Paper to be presented at the DRUID Academy Conference 2017 at University of Southern Denmark, Odense, Denmark on January 18-20, 2017

Teamwork and the structure of technological knowledge

Matteo Tubiana
Università degli Studi di Torino
Economics Department, Cognetti De Martiis
matteo.tubiana@carloalberto.org

Abstract

Teamwork and the structure of technological knowledge Matteo Tubiana, Università di Torino and BRICK, Collegio Carlo Alberto Year of Enrolment: 2014 Expected Final Date: November 2017 Email: matteo.tubiana@carloalberto.org There is increasing empirical evidence of the dominance of teams in knowledge production activities, either academic or technological. The existing literature concentrates most on the effects of this phenomenon, namely the positive and negative consequences of teamwork. Many researchers described the role of inventors networks in shaping knowledge externalities, as well as many others dug into the costs of teamwork and the effects of different internal management for creativity. There is not, instead, much work on the reasons why knowledge agents interact in order to accomplish their projects. The dominant view on the matter is the burden of knowledge hypothesis framed by Jones (2009), stating that teams arise because the number of specialists among knowledge agents rises too: no one can master the knowledge frontier alone, therefore agents team up. We buy this hypothesis, but we go a step further. Inspired by the work of Arthur (2009) on technology and complexity, we concentrate on patent data and propose a framework to test the relationship between the characteristics of technological knowledge, complexity in particular, and teamwork activity. The novelty is twofold: on a first account, we try to enter into the dynamics of knowledge creation, which is eminently interactive, evidencing the role of specific knowledge characteristics in shaping knowledge production. Secondly, we make a first step towards the implementation of the theory of creation and evolution of technological knowledge advanced by Arthur, which is the only and comprehensive one at disposal. As for the interactive nature of knowledge creation, we recall the Lane and Maxfield (1996) definition of generative relationship. They argue that, under certain circumstances, interactions among agents affect agents' interpretative structure of reality, supporting them in dealing with complex, uncertain situations. Building on the recombinant theory of knowledge production, and recalling the vast literature on the characteristics of knowledge, we assert that knowledge generation is a complex task indeed. Therefore, we describe the interactive intensity in technological knowledge production as a reaction to the complexity of the task, growing as a modus operandi. Concerning the characteristics of knowledge, we follow the literature exploiting co-occurrences of technological classes enlisted in each patent. Differently to the literature, we use this methodology to create a series of variable describing the characteristics of the technological knowledge map or space. Therefore, identifying 3-digit IPC classes as different technologies, we can track the evolution of the space throughout time. We end up with variables accounting for the internal structure of a technological domain (knowledge burden, interdependence, specialization) and its position in the space (connectivity). The

objective is to test whether these variables explain the trend in the intensity of teamwork in each technological domain. Our source of data is the CRIOS database, which is an augmented PATSTAT-based dataset. The dataset we run analysis on is, then, a panel, whose unit of analysis is technological domains, from 1989 to 2000. As for the methodology, other than standard FE estimation, we try out a fairly new methodology proposed by Bell and Jones (2015), useful to capture both within and between variation in a RE estimation, simultaneously. Even though we still have to address the heterogeneity issue in order to claim causality, the preliminary results support our hypothesis. In a few words, the more technological production gets complex within a domain, the more it happens through teamwork. This correlation is not significant between domains, when the knowledge burden plays the major role in differencing domains' teamwork intensity. Similarly, the share of company-owner and pace of growth in applications production positively matter in inducing teamwork within domains. References Arthur, B. (2015). "Complexity and the Economy", Oxford University Press. Bell, A., Jones, K. (2015). "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data". Political Science Research and Methods, 3, pp 133-153. Jones, B. F. (2009). "The burden of knowledge and the 'death of the renaissance man': Is innovation getting harder?". The Review of Economic Studies, 76(1), 283-317. Lane, D., Maxfield, R. (1996). "Strategy under complexity: Fostering generative relationships". Long Range Planning, 29(2), 215-231.

Jelcodes:O31,-

Teamwork and the structure of technological knowledge

Matteo Tubiana: Dipartimento di Economia e Statistica "Cognetti de Martiis", Università di Torino & Collegio Carlo Alberto

I. Introduction

Even though the main narratives about invention and innovation tell a story of lonely geniuses, teamwork has always been around in the field. Certainly, the pace of growth of teams in knowledge production in the last decades did accelerate a lot, calling much researchers' attention from various fields, from economics to management and psychology. We want to add a little piece of understanding to the many contributions on the topic, focusing on team working as a *modus operandi* of technological knowledge production (in this paper, we deal with patent-based data, therefore we speak of *technological* knowledge). We call it inventors' *polygamy*. We investigate the relationship between polygamy and the structural characteristics of technological knowledge, depicted as a space, map or network following Arthur (2009)'s theoretical formulation of the nature of technology.

II. The aim of the study

The aim of this explorative research is to investigate the relationships between the characteristics of the stock of technological knowledge and the increasing tendency towards teamwork in the inventive activity. In so doing, we draw upon three important past contributions, providing us the tools to frame, on the one side, the composition of the knowledge stock, on the other, the dynamics leading its reproduction to happen through team working. On the first account, we represent the stock of technological knowledge each inventor quarries on as a heterogeneous map or network, *de facto* building on the theoretical framework provided by Arthur (2009) to depict technology as a *complex* evolving system. Concerning instead the appearance of teams on the stage, we rely on the famous analysis of Jones (2009) who formalized the idea that the knowledge stock for invention, either technological or scientific, is growing in size, making harder for each inventor to reach the frontier where invention is possible. The third contribution that we will discuss, is Schmookler (1957)'s professionalization dynamics, which reflects an accurate description of the institutionalization of research – consequently of inventive – activity within firms.

As we will try to articulate throughout the paper, these approaches are not incompatible neither antithetic, instead they can be integrated in order to quest into the organizational features of technological knowledge creation, informed of the scope, size, structure and socio-institutional qualities of the technological knowledge stock. Indeed, they provide different but coherent frameworks to interpret the insurgence of teamwork in inventive activity. Jones' approach depicts knowledge creation as an individualistic problem, and the cumulability of knowledge emerges as the key explaining factor for relationships among individual inventors. Arthur's complexity framework, enriched with related contributions, instead, provides us with the ground to appreciate the intrinsic interactive nature of knowledge generation as a collective phenomenon. Finally, Schmookler's work helps to situate properly invention within a broader institutional context, in constant evolution. All of the three, together, make it possible to overcome a sharp duality between the Schumpeterian literature narratives of the lone genius against the corporate laboratory as the proper forge for path-breaking innovations, opening new paths for the exploration of the interactions between the socio-institutional environment and the complex characteristics of the knowledge stock in the organization of inventive activity.

As for the result, we find both the cumulative and complexity approaches to be confirmed by our German patent-based dataset on technology. The increasing distance of each individual from the knowledge frontier has a positive impact on the number of inventors per team, but decreasing in time for our 1989-2000 window. Next to it, our main variables catching technological knowledge complexity shows a significant role in fostering teamwork as *modus operandi* in technological knowledge production. The share of company as applicants always plays a significant and crucial role in increasing the share of team inventions in a specific technological domain.

III. Theoretical background

III.1 Literature review

The economic literature on knowledge and science devoted a big deal of attention to the existence of teams or, more broadly, to invention networks. Indeed, two methodologically distinguished but complement approaches are in place: the former takes advantage of the growing network analysis tool. The latter, instead, pertains to standard economics, which is theoretical modelling and econometrics. The two literatures partially differ in topics as well.

The network analysis approach

The network analysis approach concentrate mainly onto the knowledge flows topic. It operationalizes collaboration across various typologies of boundaries (but mainly geographical and institutional) with graph theory methodologies, in order to grasp the impact of such a graph in the, typically knowledge, system. Breschi and Lissoni (2009) proposes inventors' mobility as the key engine of knowledge diffusion via generation of co-location networks. Very similarly, Singh (2005) explores the dynamic relationship between collaboration and knowledge flows, distinguishing between intra and extra-regional interactions. Again, his findings express in favour of the predominance of inventors ties over co-location (in space or in a firm) to account for knowledge flows. Fleming, King III and Juda (2007) employs the notion of small-world networks to investigate the structure of regional collaboration networks and their impact on performances. Morescalchi et al. (2015) adds the temporal dimension to the analysis of a variety of collaboration networks pertaining knowledge creation and diffusion, particularly analysing the role of national and institutional borders (see also Miguélez, Moreno and Suriñach (2009).

The standard approach

The more methodologically standard approach, instead, is more various in topics. It embraces the knowledge flows empirical investigation (Sorenson (2005), Sorenson, Rivkin and Fleming (2006), Powell and Grodal (2005), Cassi and Zirulia (2014) for a theoretical model) but also deals directly with teams. On one side, it aims at evidencing the impact of individual characteristics on team performance (Schettino, Sterlacchini and Venturini (2013) on Italian patent data). On the other, it investigates the role of teams in the system's knowledge creation (Bercovitz and Feldman (2011)). On this venue, Crescenzi, Nathan and Rodrìguez-Pose (2016) is a prototypical enquiry on how "physical, organisational, institutional, cognitive, social, and ethnic proximities between inventors shape their collaboration decisions". Cassi and Plunket (2014) bridges the two strands addressing the proximity paradox, i.e. different kinds of proximities have diverse and even antithetical effects on the creation of networks and their subsequent performance (Boschma and Frenken (2010)). From the concept of cognitive proximity (Noteboom et al. (2007)), to that of diversity within interactions, via absorptive capacity (Cohen and Levinthal (1990)) the jump is straightforward. The connection with the idea of diversity in production goes far back to Jacobs' seminal work on the genesis of cities, in 1969. Indeed, the teamwork literature concentrates a lot on the diatribe of specialization vs diversification, homogeneity vs heterogeneity, e.g. building up a further dichotomy of lone inventor vs teams. Singh and Fleming (2010), with an innovative analysis of distribution tails, and Lee et. al. (2015) go in depth into the creativity issue (Harrison and Klein (2007) and Taylor and Greeve (2006) may be two references of a much larger debate on variety and creativity), as Malero and Palomeras (2015) does. This latter, especially, is an interesting paper double-facing Jones (2009) hypothesis, which we will go through later on. There are many other disciplines dealing with the topic of teamwork, generating many interdisciplinary cross-fertilizations. In particular, influences from psychology and management sciences are evident in works on team's internal processes of decision, division of labour, information exchange, but also on creativity (Reagans and Zuckermann (2001), Häussler and Sauermann (2014), Stewart (2006)).

What the literature partially left out is a thorough discussion on the reasons why teams pop up. Indeed, the literature studying networks' dynamics is addressing the topic of which drivers make the network to show off as it is – that is, what kind of proximities or territorially embedded characteristics favours the creation of a link between two individuals. However, they concentrate mostly on the consequences of the network formation and dynamics, for example on knowledge flows.

Such a lack is reasonable, since the possible explanations of teams' formation are many and various, ranging from the individual level to the institutional, and most of them are not really observable. This is also a reason why most of the literature on the topic employs patent data and/or publication data, even though these – especially patent data – are just partially drawing real life dynamics. However, they are widespread used sources in research activities, which makes them, in a sense, reliable in addressing specific issues in knowledge creation dynamics.

We think there is some space, in the literature, for a discussion about the connection between technological knowledge and teamwork. As briefly sketched in the literature review, it has been outlined how much networks play a role in channelling knowledge flows, and how within teams interactions may induce diversity in outcome (even if with contrasting results); but what about knowledge as a determinant of teamwork?

III.2 Theoretical basis

According to Schumpeter (1942), the introduction of innovations is becoming part of a routine within large corporations that are able to organize intra-muros the generation of technological knowledge and the eventual introduction of innovations by means of research and development activities (R&D) conducted directly within internal laboratories organized in teams guided by eminent scientists. Along these lines, the neo-Schumpeterian literature has assumed that R&D was deemed to substituting individual inventors. Schmookler (1957) criticizes, with a very simple and clear analytical procedure, the Schumpeterian idea that invention is less and less an independent, individual activity, but engineers/scientists, and teams or laboratories, dominated one. The empirical analysis grounds on the hypothesis that so called *technologists* are well-educated, experienced and *hired* workers. From both a survey and a wider demographic data analysis, Schmookler shows that

> "During the period [1900-1950] invention changed from an activity overwhelmingly dominated by independent individuals to one less overwhelmingly dominated by business enterprise. The dominance of the latter, qualitative considerations aside, is not as great as commonly assumed and amounts at the outside to no more than three-fourths of the total, measured in terms of either inventors or inventions".

Technologists are defined either as "engineers, chemists, metallurgists, and directors of research and development" or "electrical, mechanical, chemical, industrial, mining and metallurgical engineers; chemists, assayers, and metallurgists", in survey and demographic data respectively. These people are the professionals, according to the author. The connection between technologists and teams is that, from survey data, it emerges that team working is much higher among these professionals than among independent, non-graduated inventors – evidenced from a mismatch between assigned inventions and inventors. No specific explanation is advanced. Indeed, the focus of the discussion is on professionalization of the inventive activity. A spin-off argument relates to teams, mainly because the opposed thesis matches the two perspective: invention made predominantly by engineers within teams. However, the argument *in nuce* is there.

What Schmookler noticed and argued more than fifty years ago is that independent inventors are not disappearing and, most importantly, are not going to extinguish; moreover, inventive activity is indeed becoming a professionals and technicians' affair, but not at the rate popular narratives claim. As to why invention is increasingly transforming from a general attitude of problem solving towards a technical-demanding profession, the only intuition provided by the author regards the increasing dominance of chemical and electrical industries in patenting activity, fields that were and are eminently scientific. In other words, the characteristics of knowledge do not enter into the analysis. Indeed, the author says about the factors inducing the rise of technologists, that "Some of these factors are part of a poorly understood complex - reflecting a positive association between the proportion of technologists, and such things as per capita income, average educational level, and degree of industrialization, all of which tend also to be positively correlated with inventive activity".

We believe that Schmookler treatise of inventive activity guards a fundamental insight: inventors do not act freely in the social space. Institutions and cultural properties matter, and, foremost in inventive activity, companies play a fundamental role. As we will explain in the applied sections, we control for the presence and activity of companies in our analysis, but we do not articulate on the professionalization hypothesis advanced by Schmookler, as it would ask for a specific research project.

Moving forwards, research activities, either academic or commercial, are de facto mostly carried out by teams. Wutchy et al. (2007) formerly brought evidences out of academic and patent data. Citing directly from their paper:

> "For science and engineering, social sciences, and patents, there has been a substantial shift toward collective research. [..]Shifts toward teamwork in science and engineering have been suggested to follow from the increasing scale, complexity, and costs of big science. Surprisingly then, we find an equally strong trend toward teamwork in the social sciences, where these drivers are much less notable. [..]Unlike the other areas of research, single authors still produce over 90% of the papers in the arts and humanities. [..]Lastly, patents also show a rising dominance of teams."

Moreover, they built a very easy to interpret measure of relevance based on a comparison between citations received by individual authors and those received by the team, highlighting "a broad tendency for teams to produce more highly cited work than individual authors" (see also Uzzi et al. (2013)). They do not offer, however, a univocal explanation of why teamwork is arising, and that is reasonable since their fundamental study crosses a huge number of fields and activities. Wutchy et al. (2007) has not been the first work to raise the issue of teams, see for example Solla Price (1963), Zuckerman and Merton (1973) and Adams et al. (2005).

The main, most convincing and successful answer to the question "why do teams pop up" dates back to Jones (2009), in his paper about the "Death of the Renaissance Man". Jones claims a series of results, achieved mainly with a theoretical model of general equilibrium, and corroborated with basic descriptive, but very informative, regression:

> "I show that (i) the age at first invention, serving as a proxy measure for educational attainment; (ii) a measure of specialization; and (iii) team size are all increasing over time at substantial rates. [..]An informal theory of the "burden of knowledge" might suggest these effects. Innovators, when faced with greater knowledge depth, might respond through both longer educational periods and greater specialization. In cross-section, I develop a measure of "knowledge depth" and show that (iv) teamwork and (v) specialization are greater in fields with deeper knowledge. [..](vi) the average age at first invention is strikingly similar across fields and does not vary with the depth of knowledge".

Other than the micro perspective, the model account for macro stylized facts concerning the decreasing productivity of R&D investments, formerly highlighted by Jones (1995) (same surname, different name) in his critique to the "scale effect" of R&D expenditure.

Therefore, according to Jones (2009) we face an increasing teamwork activity, which he operationalize as the number of inventors behind a patent, because the knowledge necessary to reach the frontier and innovate is increasing over time (even if not steadily, indeed he accounts for Kuhn's paradigmatic revolution shrinking the amount of useful knowledge). This is the *burden of knowledge* hypothesis: in order to share the burden on their shoulder, inventors collaborate with peers, activating a process of individual specialization. Even if he does not explicitly state it, in his discourse this process takes a somewhat negative connotation. In fact, even if individual productivity logically rises in teams, given that there would not be any production at individual level otherwise, at the aggregate level the returns from R&D fall, with subsequent effects on growth.

Jones' model is very rich and ambitious, since it tries to explain both micro and macro dynamics with the very same underlying process whose crux is the career choice by potential innovator according to their income possibilities (and expected return). We do not develop such an encompassing model; rather we stick at the direct causal connection between the burden of knowledge and inventors' team size.

In this study, instead of looking at individual data about teams and their performance, we turn our attention to the technology. Much of the inspiration for this research comes from Arthur (2009). In his book, Arthur draws a theory of technology – that is a theory of the internal logics of technologies modification and evolution, and of how this mutation process links with the economy. Arthur's try is not descriptive but normative. Its fascination resides in the attempt to recognize the autonomous characteristics of technology, and only after that, pushing them towards the economic world. Technology stands up as an autonomous object of research, particularly relevant for economic matters because it constitutes the *backbone*, in Arthur's view, of the economy (whereby institutions, firms and markets are the blood and tissues). Such an argument relies on the empirical evidence that the evolution of technology is recursive, combinatory, and complex: it does not just follow a skill-demand process, even if it embraces it (Arthur, Polak (2006)). The structure of the technological space is, then, an interesting object of research per se.

According to this "Arthurian" perspective, technological creation spurs out of recombination, in a recursive fashion, of already existing technological components. This view is particularly severe towards the novel of Schumpeterian "geniality" as pivotal in new knowledge production: inventing something means discovering it in what already exists, states Arthur. This discovery happens through the scanning of the existing possible solutions to the problems impeding the implementation of the new technology. Imagination and mental association lead to combine technological components or (and), sometimes, to exploit natural phenomena with new glint.

The little piece of understanding that we want to add to the literature concerns this technological map, or space, that we operationalize via patent data. Particularly, we look at the existing link between some features of the technological map's structure and what we call *polygamy*. Let us define polygamy as the happening of an interaction during the invention process, which we detect by mean of a co-authorship in patent application data. A polygamous patent is a patent with at least two co-authoring inventors. However, what we look at is the technological space, rather than individual patents. Therefore, in this framework, polygamy is an attribute of a technological domain – that is, a fragment of the technological map. A domain may show off different degrees of polygamy, in the sense that teams or lone inventors variously drive the technological activity that characterizes it. Consequently, we look at the variable depicting the polygamy level of a domain as a variable communicating the *modus operandi* of technological knowledge creation in that domain.

From Jones' viewpoint, knowledge creation is growing complicated. From another perspective, grounded on Arthur (2009), however, generating new knowledge, both scientific and technological, has always been a *complex* task. As such, knowledge creation is a process characterized by radical uncertainty (Antonelli (2011), March and Simon (1958), Schumpeter (1947), Knight (1921)). A way for facing such an obstacle, as well as for many other social experiences qualified by complex foresight horizons, is interaction. Specifically because, in some cases, interaction may get *generative* (Lane et al. (1996) and Lane and Maxfield (1996)):

> "a relationship that can induce changes in the way the participants see their world and act in it, and even give rise to new entities"

Interaction between inventors assumes, then, the additional property of mitigating the individuals' inability to undertake successfully complex tasks, as new technological knowledge might increasingly be. The concept of polygamy conveys the idea that inventors' interactions intensity may be a strategic reaction to the complexity of technological knowledge creation in specific domains.

Summarizing, we try to observe a series of relationships between variables, which we set up, describing some structural properties of technological knowledge. As we will describe in the variables section, some of these variables account for properties of knowledge *per se*, while others, as polygamy, refers to organizational aspects of knowledge production in a domain. The reason why we develop this framework is not to state against other dominant approaches, like Jones' one, but to complement them with a wider description of technology *per se*. Sure enough, the technological map is not something directly observable in data, as are patents or publications, but it is something that we operationalize. This is why we will stick to very simple measure, in order to avoid ambiguity in interpretation of the results as much as possible.

Nevertheless, the idea of a knowledge map in empirical research is not new at all. An increasing number of authors adopts the idea that knowledge is not a homogeneous good, rather it is cumulative, heterogeneous, and in particular the outcome of a recombinatory process (Weitzman (1998), Fleming (2001), Fleming and Sorenson (2001), Sorenson, Rivkin and Fleming (2006), Saviotti (2004, 2007) or Frenken and Nuvolari (2004) among many). This literature describes such a particular good as a relational entity, eventually with a network structure, whose nodes may be individuals, patents, as well as firms and regions. Ultimately, however, each actor finds a place according to the properties of its knowledge base. In all of these works, and many others, the above-mentioned characteristics of knowledge as a network map find a confirmation. However, at the very best of our knowledge, none of them tries directly to analyse the map, thus we are the first to try.

IV. Data

IV.1 Choice of the data

We are interested in technology, which is an aspect of reality appropriately embodied by patents, beyond the full set of limitations using such source entails. However, the longstanding tradition of patent analysis guarantees the reliability of the source, and furthermore it remains the only highly available extensive repository of data about technological knowledge generation, even if not generally comprehensive. We acknowledge the limits of patent data in approximating innovation dynamics (Durio, Carota and Guerzoni (2015)) but we believe it is still a positive option for the study of invention. Moreover, we aim at drawing a picture intimately related to the *inside* of technology, rather than technology *in general*. Other than very detailed case studies on specific technologies as in Frenken, Nuvolari (2004), patent data provides the greatly valuable resource of an objective and generalizable technological classification of inventions (the International Patent Classification provided by WIPO), coherent throughout time and space, as the expanding literature on the knowledge base demonstrates. We make use of such a classification to dig into the technological knowledge stock, taking advantage of the network format the patent databases take on.

Our dataset is drawn from a PATSTAT-based database, the CRIOS DB from Bocconi (Coffano and Tarasconi (2014)), which provides augmented disambiguated data about patent activity. We limit the analysis to German patents only, which is the country with the highest number of patents. We chose the German subsample mainly because of technical reasons: we want to observe directly the characteristics of technological knowledge, thus we need a data source the more populated as possible in terms of patents for technological area (we will describe our methodology in the next sections). Germany is one of the largest producer of patents in Europe, which is the coverage of our dataset. German data, however, show trends consistent with those of the PATSTAT as a whole (the same has been verified for other major countries).

Figure 1 shows, for both left and right panel, respectively the whole PATSTAT and the Germany subset, in the first graph the trend in patent production, while in the second and third we plot the number of patents signed by teams and the average team size. Cutting off the first five or six year of the database, which are the first years of PATSTAT implementation, a quick look clearly communicates that, along with increasing absolute patents production, teamwork increases as well both in absolute number and in team size. We do observe a sharp decrease at the end of the series, but it is known that the last years in every updated version of PATSTAT must be taken carefully because of a lag in the filing process. Germany subset's trends are very consistent with the whole PATSTAT picture. Figure 2 displays general statistics about our selected source.

In this section, we will briefly describe the logic behind the aggregation process of raw patent data that brought to compile the final dataset we use. Moreover, we will describe the main variables we will use in this analysis.

As abovementioned, we want to observe the structural dynamic of the technological map, or space. According to Arthur (2009), such a space is made of technologies that, in the process of recombination, associate one with one another at different degrees. The so-built network can usefully be described in two levels: domains, which are going to be our unit of observation, and components, making up domains. In Arthur (2009)'s picture, technological domains *emerge* from the contiguity of technological components in real technologies; they are not fixed, rather vary over time (see Breitzman and Thomas (2015) or Érdi et al. (2012) for practical exercises reproducing a very similar dynamic with patent data, even if they do not directly refer to Arthur's work).

The appropriate unit of observation to catch a technological domain dynamics is the 3-digit IPC technological class. In fact, the 3-digit aggregation specification is the last step before IPC technological classes grow too much in numbers: from 121 3-digit IPCs they goes to 634 at 4-digits, 1239 at 5-digits and so on. Since we want them to represent domains, and domains are baskets of components, we think this might be a good approximation. The operation that follows is very simple: we assign each patent to as many domains as the 3-digit IPC classes it references to. We decided not to pick up a weighted or fractional assignation rule because we miss the ratio behind it. Then, we compute some indexes or simply count measures out of this patent assignation. There are two kinds of categories in which each variable falls: those related to *technological knowledge* and those related to *patents'* features. The knowledge-related measures build upon the fact that we assign to a domain each patent endowed with the set of 4-digit IPC classes it references to, which will be the technological *components* in our framework. Some knowledge-related measure reflects the position of the domain in the map; others, instead, mirror the internal composition of the domain. We posit this approach as an appropriate and viable methodology to open up a new stream of enquiries on the technological knowledge stock composition.

This paper concentrates on the relationships between some variables directly related to technological knowledge, and a patent's feature, which is polygamy. We define polygamy as the share of patents whose number of inventors is strictly bigger than one, for each domain in each point in time. As for the technological knowledge variables, we created a series of them.

Burden of knowledge (BK): the count of the backward citation tree's branches up to the third generation (Jones 2009).

Team size (Tsize): the domain average number of inventor per team.

Connectivity (Conn): it is virtually the same as an unweighted degree centrality. We computed the weighted measure as well, meaning that it becomes dependent by the number of patent produced in a certain domain, but we avoid its use so to not mix up qualitative effects with size effects.

Interdependence (Int): The notion of interdependence resembles that of 'coupling', or relates to (the inverse) of modularity (Fleming and Sorenson 2001)). The understanding of this variable is not straightforward since it heavily relies on the research on adaptive biology in Kauffman (1993). In our framework, it conveys the degree to which a change in the internal structure of a domain may be effective in creating a new useful combination of technological knowledge. The measure of interdependence is the best approximation we could think of for operationalizing technological knowledge complexity. However, we built other knowledge related measures precisely in order to grasp a width of phenomena.

Entropy (Entr): There is an increasing literature developing decomposable entropy indexes as measures of variety (Kraft, Quatraro and Saviotti (2014), Saviotti (1996)). However, we stick to the measure of so-called total variety. Technically, "the entropy index refers to the degree of randomness in the choice of technological designs as reflected by the skewness of a distribution" (Frenken and Nuvolari (2004)). We build the measure at the components level (*Class Entropy*), considering only *native* components of a domain (see Appendix).

Company share (CompS): the share of applications whose applicant is a company, against those owned by private individuals.

Age: the number of years since the domain first occurred in an application's IPC technological reference list.

Growth rate (Growth): the growth rate of applications production.

In the Appendix a more technical description of *BK*, *Conn*, *Int*, *Entr*.

V. Model and Methodology

V.1 Formulation of the model and descriptive statistics

Since what we are interested in is the interaction between knowledge complexity and polygamy, we frame our model as:

$$polygamy_{Ct} = \alpha_{Ct} + \boldsymbol{\beta}\, complexity_{Ct} + \boldsymbol{\gamma}\, controls_{Ct} + \varepsilon_{Ct}$$

The dependent variable, then, is the share of polygamous applications in a given domain at each point in time. C*omplexity* is a vector of knowledge-related covariates. Indeed, the concept of knowledge complexity is not an empirically well-defined one. There is a lot of literature on its theoretic (Arthur (2015), Antonelli (2008 and 2013)) which underlines the importance of properties like non-linearity, out-of-equilibrium dynamics, unpredictability of process outcomes and its recursive nature. However, on empirical ground, to operationalize such concepts is not straightforward. The most successful direct and explicit attempts have been in Hidalgo and Hausmann (2009)'s Method of Reflection, and Fleming and Sorenson (2001)'s Interdependence measures (drawn from Kaufmann (1993)), and both of them originated quite a bit of follow-ups. On a parallel binary, Saviotti (1988)'s adaptation of the decomposable informational entropy index to knowledge bases initiated a stream of works assimilating the idea of knowledge variety with that of complexity (Frenken and Nuvolari (2004)). None of them pretends to be comprehensive of the systemic properties of knowledge as a process. Therefore, we picked from the literature those metrics that were more suitable for our setting and easier to interpret, and we selected more than one of them, in order to convey a bundle of related but multifaceted information. Hence, the *complexity* vector is made of *interdependence*, *entropy* and *connectivity* covariates. Interdependence provides glints on the internal recombinant structure of the domain; entropy will account for the specialization patterns among the domain's native components; connectivity tells the story of the external relationships between domains.

Within the *controls* vector, instead, are a series of variable we enhance as significant in explaining for cross-sectional as well as temporal differences among observations, but that are not directly related to the structure of knowledge. They are *company share*, *age, application production's growth rate* and *burden of knowledge*. Inventors are not free to move in ether and decide at any point in time and space whether to collaborate or not, and with whom. Their endeavour, instead, is radically embedded in a social space (Antonelli and Scellato (2013)) and in an institutional environment or constraint (Cassi and Plunket (2014), Boschma (2005), Ponds et. al. (2007)). Figure 4 shows that most of patent production happens within institutions (let them be companies or public research agencies, Mazzucato (2015)), but what is interesting is that there is a clear-cut dichotomy between the tendency to polygamy in companies and that to lone invention among private individuals.

Indeed, we observe an increasing attitude towards polygamy among company-owned applications, along with a constant dominance of lone inventors among those applications signed by privates (see Figure 3 in the Appendix). This is just a picture on the surface; however, the fact that companies play a major role in polygamous patents production is a fact that deserves more attention (Nagaoka and Owan (2011)). In this work, however, we limit to control for the different concentration levels of company-applicant in a specific domain's patents production. We want to control, moreover, for the fact that domains in expansion ask for a surplus in labour force and labour intensity, due to the hurry an agent in the market for knowledge has to appropriate the most of the profit from a new technological gale. A domain that is not expanding any more, or is doing it at a very low rate, regardless of its structural characteristics, may constitute a quiet comfort zone where invention comes at a slower pace – that is, the team's productivity gain does not appear strictly necessary. On the same vain, domains might grow fast simply when they are new, or young, irrespective of their structural stature. These reasons lead us to control for *application production's growth rate* and *age*, respectively. Lastly, we check for the burden of knowledge, as computed in the former exercise. The reason why we plug-in this control is that we firmly believe it plays a pivotal role in driving technological knowledge production, as highlighted by Jones (2009). We are interested in spotting complementarities with our selected indicator for domain's knowledge complexity.

Our final dataset is made of 121 unit of observation (technological domains) assembling a strongly balanced panel of 1452 observations for 12 years, from 1989 to 2000.

VI. Econometric analysis

VI.1 Test on the burden of knowledge hypothesis

As a first exercise, we ask the newly built dataset some questions about the burden of knowledge hypothesis. This exercise mirrors two intentions: first, it introduces in a framework where domains are the subject of analysis and a patents' feature interrelates to a technological knowledge characteristic, but on a familiar theoretical ground. Second, it validates the use of such a dataset and level of analysis, given it proves successful in reproducing Jones' famous conclusions.

In Figure 5, we plot the time trends of the empirical correlation between team size and our approximate measure of the burden of knowledge. First, we note that the 1-digit IPC class reaching the highest level in team size (C) is one of the two "deepest" or "heaviest" (*C* and *G*). Looking at the differential between the initial and the final level of the series, every one of them grew. However, we did not smooth the curves in order to give the idea of the process' non-linearity.

The scatterplot of Figure 6 evidences a positive relationship between team size and the burden of knowledge. Nevertheless, it is not constant over time. It seems that the "positive" correlation is diminishing with time and getting vaguer – the scatter becomes sparser and sparser with blue and purple dots, corresponding to years from 1995 to 2000.

The very same intuition comes from a simple regression of team size against knowledge burden, in a cross section for three different years, namely 1990, 1994 and 2000.

Table 2. OLS regression

OLS regressions, corrected SEs

	Dependent variable:		
	team_size_i		
	1990	1994	2000
	(1)	(2)	(3)
burden knowledge	0.0545***	0.0227***	0.0102***
	(0.0101)	(0.0051)	(0.0024)
back citations	-0.2774***	-0.1935**	-0.1800**
	(0.0656)	(0.0861)	(0.0826)
constant	2.4283***	2.5445***	2.6755***
	(0.2533)	(0.3868)	(0.3895)
Observations	121	121	121
R2	0.2996	0.1883	0.2264
Adjusted R2	0.2877	0.1746	0.2132
Residual Std. Error	0.5037	0.5140	0.5930
F Statistic	25.2392***	13.6897***	17.2631***
Note:	*p<0.1; **p<0.05; ***p<0.01		

The estimated coefficient of the independent variable diminishes over time. With the mean number of backward citations, we control for directionality of the relationships between the knowledge burden and team size. Indeed, being *burden of knowledge* a tree size measure of backward citations, one might argue that larger teams tend to cite more. The average number of direct backward citations checks for this eventuality and contradicts it (Jones (2009)). We check for multicollinearity with the VIF test and we exclude its eventuality. The reason is that applications with very similar first generation citation patterns differs a lot in the subsequent generation schema. A pooled-OLS regression with year dummies on the panel structure gives the very same results. Concerning the reasons why the influence of the burden of knowledge on team size is shrinking, we do not have a clear explanation.

VI.2 Test of the interpretative model

In the remaining of the paper, we describe the econometric strategy we employed to give some empirical evidence of the discourses presented insofar. In this version of the working paper, we present only final estimation tables for space constraints.

A part from ordinary OLS, as a second step, we run FE regressions in order to control for fixed, time invariant unobserved effects, which is kind of a gold standard in econometrics when there is not complete certainty of having the full set of necessary controls available – that is never the case. The Hausman test results are indeed in favour of FE over RE. However, a recently published econometric paper, Bell and Jones (2015), argues against FE gold standard in economics. They state that the FE estimator predominance in empirical studies stems from a misunderstanding of the potential of RE, as well as from a misuse of the Hausman test – that is, indeed, a test of similarity between within and between effects, thus losing reliability in providing advice for choosing between FE and RE. Their argument is convincing and we think that their approach gifts us of the opportunity to improve the reliability of our estimation. Indeed, their basic critic stands on the argument that the within estimator clears out any source of variance at the higher level (considering longitudinal data with the broadest definition of hierarchical data, i.e. data made of more than one nested level). This alone might be an important drawback, because time-invariant information can represent a precious insight into the process analysed. Moreover, it may be the case that higher-level (time-invariant) entities affects lower-level (time-varying) entities. In synthesis, what they do is to correctly and explicitly account for the, so called, “heterogeneity bias” (Li (2011)) in the linear model plugging in time-invariant covariates (Mundlak (1978)), thereafter estimating the augmented equation with a RE estimator. They are rather precise in stating that “There is nothing FE-like about the model at all—it is an RE model with additional time-invariant predictors.” We adopt this methodology, including in our model specification yearly deviations from the individual domain average value of each covariate (within effect) and the domain specific means for each covariate (between effect) (Mindruta et al. (2015)). We cannot state that this is the best choice for our estimation, but we firmly believe it adds reliability to the research. After all, ‘‘all models are wrong; the practical question is how wrong do they have to be to not be useful’’ (Box and Draper (1987), Bell and Jones (2015)).

Looking at residuals plots against fitted values, it emerges that every single estimation of ours is affected by heteroscedasticity. The Breusch–Pagan test confirms this observation. Therefore, we compute robust variance-covariance matrices as Arellano suggests (Arellano (1987), Woolridge (2002)). Again, from residual plots I did not find any evidence that the functional form of the model should be different. Residuals plots against fitted values as well as partial residuals plots are not perfectly randomly distribute, but I cannot guess a specific augmentation from these outcomes. For panel estimations we perform, in addition, the Levin–Lin–Chu Unit-root test for stationarity, even if out panel is rather short; in any case, every predictor turns out stationary. Lastly, we look for time serial correlation with the Durbin–Watson test.

VI.3 Results

Table 2. Within-Between RE

PANEL REGRESSIONS, WITHIN-BETWEEN RE		
	Dependent variable:	
	Polygamy	
	Within	Between
INTERDEP	1.0036***	2.9890
	(.2843)	(1.9573)
BURDEN KNOW	.0380	.1785**
	(.0365)	(.0824)
ENTROPY CLASS	-3.1540	-5.0429*
	(4.3813)	(2.8318)
CONNECT	.1321	.0967**
	(.1141)	(.0420)
COMPANY SHARE	46.6385***	103.7574***
	(10.8077)	(9.6178)
AGE	.0824	-5.4098***
	(.2492)	(1.9439)
GROWTH	.0464**	-.5658
	(.0211)	(.3974)
CONSTANT	46.1477	
	(35.5320)	
OBSERVATIONS	1,452	
R^2	.3016	
ADJUSTED R^2	.2984	
F STATISTIC	44.3160***	

Table 2 reports the results of the within-between RE estimator, following Bell and Jones (2015)'s theoretical framework and Mindruta et al. (2015)'s applied analysis. First note that *interdependence* and *burden of knowledge* do exert their influence at the different levels, time and cross-section respectively. The other two measures of knowledge complexity, *connectivity* and *entropy*, instead, are significant only between domains, as it is for *age*. The only other variable significant only within domains is *growth*; *company share*, instead, is significant in both within and between estimators.

Our main variable of interest, that is *interdep*, is positive and significant, as well as *company share*. This latter variable, other than being highly significant, contributes enormously to the R^2. *Connectivity* loses explanatory power across time within cross-sectional groups, meaning that its domain-specific variation does not provide insights on the domain-specific variation of polygamy, but only across domains at the same point in time, across time-groups (on average). The same happens for *age*, meaning that being a newcomer raises probabilities of spotting polygamous dynamics (negative sign), but this age-effect does not change across time (it is "better" to be young than young*er*). As for the *growth rate*, it seems that growing faster highly correlates with polygamy only across time within groups. Finally, two striking outcomes are those for *burden of knowledge* and *entropy class*. The burden of knowledge has mainly, in our framework, a cross-sectional effect on polygamy, losing explanation power across time. In a sense, this traces the result of the replication exercise in the beginning of this paper, where the burden of knowledge was losing power, time passing, on determining team size. What instead is far more puzzling is the result for the entropy measure. Recall that *entropy class* computes the informational entropy index on the native components of each specific domain, i.e. low entropy means highly specialized production along set dimensions. Hence, the negative sign indicates that when a domain specializes more in one of its prototypical dimension, we observe higher polygamy rates. First, note that we do not qualify polygamy in terms of diversity within the team. However, the positive sign of *connectivity* brings about the idea that domains positioned at the middle of a multifaceted and heterogeneous network in the technological space calls in team working. Nevertheless, the two are not necessarily clashing, because *connectivity* looks outward, *entropy*, instead, inward the domain. This is as to say that domains that are on the way of specialization asks for many complementary components in order to implement ideas in such a way that they are *patentable*. Such a reactive process of connections enlargement in face of domain's specialization co-occurs with an increasing non-exchangeability among components involved in the technological, creative process (which is another definition for interdependence). We cannot test this hypothesis within this estimation framework, although it would be interesting to, but it provides a tentative explanation for the coefficients' signs.

These results suggest that structural differences across technological domains in terms of their knowledge structure are important in explaining the different levels of polygamy. The knowledge burden is a characteristic moving slowly through time, so it is reasonable that its changing across years does not affect in a significant way teamwork activities. Similarly, specialization/diversification patterns are typically path-dependent trajectories. As for connectivity, we would have expected an impact at the within level too; we should enlarge the time window of the analysis in order to check whether our expectations are disregarded because of a slow motion of *connectivity*. These results do not necessarily tell a story of a fixed technological knowledge map, rather they are compatible with a depiction of technological structural changes via paradigmatic/scientific shifts or revolutions (Kuhn (1962), Jones (2009), Arthur (2009)).

Our main variable of interest, *interdependence*, instead, plays a crucial rule in the internal evolution of each technological domain's polygamy. At an increasing complexity of the internal structure of a technological domain, the local search for improvements of knowledge components bundles – output of a recombination process along and across domains' dimensions – in the technological knowledge space gets more arduous and demanding. Simultaneously the outcome of such a process becomes even more uncertain. Our results seem to suggest that, coherently with the *generative relationship* framework, the technological milieu answers to this shift with an increasing intensity of polygamy as a *modus operandi* of technological knowledge production. Instead, and somewhat surprisingly, such *interdependence* level does not appear to be significantly correlated with variation of polygamy between domains, where the other structural characteristics of the knowledge structure dominates, as aforementioned.

Shortcomings

We do see two problems, or missing implementations, we want to address soon. First, even if the Durbin-Watson test for serial correlation does not reject the null for the within-between RE estimation, the Breusch-Godfrey/Wooldridge test (slightly) does. Hence, we propose to investigate thoroughly the issue. As mentioned, the SE are robust to heteroscedasticity and serial correlation via correction of the variance-covariance matrix (Arellano (1987), Woolridge (2002)). Second, we guess that polygamy has a path dependent nature. However, plugging-in the lagged dependent variable would make the analysis to trespass in dynamic panel models, which, because of the Nickell bias (Nickell (1981)), asks for specific methodologies for estimation, e.g. GMM.

VII. Conclusions

In this piece of work, we try to dig a little into the correlation dynamics between teamwork activities and some structural properties of technological knowledge. In order to do so, we distinguish from the literature in a number of facets:

- i) our unit of observation is the technology itself, expressed as a technological domain whose characteristics change in time (Arthur (2009));
- ii) our focus is on interactive invention, that we defined as *polygamy*, as a *modus operandi* of technological knowledge production - meaning that we do not observe the internal dynamics of teams, rather the internal dynamics of a domain's organization in knowledge production;
- iii) we try to use simultaneously different measures of knowledge structure to better approximate the concept of complexity;
- iv) we test on our framework the burden of knowledge hypothesis (Jones (2009)) and we check how it fits with the other measures of the knowledge structure.

In particular, we try to evidence the relationship between technological knowledge complexity and polygamy as a reaction for new recombinant knowledge creation. Indeed, we find that, together with the importance of the knowledge burden in explaining differences across technological domains, the pillar of our depiction of complexity – that is, interdependence (Fleming and Sorenson (2001)) – exerts a significant influence on the variation of the tendency to invent in team, within domains.

Another puzzling, but interesting result is that regarding another dimension of complexity: the specialization/variety along the native dimensions of a domain. We suggest that the negative value of our entropy index may reflect a specific dynamics: as the domains G06 ('Computing; Calculating; Counting') specializes its inventive production in, say, G06E ('Optical Computing Device'), disregarding most of all of the other G06 dimensions, the need for complementary inputs may arise in order to implement the invention. Then, we may observe G06 to connect with, say, C related domains ('Chemistry') in order to generate a viable invention with some chances to break successfully the path for becoming an innovation. There we have *connectivity* and *entropy class* significantly pushing in different directions. Teams, or, better and more generally, interactive creative processes, are at corner of this *complex* dynamics. Indeed, the story we try to tell moves a little away from the narrative of teams as the outcome of specialization trends: here, interaction is the only and easiest solution for facing search problems in a technological landscape getting rougher, for staying on the technological trajectory of specialization but retrieving the necessary heterogeneous knowledge for solving complex recombinational puzzles. Reversal, across domains we see that those domains whose variety *inward* is substantial do not connect (or tends to connect less) with other domains; in other words, when there is more variety inside, there is no urgency to look at it outside.

References

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). “Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999”. *Research policy*, *34*(3), 259-285.

Antonelli, C. (2013). "The Economic Complexity of Innovation as a Creative Response," *Department of Economics and Statistics Cognetti de Martiis. Working Papers 201326*, University of Turin.

Antonelli, C. (2011). “The Economic Complexity of Technological Change: Knowledge Interactions and Path Dependence”. In Antonelli, C. (ed.) *Handbook on the Economic Complexity of Technological Change.* 1-62. Cheltenham*:* Edward Elgar.

Antonelli, C. (2009). “The economics of innovation: from the classical legacies to the economics of complexity”. *Economics of Innovation and New Technology*, 18:7, 611-646.

Antonelli, C. (2008). “Localized technological change. Towards the economics of complexity”. Routledge, London.

Antonelli, C. (2003). “Knowledge Complementarity and Fungeability: Implications for Regional Strategy”. *Regional Studies*, 37:6-7, 595-606.

Antonelli, C., Scellato, G. (2013). “Complexity and Innovation: Social Interactions and Firm Level Productivity Growth”. *Journal of Evolutionary Economics,* 23 (1): 77-96.

Arellano, M. (1987). “Computing Robust Standard Errors for Within Group Estimators”. *Oxford Bulletin of Economics and Statistics*, 49, 431-434.

Arthur, B. (2015). “*Complexity and the Economy*”, Oxford University Press.

Arthur, W. B. (2007). “The structure of invention”. *Research policy*, *36*(2), 274-287.

Arthur, W.B. (1999). “Complexity and the Economy”. *Science*, 284, 107-109.

Arthur, W.B., Durlauf, S., Lane, D. (eds.) (1997). “*The Economy as an Evolving Complex System II”*. Reading: Addison-Wesley.

Arthur, W.B., Polak, W. (2006) “The Evolution of Technology in a Simple Computer Model,” *Complexity*, 11, 5.

Bell, A., Jones, K. (2015). “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data”. *Political Science Research and Methods*, 3, pp 133-153.

Bercovitz, J., Feldman, M. (2011). “The mechanisms of collaboration in inventive teams: Composition, social networks, and geography”. *Research Policy*, *40*(1), 81-93.

Boschma, R. (2005). “Proximity and innovation: a critical assessment”. *Regional studies*, *39*(1), 61-74.

Boschma, R., Frenken, K. (2010). “The spatial evolution of innovation networks. A proximity perspective”. *The handbook of evolutionary economic geography*, 120-135.

Breitzman, A., Thomas, P. (2015). “The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems”. *Research policy*, *44*(1), 195-205.

Breschi, S., Lissoni, F. (2009). “Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows”. *Journal of Economic Geography*, lbp008.

Carota, C., Durio, A., Guerzoni, M. (2014). “An Application of Graphical Models to the Innobarometer Survey: A Map of Firms’ Innovative Behaviour”. *Department of Economics and Statistics “Cognetti de Martiis" Working Paper Series*.

Cassi, L., Plunket, A. (2014). “Proximity, network formation and inventive performance: in search of the proximity paradox”. *The Annals of Regional Science*, *53*(2), 395-422.

Cassi, L., Zirulia, L. (2012). “Friends and rivals: modelling the social relations of inventors”. *The Rimini Centre for Economic Analysis Working Paper Series*, No. 39_12.

Coffano, M., Tarasconi, G. (2014). “CRIOS-Patstat database: sources, contents and access rules”. *Center for Research on Innovation, Organization and Strategy, CRIOS Working Paper*.

Cohen, W. M., Levinthal, D. A. (1990). “Absorptive capacity: A new perspective on learning and innovation”. *Administrative science quarterly*, 128-152.

Crescenzi, R., Nathan, M., Rodríguez-Pose, A. (2016). “Do inventors talk to strangers? On proximity and collaborative knowledge creation”. *Research Policy*, *45*(1), 177-194.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., Zalányi, L. (2013). “Prediction of emerging technologies based on analysis of the US patent citation network.” *Scientometrics*, *95*(1), 225-242.

Fleming, L. (2001). “Recombinant uncertainty in technological search”. Management Science 47, 117–132.

Fleming, L., King III, C., Juda, A. I. (2007). “Small worlds and regional innovation”. *Organization Science*, *18*(6), 938-954.

Fleming, L., Sorenson, O. (2001). “Technology as a complex adaptive system: evidence from patent data”. *Research policy*, *30*(7), 1019-1039.

Fleming, L., Sorenson, O. (2004). “Science as a map in technological search”. *Strategic Management Journal*, *25*(8-9), 909-928.

Frenken, K., Nuvolari, A. (2004). “Entropy statistics as a framework to analyse technological evolution”. *Applied evolutionary economics and complex systems*, 95-132.

Harrison, D. A., Klein, K. J. (2007). "What's the difference? Diversity constructs as separation, variety, or disparity in organizations". *Academy of management review*, *32*(4), 1199-1228.

Häussler, C., Sauermann, H. (2014). "The anatomy of teams: division of labor and allocation of credit in collaborative knowledge production". *Available at SSRN 2434327*.

Hildago, C. A., Hausmann, R. (2009). "The Building Blocks of Economic Complexity". *Proceedings of the National Academy of Science*, 106 (26): 10570-75.

Jones, B. F. (2009). "The burden of knowledge and the 'death of the renaissance man': Is innovation getting harder?". *The Review of Economic Studies*, *76*(1), 283-317.

Jones, C. I. (1995). "R&D-based models of economic growth". *Journal of political Economy*, 759-784.

Kauffman, S. (1993). "The Origins of Order". Oxford University Press, New York.

Knight, F. H. (1921). "Risk, Uncertainty, and Profit". Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Krafft, J., Quatraro, F., Saviotti, P. P. (2014). "The dynamics of knowledge-intensive sectors' knowledge base: Evidence from Biotechnology and Telecommunications". *Industry and Innovation*, 21(3), 215-242.

Kuhn, T.S. (1970[1962]), "The Structure of Scientific Revolutions", Chicago: University of Chicago Press.

Lane, D., Maxfield, R. (2005). "Ontological Uncertainty and Innovation". *Journal of Evolutionary Economics* 15 (1): 3-50.

Lane, D., Maxfield, R. (1997). "Foresight Complexity and Strategy". In *The Economy as an Evolving Complex System II* edited by Arthur, W.B., Durlauf, S.N., Lane, D.A., 169-198. Santa Fè: Westview Press.

Lane, D., Maxfield, R. (1996). "Strategy under complexity: Fostering generative relationships". *Long Range Planning*, 29(2), 215-231.

Lee, Y. N., Walsh, J. P., Wang, J. (2015). "Creativity in scientific teams: Unpacking novelty and impact". *Research Policy*, *44*(3), 684-697.

Li, G. (1985). "Robust regression in Exploring Data Tables, Trends, and Shapes", ed. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Wiley.

Li, X. (2011). "Approaches to Modelling Heterogeneity in Longitudinal Studies", Victoria University.

March, J.C., Simon, H. (1958). "*Organizations"*. New York: John Wiley and Sons.

Mazzucato, M. (2015). "The entrepreneurial state: Debunking public vs. private sector myths". Anthem Press.

Melero, E., Palomeras, N. (2015). "The Renaissance Man is not dead! The role of generalists in teams of inventors". *Research Policy*, 44(1), 154-167.

Merton, R. K., Zuckerman, H. (1973). “Institutionalized patterns of evaluation in science”. *The sociology of science*, 460-496.

Miguélez, E., Moreno, R., Suriñach, J. (2010). “Inventors on the move: Tracing inventors' mobility and its spatial distribution”. *Papers in Regional Science*, *89*(2), 251-274.

Mindruta, D., Albino Pimentel, J., Hoetker, G. (2015). “Institutions and the direction of innovative search: change and persistence between and within countries”. *HEC Paris Research Paper No. SPE-2015-1095*.

Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., Riccaboni, M. (2015). “The evolution of networks of innovators within and across borders: Evidence from patent data”. *Research Policy*, 44(3), 651-668.

Mundlak, Y. (1978). “Pooling of Time-series and Cross-section Data”. *Econometrica*, 46(1):69–85.

Nagaoka, S., Owan, H. (2011). “Exploring the Sources of Firm-level Scale Economies in R&D: Complementary assets, internal and external knowledge inflows, and inventor team size”. *RIETI Discussion Paper Series* (No. 11-E, p. 038).

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., Van den Oord, A. (2007). “Optimal cognitive distance and absorptive capacity”. *Research policy*, 36(7), 1016-1034.

Ponds, R., Van Oort, F., Frenken, K. (2007). “The geographical and institutional proximity of research collaboration”. *Papers in regional science*, 86(3), 423-443.

Powell, W. W., Grodal, S. (2005). “Networks of innovators”. *The Oxford handbook of innovation*, 56-85.

Reagans, R., Zuckerman, E. W. (2001). “Networks, diversity, and productivity: The social capital of corporate R&D teams”. *Organization science*, 12(4), 502-517.

Saviotti, P.P. (2007). “On the dynamics of generation and utilisation of knowledge: The local character of knowledge”. *Structural Change and Economic Dynamics*, 18, 387-408.

Saviotti, P.P. (2004). “Considerations about the production and utilization of knowledge”. *Journal of Institutional and Theoretical Economics*, 160, 100-121.

Saviotti, P.P. (1996), “Technological Evolution, Variety and the Economy”, Edward Elgar: Cheltenham, U.K. and Northampton, Mass.

Saviotti, P.P. (1988). “Information, variety and entropy in technoeconomic development”. *Research Policy*, 17, 89-103.

Schettino, F., Sterlacchini, A., Venturini, F. (2013). “Inventive productivity and patent quality: Evidence from Italian inventors”. *Journal of Policy Modeling*, 35(6), 1043-1056.

Schmookler, J. (1957). “Inventors past and present”. *The Review of Economics and Statistics*, 321-333.

Schumpeter, J. A. (1947). “The creative response in economic history”, *Journal of Economic History* 7(2):149-159.

Schumpeter, J.A. (1942), “Capitalism, Socialism and Democracy”. New York: Harper and Brothers.

Singh, J. (2005). “Collaborative Networks as Determinants of Knowledge Diffusion Patterns”. *Management Science*, 51(5):756-770.

Singh, J., Fleming, L. (2010). “Lone inventors as sources of breakthroughs: Myth or reality?”. *Management Science*, 56(1), 41-56.

Solla Price (de), D. J., Beaver, D. (1966). “Collaboration in an invisible college”. *American psychologist*, 21(11), 1011.

Sorenson, O. (2005). “Social Networks, Informational Complexity and Industrial Geography” in *The Role of Labour Mobility and Informal Networks for Knowledge Transfer* (pp. 79-96). Springer US.

Sorenson, O., Rivkin, J. W., Fleming, L. (2006). “Complexity, networks and knowledge flow”. *Research policy*, *35*(7), 994-1017.

Stewart, G. L. (2006). “A meta-analytic review of relationships between team design features and team performance”. *Journal of management*, 32(1), 29-55.

Taylor, A., Greve, H. R. (2006). “Superman or the fantastic four? Knowledge combination and experience in innovative teams”. *Academy of Management Journal*, 49(4), 723-740.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B. (2013). “Atypical combinations and scientific impact”. *Science*, 342(6157), 468-472.

Weitzman, M. L. (1998). “Recombinant growth”. *Quarterly journal of Economics*, 331-360.

Wooldridge, J. (2002). “Econometric Analysis of Cross-Section and Panel Data”. MIT press.

Wuchty, S., Jones, B. F., Uzzi, B. (2007). “The increasing dominance of teams in production of knowledge”. *Science*, 316(5827), 1036-1039.

Appendix

Here on a precise description of the variables used in the empirical analysis. When useful for clarity, we add some notation: *D* is the technological domain, *C* is the technological component, *y* is time (yearly), *i* is the patent.

Burden of knowledge (BK): domain average. Count of the backward citation tree’s branches up to the third generation of each patent. Redundant patents in each tree are counted once.

$$BK_{Dy} = mean(BK_i)_{Dy}$$

Connectivity (Conn): first, we computed the three-digit IPCs co-occurrences matrix *maty* for each year *y* from the beginning of PATSTAT (1977) to our last year. Note that *maty* is a symmetric matrix, thus we consider only the upper triangular matrix out of it. Then we count the non-zero element for each row, which are signaling the existence of a co-occurrence of two technological domains.

Interdependence (Int): in order to compute the value for *Interdependence* we must calculate the *Ease of Recombination* for each technological component, i.e. each IPC at 4-digits (Fleming and Sorenson (2001)). Given the degree of *Connectivity* of a component, i.e. the number of domains it is linked to via assigned patents, a domain is easier to recombine with other domains in a new patent if the number of previous patents assigned to it (*stock*) is lower, such that

$$ease_{Dy} = \frac{Conn_{Dy}}{stock_{Dy}}$$

Then we invert the average value of $ease_{Dy}$ scores for the domain they belongs, in order to compute the value of *Interdependence*.

$$Int_{Dy} = \frac{1}{mean(ease_{Cy})_{Dy}}$$

Entropy (Entr): we use the Shannon Index, which measures the uncertainty of predicting the value of a new draw, considered the distribution of the previous draws' values in a vector. Another interpretation of such a measure is the concentration over a probability distribution. We compute two measures: *Class Entropy* and *Co-occurrence Entropy*, referring respectively to a vector of native components and a vector of co-occurrences of native components. With *native*, we mean the 4-digits IPC sub-classes of a 3-digits IPC class as assigned by WIPO.

The maximum entropy for that vector divides each value of the Shannon Index so calculated, given that the maximum entropy is the natural logarithm of the number of elements (or dimensions). Therefore, if $H_{Shannon}$ is the Shannon Index, v is the vector of components or co-occurences and n_v the count of element of v,

$$Entr_{Dy} = \frac{H_{Shannon}(v_{Dy})}{\log(n_v)}$$

In the end, as explained above, we use only the first measure, *Class Entropy*.

Figures

Figure 1. PATSTAT (left) and Germany (right)

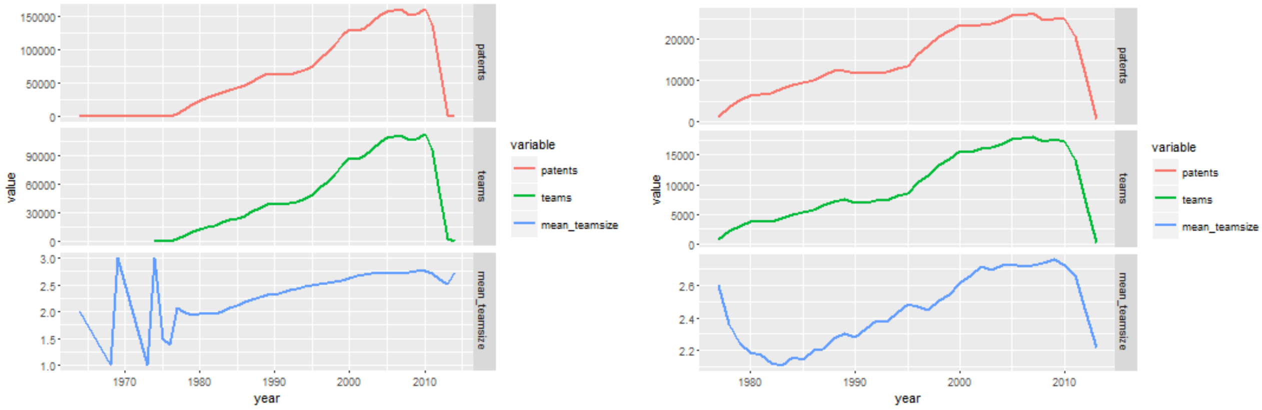
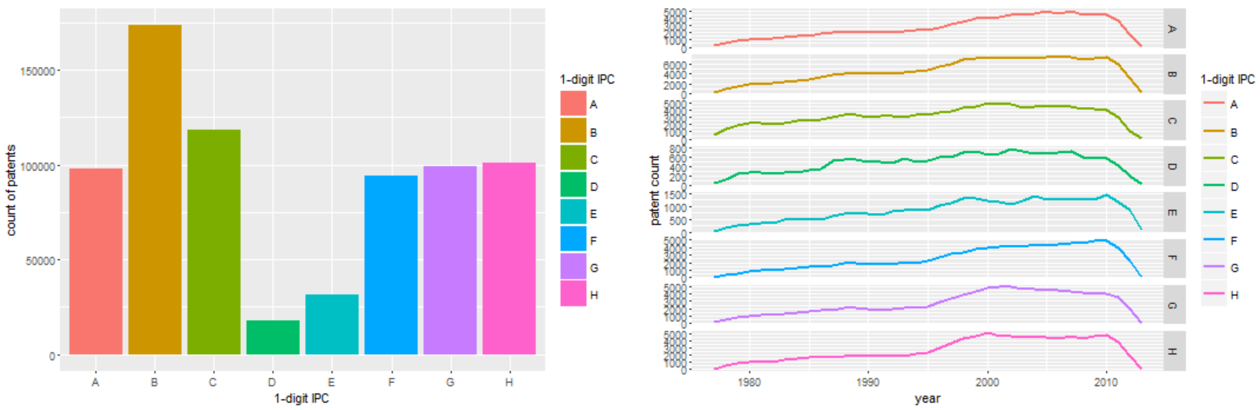


Figure 2. Distribution and time trends



Notes: 'Performing operations; Transporting' (B); 'Textiles; Papers' (D); 'Fixed Construction' (E); 'Chemistry; Metallurgy' (C); 'Mechanical Engineering; Lighting; Heating; Weapons; Blasting' (F); 'Physics' (G); 'Electricity' (H); 'Human Necessities' (A).

Figure 3

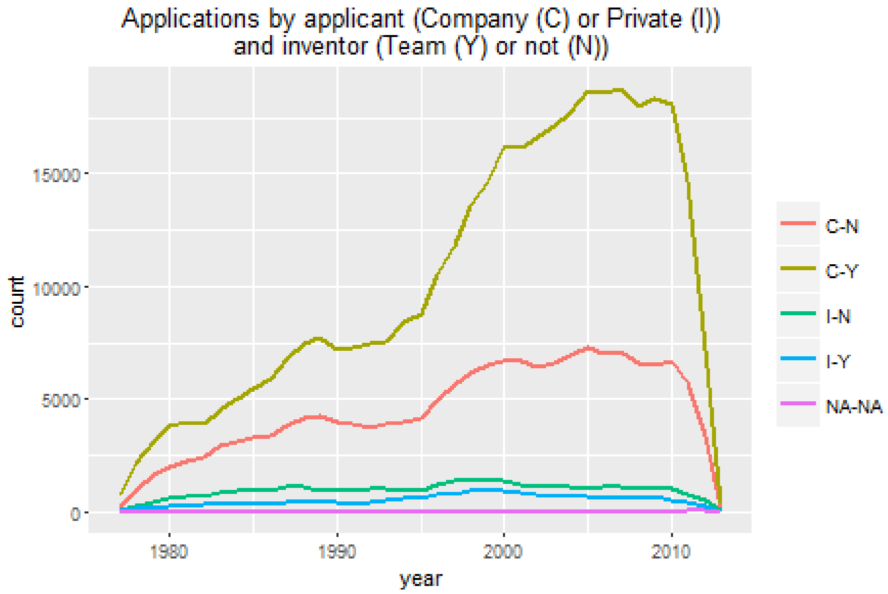


Figure 4

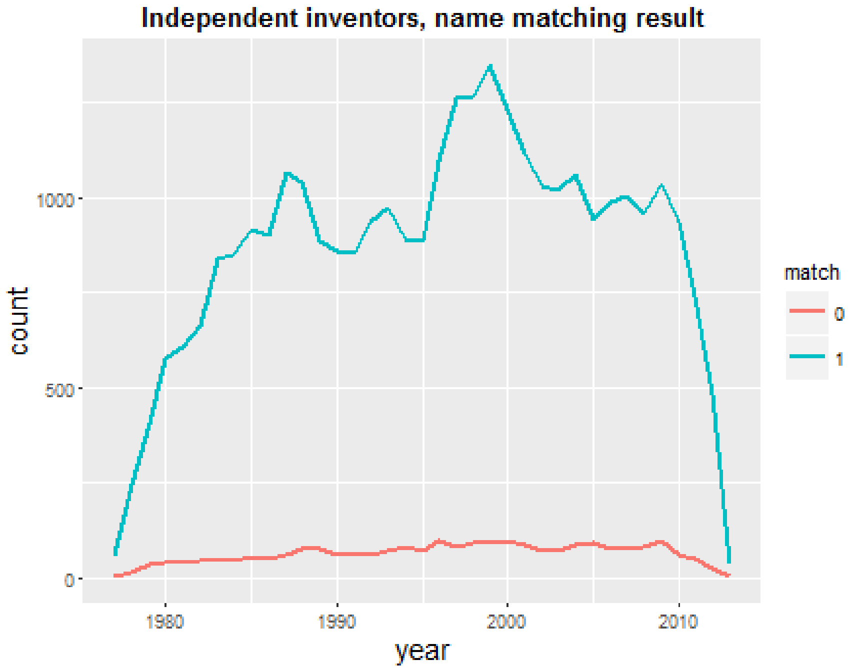


Figure 5. Burden of knowledge

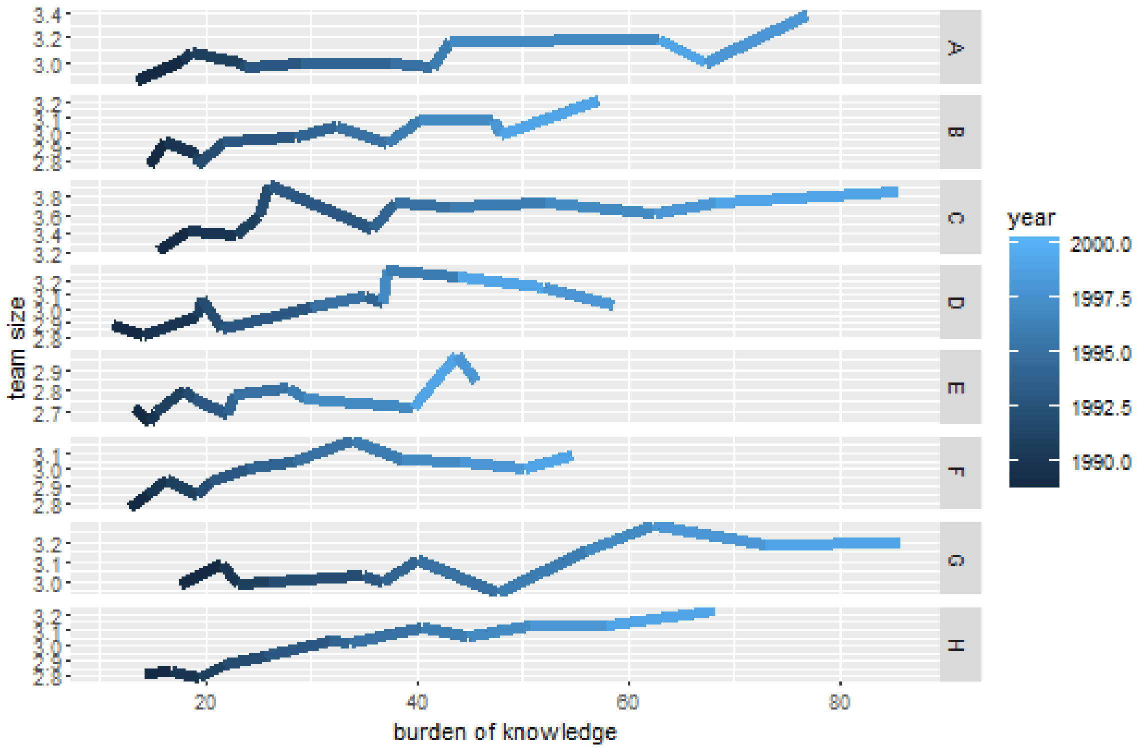


Figure 6. Burden of knowledge and team size