0.05$ for clarity of exposition. **a** Density plot of % discovered **with no** technological focus, **b** maximum advantage **with no** technological focus, **c** total number of unique knowledge units discovered **with no** technological focus, **d** density plot of % discovered **with** technological focus, **e** maximum advantage **with** technological focus, **f** total number of unique knowledge units discovered **with** technological focus

related variety debate (and the associated policy paradigms) cannot be settled without taking the (inter)regional innovation network into consideration. Harking back to Table 1, we see that several of the cells in the contingency table that are empty (i.e. these particular combinations were not explicitly discussed in the reviewed theoretical literature) have been studied using our agent-based model. Generally, though, the research question ‘given a particular geographical distribution of and interrelationship of technological knowledge, what network topology should a social planner pick to enhance innovativeness?’ has no simple answer. A short answer revolves around the fact access to related variety but particularly also ‘incidentally related’ knowledge is crucial for long term technological progress. There are several combinations of a spatial knowledge distribution and a network topology that inhibit access to such incidentally related knowledge. Not only regional networks in case regions have specialized knowledge or related variety, but also smart (interregional) networks with agents in specialized regions, or regions with specialized knowledge or related variety connected by gatekeepers that connect on geographical rather than technological basis. That said, although diverse regions are generally performing relatively well, notably because agents have access to ‘incidentally related’ knowledge in each of the studied network topologies, performance is greatly enhanced by increasing the number of relationships (e.g. through triadic closure) and access to knowledge outside the region (notably through gatekeepers).

Our main conclusions and this short answer do rest on the premise that the knowledge graph truly is as calibrated. However, in the long answer to the research question found below, we elaborate also on how the significance of collaboration with firms

in other fields and in other regions depends on *structural features of the underlying technological knowledge*. First of all, strictly regional networks and thereby industrial district policies are to be dismissed: strictly regional networks are almost always outperformed by the other two topologies, irrespective of the spatial knowledge distribution, if not because of hampered access to (incidentally) related knowledge, then simply because the total pool of knowledge available to agents is smaller. Fundamental analysis in Sect. 5.4 showed this also is the case for regional networks if the technological knowledge field is (mostly) self-referential, i.e. if the co-occurrence matrix is (nearly) the identity matrix. Moreover, smart networks are in turn doing about as well as but are a few times outperformed by the gatekeeper networks. Given the high number of backward citations and frequent cross-references between knowledge fields, access to knowledge in other fields is crucial. Persistent innovativeness requires access to merely ‘incidentally’ related knowledge and smart networks provide this less often than do gatekeeper networks. For high connectivity scenarios, this is because gatekeeper networks have ‘weak’ ties to random knowledge pools, while smart networks have mostly ties based on technological co-occurrence probabilities, thus overlooking the importance of incidentally related knowledge for path-breaking or rather subgraph unlocking.

For the same reason, the performance of the recently advocated ‘regional branching’ policy paradigm (in which regions merely host a mix of ‘related variety’) is performing sub par. As agents cannot access incidentally related knowledge, particular subgraphs in the knowledge graph are not unlocked such that innovativeness ultimately tapers off. We coined the term for ‘high-level’ lock-in, which distinguishes stifling of technological progress because agents are limited to *specialized* knowledge of that because agents have access to related variety for branching, but not to incidentally related knowledge required to unlock alternative subgraphs of knowledge. Interestingly, we also find that the ‘smart specialization policy’ followed by, e.g. the European Commission, has a drawback for the same reason, namely that there is limited access to incidentally related knowledge. Not surprisingly, the smart specialization policy performs better if there are more relationships and when there is triadic closure that connects partners that are otherwise not collaborating.

We find that the gatekeeper network has remarkably high advancedness of and total number of discoveries, across all connection settings (k and d), but saliently for cases with few connections (low k or d). Analysis of the knowledge repository of the gatekeepers reveals that they accumulate an extensive and relatively diverse set of knowledge and thus are the motor for discoveries in their principal region, in connected regions, but indirectly also in other regions. Fundamental analysis in Sect. 5.4 on the moderating role of the knowledge structure reveals that the gatekeeper network topology is generally performing better than the smart and regional network topologies. However, if the knowledge structure is nearly uniform ($\zeta \uparrow 1/m$), the smart network outperforms the gatekeeper network if regions are highly specialized and gatekeepers have few and mostly geographically proximate relationships. Moreover, the smart network topology has a consistent performance across the various spatial distributions of knowledge. The general implication here is that technological relatedness should prevail over geographical relatedness in constructing (inter)regional relationships.

Apart from the sweeping conclusions on the type of network topology that a social planner should prefer in conjunction to the geographical distribution of knowledge (and a particular knowledge structure, which we presume to have calibrated correctly), there are a few general recommendations we can make. As a first general recommendation, it should be noted that technological knowledge is fairly self-referential such that buffering against inefficient interaction across knowledge fields is commendable. Finding a balance between collaborating with immediate peers, to parties active in frequently related fields, and ultimately also parties only incidentally related fields seems to be a challenge. The simulations reveal that gatekeepers do not only provide this buffer, but create short paths, provide triadic closure, and provide rich knowledge repository, and are as such a valuable asset in innovation networks.

An important observation we did was that ‘technological focus’ of agents strongly impedes technological progress, not only of the individual agent but also of the region and thereby the world. This ‘technological focus’ may be argued to be the flip side of the medal when it comes to the benefits of vertical specialization. We argue that social planners should make sure that firms should also care about research when it merely applies its current knowledge outside its field of expertise. The social planner could stimulate applied research, install research valorization centers, stimulate horizontal diversification of firms, etc.

As discussed before, there are several caveats to using patents in general and using patents in the way we did. Apart from fairly limited expositions in the history of technology literature, there are, to our knowledge, no serious alternatives available that we can use to calibrate knowledge graphs. This while we need such a knowledge graph with a real-world structure to address research questions as ours in this paper. That said, we are aware that the method proposed in the paper for empirical calibration and also the level of classification codes that we have used (at the patent section level) may well be improved. For one, we can imagine that a higher granularity of knowledge is desired, e.g. to the level of subsections, class, or even subclass.

However, this paper is, in our view, a first step in a potentially fruitful research line that addresses research questions that require the control that agent-based models can provide, yet are not purely fundamental with regard to actual technology/knowledge search processes. In our view, a model such as ours thus moves research in the direction of providing more specific and yet externally valid implications for social planners and policy makers when it comes to stimulating innovativeness.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A: Average Minimum Shortest Paths Between Agents by Specialization

To assess the chance to make particular combinations of knowledge units, the topologies and knowledge distributions in case of each of the 3×3 scenarios in technology

		REGIONAL								GATEKEEPER								SMART							
		A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H
SPECIALIZED	A	1.00 [99]	-	-	-	-	-	-	-	1	1	2	3	4	3	2	1	1.00 [85]	2.19 [57]	1.19 [75]	2.62 [60]	3.93 [44]	2.89 [50]	2.45 [54]	3.52 [50]
	B	-	1.00 [99]	-	-	-	-	-	-	1	1	1	2	3	4	3	2	2.19 [57]	1.00 [87]	1.53 [62]	1.47 [57]	2.07 [47]	1.62 [57]	1.93 [58]	2.44 [46]
	C	-	-	1.00 [98]	-	-	-	-	-	2	1	1	1	2	3	4	3	1.19 [75]	1.53 [62]	1.00 [88]	1.71 [64]	3.14 [46]	2.39 [55]	1.82 [58]	2.18 [51]
	D	-	-	-	1.00 [98]	-	-	-	-	3	2	1	1	1	2	3	4	2.62 [60]	1.47 [57]	1.71 [64]	1.00 [79]	4.14 [44]	3.36 [47]	3.20 [50]	3.78 [44]
	E	-	-	-	-	1.00 [98]	-	-	-	4	3	2	1	1	1	2	3	3.93 [44]	2.07 [47]	3.14 [46]	4.14 [44]	1.00 [82]	2.44 [46]	3.53 [47]	4.07 [38]
	F	-	-	-	-	-	1.00 [98]	-	-	3	4	3	2	1	1	1	2	2.89 [50]	1.62 [57]	2.39 [55]	3.36 [47]	2.44 [46]	1.00 [81]	3.13 [51]	2.74 [46]
	G	-	-	-	-	-	-	1.00 [98]	-	2	3	4	3	2	1	1	1	2.45 [54]	1.93 [58]	1.82 [58]	3.20 [50]	3.53 [47]	3.13 [51]	1.00 [82]	1.21 [65]
	H	-	-	-	-	-	-	-	1.00 [99]	1	2	3	4	3	2	1	1	3.52 [50]	2.44 [46]	2.18 [51]	3.78 [44]	4.07 [38]	2.74 [46]	1.21 [65]	1.00 [85]
DIVERSIFIED	A	-	1.05 [12]	1.05 [12]	1.09 [12]	1.05 [12]	1.06 [12]	1.06 [12]	1.16 [12]	2.11	1.09	1.06	1.05	1.04	1.06	1.06	1.08	1.00 [83]	1.74 [57]	1.17 [71]	2.43 [56]	3.24 [43]	3.14 [42]	2.64 [53]	3.53 [44]
	B	1.05 [12]	-	1.07 [12]	1.07 [12]	1.08 [12]	1.10 [12]	1.09 [12]	1.12 [12]	1.09	2.27	1.14	1.08	1.05	1.13	1.12	1.07	1.74 [57]	1.00 [81]	1.36 [61]	1.77 [58]	2.12 [49]	1.74 [51]	1.77 [58]	2.23 [47]
	C	1.05 [12]	1.07 [12]	-	1.05 [12]	1.12 [12]	1.10 [12]	1.06 [12]	1.08 [12]	1.06	1.14	2.27	1.05	1.03	1.1	1.17	1.09	1.17 [71]	1.36 [61]	1.00 [84]	1.68 [60]	2.79 [48]	2.34 [52]	1.86 [48]	2.58 [43]
	D	1.09 [12]	1.07 [12]	1.05 [12]	-	1.09 [12]	1.09 [12]	1.07 [12]	1.11 [12]	1.05	1.08	1.05	2.12	1.05	1.08	1.08	1.1	2.43 [56]	1.77 [58]	1.68 [60]	1.00 [79]	4.20 [45]	3.53 [48]	3.13 [45]	3.81 [40]
	E	1.05 [12]	1.08 [12]	1.12 [12]	1.09 [12]	-	1.05 [12]	1.13 [12]	1.04 [12]	1.04	1.05	1.03	1.05	2.1	1.08	1.04	1.08	3.24 [43]	2.12 [49]	2.79 [43]	4.20 [39]	1.00 [78]	3.10 [41]	3.28 [43]	4.08 [34]
	F	1.06 [12]	1.10 [12]	1.10 [12]	1.09 [12]	1.05 [12]	-	1.04 [12]	1.08 [12]	1.06	1.13	1.1	1.08	1.08	2.28	1.15	1.12	3.14 [42]	1.74 [58]	2.34 [48]	3.53 [45]	3.10 [44]	1.00 [77]	2.79 [45]	2.86 [42]
	G	1.06 [12]	1.09 [12]	1.06 [12]	1.07 [12]	1.13 [12]	1.04 [12]	-	1.04 [12]	1.06	1.12	1.17	1.08	1.04	1.15	2.27	1.07	2.64 [53]	1.77 [56]	1.86 [52]	3.13 [48]	3.28 [45]	2.79 [44]	1.00 [45]	1.25 [59]
	H	1.16 [12]	1.12 [12]	1.08 [12]	1.11 [12]	1.04 [12]	1.08 [12]	1.04 [12]	-	1.08	1.07	1.09	1.1	1.08	1.12	1.07	2.2	3.53 [44]	2.23 [47]	2.58 [43]	3.81 [40]	4.08 [34]	2.86 [42]	1.25 [59]	1.00 [85]
REL VARIETY	A	1.01 [59]	1.24 [05]	1.39 [05]	1.44 [05]	1.60 [05]	1.48 [05]	1.28 [05]	1.37 [05]	1.24	1.22	1.66	1.76	1.81	1.7	1.75	1.34 [99]	1.00 [89]	2.29 [57]	1.21 [75]	2.89 [60]	3.70 [42]	3.22 [48]	2.46 [49]	3.64 [41]
	B	1.24 [05]	1.02 [58]	1.50 [03]	1.25 [04]	1.34 [05]	1.26 [06]	1.32 [06]	1.41 [06]	1.22	1.17	1.22	1.64	1.71	1.85	1.76	1.6	2.29 [57]	1.00 [85]	1.37 [61]	2.09 [56]	2.11 [47]	1.52 [56]	1.69 [46]	2.39 [46]
	C	1.39 [05]	1.50 [03]	1.06 [55]	1.16 [06]	1.16 [04]	1.42 [04]	1.24 [06]	1.46 [06]	1.66	1.22	1.21	1.31	1.74 [99]	1.64 [97]	1.93 [91]	1.87 [95]	1.21 [75]	1.37 [61]	1.00 [96]	1.54 [65]	2.91 [42]	1.92 [51]	1.94 [53]	2.44 [50]
	D	1.44 [05]	1.25 [04]	1.16 [06]	1.04 [57]	1.30 [05]	1.30 [04]	1.24 [06]	1.27 [06]	1.76	1.64	1.31	1.23	1.32	1.53	1.63	1.65	2.89 [58]	2.09 [56]	1.54 [65]	1.00 [84]	4.27 [35]	3.43 [49]	3.29 [48]	3.32 [44]
	E	1.60 [05]	1.34 [04]	1.38 [05]	1.30 [05]	1.02 [59]	1.29 [05]	1.36 [06]	1.39 [05]	1.81	1.71	1.74	1.32	1.21	1.29	1.61	1.78	3.70 [42]	2.11 [47]	2.91 [42]	4.27 [35]	1.00 [82]	2.28 [37]	3.55 [35]	4.01 [31]
	F	1.48 [05]	1.26 [06]	1.42 [04]	1.30 [04]	1.29 [05]	1.01 [60]	1.34 [04]	1.52 [06]	1.7	1.85	1.64 [96]	1.53	1.29	1.28	1.28	1.59	3.22 [48]	1.52 [56]	2.92 [51]	3.43 [49]	2.28 [37]	1.00 [79]	2.53 [48]	2.96 [46]
	G	1.28 [05]	1.32 [06]	1.24 [06]	1.24 [06]	1.36 [06]	1.34 [06]	1.01 [06]	1.26 [06]	1.75	1.76	1.93 [99]	1.63 [91]	1.61 [95]	1.28 [96]	1.19 [96]	1.32	2.46 [49]	1.69 [46]	1.94 [53]	3.29 [48]	3.55 [43]	2.53 [35]	1.00 [48]	1.18 [77]
	H	1.37 [05]	1.41 [06]	1.46 [04]	1.27 [06]	1.39 [05]	1.52 [06]	1.26 [06]	1.01 [56]	1.34	1.6	1.87 [95]	1.65 [99]	1.78 [97]	1.59 [98]	1.32 [92]	1.35	3.64 [41]	2.39 [46]	2.44 [50]	3.32 [44]	4.01 [31]	2.96 [46]	1.18 [56]	1.00 [91]

Fig. 15 Matrix of the pairwise average minimum shortest paths by field for each of the 3×3 scenarios, for $k = 1$. Between square brackets the percentage of paths between agents specialized in the two fields

discovery, the shortest path length to agents with (other) specializations is computed. Given that the topologies, generally, contain asymmetries, the various agents active in one and the same field may have different shortest path lengths to other agents. As the depth d of relationships of agents should be equal or higher than the *minimum* shortest path length for agents to be able to make combinations, the latter is reported. The tables in Figs. 15 and 16 report the average minimum shortest path length of 100 different network and knowledge distribution instances for m times m pairwise combinations between fields. To compare the inherent chances of making combinations across the 3×3 scenarios, we provide the matrix of average minimum shortest path lengths of the pairwise combinations of fields for each of these scenarios. The average minimum shortest path length of pairs of fields is used in explaining differences in the average maximum advancedness, the cumulative number of unique discoveries, and density of discoveries.

		REGIONAL								GATEKEEPER								SMART							
		A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H
SPECIALIZED	A	1	-	-	-	-	-	-	-	1	1	1.23	1.2	1.16	1.21	1.24	1	1	1.14	1	1.38	1.77	1.63	1.38	1.92
	B	-	1	-	-	-	-	-	-	1	1	1	1.18	1.17	1.21	1.2	1.23	1.14	1	1.03	1.04	1.14	1.03	1.1	1.21
	C	-	-	1	-	-	-	-	-	1.23	1	1	1	1.16	1.21	1.19	1.26	1	1.03	1	1.03	1.45	1.32	1.23	1.26
	D	-	-	-	1	-	-	-	-	1.2	1.18	1	1	1	1.19	1.19	1.19	1.38	1.04	1.03	1	1.96	1.66	1.62	1.94
	E	-	-	-	-	1	-	-	-	1.16	1.17	1.16	1	1	1	1.17	1.16	1.77	1.14	1.45	1.96	1	1.27	1.58	1.98
	F	-	-	-	-	-	1	-	-	1.21	1.21	1.21	1.19	1	1	1	1.25	1.63	1.03	1.32	1.66	1.27	1	1.29	1.46
	G	-	-	-	-	-	-	1	-	1.24	1.2	1.19	1.19	1.17	1	1	1	1.38	1.1	1.23	1.62	1.58	1.29	1	1
	H	-	-	-	-	-	-	-	1	1	1.23	1.26	1.19	1.16	1.25	1	1	1.92	1.21	1.26	1.94	1.98	1.46	1	1
DIVERSIFIED	A	-	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.09 [13]	1.00 [13]	1.00 [13]	1.93	1	1	1	1	1	1	1	1	1.18	1	1.46	1.59	1.62	1.26	1.86
	B	1.00 [13]	-	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.07 [13]	1.00 [13]	1	2.19	1	1	1	1	1	1	1.18	1	1.01	1.07	1.13	1.03	1.06	1.28
	C	1.00 [13]	1.00 [13]	-	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.10 [13]	1	1	2.18	1	1	1	1	1	1	1.01	1	1.01	1.41	1.3	1.16	1.25
	D	1.00 [13]	1.00 [13]	1.00 [13]	-	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1	1	1	2	1	1	1	1.01	1.46	1.07	1.01	1	1.96	1.73	1.68	1.86
	E	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	-	1.00 [13]	1.00 [13]	1.00 [13]	1	1	1	1	1.92	1	1	1.01	1.59	1.13	1.41	1.96	1	1.16	1.6	1.94
	F	1.09 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	-	1.00 [13]	1.00 [13]	1	1	1	1	1	2.21	1	1	1.62	1.03	1.3	1.73	1.16	1	1.26	1.44
	G	1.00 [13]	1.07 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	-	1.00 [13]	1	1	1	1	1	1	2.16	1.01	1.26	1.06	1.16	1.68	1.6	1.26	1	1.01
	H	1.00 [13]	1.00 [13]	1.10 [13]	1.00 [13]	1.00 [13]	1.00 [13]	1.00 [13]	-	1	1	1	1.01	1.01	1	1.01	2.03	1.86	1.28	1.25	1.86	1.94	1.44	1.01	1
REL VARETY	A	1.01 [59]	1.05 [05]	1.02 [05]	1.05 [05]	1.06 [05]	1.00 [05]	1.02 [05]	1.03 [05]	1	1.11	1.23	1.17	1.18	1.18	1.23 [99]	1.17	1	1.16	1	1.31	1.88 [99]	1.68 [99]	1.35 [98]	1.84
	B	1.05 [05]	1.00 [59]	1.05 [04]	1.02 [04]	1.00 [06]	1.03 [06]	1.03 [06]	1.06 [06]	1.11	1	1.1	1.19	1.23	1.27 [98]	1.19 [98]	1.24	1.16	1	1.02	1.02	1.09 [97]	1.01 [95]	1.03 [95]	1.14
	C	1.02 [05]	1.05 [04]	1.00 [56]	1.02 [06]	1.00 [04]	1.02 [04]	1.07 [04]	1.05 [04]	1.23	1.1	1	1.23 [99]	1.25 [97]	1.23 [96]	1.22 [91]	1.25 [95]	1	1.02	1	1.04	1.54 [99]	1.39 [97]	1.11 [95]	1.26
	D	1.05 [05]	1.02 [04]	1.02 [06]	1.00 [58]	1.00 [05]	1.02 [04]	1.00 [06]	1.00 [06]	1.17	1.19 [99]	1.23 [99]	1	1.2	1.19	1.21 [95]	1.22 [99]	1.31	1.02	1.04	1	1.90 [99]	1.70 [97]	1.70 [95]	1.89 [99]
	E	1.06 [05]	1.00 [06]	1.00 [04]	1.00 [05]	1.00 [61]	1.04 [05]	1.00 [06]	1.00 [05]	1.18	1.23 [97]	1.25 [96]	1.2	1	1.16	1.22 [95]	1.23 [97]	1.88	1.09	1.54 [99]	1.90 [99]	1	1.16 [97]	1.59 [96]	1.77
	F	1.00 [05]	1.03 [06]	1.02 [04]	1.02 [06]	1.04 [05]	1.00 [61]	1.08 [04]	1.03 [06]	1.18	1.27 [96]	1.23 [96]	1.19	1.16	1	1.18 [96]	1.22 [98]	1.68 [99]	1.01 [97]	1.39 [97]	1.70 [97]	1.16 [97]	1.00 [99]	1.33 [93]	1.36 [96]
	G	1.02 [05]	1.03 [06]	1.07 [06]	1.00 [06]	1.00 [06]	1.08 [04]	1.00 [58]	1.02 [06]	1.23	1.19 [99]	1.22 [98]	1.21 [91]	1.22 [95]	1.18 [96]	1.01 [96]	1.21 [92]	1.35 [98]	1.03 [95]	1.11 [95]	1.70 [95]	1.59 [96]	1.33 [93]	1.00 [97]	1.00 [95]
	H	1.03 [05]	1.06 [06]	1.05 [04]	1.00 [06]	1.00 [05]	1.03 [06]	1.02 [05]	1.00 [56]	1.17	1.24 [95]	1.25 [99]	1.22 [97]	1.23 [98]	1.22 [92]	1.21 [92]	1	1.84	1.14	1.26 [99]	1.89 [99]	1.77 [96]	1.36 [95]	1.00 [95]	1

Fig. 16 Matrix of the pairwise average minimum shortest paths by field for each of the 3×3 scenarios, for $k = 3$. Between square brackets the percentage of paths between agents specialized in the two fields

Appendix B: Simulation Onset Particularities Due to Low Occurrence

From the recursive algorithm described in Sect. 3.4, it is clear that the number of knowledge units in the first few tiers is relatively low and only gradually grows. The number of knowledge units in these first few tiers is of course particularly low for fields with a relatively low occurrence rate \mathcal{F} . In Fig. 1, we see that particularly fields *D*, *E*, and *F* have a low occurrence rate. As such, in various simulation runs it does occur that the first few tiers have *no* knowledge units in either of these fields. Consequently, in the case agents are technologically focused, there are agents that will not be engaged in research and development. Moreover, in expanding the knowledge graph, knowledge

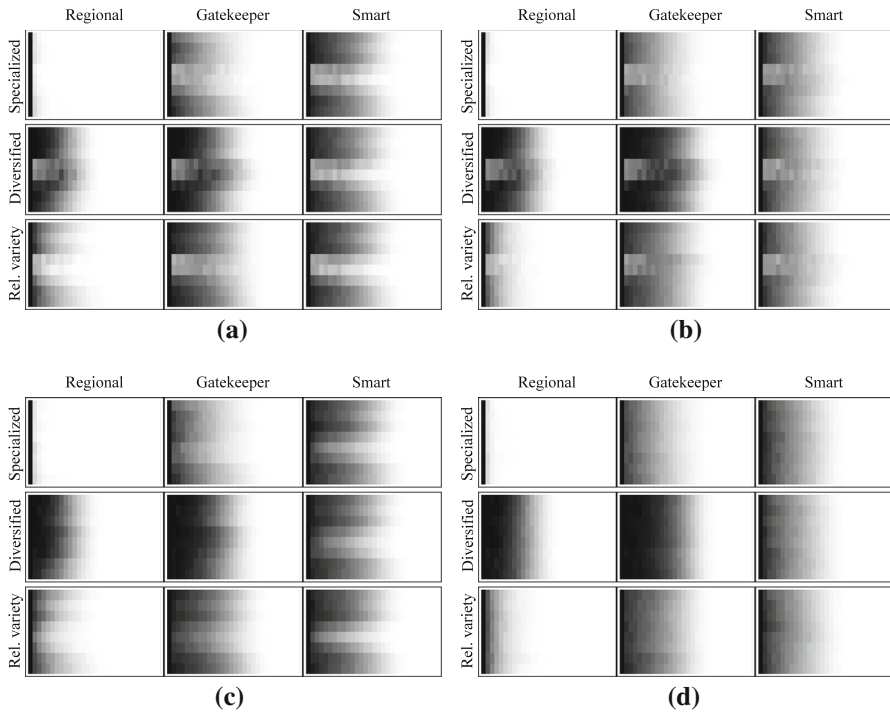


Fig. 17 Density plots of the average percentage of knowledge discovered at $T = 1000$ for 100 simulation runs (for connectivity settings $k = 2$ and $d = 1$) for the four combinations of using the empirical (co-) occurrence rates or rather the controlled rates (i.e. $\mathcal{F} = (1/m, \dots, 1/m)$ and $\mathcal{C}(\zeta = 0.04)$). **a** Empirical occurrence \mathcal{F} , empirical co-occurrence \mathcal{C} , **b** empirical occurrence \mathcal{F} , controlled co-occurrence \mathcal{C} with $\zeta = 0.04$, **c** uniform occurrence \mathcal{F} , empirical co-occurrence \mathcal{C} , **d** uniform occurrence \mathcal{F} , controlled co-occurrence \mathcal{C} with $\zeta = 0.04$

units in other fields have no ancestors in these fields to cite. Ultimately, though, with the increase in total number N_t of knowledge units per tier grows and the chances grow rapidly that there are knowledge units in these ‘rare’ fields. However, it is clear that the simulation results thus feature ‘onset behavior’ due to initialization of the simulation during early periods that gradually vanishes over simulation time. Although this particular ‘onset behavior’ is not an anomaly of our operationalization, we have decided to alter the occurrence rates to completely uniform, i.e. $\mathcal{F} = (1/m, \dots, 1/m)$, such that the simulation results differ purely due to the differences in the co-occurrence matrix \mathcal{C} . By doing so, we focus purely on the effect of the relationships between fields as captured in matrix \mathcal{C} and not the occurrence rates \mathcal{F} . Comparing the average percentages of discovered knowledge at the end of simulation runs (here for $k = 2$ and $d = 1$, but the results remain structurally the same for other settings) for cases with both the empirical and uniform occurrence rate (i.e. comparing Fig. 17a with Fig. 17c), we see that the ‘irregularities’ in the early tiers on the left-hand side are gone. The effects of, in addition, also controlling the knowledge graph by using the co-occurrence matrix $\mathcal{C}(\zeta = 0.04)$ can be seen by comparing these figures with Fig. 17b, d.

References

- Arthur, W. (2009). *The nature of technology: What it is and how it evolves*. New York: Free Press.
- Arthur, W., & Polak, W. (2006). The evolution of technology within a simple computer model. *Complexity*, 11, 23–31.
- Asheim, B., Boschma, R., & Cooke, P. (2011). Constructing regional advantage: Platform policies based on related variety and differentiated knowledge bases. *Regional Studies*, 45(7), 893–904.
- Asheim, B., & Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. *Research Policy*, 34, 1173–1190.
- Auerswald, P., Kauffman, S., Lobo, J., & Shell, K. (2000). The production recipes approach to modeling technological innovation: An application to learning by doing. *Journal of Economic Dynamics and Control*, 24(3), 389–450.
- Autant-Bernard, C., Fadairo, M., & Massard, N. (2013). Knowledge diffusion and innovation policies within the European regions: Challenges based on recent empirical evidence. *Research Policy*, 42, 196–210.
- Balland, P. A., Rigby, D., & Boschma, R. (2015). The technological resilience of us cities. *Cambridge Journal of Regions, Economy and Society*, 8(2), 167–184.
- Basalla, G. (1988). *The evolution of technology*. Cambridge: Cambridge University Press.
- Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1), 31–56.
- Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, 38, 318–337.
- Boschma, R. (2005). Proximity and innovation. A critical assessment. *Regional Studies*, 39(1), 61–74.
- Boschma, R. (2011). Regional branching and regional innovation policy. In K. Kourtit, P. Nijkamp, & R. R. Stough (Eds.), *Drivers of innovation, entrepreneurship and regional dynamics* (pp. 359–368). Berlin: Springer.
- Boschma, R. (2014). Constructing regional advantage and smart specialisation: Comparison of two European policy concepts. *Scienze Regionali*, 13, 51–68.
- Bourdieu, P. (1980). Le capital sociale: Notes provisoires. *Actes de la Recherche en Sciences Sociales*, 31, 2–3.
- Breschi, S., & Lenzi, C. (2015). The role of external linkages and gatekeepers for the renewal and expansion of us cities' knowledge base, 1990–2004. *Regional Studies*, 49(5), 782–797.
- Bristow, G. (2010). Resilient regions: Re-'place'ing regional competitiveness. *Cambridge Journal of Regions, Economy and Society*, 3, 153–167.
- Burt, R. (1992). *Structural holes: The social structure of competition*. Cambridge: Harvard University Press.
- Camagni, R., & Capello, R. (2013). Regional innovation patterns and the EU regional policy reform: Toward smart innovation policies. *Growth and Change*, 44(2), 355–389.
- Capaldo, A., & Petruzzelli, A. (2014). Partner geographic and organizational proximity and the innovative performance of knowledge creating alliances. *European Management Review*, 11(1), 63–84.
- Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of us state-level patenting. *Regional Studies*, 49(5), 767–781.
- Chen, S., & Chie, B. (2006). A functional modularity approach to agent-based modeling of the evolution of technology. In A. Namatame, T. Kaizouji, & Y. Aruka (Eds.), *The complex networks of economic interactions* (pp. 165–178). Berlin: Springer.
- Chie, B. T., & Chen, S. H. (2013). Non-price competition in a modular economy: An agent-based computational model. *Economia Politica: Journal of Analytical and Institutional Economics*, 3, 273–300.
- Childe, V. (1936). *Man makes himself*. London: New Thinker's Library.
- Christopherson, S., Michie, J., & Tyler, P. (2010). Regional resilience: Theoretical and empirical perspectives. *Cambridge Journal of Regions, Economy and Society*, 3(1), 3–10.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge: Harvard University Press.
- Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change*, 10(4), 945–974.
- Cowan, R., & Jonard, N. (2007). Structural holes, innovation and the distribution of ideas. *JEIC*, 2(2), 93–110.
- Cumbers, A., & MacKinnon, D. (2004). Introduction: Clusters in urban and regional development. *Urban Studies*, 41(5,6), 959–969.
- Feldman, M. (1994). *The geography of innovation*. Boston: Kluwer Academic Publishers.

- Feldman, M., & Audretsch, D. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2), 409–429.
- Foray, D. (2014). *Smart specialisation: Opportunities and challenges for regional innovation policy*. London: Routledge.
- Foray, D., David, P., & Hall, B. (2009). Smart specialisation. The concept. Knowledge economists policy brief 9. Expert group on Knowledge for growth.
- Foray, D., David, P., & Hall, B. (2011). *Smart specialisation from academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation*. Tech. rep. EPFL.
- Frenken, K. (2006). *Innovation, evolution and complexity theory*. Cheltenham: Edward Elgar.
- Frenken, K., van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41, 685–697.
- Gertler, M. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3, 75–99.
- Gilbert, N., Pyka, A., & Ahrweiler, P. (2001). Innovation networks: A simulation approach. *Journal of Artificial Societies and Social Simulation*, 4(3). <http://www.soc.surrey.ac.uk/JASSS/4/3/8.html>
- Glaeser, E., Kallal, H., Scheinkman, J., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126–1152.
- Graf, H. (2011). Gatekeepers in regional networks of innovators. *Cambridge Journal of Economics*, 35(1), 173–198.
- Graf, H., & Krüger, J. (2011). The performance of gatekeepers in innovator networks. *Industry and Innovation*, 18(1), 69–88.
- Grant, R., & Baden-Fuller, C. (1995). A knowledge-based theory of inter-firm collaboration. *Academy of Management Proceedings*, 1, 17–21.
- Grant, R., & Baden-Fuller, C. (2004). A knowledge accessing theory of strategic alliances. *Journal of management studies*, 41(1), 61–84.
- Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4), 501–522.
- Hagedoorn, J. (2002). Inter-firm R&D partnerships: An overview of major trends and patterns since 1960. *Research Policy*, 31, 477–492.
- Hall, B., Helmers, C., & von Graevenitz, G. (2015). *Technology entry in the presence of patent thickets*. Tech. rep.
- Hall, B., Helmers, C., von Graevenitz, G., & Rosazza-Bondibene, C. (2012). *A study of patent thickets*. Report.
- Hassink, R. (2005). How to unlock regional economies from path dependency? *European Planning Studies*, 13(4), 521–535.
- Johnson, B., Lorenz, E., & Lundvall, B. (2002). Why all this fuss about codified and tacit knowledge? *Industrial and Corporate Change*, 11(2), 245–262.
- Karlsson, C., & Stough, R. (Eds.). (2005). *Industrial clusters and inter-firm networks*. Cheltenham: Edward Elgar Publishing.
- Korber, M., & Paier, M. (2014). R&D networks and regional knowledge production: An agent-based simulation of the Austrian competence centres programme. *Economy of Region*, 2, 264–275.
- Korhonen, J., & Kasmire, J. (2013). *Adder: A simplified model for simulating the search for innovations*. Working paper presented at druid.
- MacKinnon, D., Cumbers, A., & Chapman, K. (2002). Learning, innovation and regional development: A critical appraisal of recent debates. *Progress in Human Geography*, 26(3), 293–311.
- Martin, R., & Sunley, P. (2011). Conceptualizing cluster evolution: Beyond the life cycle model? *Regional Studies*, 45(10), 1299–1318.
- Maskell, P., & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics*, 23, 167–185.
- Menzel, M. P., & Fornahl, D. (2010). Cluster life cycles—Dimensions and rationales of cluster evolution. *Industrial and Corporate Change*, 19(1), 205–238.
- Morone, P., & Taylor, R. (2010). *Knowledge diffusion and innovation: Modelling complex entrepreneurial behaviours*. Cheltenham: Edward Elgar Publishing.
- Neffke, F., Henning, M., Boschma, R., Lundquist, K., & Olander, L. (2011). The dynamics of agglomeration externalities along the life cycle of industries. *Regional Studies*, 45, 49–65.
- Paci, R., & Usai, S. (2000). The role of specialization and diversity externalities in the agglomeration of innovative activities. *Rivista Italiana degli Economisti*, 2, 237–268.

- Padgett, J., Lee, D., & Collier, N. (2003). Economic production as chemistry. *Industrial and Corporate Change*, 12(4), 843–877.
- Pinch, S., Henry, N., Jenkins, M., & Tallman, S. (2003). From ‘industrial districts’ to ‘knowledge clusters’: A model of knowledge dissemination and competitive advantage in industrial agglomerations. *Journal of Economic Geography*, 3, 373–388.
- Polanyi, M. (1966). *Tacit knowing*. New York: Doubleday & Co.
- Pyka, A. (2002). Innovation networks in economics: From the incentive-based to the knowledge-based approaches. *European Journal of Innovation Management*, 5(3), 152–163.
- Pyka, A., & Küppers, G. (Eds.). (2002). *Innovation networks*. Cheltenham: Edward Elgar Publishing.
- Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. *Management Science*, 49(6), 751–766.
- Saxenian, A. (1996). Inside-out: Regional networks and industrial adaptation in Silicon Valley and Route 128. *Cityscape: A Journal of Policy Development and Research*, 2(2), 41–60.
- Spencer, J. (2003). Global gatekeeping, representation, and network structure: A longitudinal analysis of regional and global knowledge-diffusion networks. *Journal of International Business Studies*, 34(5), 428–442.
- Tödtling, F., & Trippel, M. (2005). One size fits all? Towards a differentiated regional innovation policy approach. *Research Policy*, 34, 1203–1219.
- van der Panne, G. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics*, 14(5), 593–604.
- Vermeulen, B., & Paier, M. (Eds.). (2016). *Innovation networks for regional development. Concepts, case studies, and agent-based models*. Berlin: Springer.
- Vermeulen, B., & Pyka, A. (2014a). The effects of supraregional innovation and production collaboration on technology development in a multiregional world: A spatial agent-based model study. *Lecture Notes in Computer Science*, 8751, 698–707.
- Vermeulen, B., & Pyka, A. (2014b). Technological progress and effects of (supra)regional innovation and production collaboration. An agent-based model simulation study. In *Proceedings of IEEE conference on computational intelligence for financial engineering and economics (CIFER)*.
- Vermeulen, B., Pyka, A., La Poutre, J., & de Kok, A. (2016a). Capability-based governance patterns over the product life-cycle. An agent-based model. *Journal of Economic Interaction and Coordination*, 1–39. <https://doi.org/10.1007/s11403-016-0184-x>.
- Vermeulen, B., Zvarich, I., & Messmer, B. (2016b). A descriptive statistics exploration of spatio-temporal patterns in inventive activities in the pharmaceutical industry. In B. Vermeulen & M. Paier (Eds.), *Innovation networks for regional development: Concepts, case studies, and agent-based models*. Cham: Springer.
- Woodard, C., & Clemons, E. (2012). Modeling technology evolution using generalized genotype–phenotype maps. In *Proceedings of the 14th annual genetic and evolutionary computation conference* (pp. 323–330).