

# POLITECNICO DI TORINO

Corso di Laurea Magistrale  
in Ingegneria Gestionale

**ATTIVITÀ DI INNOVAZIONE TECNOLOGICA E RICADUTE DI  
CONOSCENZA: ANALISI EMPIRICA DELLE RETI DI COLLABORAZIONE  
BREVETTUALE NEI SISTEMI LOCALI DEL LAVORO ITALIANI**



## **Relatori**

Prof. Luigi Buzzacchi  
Prof. Antonio De Marco

## **Candidato**

Stefano Laterza

Dicembre 2022

## Sommario

La letteratura economica mostra chiaramente come i brevetti possono non solo essere considerati come una buona approssimazione dell'attività inventiva, ma anche un mezzo per studiare i flussi di conoscenza all'interno della società. Pertanto, a partire dal filone di ricerca che studia le variabili che favoriscono la produzione brevettuale, questo elaborato ha l'obiettivo di valutare un duplice aspetto: da un lato si costruisce la rete delle collaborazioni brevettuali studiandone le principali proprietà; dall'altro si analizza la varietà della struttura industriale del territorio scelto. Utilizzando i dati dei brevetti depositati presso l'Ufficio Italiano Brevetti e Marchi negli ultimi venti anni, un primo contributo di questo lavoro consiste nella scelta dell'area geografica oggetto di studio: la partizione funzionale dei sistemi locali del lavoro italiani. Si analizza quindi la rete di co-assegnazioni brevettuali, da cui appare già chiaro quali siano le aree del Paese maggiormente innovative e che rappresenta un primo risultato del lavoro di tesi. Infine, si realizza un modello di regressione lineare che valuti come le variabili di rete e la struttura industriale possano influenzare l'attività inventiva italiana.

# Indice

<b>Elenco delle figure</b>	4
<b>Elenco delle tabelle</b>	5
<b>1 Introduzione</b>	7
<b>2 Rassegna della letteratura</b>	9
<b>3 Dati</b>	15
3.1 Diritti di proprietà intellettuale . . . . .	15
3.1.1 Brevetti . . . . .	15
3.2 Aree funzionali . . . . .	20
3.2.1 Sistemi locali del lavoro . . . . .	22
3.3 Dati empirici . . . . .	26
3.4 Il campione analizzato . . . . .	27
3.4.1 Analisi temporale . . . . .	28
3.4.2 Analisi geografica . . . . .	30
3.4.3 Analisi per classi tecnologiche . . . . .	32
<b>4 Metodologia</b>	37
4.1 Le co-assegnazioni brevettuali . . . . .	37
4.2 Costruzione delle variabili . . . . .	42
4.2.1 Brevetti pro-capite . . . . .	42
4.2.2 Densità . . . . .	42
4.2.3 Coesione interna . . . . .	42
4.2.4 Transitività . . . . .	43
4.2.5 Quota di richiedenti rispetto alla popolazione . . . . .	43
4.2.6 Collegamenti esterni . . . . .	44
4.2.7 Struttura industriale: la <i>variety</i> . . . . .	45
4.2.8 Variabili di controllo . . . . .	46
4.3 Le proprietà della rete . . . . .	47

<b>5</b>	<b>Analisi econometrica</b>	<b>57</b>
5.1	Modello di regressione lineare . . . . .	57
5.2	Specificazione del modello . . . . .	61
5.3	Analisi descrittive . . . . .	62
5.4	Risultati . . . . .	64
<b>6</b>	<b>Conclusione e sviluppi futuri</b>	<b>69</b>
	<b>Riferimenti bibliografici</b>	<b>73</b>
<b>A</b>	<b>Esempio di brevetto</b>	<b>79</b>
<b>B</b>	<b>Pulizia dati (Parte I)</b>	<b>81</b>
<b>C</b>	<b>Pulizia dati (Parte II)</b>	<b>85</b>
<b>D</b>	<b>Classificazione IPC: sezione A</b>	<b>89</b>
<b>E</b>	<b>Classificazione IPC: sezione B</b>	<b>91</b>
<b>F</b>	<b>Algoritmo di creazione della <i>Edge list</i></b>	<b>93</b>
<b>G</b>	<b>Algoritmo di costruzione dei regressori</b>	<b>95</b>
<b>H</b>	<b>Analisi di robustezza dei modelli econometrici</b>	<b>101</b>

# Elenco delle figure

3.1	Invenzioni, innovazioni, brevetti . . . . .	19
3.2	Configurazione territoriale dei sistemi locali del lavoro nel 2011 . . . . .	25
3.3	Andamento temporale domande brevettuali vs brevetti approvati . . . . .	29
3.4	I sll con attività brevettuale più intensa . . . . .	31
3.5	I sll con attività brevettuale pro capite più intensa . . . . .	31
3.6	Produzione brevettuale delle 8 sezioni (classificazione IPC) . . . . .	33
3.7	Brevetti associati alle principali classi (sezioni A e B, classificazione IPC) . . . . .	33
3.8	Produzione brevettuale rispetto alle classi tecnologiche . . . . .	35
4.1	Esempio di grafo non orientato (6 nodi, 7 archi) . . . . .	39
4.2	I sistemi locali con il maggior numero di collaborazioni brevettuali . . . . .	39
4.3	Brevetti vs collaborazioni . . . . .	40
4.4	Andamento temporale delle collaborazioni brevettuali . . . . .	40
4.5	Analisi per sezioni tecnologiche delle collaborazioni brevettuali . . . . .	41
4.6	<i>Open triad</i> e <i>closed triad</i> a confronto . . . . .	43
4.7	Analisi temporale della densità della rete dei principali sll . . . . .	49
4.8	Analisi temporale del coefficiente assortativo a tre diversi livelli spaziali . . . . .	50
4.9	Rete con <i>support</i> pari a 1, ma coefficiente di <i>clustering</i> molto minore di 1 . . . . .	52
4.10	Analisi temporale del livello massimo di coesione interna dei principali sll . . . . .	53
4.11	Rete delle collaborazioni brevettuali nel 2018 (filtro sul grado) . . . . .	55
4.12	Rete delle collaborazioni brevettuali nel 2018 (filtro su grado e sll di Milano) . . . . .	56
5.1	La generica retta di regressione lineare . . . . .	58
5.2	Distribuzione degli errori attorno alla retta di regressione . . . . .	58
B.1	Analisi temporale dei valori nulli per il codice fiscale dei richiedenti . . . . .	81
B.2	Analisi spaziale dei valori nulli per il codice fiscale dei richiedenti . . . . .	82
B.3	Analisi per classi tecnologiche dei valori nulli per il codice fiscale dei richiedenti . . . . .	83
C.1	Analisi temporale dei valori nulli per codice del sll . . . . .	85
C.2	Confronto distribuzione brevettuale nel tempo.1 . . . . .	86
C.3	Confronto distribuzione brevettuale nel tempo.2 . . . . .	86
C.4	Confronto distribuzione brevettuale nel tempo.3 . . . . .	87
C.5	Confronto distribuzione brevettuale nel tempo.4 . . . . .	87

# Elenco delle tabelle

3.1	Principi per la costruzione dei SL a livello europeo . . . . .	21
3.2	Sistemi locali del lavoro per regione . . . . .	23
3.3	I sistemi locali del lavoro di grandi dimensioni . . . . .	24
3.4	Numero di brevetti per regione . . . . .	30
4.1	Densità della rete per ogni anno e sll disponibili: un estratto . . . . .	48
4.2	Densità della rete dei principali sll . . . . .	49
4.3	Misura di transitività dei principali sll . . . . .	51
5.1	Statistiche descrittive . . . . .	62
5.2	Matrice di correlazione . . . . .	62
5.3	Modelli di regressione . . . . .	64
H.1	Modelli di regressione . . . . .	101



# Capitolo 1

## Introduzione

In un contesto economico sempre più globale, in rapida evoluzione e fortemente competitivo, appare logico immaginare che la *conoscenza*, nel suo significato più generale e in quanto risorsa scarsa, possa garantire di trarre un vantaggio competitivo notevole. A partire da questa intuizione, nella seconda metà del secolo scorso, l'economista e saggista Peter Ducker coniò la locuzione *economia della conoscenza*, facendo così riferimento all'utilizzo delle informazioni per generare valore, con particolare attenzione a natura, creazione, diffusione, trasformazione, trasferimento e utilizzo della *conoscenza* in ogni suo aspetto.

Formalizzata questa assunzione, la letteratura economica ha cercato di concretizzare questi concetti, altrimenti validi solamente ad un livello di astrazione molto elevato. Innanzitutto si è individuato ciò che avrebbe potuto *mappare* al meglio la *conoscenza* di un territorio: l'attività inventiva, in ogni sua forma. Di qui il tentativo (vano) di trovare una grandezza quantitativa (univoca) che potesse identificare in modo esaustivo il concetto ampio di attività inventiva.

Accertato che non esista una misura assolutamente migliore delle altre e che invece la scelta debba essere puntualmente adattata al contesto e alla domanda di ricerca che si intende studiare, un particolare filone di ricerca si è interrogato sulle proprietà di queste *informazioni* e, in particolare, sul modo in cui queste siano trasmesse da un agente economico a un altro. L'intuizione secondo cui è possibile che ciò accada (i.e., *esternalità di conoscenza*), insieme ad un contesto macroeconomico sempre più interconnesso, hanno ispirato centinaia di articoli scientifici e contribuito affinché questa branca di ricerca proliferasse.

Questo elaborato, pertanto, guidato in particolare dalla letteratura economica degli ultimi vent'anni, ha l'obiettivo di adattare le ipotesi, la metodologia e i risultati empirici già documentati in letteratura utilizzando un dataset fornito dall'Ufficio Italiano Brevetti e Marchi (UIBM). L'obiettivo del lavoro di tesi consiste nell'analisi delle principali determinanti dell'attività inventiva italiana e dei *flussi di conoscenza* che la regolano, scegliendo come unità geografica di analisi la partizione funzionale dei sistemi locali del lavoro (sll). L'elaborato di tesi è così strutturato: nel Capitolo 2 si mostra una puntuale rassegna della letteratura, dalla nascita dell'espressione *economia della conoscenza* agli articoli di riferimento più recenti; nel Capitolo 3, dopo una breve descrizione dei diritti di proprietà



intellettuale e dei sll, si mostrano alcune analisi descrittive del campione analizzato; il Capitolo 4 ha uno scopo duplice: da un lato si studiano le principali proprietà delle reti costruite, dall'altro si presentano i regressori utilizzati nei modelli e il modo in cui sono stati calcolati; nel Capitolo 5 si commentano i risultati dell'analisi econometrica; l'elaborato si chiude con un capitolo conclusivo, all'interno del quale si sintetizzano i risultati raggiunti, i limiti della ricerca e i punti ancora aperti.

## Capitolo 2

# Rassegna della letteratura

Con l'espressione *economia della conoscenza*, resa celebre da Peter Ducker nel suo libro *The Age of Discontinuity* (1969), si fa riferimento ad un sistema economico all'interno del quale la conoscenza rappresenta una risorsa fondamentale e in grado di generare un vantaggio competitivo notevole per un Paese e, più in particolare, per un'impresa. Nel suo significato più generale, la conoscenza può essere definita come l'insieme delle informazioni e delle abilità che gli individui utilizzano per risolvere problemi, prendere decisioni e apprendere informazioni dall'esterno (Doring & Schnellenbach, 2006). In un primo momento, la letteratura neoclassica ha modellato la conoscenza in maniera statica, definendola semplicemente come un fattore di *input* all'interno della generica funzione di produzione  $Y = f(L)$ <sup>1</sup>. In particolare, secondo la visione neoclassica, definito il livello di conoscenza, considerata una variabile esogena, è possibile conseguire una crescita del reddito pro capite solo attraverso un incremento del fattore di *input*, grazie al quale sarà possibile migliorare l'efficienza tecnologica del processo produttivo in questione. Successivamente, a partire dal contributo proposto da Arrow (1971), si è sviluppata una nuova teoria economica che tenta di rendere endogena la crescita di conoscenza, in contrasto con la teoria neoclassica. Infatti, l'idea secondo cui sia possibile raggiungere un progresso tecnologico senza incrementare i fattori di input della funzione di produzione, ma solamente tramite meccanismi di *learning by doing*, consente di spiegare l'assenza di produttività marginale decrescente, aspetto cardine della teoria neoclassica. Tuttavia, appare evidente come la presenza di una crescita endogena di conoscenza non sia necessariamente correlata a fenomeni di *esternalità di conoscenza*. Nel linguaggio economico il termine *esternalità* è utilizzato per descrivere gli effetti, positivi o negativi, che un agente economico (e.g., individuo, impresa, etc.) esercita sulla produzione o sul benessere di un altro, al di fuori delle transazioni di mercato. Nel caso in esame, pertanto, qualora si assuma che la conoscenza non sia condivisa con altri attori economici, i benefici maturati grazie ad economie di apprendimento riguarderanno solamente le imprese coinvolte da questi meccanismi; diversamente, l'effetto positivo derivante da economie esterne coinvolgerà una realtà industriale più ampia. Nel

---

<sup>1</sup>Y rappresenta l'*output* della funzione di produzione, la quantità massima di prodotto ottenibile; L indica la quantità dei singoli fattori di *input* utilizzati come, ad esempio, il livello di conoscenza.

primo caso, la conoscenza è modellata come un bene privato e in quanto tale destinato ad un utilizzo esclusivo di alcuni individui; questo approccio al problema assume quindi che differenze nel grado di sviluppo di un territorio siano da imputare a diversi livelli di conoscenza iniziali. Il secondo modello, invece, definisce la conoscenza un bene con caratteristiche di non rivalità e non escludibilità e, pertanto, un bene pubblico (Romer, 1986). Entrambe le assunzioni risultano alquanto estreme e sembra invece sensato affermare che esistono delle esternalità di conoscenza, ma che i processi di diffusione risultano incompleti e limitati nel tempo e nello spazio (Jaffe et al., 1993). Appare a questo punto di estrema importanza capire quanto questi meccanismi siano diffusi e in modo particolare quali le variabili coinvolte. Certamente, la più recente letteratura evidenzia in modo chiaro come l'approccio neoclassico sia ormai superato: non è corretto soffermarsi solamente su scelte ottimali individuali, ma risulta necessario valutare gli agenti in un contesto più ampio, influenzato da decisioni e condizioni terze. Più chiaramente, i risultati individuali non dipendono solamente dal proprio livello di conoscenza tecnologica, ma anche da quanto l'agente sia inserito in un contesto regionale e industriale che gli consenta di accedere al livello di conoscenza esistente (Doring & Schnellenbach, 2006).

Un primo approccio al problema proposto dalla letteratura consiste nel misurare quanto sia estesa l'esternalità di conoscenza e quindi il suo impatto sull'attività inventiva dell'area territoriale scelta come unità di analisi. È importante studiare il livello di innovazione di un territorio poiché risulta strettamente legato alle *performance* economiche e al vantaggio competitivo che si generano in quell'area. L'idea alla base è che imprese relativamente vicine possano beneficiare maggiormente di meccanismi di *spillover* rispetto a realtà industriali geograficamente più distanti. La difficoltà che emerge in analisi di questo tipo consiste principalmente nel distinguere in maniera inequivocabile fenomeni di *clustering* conseguenza di *spillover* da quelli dovuti a meccanismi di *sharing* o *matching*<sup>2</sup> o, ancora, già esistenti a causa di vantaggi naturali tipici di alcune aree geografiche. Articolo pioniero in questo senso è quello di Jaffe et al. (1993), che utilizzano le citazioni brevettuali e la loro distribuzione spaziale come approssimazione dei flussi di conoscenza esistenti. In particolare, da un lato si calcola la probabilità che un brevetto ne citi un altro dello stesso periodo, appartenente alla medesima categoria tecnologica e soprattutto nella stessa area geografica; dall'altro si calcola la probabilità che due brevetti abbiano in comune la stessa area geografica. Quando la prima probabilità risulta in maniera significativa superiore alla seconda, si può affermare che esistano *spillover* di conoscenza localizzati in una precisa area geografica e eventualmente valutare come cambia questa probabilità all'aumentare della distanza. Appare evidente come in questa tipologia di analisi sia fondamentale scegliere correttamente l'unità territoriale di riferimento.

Una seconda branca della letteratura scientifica si occupa invece di stimare, tramite opportuni modelli di regressione, in che modo l'attività inventiva (i.e., variabile dipendente) di una certa area geografica sia influenzata da alcuni fattori, in base agli obiettivi della domanda di ricerca. In particolare, scelta l'unità geografica di riferimento, in letteratura è presente un dibattito su quale variabile dipendente utilizzare nel modello. Una delle scelte

---

<sup>2</sup>Si tratta di processi alternativi e responsabili di fenomeni di *clustering*; per una trattazione più approfondita si veda Duranton e Puga (2004)

più comuni in tal senso consiste nell'utilizzare le spese di R&D (misura di *input*) come misura dell'attività inventiva. Tuttavia, la difficoltà nel reperire questi dati a livello locale, la ricerca di maggiore granularità e la volontà di utilizzare una misura diretta dei risultati del processo innovativo, ha indotto molti autori a utilizzare la produzione brevettuale in alternativa alle spese in R&D. Alcuni autori si sono interrogati su quale delle due misure risulti più appropriata. Secondo Pavitt (1982), entrambe rappresentano una buona approssimazione della *performace* innovativa, ma sono caratterizzate da alcune imperfezioni. Ad esempio, alcuni lavori che avevano utilizzato i brevetti (misura di *output*) come *proxy* dell'attività inventiva avevano erroneamente constatato che la produzione innovativa fosse negativamente influenzata dalla dimensione dell'impresa. Utilizzare invece le spese in R&D aveva condotto a risultati opposti. Risultati così diversi sono giustificati dal fatto che le imprese di piccole dimensioni mostrano un tasso di brevettazione per unità di spesa in R&D superiore rispetto alle imprese più grandi. Allo stesso modo, Pavitt (1982) mostra come sia opportuno impostare queste riflessioni tenendo anche in considerazione quale sia il settore industriale oggetto di analisi. Il suo studio mostra infatti come nel settore chimico, elettrico ed elettronico utilizzare come indicatore le spese in R&D mostri risultati migliori e, allo stesso modo, utilizzando statistiche basate sui brevetti per la produzione di macchinari non elettrici e per l'industria metallurgica. Pertanto, nonostante questi studi assumano una correlazione positiva tra queste due misure, riconoscono nel contempo che una non sia la perfetta alternativa dell'altra. Infatti, è possibile che un'attività di ricerca e sviluppo non sia protetta da un brevetto e si preferisca quindi mantenere il segreto industriale o che si tratti di una invenzione non brevettabile: in questo caso utilizzare le spese in R&D come misura è certamente più indicato. D'altro canto, come evidenziato in OECD (2009), può accadere che l'attività inventiva sia descritta solamente dai diritti di proprietà intellettuale, che verosimilmente garantiscono alcuni vantaggi. Innanzitutto, i brevetti descrivono nel dettaglio alcuni settori che altrimenti non disporrebbero di una documentazione adeguata. Inoltre, i dati brevettuali sono pubblici, facilmente accessibili, strutturati (o semi-strutturati), ricchi di informazioni dettagliate (e.g., descrizione, settore tecnologico, indirizzo di residenza degli inventori, richiedenti, etc.). Infine, le operazioni preliminari svolte dagli uffici competenti consentono di accedere a queste informazioni ad un costo relativamente basso, se comparato con forme alternative di raccolta dati (e.g. sondaggi, questionari, interviste).

Tuttavia, le perplessità circa l'effettivo valore del brevetto depositato non mancano, in quanto c'è evidenza empirica sul fatto che la distribuzione del valore associato ai brevetti è molto asimmetrica: può variare in base al Paese in cui viene rilasciato e al settore tecnologico di riferimento. Tipicamente, in ambito farmaceutico vi sono regole specifiche che è necessario osservare, date da circostanze tecniche tipiche del settore. Una di queste particolarità è rappresentata dalla possibilità di ottenere un prolungamento della protezione brevettuale (*Supplementary Protection Certificate*) al fine di recuperare il tempo che necessariamente intercorre tra il deposito della domanda di brevetto e l'effettiva commercializzazione del farmaco. In questo ambito il brevetto rappresenta uno strumento molto valido e che ha il duplice obiettivo di stimolare la produzione innovativa del settore e ripagare l'investimento in ricerca e sviluppo delle aziende farmaceutiche garantendo loro l'esclusivo sfruttamento commerciale e industriale dell'invenzione. Accanto alle peculiarità di settore, il tema della protezione della proprietà intellettuale è ulteriormente

complicato dalla possibilità di scegliere degli strumenti alternativi ai brevetti, come indicato nel paragrafo 3.1. Gambardella et al. (2008), ad esempio, analizzano un ampio dataset di brevetti per studiarne l'effettivo valore.<sup>3</sup> I risultati empirici mostrano che il valore medio di questi brevetti è di circa 3 milioni di euro, mentre la mediana è pari a 400 mila euro. Questa discrepanza è giustificata dal fatto che la media risente delle code della distribuzione, confermandone il carattere altamente distorto. Alla luce di queste evidenze e con la consapevolezza che non esista una misura perfetta (Pavitt (1982)) si sceglie qui di utilizzare i brevetti come misura di riferimento. Il brevetto costituisce infatti il diritto di escludere i terzi dall'utilizzo di una nuova invenzione, che sia nuova, ovvero non presente nello stato della tecnica esistente e non ovvia, pertanto un risultato che una persona con ordinaria abilità in quel settore non potrebbe facilmente dedurre (Katila, 2000). Inoltre, poiché è possibile brevettare durante diverse fasi del processo di ricerca e sviluppo, è possibile considerare i brevetti non solo come *output* delle spese in R&D, ma anche come *input* dell'innovazione: questo definisce il brevetto come *ponte* tra i dati di R&D e quelli della *performance* innovativa (OECD, 2009). Esistono quindi le condizioni per considerare i brevetti come una buona approssimazione dei processi innovativi e, in ultima analisi, "migliore delle spese in R&D" (Griliches et al., 1986).

In questa direzione, un ampio filone di ricerca analizza il rapporto tra la rete sociale di un territorio e la sua produzione innovativa. In particolare, questa rete è costruita utilizzando le collaborazioni brevettuali: in alcuni casi, infatti, il deposito della domanda di brevetto per invenzione è effettuato da più individui (persone fisiche o imprese), appartenenti o meno alla stessa area geografica. I nodi del grafo rappresentano solitamente gli inventori o le imprese, mentre ogni arco descrive la singola collaborazione brevettuale. A partire dal *network* così realizzato, si costruiscono le principali variabili di rete e quindi il loro effetto sulla variabile dipendente. Recentemente, gli studi che utilizzano la teoria dei grafi per descrivere l'impatto che le proprietà della rete hanno su un *cluster* sono in forte crescita (Breschi & Malerba, 2005). Ciascun modello focalizza la propria analisi su alcune delle possibili micro e macro proprietà di una rete (per una tassonomia completa si faccia riferimento a Jackson et al. (2017)). Uno dei primi contributi alla discussione è stato quello di Lobo e Strumsky (2008), che analizzando le co-assegnazioni brevettuali, si domandano cosa influenzi maggiormente la produzione brevettuale delle aree metropolitane negli Stati Uniti. In particolare, la concentrazione degli inventori (*agglomeration economies*) risulta avere un impatto maggiore rispetto alla struttura della rete, che comunque rimane nel complesso rilevante. Gli autori sottolineano che i risultati parzialmente sorprendenti circa la scarsa influenza delle variabili di rete sull'attività inventiva necessitano di alcune precisazioni. Innanzitutto, la rete di collaborazioni brevettuali non considera tutti i contributi al processo innovativo: ci sono ad esempio dipendenti che partecipano alle attività di R&D, anche se non presenti nella lista degli inventori di brevetti. Inoltre, l'analisi

---

<sup>3</sup>Si tratta di dati raccolti tramite un questionario e relativi a più di 9000 brevetti con anno di priorità 1993-1997 e rilasciati dall'EPO, l'Ufficio Brevetti Europeo. Le informazioni associate a ciascun brevetto includono l'indirizzo di residenza del primo inventore presente in elenco e residente in Danimarca, Francia, Germania, Ungheria, Italia, Olanda, Regno Unito o Spagna.

condotta non distingue l'attività brevettuale in base alla complessità tecnologica. Suddividere la popolazione di brevetti tenendo in considerazione questo fattore consentirebbe di valutare reti diverse e quindi risultati potenzialmente più incoraggianti. Più recentemente, Breschi e Lenzi (2016) hanno condotto uno studio focalizzandosi, in aggiunta alla concentrazione degli inventori, anche su una tipica variabile di rete, definita in letteratura *social proximity*. L'analisi empirica conferma che questa variabile, modellata con una misura di transitività, influenza positivamente l'attività inventiva delle 331 città degli Stati Uniti prese in considerazione. Inoltre, rispetto ai precedenti contributi, gli autori prestano maggiore attenzione ai collegamenti esterni (i.e., collaborazioni brevettuali con almeno un nodo esterno al territorio scelto come unità di analisi), verificando anche in questo caso una relazione positiva con la variabile dipendente del modello di regressione.

Una secondo filone di letteratura, invece, indaga quale struttura industriale contribuisca maggiormente all'innovazione di un territorio. In questo senso, di fondamentale importanza risultano i lavori di Marshall (1890), Arrow (1971) e Romer (1990), i quali identificano una tipologia di esternalità di conoscenza (di MAR, da Marshall-Arrow-Romer) favorita da contesti in cui prevale la specializzazione di molte imprese nello stesso settore. Pertanto, secondo questo modello, gli *spillover* si verificano all'interno del medesimo settore industriale, ad esempio grazie a un dipendente che sceglie di cambiare impresa, pur continuando a lavorare nello stesso ambito industriale. Di parere opposto è invece Jacobs (2016), secondo la quale è invece possibile che la conoscenza originata da settori diversi possa essere condivisa e che sia proprio la diversità il fattore determinante affinché siano presenti dei flussi di conoscenza. In particolare, secondo l'antropologa statunitense, questa teoria giustifica per quale motivo così tante persone scelgano di vivere in città e non in aree rurali, che richiedono un costo della vita decisamente inferiore. Il vantaggio, in termini di costi-benefici, è da ritrovare nella possibilità di apprendere conoscenze diverse, che in ultima analisi consentono di migliorare la propria posizione lavorativa e quindi il proprio salario. Riprendendo questo dibattito, nel lavoro di Frenken et al. (2007) si afferma che gli *spillover di conoscenza* presentati da Jacobs sono verosimilmente presenti principalmente tra settori che condividono tecnologie simili, in modo da poter comunicare efficacemente. Con l'obiettivo di misurare questa affinità, viene introdotto il concetto di *related variety*, che d'ora in avanti si traduce con il termine *varietà correlata*, definita come l'entropia dell'occupazione tra i settori industriali a 5 cifre, all'interno di ciascuna classe a 2 cifre secondo la *Standard Industrial Classification*<sup>4</sup>. Inoltre, con l'obiettivo di valutare l'impatto della diversità tra settori sulla crescita economica, viene aggiunta una seconda variabile, definita *unrelated variety* (da qui in avanti *varietà non correlata* e calcolata come l'entropia ad un minor livello di dettaglio (2 cifre, secondo la medesima classificazione). La recente letteratura indica un effetto positivo della varietà correlata sulla crescita economica e occupazionale a livello regionale (Boschma & Iammarino, 2009; Frenken et al., 2007), e, in particolare, sulla produzione brevettuale (Aarstad et al., 2016; Castaldi et al.,

---

<sup>4</sup>Introdotta nel 1937, è stata sostituita nel 1997 dalla classificazione NAICS (*North American Industry Classification System*), ma è tuttora utilizzata da alcune agenzie governative. I codici SIC sono numeri di 4 cifre (cosiddetti *digits*), che raggruppano i settori in una gerarchia di sottocategorie per progressivo dettaglio della tipologia di prodotto o produzione.

2015). Meno chiaro risulta essere il ruolo della varietà non correlata sull'attività inventiva, ma in ultima analisi prevale l'idea secondo cui anche questa variabile abbia un impatto positivo sulla produzione brevettuale (Tavassoli & Carbonara, 2014) e più in particolare sul numero dei cosiddetti *superstar patents* (Castaldi et al., 2015), ovvero brevetti che risultano di notevole interesse per lo sviluppo di una nuova tecnologia, proprietà osservabile ad esempio attraverso il numero di citazioni ricevute.

Negli ultimi anni sono emersi i primi contributi che tentano di valutare congiuntamente i due aspetti e, in particolare, l'interazione tra le variabili di rete e la varietà industriale, secondo quanto definito sopra. Risultati incoraggianti giungono ad esempio dai lavori empirici di Ebersberger et al. (2014) e Rigby (2018), che rendono robusta la teoria secondo cui l'interazione rafforza l'attività inventiva del territorio. In questa direzione, questo lavoro di tesi è ispirato dall'articolo di Innocenti et al. (2020), che ha l'obiettivo di posizionarsi in uno spazio ancora poco esplorato, ma altamente significativo per capire le dinamiche che influenzano i flussi di conoscenza e la crescita economica territoriale. La loro ricerca sfrutta i dati delle collaborazioni brevettuali tra imprese nel periodo 2001-2016, utilizzando tre intervalli di 5 anni e scegliendo come unità territoriale le province italiane. L'analisi empirica consiste nella costruzione di un modello di regressione che valuta come l'attività inventiva di una provincia e in un certo arco temporale è al contempo influenzata dalla composizione del *network* e dalla struttura industriale. In primo luogo, l'articolo rafforza la teoria secondo cui la struttura della rete sociale, qui rappresentata dalle collaborazioni brevettuali italiane, influisce sulla produzione innovativa. Questo viene provato utilizzando le più comuni proprietà delle reti sociali come regressori (i.e., densità, misure di transitività, numero di collegamenti esterni alla rete, concentrazione degli inventori). Nel modello successivo, sono inserite le variabili che descrivono la struttura industriale e nello specifico, riprendendo la proposta di Frenken et al. (2007), la varietà correlata e quella non correlata; con parziale sorpresa, solamente la seconda risulta significativa. Infine, in aggiunta alle principali variabili di controllo che normalmente vengono inserite nei modelli economici di crescita, seguono dei modelli in cui è verificata l'entità dell'interazione tra le variabili di rete e quelle di settore. L'elaborato di tesi intende quindi svolgere un lavoro empirico simile, provvedendo da un lato ad aggiungere ulteriori dati al dibattito sul tema (e.g. ruolo della struttura del *network*, *agglomeration economies*, struttura industriale), dall'altro apportando opportune modifiche (e.g., utilizzo dei sistemi locali del lavoro come unità geografica di riferimento) con l'auspicio di ottenere *output* interessanti e offrire nuovi spunti di riflessione per i futuri lavori di ricerca.

# Capitolo 3

## Dati

### 3.1 Diritti di proprietà intellettuale

Il concetto di diritto di proprietà industriale, regolato dall’omonimo codice (decreto legislativo, 10/02/2015) è da alcuni considerato un caso particolare del più ampio e rappresentato diritto di proprietà intellettuale, a cui si riferiscono le norme che regolano ad esempio il diritto d’autore. L’articolo 1 del “Codice della proprietà industriale” (2022) stabilisce cosa sia incluso nell’espressione *proprietà industriale* e, in particolare, i marchi e altri segni distintivi come nomi commerciali o domini internet, indicazioni geografiche, denominazioni di origine, disegni e modelli, invenzioni, modelli di utilità, topografie di prodotti a semiconduttori, segreti commerciali e nuove varietà vegetali. Il Codice distingue due categorie di diritti di proprietà industriale: quelli titolati e quelli non titolati. La prima categoria di diritti è costituita tramite un processo di brevettazione (invenzioni, modelli di utilità, nuove varietà vegetali) o di registrazioni (marchi, disegni e modelli, topografie dei prodotti a semiconduttori). Prima di poter acquisire un diritto titolato, è quindi necessario che siano effettuati degli esami formali delle domande pervenute. Queste verifiche sono eseguite da appositi Uffici e, a seconda del caso, dall’Ufficio Italiano Brevetti e Marchi (UIBM), dall’*European Union Intellectual Property Office* (EUIPO) o dall’Ufficio dei Brevetti Europeo (EPO). Questi uffici svolgono quindi il duplice obiettivo di garantire i requisiti di brevettabilità e di fornire ai futuri inventori (e ai loro aventi causa) un database sufficientemente ricco per capire cosa è già presente nello stato della tecnica ed evitare quindi di spendere inutilmente delle risorse per realizzare un prodotto che non avrà il requisito di novità e non potrà quindi essere protetto da un brevetto.

#### 3.1.1 Brevetti

Tra i metodi per proteggere la proprietà intellettuale, certamente le invenzioni certificate da un processo di brevettazione risultano le più interessanti per lo scopo di questo lavoro di tesi.

L’invenzione è definita come attività creativa che consiste nella soluzione originale di un problema tecnico e brevettarla significa contrassegnare questo prodotto innovativo con un brevetto per assicurarne la priorità e l’esclusiva. Le invenzioni oggetto di brevetto possono



far parte di qualunque settore della tecnica (meccanica, chimica, etc.). Teorie scientifiche, metodi matematici, software non sono classificati come invenzioni e quindi non risultano brevettabili. Affinché l'Ufficio di riferimento accolga la domanda di brevettazione, è necessario che l'invenzione rispetti quattro requisiti:

- novità;
- attività inventiva;
- applicazione industriale;
- liceità.

L'invenzione deve quindi innanzitutto avere carattere di novità, non essendo compresa nella stato della tecnica esistente, ovvero tutto ciò che è stato reso accessibile al pubblico prima della data di deposito della domanda di brevetto, mediante descrizione scritta, orale o tramite un mezzo di comunicazione alternativo. In secondo luogo, l'invenzione implica attività inventiva se, per una persona esperta del settore non risulta in modo evidente dallo stato della tecnica. Inoltre, l'oggetto in questione deve poter essere fabbricato o utilizzato in qualunque genere di industria (non si esclude un prodotto artigianale, ma è necessario che esista una ripetibilità del processo di fabbricazione per un numero non finito di volte con risultati che siano costanti). Infine, non possono essere oggetto di brevettazione le invenzioni la cui attuazione risulti contraria all'ordine pubblico o al buon costume (questo aspetto trova spesso applicazione nel campo biotecnologico). Qualora l'invenzione in questione rispetti tutti i requisiti, l'Ufficio di riferimento accetta la richiesta di brevettazione e i diritti che sorgono con la concessione del brevetto decorrono dalla data in cui la domanda con descrizione, disegni e rivendicazioni (i.e., ciò che effettivamente il brevetto protegge) è resa accessibile al pubblico, normalmente 18 mesi dalla data di deposito o di priorità (cfr. Art. 4, "Codice della proprietà industriale" (2022)). Qualora la richiesta di brevettazione riceva un riscontro favorevole, il brevetto concesso contiene una serie di informazioni utili a identificarlo e a distinguere i diritti che da quel momento spettano al richiedente (un esempio della prima pagina di un brevetto depositato presso l'EPO è disponibile in Appendice A). In particolare, la prima pagina del brevetto mostra alcune informazioni, tra cui le principali sono:

- la data del deposito, la priorità e la data di pubblicazione;
- il richiedente, l'inventore e il titolare;
- il numero associato alla richiesta del brevetto;
- il titolo e un riassunto.

Il documento è inoltre composto dalla descrizione del prodotto e dall'elenco delle rivendicazioni, che stabiliscono l'ambito di protezione garantito; il brevetto è quindi completato dalla lista dei disegni. Di particolare interesse risulta essere la prima pagina del documento in quanto, oltre ad alcune delle informazioni già citate, contiene un codice che inserisce il brevetto all'interno di una precisa classe tecnologica. La classificazione internazionale

dei brevetti (IPC<sup>1</sup>), utilizzata in questo lavoro è costituita da 8 sezioni, divise in 131 classi e in 645 sottoclassi. La scomposizione prosegue poi ad un livello di dettaglio ancora maggiore con circa 70.000 gruppi. Per semplicità di trattazione, si descrivono solamente le 8 sezioni, rimandando i ragionamenti sulle altre categorie alla parte di analisi dei dati e costruzione del modello econometrico. Le 8 sezioni sono quindi così costruite:

- 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*;
- H: *electricity*.

Il brevetto di invenzione ha una durata pari a 20 anni, non può essere rinnovato, nè può essere prorogata la data di scadenza. La *ratio* della norma coniuga la protezione garantita all'inventore, e quindi la possibilità di sfruttare economicamente il brevetto, con la garanzia che la società possa beneficiare, dopo un periodo limitato e fisso, della nuova invenzione. Il sistema brevettuale così costruito ha infatti l'obiettivo primario di incoraggiare gli investimenti in ricerca e sviluppo per favorire, in ultima analisi, il progresso tecnologico e scientifico della società tutta. La concessione del brevetto determina la nascita di due tipologie di diritti: i diritti patrimoniali, alienabili, che conferiscono al titolare la facoltà esclusiva di attuare l'invenzione e di trarne profitto, e il diritto morale, acquisito per il solo fatto di aver realizzato l'invenzione, non trasferibile, in ultima analisi costituisce il diritto di essere riconosciuto autore dell'invenzione. Proprio in virtù dell'esistenza di diritti patrimoniali legati al brevetto, esistono degli appositi contratti per gestirne l'utilizzo e i profitti che ne derivano. Una prima possibilità prevede la cessione del brevetto ad un secondo soggetto (cessionario) sulla base di un corrispettivo, spesso non facile da determinare; in alternativa, il proprietario non si spoglia della titolarità del brevetto, ma concede il suo utilizzo ad un terzo tramite contratti di licenza. Contratti di questo tipo

---

<sup>1</sup>*International Patent Classification*, la classificazione internazionale dei brevetti IPC rappresenta il sistema più utilizzato a livello internazionale per i brevetti e i modelli d'utilità per definire in maniera univoca il settore tecnologico di riferimento di un'invenzione, superando in questo modo le possibili incongruenze tra i diversi uffici nazionali. Istituita a seguito dell'Accordo di Strasburgo del 1971, è strutturata in modo gerarchico e suddivide le tecnologie brevettabili in otto sezioni (A - H), a loro volta distribuite in livelli sempre più dettagliati (sottosezioni, classi, sottoclassi, gruppi e sottogruppi). Viene aggiornata periodicamente: dal 2006 è in vigore l'ottava versione, che contiene circa 70.000 voci. La IPC non viene utilizzata solo per classificare e ricercare brevetti, ma anche pubblicazioni, articoli scientifici e testi tecnici in generale, al fine di valutare lo stato della tecnica in un particolare settore.

possono eventualmente prevedere delle clausole che indicano una licenza esclusiva, eventualmente ristretta a livello territoriale e/o di settore industriale. Concedere la licenza del brevetto consente di diffondere il prodotto anche in mercati che il titolare non è in grado di raggiungere o che richiedono dei costi così elevati da scoraggiarne l'iniziativa.

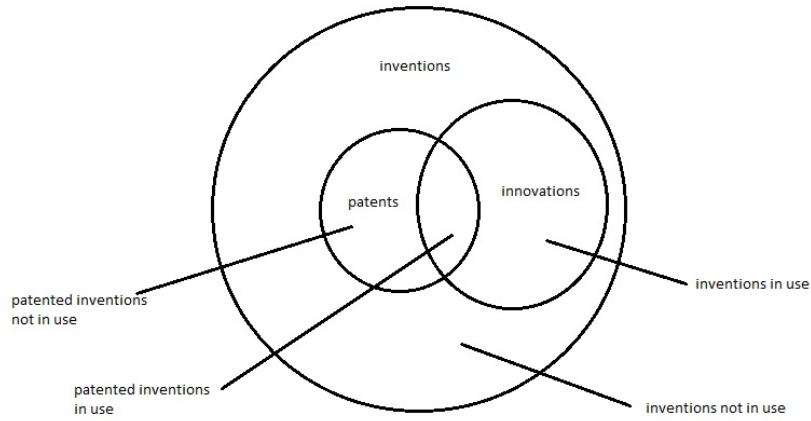
Un'altra questione delicata e riguarda le invenzioni dei dipendenti o dei ricercatori delle università (per maggiore dettaglio, si rimanda agli articoli 64 e 65 del CPI). In generale, è infatti possibile e tutt'altro che insolito che il processo di brevettazione coinvolga più soggetti: l'inventore (eventualmente più di uno), necessariamente una persona fisica e il titolare (o più di uno), persona fisica o giuridica che sfrutterà economicamente il brevetto. Pur essendo vero che le due figure possono coincidere, molto spesso l'inventore, per evitare di sostenere le spese di brevettazione, può cedere a terzi, dietro compenso, il suo diritto a sfruttare l'invenzione. Gli articoli 64 e 65 del CPI distinguono i casi specifici in cui l'invenzione sia effettuata da un dipendente legato ad un'impresa da un contratto di lavoro dalla fattispecie in cui sia un ricercatore di un'università o di un ente pubblico di ricerca a figurare come inventore. Sia che l'inventore depositi in prima persona la domanda di brevettazione, sia che un terzo, ottenuta la titolarità, si sostituisca a lui, la figura che in questa fase si interfaccia con l'Ufficio competente è chiamata richiedente. Come anticipato, è possibile che esista più di un richiedente; in assenza di annotazioni diverse sul modulo di domanda, le quote di titolarità sono divise equamente tra i richiedenti. In particolare, questa fattispecie si può presentare in tutti quei casi in cui più inventori, potenzialmente provenienti da località diverse, collaborano alla stessa invenzione. La letteratura economica che studia le esternalità di conoscenza ha negli ultimi anni focalizzato la propria attenzione sulle reti di collaborazioni brevettuali, ritenendo che possano fornire un contributo interessante sulle variabili che favoriscono questi *spillover*. Inoltre, uno specifico filone di ricerca si occupa di costruire dei modelli statistici per capire quali siano i fattori che maggiormente influenzano le *performance* innovative di un territorio. Per fare questo, spesso l'attività brevettuale è utilizzata come *proxy* dell'attività inventiva. Tale scelta non trova parere unanimemente favorevole tra i ricercatori e in letteratura sono presenti scelte diverse. Escludendo misure prevalentemente qualitative (i.e., interviste o questionari) e che in quanto tali nascondono delle debolezze rispetto all'oggettività dei giudizi e in quanto la raccolta dati risulta molto onerosa in termini di tempo e risorse, l'alternativa spesso proposta consiste nell'utilizzare le spese in R&D. Partendo dalla proposta che i brevetti rappresentino una migliore approssimazione dell'attività inventiva (Griliches et al., 1986), la scelta di utilizzare in maniera così diffusa la produzione brevettuale è in ultima analisi giustificata da alcuni vantaggi garantiti dai brevetti (Archibugi, 1992).

Innanzitutto, i brevetti rappresentano l'output del processo inventivo e in particolare delle invenzioni che ci si attende possano avere un impatto economico rilevante. Inoltre, richiedere la protezione brevettuale comporta dei costi non indifferenti, lasciando quindi ipotizzare che le domande presentate siano associate a progetti con valore potenziale rilevante e pertanto in grado di coprire i costi inizialmente sostenuti. In aggiunta, i brevetti garantiscono una quantità di informazioni ampiamente distribuita rispetto al tempo, allo spazio e al campo tecnologico di riferimento. Infine, l'accesso a queste informazioni è relativamente semplice ed economico.

Cionondimeno, la relazione tra brevetti, invenzioni e innovazioni risulta alquanto complessa. Come mostrato in figura 3.1, infatti, solo alcune delle invenzioni totali sono brevettate

e solo una porzione di invenzioni rappresenta un'innovazione rilevante (Basberg, 1987).

Figura 3.1: Invenzioni, innovazioni, brevetti



Alcuni studi empirici hanno mostrato che le imprese depositano la domanda di brevetto per il 66-87 % dei prodotti brevettabili presenti nel loro portafoglio. Questi dati apparentemente incoraggianti sono però da maneggiare con cura, poiché non considerano le invenzioni non brevettabili e quelle realizzate da centri di ricerca e università, che per loro natura hanno minore incentivo a presentare richiesta di brevetto. Inoltre, la seconda questione riguarda la quota di brevetti che diventano effettivamente innovazioni. I risultati mostrano dati incoraggianti, in quanto circa la metà dei brevetti è poi utilizzato in applicazioni concrete (Archibugi, 1992).

In questo lavoro di tesi si utilizzano i dati relativi all'attività brevettuale italiana e in particolare un dataset estratto dall'UIBM, che sarà descritto con maggiore dettaglio nei prossimi paragrafi.

## 3.2 Aree funzionali

È noto come nelle città europee sia presente il 70 per cento della popolazione complessiva, così come 2/3 del PIL. È quindi evidente l'importanza di studiare con attenzione queste realtà e i territori circostanti. In particolare, negli ultimi decenni è cresciuta la necessità di descrivere in modo più approfondito le aree urbane, prendendo in considerazione anche la complessità delle relazioni spaziali tra i vari centri amministrativi. Parallelamente, e per rispondere alle nuove esigenze del mercato, la produzione statistica ha innovato i propri strumenti, garantendo un'ampiezza tematica, territoriale e di informazione senza precedenti. La granularità delle informazioni a disposizione unita alla velocità con cui questi dati sono acquisiti pongono nuove sfide per gli Istituti Nazionali di Statistica. A partire dal 1971, l'Istat innova il proprio Censimento, introducendo i quesiti sul pendolarismo dal luogo di residenza a quello di lavoro. L'obiettivo dell'Istituto consiste quindi nel superare le tradizionali partizioni amministrative che non descrivono correttamente il "Paese reale". Pertanto, utilizzando le informazioni sugli spostamenti dei cittadini per motivi di studio o lavoro da una città all'altra, si costruiscono i primi sistemi locali (SL). Una geografia così realizzata consente di comprendere meglio le dinamiche socio-economiche del territorio in quanto, a differenza delle partizioni amministrative, risulta dall'organizzazione autonoma degli individui e non da eventi storici passati o dalla morfologia del territorio. Avendo come modello di riferimento la scuola americana che già nella prima metà del secolo scorso iniziava a definire delle nuove partizioni (i.e., *Metropolitan Statistical Areas*), l'Istat decide quindi di adottare una misura di questo tipo in maniera continuativa e nella convinzione che una partizione funzionale del territorio sia fondamentale per capire dove investire e come massimizzarne i risultati. Il concetto di sistema locale si fonda su alcune ipotesi chiave (Istat, 2015):

1. gli spostamenti quotidiani delle persone esauriscono le loro relazioni sociali ed economiche;
2. gli spostamenti tra luogo di residenza e luogo di lavoro sono una buona approssimazione degli spostamenti legati al tempo libero, alla gestione familiare, etc.;
3. l'auto-organizzazione delle relazioni definisce delle realtà *auto-contenute* e *integrate*, cioè fortemente interconnesse al loro interno e relativamente impermeabili ai flussi esterni.

Inoltre, con l'obiettivo di rendere omogenee le varie misure a livello europeo, Eurostat ha elaborato dei principi comuni per la costruzione dei SL, come in tabella 3.1. Da un punto di vista tecnico, i sistemi locali sono aree funzionali costruite aggregando le aree elementari scelte come riferimento (nel caso italiano i comuni) tramite un algoritmo che soddisfi i vincoli di dimensione e di auto-contenimento della domanda e dell'offerta di lavoro. Con l'obiettivo di definire meglio l'algoritmo utilizzato dall'Istat, si definiscono le seguenti grandezze:

- $f_{hk}$ , numero di pendolari che si recano dalla località di residenza  $h$  alla località di lavoro  $k$ ;
- $R_i$ , il numero di occupati residenti nella località  $i$ -esima;

- $W_i$ , il numero dei posti di lavoro nella località  $i$ -esima;
- $SS\_SC = \frac{f_{ii}}{R_i}$ , auto-contenimento lato offerta, definito dal rapporto tra gli occupati che risiedono e lavorano in una località  $i$  e il numero dei residenti totali;
- $DS\_SC = \frac{f_{ii}}{W_i}$ , auto-contenimento lato domanda, calcolato rispetto al totale dei posti di lavoro;

Tabella 3.1: Principi per la costruzione dei SL a livello europeo

Principio	Descrizione
Scopo	Ciascun sistema locale rappresenta un mercato del lavoro
Rilevanza	I SL permettono di diffondere informazione statistica affidabile e confrontabile
Completezza	I SL sono una partizione dell'intero territorio nazionale
Contiguità	Ciascun SL è costituito da un insieme di comuni contigui
Autonomia	I SL massimizzano l'auto-contenimento dei flussi
Omogeneità	I SL non sono troppo estesi territorialmente o troppo numerosi in termini di occupati
Coerenza	Ciascun SL è costituito da un insieme di comuni non frazionati
Conformità	I SL possono rispettare i confini amministrativi
Flessibilità	Il metodo per la creazione dei SL deve essere performante in regioni diverse

L'algoritmo è costruito in modo da gestire il *trade-off* tra la dimensione del sistema locale (misurata con il numero degli occupati residenti) e l'auto-contenimento lato domanda e offerta. Definiti i parametri minimi e target di queste due grandezze, si identifica una funzione di validità che definisce i livelli di ampiezza dimensionale e auto-contenimento che ogni *cluster* deve rispettare. A questo punto si imposta un procedimento iterativo che assegna ogni unità elementare ad un sistema locale, verificando ad ogni passo la bontà dell'assegnazione ed eventualmente disaggregando le singole unità spaziali qualora non siano più rispettati i vincoli stabiliti. Il *trade-off* tra ampiezza dimensionale e auto-contenimento si traduce nella costruzione di sistemi locali con un livello di auto-contenimento *accettabile* tanto più basso quanto maggiore risulti la loro dimensione (Istat, 2015).

L'ampia disponibilità di dati garantita dagli Istituti Nazionali di Statistica e la riconosciuta importanza di una partizione territoriale di questo tipo a livello governativo (L.317/1991, L.144/1999, DPEF 1999) e di ricerca (Breschi & Lenzi, 2016; Fleming et al., 2007; Van der Wouden & Rigby, 2019) rendono robusta la scelta di utilizzare i SL come area territoriale oggetto di studio in questo lavoro di tesi. Senza dubbio utilizzare una partizione

---

amministrativa secondo il sistema NUTS<sup>2</sup> presenterebbe il notevole vantaggio dell'ampia disponibilità e qualità dei dati. Tuttavia, si è ormai raggiunta la consapevolezza che i confini amministrativi sono il risultato di processi storico-politici che nulla hanno a che fare con l'attuale quadro socio-economico che si intende studiare. Esempio chiaro di questo fenomeno è rappresentato dai grandi centri urbani, che nonostante abbiano un numero di residenti relativamente basso, sono quotidianamente raggiunti da impiegati che risiedono nelle zone circostanti ma si spostano nelle grandi metropoli per motivi di lavoro. Un filone della letteratura economica sceglie quindi di rinunciare ad una grande quantità di dati disponibili per le divisioni amministrative, preferendo una partizione funzionale che sia in grado di valutare con maggiore precisione fenomeni socio-economici rilevanti come il livello di disoccupazione o la forza lavoro disponibile in una regione (Istat, 2015). Nonostante questa scelta trovi ampio riscontro in letteratura, è necessario considerare un'ulteriore criticità che lavori di questo tipo devono gestire con attenzione. Nella costruzione di modelli di regressione che utilizzino questa partizione funzionale, infatti, risulta spesso necessario aggiungere le variabili di controllo utilizzate nei principali modelli di crescita economica (textciteinnocenti2020knowledge). Poiché spesso questi dati sono disponibili al più per i livelli NUTS 2 (regioni) o NUTS 3 (province), bisogna riscalarli a livello delle partizioni funzionali e valutare i risultati ottenuti con cautela.

### 3.2.1 Sistemi locali del lavoro

In occasione dei censimenti generali della popolazione e delle abitazioni e dei censimenti generali dell'industria e dei servizi, l'Istat pubblica un elenco di dati a livello di sistema locale del lavoro (SLL). Le informazioni più recenti a nostra disposizione si riferiscono al 2011 e si differenziano in modo evidente dalle precedenti partizioni. Il numero e quindi la geografia dei SLL sono infatti rapidamente e costantemente cambiati negli ultimi decenni: 955 SLL nel 1981, 784 nel 1991, 686 nel 2001, 611 nel 2011. In tabella 3.2 e in figura 3.2 si mostrano i dati relativi all'ultimo censimento attualmente reso noto dall'Istat e riferito al 2011. In particolare, il Sud risulta l'area con il maggior numero di sistemi locali (171, 28% del totale). A livello regionale, la Sicilia è la regione a cui è attribuito il maggior numero di sll (l'assegnazione di un sll ad una regione avviene in base all'appartenenza del comune che dà il nome al sll), mentre Molise e Valle d'Aosta sono le regioni con il minor numero di partizioni. Per definizione, alcune aree sono costruite in modo da superare i confini regionali e provinciali. È in particolare il caso di Lombardia e Emilia-Romagna, i cui sll superano quasi la metà delle volte i confini di una singola provincia (rispettivamente in 24 e 18 occasioni). Diversamente, Valle d'Aosta, Puglia, Sicilia e Sardegna sono composte da sistemi locali i cui confini non oltrepassano quelli regionali. Escludendo i motivi prettamente geografici come nel caso delle isole, singolare risulta il

---

<sup>2</sup>La nomenclatura delle unità territoriali statistiche (dal francese: *Nomenclature des unités territoriales statistiques*) identifica la ripartizione del territorio dell'Unione europea a fini statistici. Ideata dall'Eurostat nel 1988 tenendo come riferimento di base l'unità amministrativa locale, da allora è la principale regola per la redistribuzione territoriale dei fondi strutturali dell'UE, fornendo uno schema unico di ripartizione geografica, a prescindere dalle dimensioni territoriali delle unità amministrative.

caso pugliese, che è l'unica regione di dimensione medio-grande (44 SLL, 7.2 % del totale) che non presenta sistemi locali che oltrepassino i confini regionali.

Tabella 3.2: Sistemi locali del lavoro per regione

Area geografica	Totale	Di cui multi-regionali	Di cui multi-provinciali	% sul totale nazionale
Piemonte	36	5	19	5,9
Valle d'Aosta	5			0,8
Lombardia	51	4	24	8,3
Prov. Bolzano	12	1	2	2,0
Prov. Trento	14	2	3	2,3
Veneto	43	5	17	7,0
Friuli-Venezia-Giulia	11	2	4	1,8
Liguria	14	4	5	2,3
Emilia-Romagna	39	9	18	6,4
Toscana	48	3	14	7,9
Umbria	14	3	4	2,3
Marche	25	3	11	4,1
Lazio	18	3	8	2,9
Abruzzo	18	3	7	2,9
Molise	5	3	4	0,8
Campania	46	3	13	7,5
Puglia	44		3	7,2
Basilicata	14	2	4	2,3
Calabria	44	1	6	7,2
Sicilia	71		11	11,6
Sardegna	39		9	6,4
Nord-ovest	106	13	48	17,3
Nord-est	119	19	44	19,5
Centro	105	12	37	17,2
Sud	171	12	36	28,0
Isole	110	0	20	18,0
<b>ITALIA</b>	<b>611</b>	<b>56</b>	<b>185</b>	<b>100,0</b>

Il numero di partizioni presente in ciascuna regione o macro-area del Paese dipende da molteplici fattori, tra cui i principali:

- le infrastrutture, che possono favorire o disincentivare gli spostamenti;
- la morfologia di un territorio, che può a volte rappresentare una barriera naturale a fenomeni di pendolarismo;
- il contesto socio-economico e in primis le opportunità lavorative;
- la presenza di grandi centri urbani che tendono a *inglobare* le realtà adiacenti più piccole.

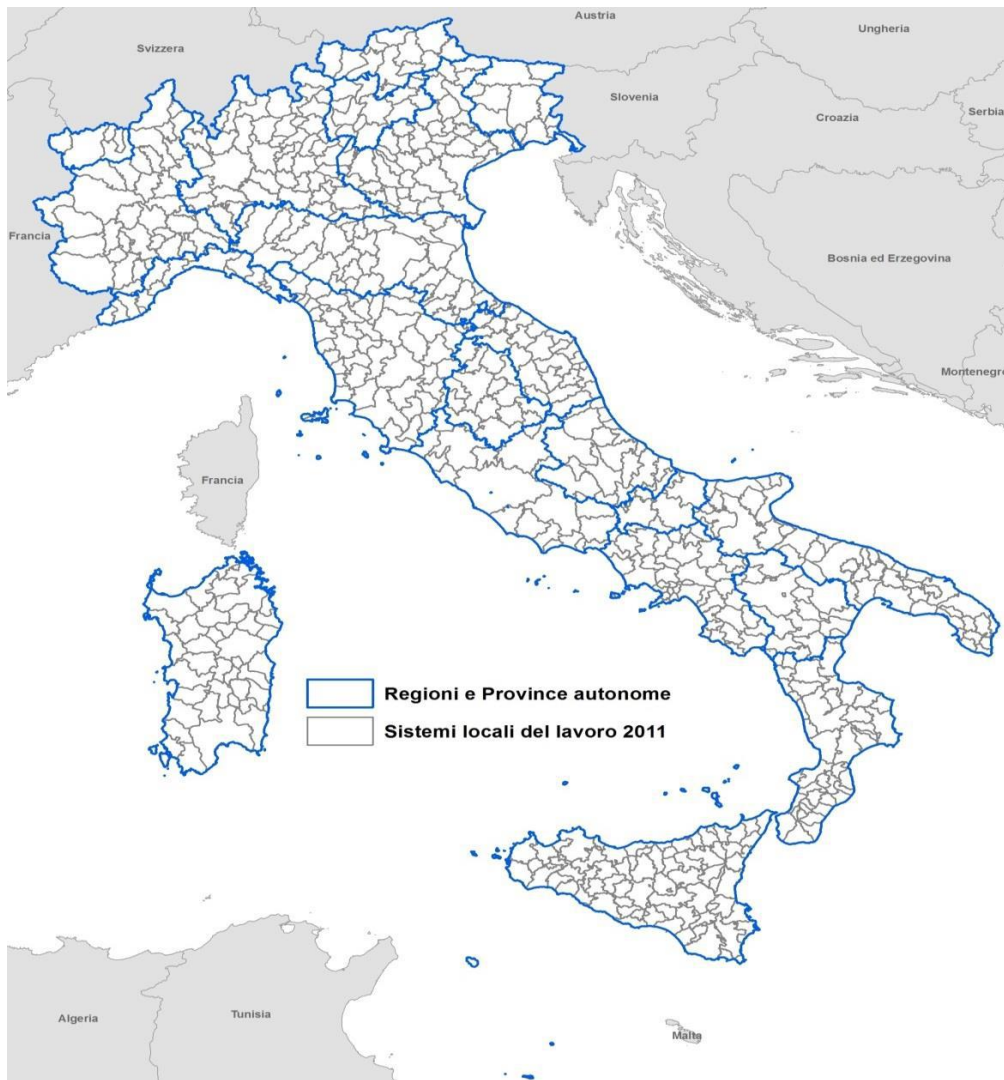


Con particolare riferimento agli ultimi due punti, risulta interessante focalizzarsi proprio sui sistemi locali più grandi e che rappresentano le aree più produttive del Paese. In questa direzione, in tabella 3.3 si mostrano i dati relativi ai sll con una popolazione residente superiore a 500.000 unità e, in particolare, si confrontano queste informazioni con il censimento del 2001. Questo tipo di analisi, pur circoscritta a soli 17 sll, consente di valutare come le principali partizioni italiane si evolvano nel tempo. A testimonianza dell'importanza di queste aree, nel decennio 2001-2011, in termini relativi, la crescita della popolazione nei 17 grandi sll supera di 5,2 punti percentuali la media nazionale (+9,5% vs +4,3%). Il sll che si è espanso maggiormente è quello di Busto Arstizio (+86,9% rispetto alla popolazione residente), seguito da Bergamo (+52,8%) e Como (+35%). Ad eccezione di Genova, Palermo e Verona, gli altri sll crescono sia in termini di popolazione residente che di superficie occupata. La tendenza evidenziata dalla tabella 3.3 non trova invece riscontro tra i sll di minori dimensioni: soprattutto nel Mezzogiorno e prevalentemente tra le aree di piccole dimensioni, si registra un calo della popolazione residente nel periodo 2001-2011. Un'analisi più approfondita e che consideri anche i dati sul mercato del lavoro, nonché i diversi valori di auto-contenimento (lato domanda e offerta) associati a ciascun sll, è disponibile nel report Istat (2014).

Tabella 3.3: I sistemi locali del lavoro di grandi dimensioni

SLL	N. di comuni	Popolazione 2011	Var. % Popolazione	Var. % Superficie
Torino	42	172.108	11,0	71,1
Genova	31	681.097	-5,0	-2,9
Busto Arstizio	53	623.023	86,9	130,7
Como	99	535.951	35,0	38,0
Milano	174	3.685.101	7,0	9,9
Bergamo	123	802.731	52,8	73,0
Verona	23	458.940	-9,9	-31,2
Venezia	19	606.002	3,7	0,5
Padova	52	664.591	15,9	3,5
Bologna	40	847.058	19,0	22,0
Firenze	18	687.304	2,4	-7,6
Roma	89	3.479.572	7,7	11,0
Napoli	58	2.510.848	2,7	19,6
Bari	20	737.008	2,6	1,0
Palermo	18	880.046	0,5	-9,7
Catania	22	676.742	18,8	11,7
Cagliari	42	504.580	2,1	16,4
<b>ITALIA</b>	<b>8.092</b>	<b>59.433.744</b>	<b>4,3</b>	<b>-</b>

Figura 3.2: Configurazione territoriale dei sistemi locali del lavoro nel 2011



### 3.3 Dati empirici

L'elaborato di tesi utilizza come database di partenza l'insieme dei brevetti depositati presso l'UIBM dal 1994 al 2019. In particolare, una prima tabella, *patents*, è composta dai seguenti campi (si riportano quelli di maggiore interesse):

- codice univoco identificativo del brevetto;
- data in cui è depositata la richiesta di brevettazione;
- stato di avanzamento della domanda di brevetto;
- data in cui la domanda di brevetto viene accolta, qualora questo avvenga; altrimenti, la cella di riferimento risulta vuota.

Questa tabella presenta 230.636 diversi brevetti. La chiave primaria<sup>3</sup> è rappresentata dal codice univoco identificativo del brevetto. Ulteriori informazioni sono ricavate invece dalla tabella *applicants*; per ogni codice brevetto si conoscono i seguenti dati dei richiedenti:

- cognome;
- nome;
- codice fiscale;
- ruolo, in particolare, *owner*, ovvero il titolare del brevetto o *applicant*; si noti che le due funzioni spesso coesistono nel medesimo soggetto (persona fisica o giuridica, nel caso in cui sia un'impresa a depositare la domanda di brevetto);
- nazionalità;
- provincia;
- codice univoco del sll;
- nome della città;
- codice univoco della città.

Questa seconda tabella presenta 273.578 righe, ciascuna contenente i dati di un richiedente. A differenza della precedente, qui è possibile che lo stesso codice brevetto sia duplicato: questo accade poiché capita frequentemente che la medesima domanda sia congiuntamente depositata da più richiedenti. Questa fattispecie, definita *collaborazione brevettuale*, rappresenta un punto cardine di questo elaborato, in quanto nella letteratura economica di riferimento è utilizzata per approssimare possibili esternalità di conoscenza. Il dataset è quindi completato dalla lista dei codici IPC associati a ciascun brevetto e dall'elenco delle aree geografiche di riferimento (a livello di regione, provincia, comune e sll).

---

<sup>3</sup>Una speciale colonna della tabella del database relazionale (o combinazione di colonne) designata per identificare in modo univoco ogni record della tabella

## 3.4 Il campione analizzato

Con l'obiettivo di descrivere e analizzare un campione coerente con gli obiettivi prefissati, si procede con una prima pulizia del dataset di partenza. Con riferimento alla tabella *applicants*, un passaggio fondamentale per la costruzione della rete di collaborazioni brevettuali consiste nel filtrare i dati mantenendo solamente i record con il codice fiscale dei richiedenti non nullo. È infatti necessario conoscere questo dato perché rappresenta il nodo della rete che sarà successivamente realizzata. Poiché questa modifica coinvolgerebbe migliaia di righe, si verifica che i dati siano cancellati in maniera omogenea rispetto al tempo (anni), spazio (area geografica di riferimento) e tecnologia (codici IPC).<sup>4</sup> Si può pertanto procedere con la pulizia del dataset secondo l'intenzione iniziale. Si prosegue quindi a eliminare tutte le righe con un valore nullo per il campo *nazionalità*. Si prosegue eliminando i record con codice del sll nullo e riferiti a brevetti italiani: con questa accuratezza sono conservate le informazioni relative a richiedenti stranieri, che rappresentano in verità una quota minima del totale (3134 righe su un totale di 273.578, poco più dell'1%). Anche in questo caso è necessario verificare preliminarmente di eliminare i record in maniera uniforme rispetto alle classi tecnologiche e agli anni di riferimento.

Alla luce di quanto descritto<sup>5</sup>, si sceglie di eliminare dalla trattazione le richieste di brevetto pervenute negli anni 2015, 2016 e 2017. Nei rimanenti anni compresi nel periodo 1994-2018, si procede ad eliminare i record con codice *llma* mancante. Questo passaggio si è reso necessario in quanto, diversamente, non sarebbe possibile costruire la rete di collaborazioni brevettuali che associ a ciascun richiedente un SLL.

Inoltre, si sottolinea che la tabella *applicants* è costruita in modo tale da memorizzare il ruolo del richiedente in occasione del deposito della domanda di brevetto. Qualora, come anticipato, un individuo figuri sia come *owner* che come *applicant*, saranno presenti due righe identiche, fatta eccezione per il campo *ruolo*. Poiché si tratta evidentemente del medesimo brevetto e non si può configurare come collaborazione brevettuale, le righe duplicate sono eliminate.

Infine, con l'obiettivo di certificare la qualità dell'attività inventiva prodotta nei sistemi locali, si procede a filtrare il dataset di partenza conservando solamente i brevetti la cui domanda depositata risulti *concessa, titolata, trasformata, convertita o in attesa di attestato*, eliminando in ultima analisi le domande non approvate dall'UIBM.

A seguito di queste verifiche preliminari, il campione disponibile è costituito da migliaia di record, ciascuno dei quali associa ad ogni codice brevetto le informazioni temporali, spaziali e tecnologiche che sono state illustrate nella sezione precedente.

Prima di procedere con le analisi descrittive del dataset, è però necessario introdurre un nuovo elemento: i *pesi* relativi a ciascun brevetto. Infatti, nel caso in cui il brevetto sia condiviso da due o più inventori, il database è costruito in modo tale da duplicare la riga in questione, che presenterà lo stesso codice brevetto e le informazioni relative all'ulteriore inventore. Come mostrato in OECD (2009), la crescente tendenza a collaborare da parte degli inventori ha indotto i ricercatori a introdurre un nuovo metodo di conteggio

---

<sup>4</sup>si veda Appendice B

<sup>5</sup>si veda Appendice C

dei brevetti, *fractional counts*, contrapposto al tradizionale *whole counts*. Detto che non esiste una scelta metodologica in assoluto migliore poiché la preferenza è motivata dagli obiettivi del lavoro di ricerca, si sceglie qui di utilizzare il metodo *fractional* in quanto in questo modo si evita un doppio conteggio ed è possibile aggregare in maniera credibile i dati a diversi livelli di dettaglio. Tuttavia, si precisa che una scelta di questo tipo espone alla possibilità di valutare diversamente i brevetti realizzati da inventori multipli da quelli depositati da un singolo richiedente. Poiché questa fattispecie non è rilevante per la costruzione di gran parte dei regressori del modello (che dipendono invece dalla rete delle collaborazioni brevettuali), pur consapevoli dei limiti di questa misura, si sceglie di adottare un sistema di *pesi*, che si dettaglia di seguito. Si provvede a identificare le righe duplicate per il codice identificativo del brevetto e ad associare un *peso* al brevetto tale per cui, ad esempio, in presenza di due inventori ciascuna riga sarà conteggiata moltiplicando per un fattore 0.5 (0.33 nel caso di tre inventori, etc.). Un problema simile si riscontra per i codici IPC associati a ciascun brevetto. Infatti, non è raro che a un maggior livello di dettaglio un brevetto sia associato a due o più codici IPC. È quindi opportuno *pesare* correttamente questi dati, applicando lo stesso concetto.

### 3.4.1 Analisi temporale

Una prima riflessione consiste nel valutare l'andamento dell'attività inventiva nel periodo di riferimento. Negli elaborati che utilizzano i brevetti come unico indicatore per valutare l'attività inventiva di un Paese, una delle scelte fondamentali consiste nel collocare temporalmente il brevetto, che attraversa diversi step intermedi prima di ottenere l'approvazione. Le principali date possono essere così riassunte:

1. *priority date*: la prima richiesta di brevettazione, ovunque nel mondo, il momento più vicino all'invenzione;
2. *application date*: la data in cui è depositata la domanda presso uno specifico ufficio;
3. *publication date*: solitamente dopo 18 mesi dalla *priority date*, riflette il momento in cui l'invenzione è divulgata al pubblico;
4. *grant date*: è il momento finale dell'intero processo, a partire dal quale sorgono in capo ai richiedenti (secondo i meccanismi già descritti) tutti i diritti conferiti con l'approvazione del brevetto. Il tempo necessario per completare la procedura varia in base all'ufficio competente e, in particolare, dai 3 ai 10 anni.

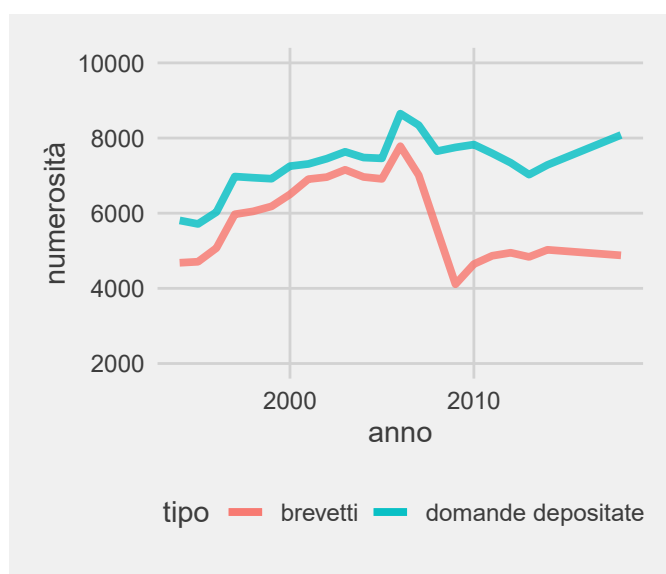
In generale, non esiste uno strumento migliore di altri per collocare nel tempo un'invenzione ed è invece opportuno adattare la scelta agli obiettivi dell'elaborato. Tuttavia, poiché spesso si cerca di valutare la *performance* innovativa di un territorio, la *priority date*, in quanto temporalmente prossima alla data effettiva dell'invenzione, risulta la scelta più appropriata. Non disponendo di questo dato, in questo lavoro si sceglie di utilizzare l'*application date* per i seguenti motivi:

- questa data è certamente da preferire rispetto alla *grant date* in quanto temporalmente più vicina all'invenzione;

- essendo i brevetti presenti nel dataset tutti depositati presso l’UIBM, si supera la difficoltà nel confrontare dei brevetti appartenenti a due Paesi diversi. Infatti, esiste un lag temporale tra la prima *application date* e quelle effettuate per lo stesso brevetto, ma in un altro Paese. In ultima analisi, in questo modo invenzioni temporalmente distanti sarebbero erroneamente attribuite allo stesso periodo.

In figura 3.3 si confronta l’andamento nel tempo delle domande depositate presso l’UIBM con i brevetti che hanno invece superato con successo le verifiche dell’ufficio competente.

Figura 3.3: Andamento temporale domande brevettuali vs brevetti approvati



Se nei primi anni il trend delle due curve appare sostanzialmente sovrapponibile (al netto di una naturale traslazione verso il basso di una delle due, che non include le domande di brevetto rifiutate), a partire dal 2008 si riscontra un crollo del numero di brevetti. Se per la parte finale del campione questa differenza così marcata trova una parziale giustificazione nell’aumento delle domande di brevetto che sono ancora al vaglio dell’UIBM, ma che ai fini della trattazione sono comunque escluse, lo stesso ragionamento non può essere applicato per la parte centrale del grafico. Questa anomalia è invece spiegata da un aumento delle domande di brevetto rifiutate dall’esaminatore. In particolare, diversamente da quanto accadeva fino a quel momento, l’ufficio italiano ha iniziato a richiedere *ex-ante* la consulenza dell’EPO per giudicare il contenuto della domanda e determinarne quindi la brevettabilità. Questo diverso approccio ha determinato un imponente aumento del tasso di domande rifiutate da parte dell’UIBM. Poiché questo fenomeno risulta uniformemente distribuito sia rispetto ai sistemi locali che ai codici IPC (a livello delle sezioni), si ritiene di poter conservare questi dati e procedere nella trattazione.

### 3.4.2 Analisi geografica

Una seconda valutazione possibile riguarda la distribuzione spaziale dell'attività inventiva a diversi livelli. La tabella 3.4, ad esempio, mostra il numero di brevetti associato a ciascuna regione.

Tabella 3.4: Numero di brevetti per regione

Regione	N. brevetti	N. brevetti (%)	N. brevetti per 100k abitanti
Abruzzo	1728.86	1.37	135.73
Basilicata	209.61	0.16	38.81
Calabria	822.90	0.65	44.61
Campania	1748.30	1.38	31.27
Emilia Romagna	21504.03	17	485.21
Friuli Venezia Giulia	4626.78	3.66	386.43
Lazio	7304.76	5.78	127.81
Liguria	2657.48	2.1	176.29
Lombardia	32370.07	25.61	324.83
Marche	4784.03	3.78	321.12
Molise	215.75	0.17	74.19
Piemonte	12035.00	9.52	283.02
Puglia	1662.51	1.31	42.49
Sardegna	663.68	0.52	42.02
Sicilia	1736.98	1.37	36.17
Toscana	8912.95	7.05	242.44
Trentino Alto Adige	2061.55	1.63	191.25
Umbria	1381.90	1.09	160.76
Valle d'Aosta	125.36	0.09	101.64
Veneto	19842.35	15.69	408.73

In termini assoluti, le regioni che vantano il più alto numero di brevetti sono Lombardia (32370.07), Emilia Romagna (21504.03) e Veneto (19842.35), confermando quindi una forte relazione tra attività inventiva ed economica: infatti, sono le aree, insieme alla regione Lazio, che contribuiscono in misura maggiore al PIL nazionale. Inoltre, con l'obiettivo di ottenere una stima più accurata della distribuzione spaziale, si calcola il numero di brevetti per 100 mila residenti, per ogni regione. Se da un lato i dati confermano i risultati già discussi, dall'altro bisogna sottolineare il peso dell'attività inventiva di regioni come il Friuli Venezia Giulia (386.43) e le Marche (321.12) le quali, valutate rispetto alla popolazione residente, mostrano dei risultati migliori di regioni come il Piemonte (283.02) e il Lazio (127.81). In entrambe le valutazioni numeriche il Mezzogiorno rivela i dati peggiori: il numero di brevetti per 100 mila abitanti mostra i valori più bassi in Calabria, Puglia, Sardegna, Basilicata e Sicilia, con soli 36.17 brevetti ogni 100 mila residenti.

Ad un livello di dettaglio maggiore, in figura 3.4 sono elencati i sistemi locali con attività brevettuale in assoluto più intensa.

Figura 3.4: I sls con attività brevettuale più intensa

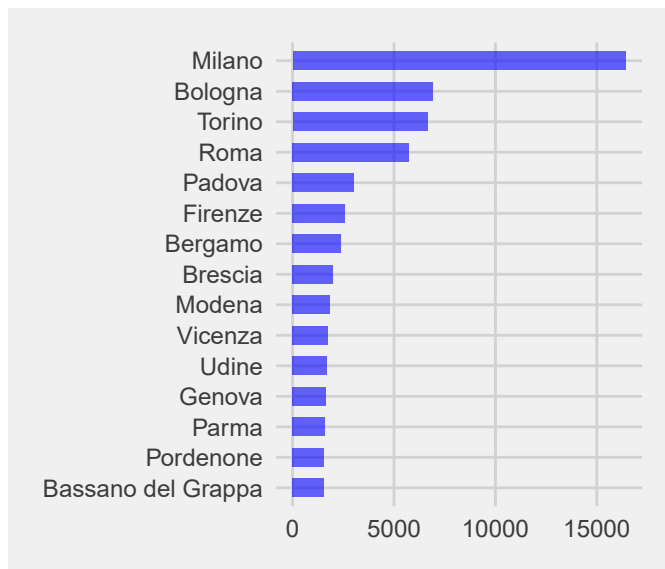
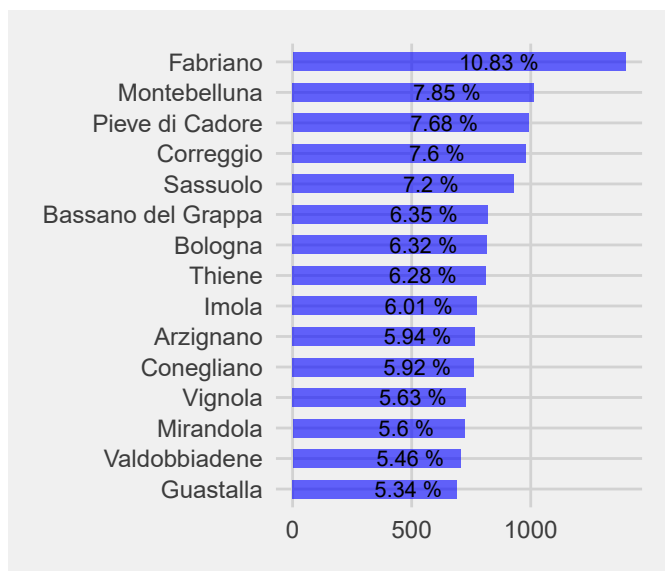


Figura 3.5: I sls con attività brevettuale pro capite più intensa



Questo grafico è evidentemente influenzato dalla dimensione di alcune città metropolitane. Infatti, se alcuni sistemi locali come quelli di Milano e Bologna confermano i trend già



descritti, sono qui presenti delle aree geografiche (come quella associata alla città metropolitana di Roma) non tanto grazie ad una solida produzione brevettuale, quanto invece per via della grandezza, in termini di popolazione residente e quindi potenziali inventori, di questi territori. A maggior ragione, l'assenza in questo elenco delle più popolate realtà del Sud conferma la tendenza negativa di queste aree. La classifica è invece completata da sistemi locali situati principalmente in Emilia Romagna e nel Nord-est del Paese.

In aggiunta, in figura 3.5 si effettua la medesima valutazione ma eliminando la distorsione dovuta alla presenza di importanti differenze nella popolazione residente tra i diversi sistemi locali. Nonostante apparentemente non interessante in quanto, ad eccezione di Bologna, sono elencati solo di piccole dimensioni, questo grafico comunica una preziosa informazione: le aree presenti in questa classifica corrispondono infatti alle regioni dell'Emilia Romagna e del Veneto, confermandosi quindi traino fondamentale dell'economia del Paese. La presenza di sistemi locali di piccole dimensioni è giustificata dalla metrica con cui è stato costruito l'indicatore, che premia le realtà con un livello di popolazione residente più basso e, in secondo luogo, dalla forte presenza industriale in queste aree, tradizionalmente lontane dai più grandi centri abitati.

### 3.4.3 Analisi per classi tecnologiche

Definito il contesto temporale e spaziale del dataset, è certamente interessante valutare l'esistenza di settori tecnologici con un tasso di produzione brevettuale superiore alla media. Come anticipato, ciascun brevetto è classificato secondo l'IPC, suddivisa secondo diversi livelli di dettaglio. In figura 3.6 è presentato un primo istogramma che mostra come la produzione brevettuale dell'intero dataset sia fortemente sbilanciata a favore delle prime due sezioni, A e B, che nella classificazione originale riguardano, rispettivamente, *Human necessities* e *Performing operators; transporting*. Poiché queste categorie rappresentano più del 50% dei brevetti del campione, appare sensato approfondire, in particolare, quali sottosezioni contribuiscano maggiormente a formare questo dato.

In figura 3.7 sono quindi elencate, in ordine decrescente, le classi con il più alto numero di brevetti associati. Pur non completo per ragioni grafiche, il campione descritto con questo diagramma a barre rappresenta più del 95% dei brevetti del dataset (sezioni A e B). In appendice, per una più completa interpretazione dei grafici, si mostra l'elenco delle classi appartenenti alle sezioni A e B, con relativa descrizione del campo tecnologico di riferimento. Le prime quattro classi, rappresentative di quasi la metà delle categorie comprese in queste due sezioni, risultano essere:

- A61, *Medical or veterinary science; hygiene;*
- B65, *Conveying; packing; storing; handling thin or filamentary materia;*
- A47, *Furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in genera;*
- B60, *Vehicles in general.*

I risultati ottenuti sono in linea con le aspettative, in quanto i settori elencati condividono un fatturato non indifferente e un ruolo decisivo della produzione innovativa come

Figura 3.6: Produzione brevettuale delle 8 sezioni (classificazione IPC)

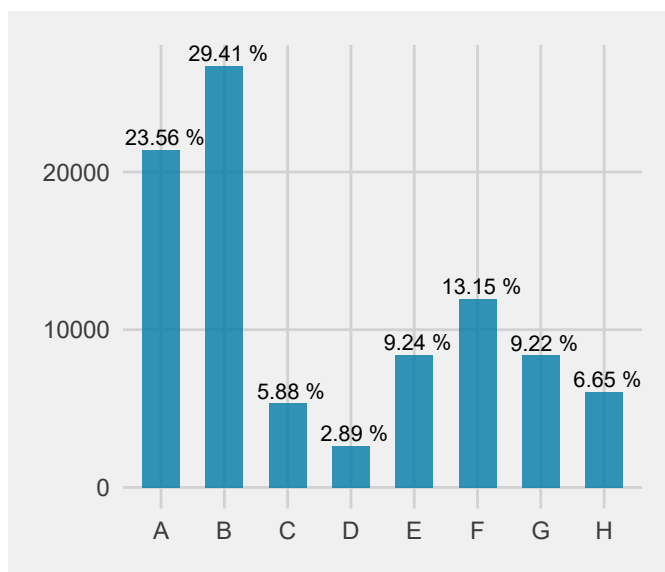
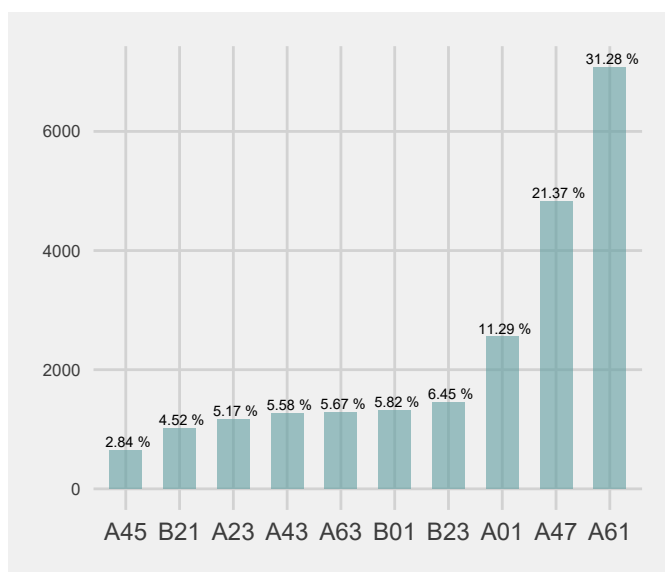


Figura 3.7: Brevetti associati alle principali classi (sezioni A e B, classificazione IPC)



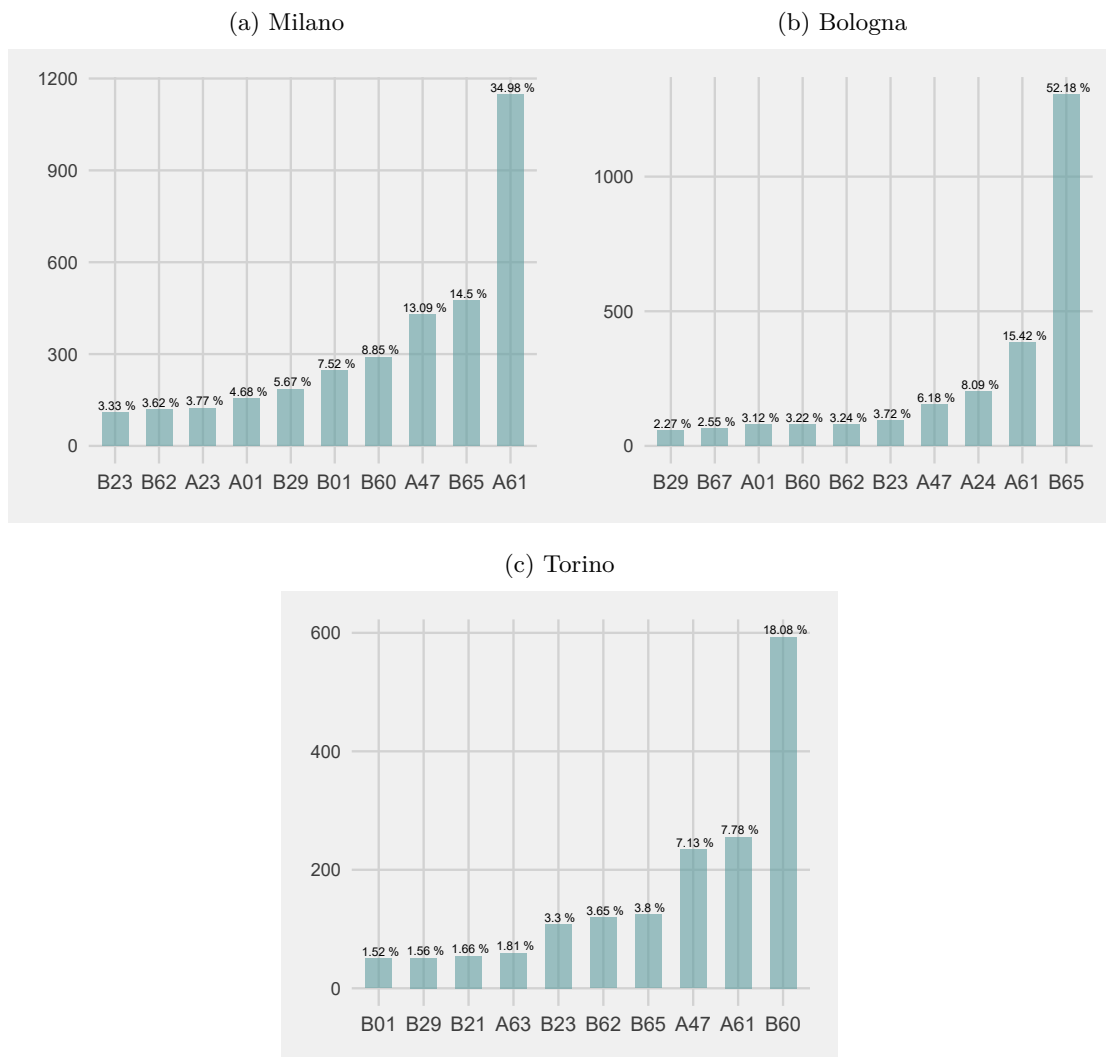
vantaggio competitivo rispetto agli altri operatori del mercato.

Poiché l'elaborato è realizzato con l'obiettivo di valutare congiuntamente l'importanza dell'aspetto spaziale e di quello tecnologico nella produzione innovativa, seppur a grandi linee per ragioni grafiche, è certamente utile descrivere la distribuzione tecnologica dei brevetti all'interno dei principali sistemi locali. I sili che in assoluto vantano la maggiore produzione brevettuale sono, in ordine decrescente, quelli associati alle città di Milano,

Bologna e Torino. In linea con il trend generale, anche in questo caso le prime due sezioni, A e B, appaiono quelle con il più alto numero di brevetti associati. In particolare, rispetto al valore di riferimento dell'intero campione (52.96%), per i tre sistemi locali che si sceglie di studiare, la percentuale di brevetti associati alle prime due sezioni si attesta, rispettivamente, al 42.39%, 65.26% e 48.94%.

Alla luce di questi dati, risulta opportuno focalizzare l'attenzione su queste due sezioni e sulle relative classi. In particolare, in figura 3.8 si confrontano i dati dei tre sistemi locali. Pur con numeri diversi, globalmente i sll di Milano e Torino rispecchiano l'andamento generale; il sistema locale associato al capoluogo piemontese presenta un picco di produzione brevettuale per la classe B60, *Vehicles in general*, in linea con la tradizione manifatturiera del territorio. Per quanto riguarda il sll di Bologna, spicca l'elevato numero di brevetti per la classe B65, *Conveying; packing; storing; handling thin or filamentary materia* e si aggiunge, alle già citate classi tecnologiche, A24 (*Tobacco; cigars; cigarettes; smokers' requisites*), con circa 200 brevetti e al terzo posto di questa classifica.

Figura 3.8: Produzione brevettuale rispetto alle classi tecnologiche





## Capitolo 4

# Metodologia

Diversamente da quanto stabilito dalla teoria neoclassica, secondo cui il sistema economico può essere interpretato valutando l'insieme dei comportamenti dei singoli individui, è ormai appurato che le azioni degli operatori devono essere considerate in relazione al contesto in cui avvengono. Pertanto, non è possibile affermare che il comportamento degli individui sia determinato solamente da un sistema di prezzi, poiché questo si verifica solamente in un mercato perfettamente concorrenziale, che tuttavia scompare con la presenza di un minimo di potere di mercato da parte di uno o più agenti. Successivamente, grazie al contributo della teoria dei giochi, una nuova teoria economica ha cercato di fornire un modello credibile anche in caso di concorrenza imperfetta. Tuttavia, anche in questo caso, alcune assunzioni hanno trovato scarso riscontro in applicazioni reali: la prima, presuppone che ciascun individuo possieda una quantità di informazioni tali per cui sia in grado di conoscere (e quindi prevedere) il comportamento degli altri operatori; la seconda teorizza che ciascun agente sia potenzialmente *raggiungibile*, anche in contesti di notevoli dimensioni. Definire un ambiente di natura spaziale o tecnologica al di fuori del quale non sia possibile per un agente economico operare, costituisce un passo fondamentale nella teoria economica e, in particolare, nell'economia della conoscenza. Alla luce di questo modello, risulta di non trascurabile importanza capire come sia formata questo *network* e quali proprietà abbia (Michael & Battiston, 2009).

### 4.1 Le co-assegnazioni brevettuali

Nel filone di ricerca che si occupa di studiare le esternalità di conoscenza e i meccanismi che le determinano, l'idea alla base è che agenti relativamente *vicini* (da un punto di vista principalmente geografico, ma non solo) possano beneficiare di *spillover* di conoscenza, che si traducono ad esempio in un aumento di produttività o di attività inventiva nel territorio oggetto di analisi. In altri termini, la possibilità di condividere delle informazioni con un altro operatore nel mercato, consente di ottenere un *output* maggiore, a parità di fattori di *input* nella generica funzione di produzione  $Y = f(L)$ . Questa teoria economica si contrappone al concetto di *non-fungible knowledge* espresso da Hidalgo (2022). Secondo

l'autore cileno, questa proprietà possiede profonde implicazioni, matematiche e concettuali. Limitando qui l'analisi agli aspetti puramente concettuali, un'idea di *conoscenza* così definita implica innanzitutto che questa non possa subire dei processi di aggregazione e che, in ultima analisi, non sia possibile affermare che *one plus one knowledge is two knowledges*. Ne deriva che questo concetto rinforza ulteriormente l'idea di barriere alla diffusione della conoscenza. Infatti, all'interno di un sistema economico così costruito, non basta *accumulare conoscenza*, bensì è necessario che siano presenti esattamente gli *elementi di conoscenza* funzionali e complementari al contesto analizzato.

Uno dei metodi più utilizzati in letteratura consiste nell'approssimare la condivisione di conoscenza tramite le co-assegnazioni brevettuali: è infatti comune che le imprese sviluppino congiuntamente un brevetto e che quindi la sua paternità sia condivisa tra più agenti. Lo sviluppo di questi modelli trova le sue radici nella teoria dei grafi, la disciplina che si occupa di studiare i grafi, oggetti in grado di rappresentare e descrivere una grande quantità di situazioni. Un grafo  $G$  è una coppia di elementi,  $G(V,E)$ , dove  $V(G)$  è un elenco di nodi o vertici e  $E(G)$  un insieme di archi. Tali connessioni possono essere:

- non orientate, dotate di una direzione, ma non di un verso;
- orientate, dotate di una direzione e di un verso;
- *pesate*, quando ad ogni collegamento è associato un *peso*, tipicamente un valore numerico.

Con l'obiettivo di costruire una rete che possa modellare le collaborazioni brevettuali nei sll italiani, i grafi realizzati rispettano le seguenti caratteristiche:

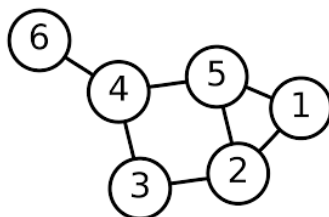
1. ciascun nodo rappresenta un inventore, identificato in maniera univoca dal codice fiscale identificativo del richiedente;
2. ciascun arco rappresenta un brevetto depositato congiuntamente presso l'UIBM da due o più richiedenti.

Dunque, utilizzando il campione costruito a valle delle operazioni preliminari sul dataset, si implementa un algoritmo per costruire la rete di co-assegnazioni. Il codice che di seguito si va a dettagliare è stato eseguito utilizzando R studio<sup>1</sup> (in appendice il codice implementato), software open source di analisi statistica. Come primo passo, si estraggono i record che abbiano almeno un duplicato all'interno del database, eliminando quindi i dati relativi a brevetti depositati da un unico richiedente. In secondo luogo, si provvede ad estrarre l'elenco dei codici fiscali, ciascuno contato una sola volta: questo vettore costituisce quindi l'insieme dei nodi del grafo. A questo punto, si crea una *edge list*, ovvero una lista di nodi tra loro legati da una richiesta di brevetto comune. É possibile ora costruire un grafo non orientato, che descriva la rete di co-assegnazioni del database a disposizione.

---

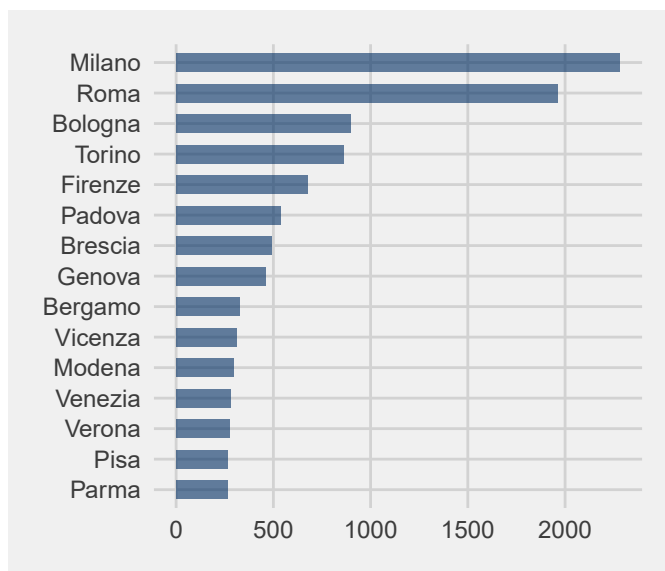
<sup>1</sup>RStudio è un ambiente di sviluppo integrato relativo al software madre R. Linguaggio di programmazione open-source, grazie ad opportune librerie consente di elaborare dati e effettuare analisi statistiche complesse.

Figura 4.1: Esempio di grafo non orientato (6 nodi, 7 archi)



Da un punto di vista metodologico, si costruisce per ogni anno il *network* delle collaborazioni brevettuali secondo un algoritmo che memorizza un collegamento (i.e., arco) tra inventori (i.e., nodi) per ogni uguaglianza tra i codici brevettuali univoci registrata. Inoltre, la logica del procedimento evita che lo stesso arco possa essere duplicato per lo stesso codice brevetto: qualora esista già per il codice brevetto *id1* un arco tra il nodo *i* e il nodo *j*, l'algoritmo non memorizza la relativa permutazione, ma prosegue l'iterazione. A questo punto, si ricavano dei grafi parziali, ciascuno riferito a un preciso sistema locale. Per svolgere questa operazione si utilizza la libreria *igraph* e in particolare la funzione *induced.subgraph*, che genera un grafo con tutti i nodi associati al sll per il periodo oggetto di valutazione e i soli archi che hanno il vertice di partenza e quello di arrivo nella lista dei nodi del sottografo indotto. In figura 4.2, si presentano i sistemi locali con il maggior numero di collaborazioni brevettuali nel periodo considerato.

Figura 4.2: I sistemi locali con il maggior numero di collaborazioni brevettuali



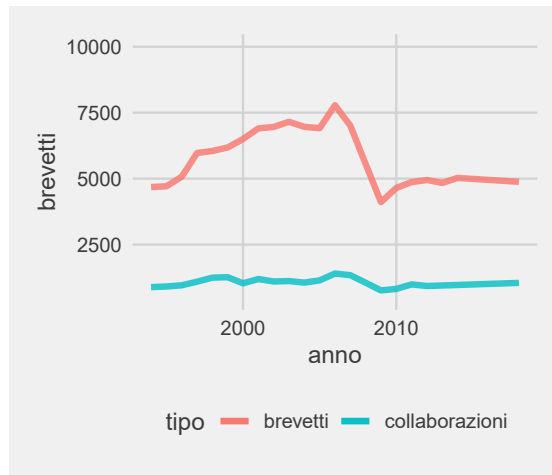
I risultati non si discostano in maniera evidente da quelli proposti in figura 3.4: ad eccezione di Pisa, Venezia e Verona, infatti, la classifica dei 15 sll con il maggior numero



di collaborazioni brevettuali è in linea con i dati delle domande di brevetti accettate dall'UIBM. A tal proposito, si conferma il trend secondo cui la maggior parte dell'attività inventiva italiana è concentrata in specifiche partizioni funzionali del territorio (Paci & Usai, 2000).

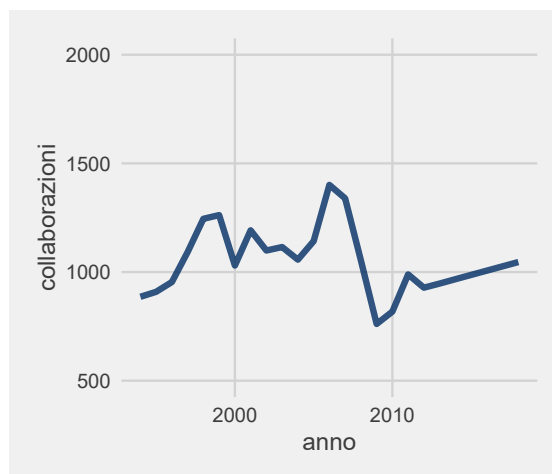
Si confronta ora l'andamento temporale delle richieste brevettuali approvate con il numero di collaborazioni attive nello stesso periodo (figura 4.3). Il grafico mostra come il

Figura 4.3: Brevetti vs collaborazioni



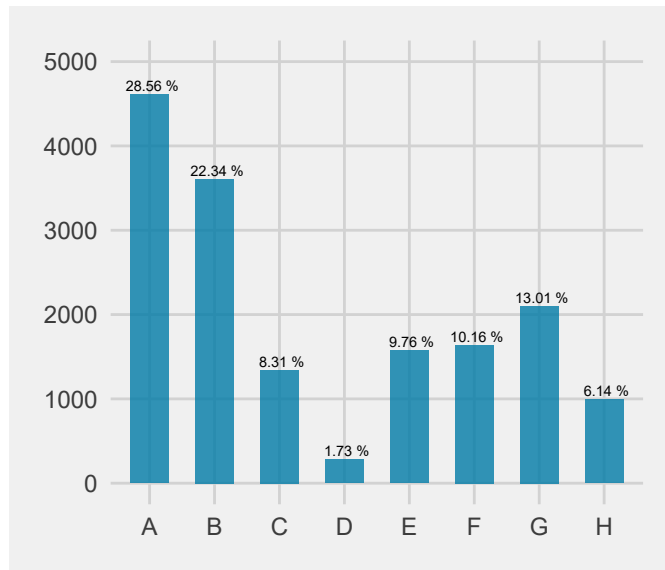
numero delle co-assegnazioni sia come atteso significativamente e costantemente inferiore al numero di brevetti approvati nello stesso anno, con una differenza tra le due categorie che raggiunge il valore massimo nel 2006 (6382 unità). Tuttavia, per poter analizzare correttamente il trend temporale delle collaborazioni, è necessario adottare una scala che possa consentire una migliore visualizzazione, che si può apprezzare in figura 4.4.

Figura 4.4: Andamento temporale delle collaborazioni brevettuali



In particolare, si registra una brusca inversione di tendenza a partire dal 2006, anno in cui l'UIBM ha modificato i criteri di valutazione delle richieste brevettuali. Infine, si completa l'analisi di questo sottocampione con una valutazione delle classi tecnologiche prevalentemente presenti, secondo la classificazione IPC e a livello delle sezioni (figura 4.5). I risultati sono complessivamente in linea con quanto proposto in figura 3.6, confermando quindi l'importanza relativa delle sezioni A e B.

Figura 4.5: Analisi per sezioni tecnologiche delle collaborazioni brevettuali



## 4.2 Costruzione delle variabili

Costruito il campione, si introducono le variabili utilizzate nel modello di regressione. Dopo la definizione della variabile dipendente, si esplicitano i regressori, con riferimento puntuale alla letteratura economica e calcolati per ciascun periodo  $t$  (i.e., anno) e per ogni  $i$ -esimo sll. Il capitolo si conclude con la descrizione delle variabili di controllo.

### 4.2.1 Brevetti pro-capite

In virtù delle riflessioni presentate nei capitoli precedenti, si sceglie di utilizzare la produzione brevettuale come variabile dipendente del modello. In particolare, per ciascun anno sono estratte le domande di brevetto che sono state valutate positivamente dall'UIBM. Infine, si sceglie di normalizzare questi dati rispetto alla popolazione residente. Il risultato di queste trasformazioni consiste in un insieme di vettori (i.e., uno per ogni anno analizzato), ciascuno con 611 valori, che indicano il numero di brevetti pro-capite associati al sistema locale di riferimento.

### 4.2.2 Densità

Il primo aspetto considerato nella maggior parte dei lavori che si occupano di studiare le reti sociali è certamente il livello di connessione del grafo, ovvero capire quanti vertici sono tra loro collegati rispetto al numero totale di archi possibili. La misura che meglio descrive questo concetto è la densità del grafo, che matematicamente è così espressa:

$$D_{it} = \frac{e_{it}}{(N_{it} * (N_{it} - 1)/2)}, \quad (4.1)$$

dove  $e_{it}$  rappresenta il numero di archi esistenti e  $N_{it}$  il numero di nodi: infatti, fissato il numero di vertici  $N$ , sono possibili  $(N*(N-1)/2)$  collegamenti. Per costruzione, questa misura varia nell'intervallo  $[0,1]$  e a valori più alti dell'indice corrisponde una rete altamente connessa.

In un primo momento la letteratura economica ha ritenuto che all'interno di una rete con una densità elevata fosse più semplice il trasferimento di informazioni e di conoscenza. Se questa convinzione trova ancora riscontro nei più recenti lavori sul tema (Breschi & Lenzi, 2016), alcuni lavori empirici hanno invece evidenziato un trend opposto: un *network* altamente connesso può avere un impatto negativo sull'attività inventiva del territorio. Secondo Lobo e Strumsky (2008) e Innocenti et al. (2020), infatti, il rischio è che circoli un'informazione ormai ridondante e non più utile per migliorare i processi innovativi esistenti.

### 4.2.3 Coesione interna

Una delle più comuni misure di rete a livello locale è senza dubbio la coesione interna. Questo parametro valuta l'esistenza di sottogruppi di individui altamente connessi tra loro. Secondo la proposta di Wasserman, Faust et al. (1994), questa variabile può essere modellata usando la decomposizione *k-cores*: si applica cioè un algoritmo in grado di

decomporre il grafo in sottogruppi tali per cui ciascun sottinsieme sia composto solamente da nodi connessi ad un certo numero  $k$  di altri nodi. L'idea di Innocenti et al. (2020), qui ripresa, è di calcolare il livello di coesione interna di un grafo come segue:

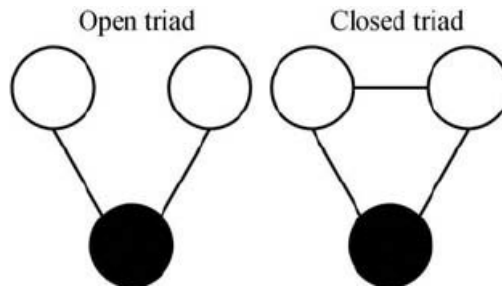
$$C_{it} = \max(k - cores_{it}), \quad (4.2)$$

concentrando quindi l'analisi sul numero massimo di collegamenti ripetuti nel grafo. Confermando i lavori empirici di Fornahl et al. (2011) e Innocenti et al. (2020), ci si attende un impatto positivo di questa variabile sull'attività inventiva.

#### 4.2.4 Transitività

Un'altra proprietà attraverso cui è possibile valutare una porzione di rete è una misura di prossimità sociale o, in altri termini, di transitività. In un grafo non orientato questo coefficiente corrisponde al rapporto tra le coppie di nodi che condividono un legame con uno stesso nodo (i.e., i *triangoli Closed triad* in figura 4.6) e l'insieme delle triplete di nodi esistenti (i.e., *Open triad*).

Figura 4.6: *Open triad* e *closed triad* a confronto



La fiducia tra individui prossimi nel grafo riveste qui un ruolo decisivo nella condivisione di conoscenza. In letteratura si evidenzia un impatto positivo di questa variabile sui processi innovativi (Breschi & Lenzi, 2016). Nel contributo di Broekel e Boschma (2012), pur riconoscendo l'importanza di questo fattore per spiegare la trasmissione di conoscenza all'interno del settore aeronautico olandese, si sottolinea come il beneficio marginale garantito da questa variabile sia decrescente con l'aumentare del suo valore. In altri termini, una prossimità sociale eccessivamente elevata non garantisce un significativo aumento del livello di attività inventiva del territorio.

#### 4.2.5 Quota di richiedenti rispetto alla popolazione

Un'altra proprietà spesso analizzata da questo filone di ricerca è senza dubbio la concentrazione degli inventori nel territorio. In questo lavoro, in particolare, si calcola la quota dei richiedenti (i.e., *applicants*) presenti rispetto alla popolazione residente nell' $i$ -esimo sll nell'ultimo censimento disponibile:

$$SH_{it} = \frac{applicants_{it}}{(Population_i)} \quad (4.3)$$

Non è tuttavia chiaro se questa misura possa risultare più importante, ai fini della produzione innovativa, rispetto alla struttura di rete del territorio. Pur concordi sull'idea che una concentrazione più elevata favorisca i processi innovativi, alcuni autori ritengono che siano altre le proprietà da attenzionare, in Breschi e Lenzi (2016) la prossimità sociale, mentre altri documentano una tendenza opposta, come Lobo e Strumsky (2008) e Innocenti et al. (2020): questi studi sottolineano il peso relativo dei processi di *agglomeration economies*<sup>2</sup>.

#### 4.2.6 Collegamenti esterni

Riprendendo l'idea che una rete eccessivamente connessa determini un *network* stagnante, per estensione è possibile immaginare che lo stesso possa accadere nel caso di flussi di conoscenza limitati ai confini territoriali e, in particolare, ai singoli sistemi locali. La letteratura economica offre in questo senso buoni risultati, confermando che le connessioni esterne al sistema di riferimento consentono di sviluppare tecnologie che sarebbero altrimenti rimaste sconosciute. Di particolare interesse appare il lavoro di Laursen e Salter (2006), i quali introducono due forme di connessioni esterne, che misurano rispettivamente l'"ampiezza" (i.e., *Breadth*) e la "profondità" (i.e., *Depth*) di questi collegamenti. In questo lavoro la prima variabile è calcolata come rapporto tra i collegamenti esterni (i.e., *ext*) alla rete del sistema locale *i*-esimo e il numero totale di collegamenti di quel SLL:

$$BH_{it} = \frac{ext_{it}}{(e_{it})} \quad (4.4)$$

Nel secondo caso, invece, è necessario valutare quanto questi collegamenti siano concentrati verso specifiche aree. Servendosi quindi della definizione dell'indice *HHI*<sup>3</sup> si calcola la "quota di collaborazione" (*LS*) della rete associata all'*i*-esimo sistema locale rispetto al totale dei collegamenti esterni. La "profondità" delle connessioni esterne alla rete sarà quindi calcolata come segue:

$$DH_{it} = \sum_{k=0}^n \frac{(LS_k)_{it}^2}{(BH_{it})} \quad (4.6)$$

Il più recente studio che applica queste due variabili per studiare l'attività inventiva di un territorio è quello di Innocenti et al. (2020): entrambe le misure mostrano un risultato significativo nel modello di regressione, confermando che una rete che abbia dei collegamenti con l'esterno risulti più innovativa e, in ultima analisi, vanti una produttività maggiore.

<sup>2</sup>Nella letteratura economica, locuzione utilizzata per indicare ad esempio, la concentrazione degli inventori in un territorio, proprietà spesso analizzata

<sup>3</sup>Indice di Herfindahl–Hirschman, solitamente utilizzato per calcolare, attraverso le quote di mercato in possesso di ciascun agente, il livello di concorrenza di specifici settori:

$$HHI = \sum_{k=0}^n (MS_k)^2, \quad (4.5)$$

dove  $MS_k$  rappresenta la quota di mercato di un agente.

### 4.2.7 Struttura industriale: la *variety*

Inserendosi nel complesso dibattito su quale tipologia di esternalità sia prevalente (i.e., in ultima analisi se quella di Marshall-Arrow-Romer o quella di Jacobs), il lavoro innovativo proposto da Frenken et al. (2007) sottolinea come le esternalità di conoscenza studiate da Jacobs siano più efficaci tra settori che condividono tecnologie simili. È quindi introdotta una nuova variabile, in grado di catturare questo fenomeno: la *related variety* (i.e., varietà correlata). Lo studio è quindi completato dall'analisi di una seconda misura, la varietà non correlata, che invece valuta l'impatto della diversità tra settori sull'attività inventiva del territorio. Successivamente, diversi lavori hanno ripreso queste definizioni per le proprie analisi empiriche. Di particolare interesse per questo elaborato risulta lo studio di Castaldi et al. (2015) in quanto, utilizzando dati brevettuali statunitensi del periodo 1977-1999, indagano il rapporto tra le due variabili e l'*output* innovativo di una regione. I risultati mostrano una relazione positiva per entrambe le ipotesi, garantendo ulteriormente circa l'adeguatezza di queste misure. Se immaginare il beneficio di elevati livelli di varietà correlata appare più semplice e intuitivo, l'ipotesi degli autori è che spesso la varietà non correlata risulti perfino più importante. Infatti, anche se si tratta di un fenomeno più raro, quando tecnologie precedentemente tra loro non correlate subiscono un processo di trasformazione tale da essere tra loro legate, l'invenzione che ne deriva rappresenta il più delle volte una scoperta fondamentale. È il caso del settore dell'autoveicolo, dove le tecnologie del motore, della bicicletta e della carrozza, che non condividevano informazioni rilevanti, sono diventate durante il processo di sviluppo dell'automobile fortemente connesse.

In letteratura, queste variabili sono tradizionalmente definite con delle misura di entropia, lasciando sottintese alcune ipotesi:

1. una misura di entropia è in grado di catturare la varietà valutando l'*incertezza* della distribuzione di probabilità (Castaldi et al., 2015);
2. l'entropia generale può essere scomposta a diversi livelli di aggregazione (Frenken et al., 2007).

In termini generali, definito  $E_f$  l'evento che in una regione si brevetti nel campo tecnologico  $f$  e definita  $p_f$  la probabilità che questo evento si verifichi, il livello generale di entropia è:

$$H = \sum_{f=0}^n p_f \ln\left(\frac{1}{p_f}\right), \quad (4.7)$$

che varia in un intervallo  $[0, \ln(p)]$ ;  $H$  ha valore 0 se per un certo  $\text{textit}p_f$  è pari a 1 e pari a zero negli altri casi. Applicando questi concetti all'analisi brevettuale, si tratta del caso in cui all'interno di un certo territorio siano presenti brevetti appartenenti ad un'unica classe tecnologica; al contrario,  $H$  ha valore massimo quando tutti i valori  $p_f$  risultano tra loro uguali.

Definito quindi il contesto in cui si intende operare e le fonti di riferimento, si introducono le variabili da utilizzare per descrivere la struttura industriale del dataset:

$$UV_{it} = \sum_{s=1}^8 p_{s,it} \ln\left(\frac{1}{p_{s,it}}\right) \quad (4.8)$$

$$SRV_{it} = \sum_{c=1}^{131} p_{c, it} \ln\left(\frac{1}{p_{c, it}}\right) - \sum_{s=1}^8 p_{s, it} \ln\left(\frac{1}{p_{s, it}}\right) \quad (4.9)$$

$$RV_{it} = \sum_{sc=1}^{645} p_{sc, it} \ln\left(\frac{1}{p_{sc, it}}\right) - \sum_{c=1}^{131} p_{c, it} \ln\left(\frac{1}{p_{c, it}}\right) \quad (4.10)$$

L'equazione 4.8 rappresenta la varietà non correlata, definita come misura di entropia a livello delle 8 sezioni tecnologiche (A,B, ..., H) della classificazione IPC. Un valore elevato di questo parametro indica un notevole livello di diversificazione, rispetto a queste categorie, tra i brevetti rilasciati dall'UIBM.

La seconda misura (4.9) è una variante proposta da Castaldi et al. (2015), definita *semirelated variety*, che è possibile determinare grazie al teorema di decomposizione dell'entropia presentato da Theil (1972). Lo scopo che si vuole raggiungere inserendo questa variabile è capire quanto sia importante la varietà industriale di un territorio a un livello di dettaglio intermedio (i.e., 3 digit e quindi in corrispondenza delle classi tecnologiche) per spiegarne la *performance* innovativa.

Il terzo parametro (4.10) è invece la varietà correlata, calcolata seguendo il medesimo teorema, ma ad un livello di dettaglio maggiore (i.e., 4 digit).

Si precisa che per i primi anni (1994-1999) non sono presenti le informazioni sul codice IPC associato a ciascun brevetto e risulta quindi impossibile calcolare queste variabili. Ad eccezione di pochi casi, circoscritti solitamente a sistemi locali di piccole dimensioni, i dataset costruiti sono privi di valori nulli e consentono quindi un'analisi puntuale della struttura industriale italiana.

Alla luce dei risultati offerti dalla letteratura (Aarstad et al., 2016; Castaldi et al., 2015; Frenken et al., 2007; Innocenti et al., 2020; Tavassoli & Carbonara, 2014), ci si attende che queste variabili abbiano un impatto significativo sul numero di brevetti pro-capite, variabile dipendente del modello.

#### 4.2.8 Variabili di controllo

Il modello econometrico è quindi arricchito da una serie di variabili di controllo, seguendo l'impostazione del lavoro di Innocenti et al. (2020). Una prima misura è il *gross domestic product (gdp)* [Mln € per 100k abitanti], disponibile a livello provinciale per ciascun anno di nostro interesse e utilizzabile come buona approssimazione dello sviluppo economico generale di un' area geografica. Si procede quindi a riscaldare questo parametro utilizzando i livelli di popolazione dei 611 sll. Si integra quindi con due ulteriori variabili, disponibili in Eurostat ma solamente per il livello NUTS 2, che mostrano rispettivamente informazioni sul *capitale umano*, definite dalla percentuale di residenti con un'istruzione di livello universitario, e le *spese in ricerca e sviluppo* [Mln € per 100k abitanti], riscaldate anche in questo caso rispetto alla popolazione residente in ciascun sistema locale. Infine, grazie ai dati offerti dall'Istat, si aggiunge un controllo per la *densità di popolazione* di ogni sll.

## 4.3 Le proprietà della rete

Dopo aver definito il campione oggetto di studio e le principali variabili che saranno utilizzate per le analisi econometriche, un primo risultato interessante dell'elaborato consiste nell'esaminare alcune di queste proprietà per evidenziarne le principali caratteristiche. In questa direzione, è opportuno valutare innanzitutto la struttura delle reti di collaborazioni brevettuali. Con riferimento alla tassonomia proposta da Jackson et al. (2017), è utile dividere le caratteristiche di un *network* in due categorie:

- le proprietà ad un livello *macro*, *global* o *aggregato*;
- le proprietà ad una scala *micro*, *locale* o *individuale*.

Nonostante non sia semplice effettuare una distinzione netta, è necessario approcciare le due categorie in maniera differente. Infatti, oltre ad applicare criteri metodologici diversi per la raccolta e l'analisi dei dati, queste due classi sono concettualmente diverse. Applicando questa intuizione ad un'analisi che studia la diffusione di un'infezione, le proprietà *macro* consentono di valutare complessivamente il processo di contagio, quelle *micro*, invece, focalizzano l'attenzione sui singoli individui o, al più, su piccole porzioni di rete e sulla loro importanza relativa rispetto al processo globale. Pertanto, con l'obiettivo di rappresentare in modo completo le reti di co-assegnazione brevettuale e impiegando i dati ottenuti dalla costruzione dei regressori, si sceglie di approfondire le seguenti proprietà:

- Proprietà *macro*:
  1. densità della rete;
  2. coefficiente assortativo della rete.
- Proprietà *micro*:
  1. coefficiente di *clustering* o misura di transitività;
  2. coesione interna, metodo *k-cores*;
  3. misura di centralità, in particolare quella che in letteratura è definita *betweenness centrality*.

Come sottolineato anche da Jackson et al. (2017), poiché le reti sociali sono un oggetto estremamente complesso, esistono proprietà alternative che potrebbero essere prese in considerazione. La scelta è dunque sostanzialmente così giustificata:

- le proprietà citate risultano essere le più comuni nelle analisi delle reti sociali e pertanto quelle da cui ci si attende i risultati più significativi;
- alcune possibili variabili (e.g., omofilia tra nodi rispetto ad alcune caratteristiche, come genere o etnia) non sono ritenute utili ai fini della presente trattazione;
- alcune proprietà sono tra loro strettamente legate: la distribuzione dei *gradi* dei singoli nodi è ad esempio sintetizzata dalla misura di densità, che rappresenta in ultima analisi una *media* dei *gradi* dell'intera rete.



Una delle prime caratteristiche che emerge negli studi delle reti sociali è senza dubbio la densità che, come definito nell'equazione 4.1, è una misura calcolata sulla totalità della rete come rapporto tra i collegamenti presenti e quelli possibili. Poiché obiettivo dell'analisi econometrica sarà valutare l'andamento della produzione brevettuale italiana secondo il duplice aspetto temporale (anni) e spaziale (sll), si costruisce un dataset con tante righe quanti sono i sistemi locali oggetto di analisi e tante colonne quanti sono gli anni di riferimento. In tabella 4.1 si mostra un estratto di questa tabella: il dataset completo è costituito da 611 righe e 22 colonne, che fanno rispettivamente riferimento all'elenco dei sistemi locali italiani e al periodo scelto per queste analisi (i.e., 1994-2018, fatta eccezione per il triennio 2015-2017, come spiegato nel capitolo precedente).

Tabella 4.1: Densità della rete per ogni anno e sll disponibili: un estratto

	1994	1995	1996	1997	1998	1999	2000	2001	...
Chieri	0.00	0.00		0.00	0.00	0.00	0.00	0.17	...
Ivrea	0.20	0.20	0.17	0.17	0.00		0.33	0.17	...
Pinerolo	0.10	0.00	0.13		0.11	0.00	2.00	0.17	...
Rivarolo Canavese			0.00	0.00			0.17		...
Susa									...
Torino	0.04	0.04	0.02	0.02	0.03	0.02	0.03	0.03	...
Borgosesia				0.10	0.30	0.00			...
Santhià	0.00	0.00			0.13		0.00	0.00	...
Vercelli									...
Borgomanero	0.00	0.00	0.13		0.00	0.00		0.00	...

Si chiarisce che i valori *0.00* in tabella 4.1 si riferiscono a reti in cui è presente almeno un nodo, ma nessun arco; i campi nulli sono invece relativi a sistemi locali con nessun nodo presente (i.e., nessun richiedente associato a quel sll che abbia brevettato durante quello specifico anno). Come appare già nel breve estratto presentato, la maggior parte dei campi di questa matrice è composta da zeri (15.34 %) o da valore nulli (67.57 %): significa che più dell'80 % delle reti costruite non comprende alcun arco al suo interno. Questi dati sono certamente influenzati dalla scelta dell'unità territoriale di riferimento: infatti, preferire un livello di dettaglio maggiore (i.e., i sll) ad una partizione diversa, come ad esempio quella regionale, determina inevitabilmente delle reti più piccole e con meno collegamenti. Per questi motivi si sceglie in primo luogo di restringere l'analisi ai sll che, nel periodo temporale considerato, abbiano una densità non nulla almeno nel 50 % dei casi (nel periodo temporale considerato). Il risultato di questa partizione genera un dataset composto da 57 sistemi locali, principalmente associati alle più grandi realtà urbane del Paese. Infatti, tra le 30 città più popolose d'Italia (dati Istat) solo 7 non sono presenti in questa partizione (i.e., Napoli, Messina, Taranto, Reggio Calabria, Ravenna, Foggia e Salerno, in ordine decrescente per numero di residenti). Ancora una volta, si registra un dato negativo principalmente associato alle città del Mezzogiorno, sia per una produzione brevettuale non sufficiente, sia per l'assenza di una rete solida di collaborazioni brevettuali. Con l'obiettivo di poter rappresentare un grafico comprensibile e che rappresenti reti significative, si effettua una ulteriore partizione, conservando unicamente

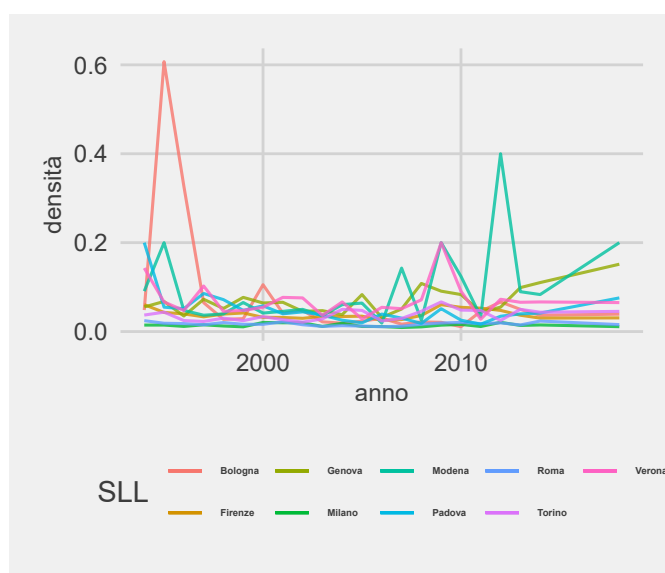
i dati dei sistemi locali con valori sempre non nulli negli anni considerati. In tabella 4.2 sono mostrati i risultati di questa operazione, qui con arrotondamento alla seconda cifra decimale più vicina.

Tabella 4.2: Densità della rete dei principali sl

	Bologna	Firenze	Genova	Milano	Modena	Padova	Roma	Torino	Verona
1994	0.05	0.06	0.06	0.01	0.09	0.20	0.02	0.04	0.14
1995	0.61	0.04	0.07	0.01	0.20	0.05	0.02	0.04	0.07
1996	0.33	0.04	0.04	0.01	0.05	0.05	0.02	0.02	0.05
1997	0.07	0.03	0.07	0.02	0.04	0.09	0.02	0.02	0.10
1998	0.03	0.04	0.05	0.01	0.04	0.07	0.02	0.03	0.05
1999	0.03	0.04	0.08	0.01	0.07	0.05	0.02	0.02	0.05
2000	0.11	0.03	0.06	0.02	0.04	0.06	0.02	0.03	0.05
2001	0.04	0.03	0.07	0.02	0.05	0.04	0.02	0.03	0.08
2002	0.05	0.03	0.05	0.02	0.05	0.04	0.02	0.02	0.08
2003	0.02	0.03	0.05	0.01	0.04	0.04	0.01	0.03	0.04
2004	0.02	0.03	0.04	0.02	0.06	0.03	0.01	0.05	0.07
2005	0.02	0.03	0.08	0.01	0.06	0.02	0.01	0.05	0.03
2006	0.03	0.03	0.03	0.01	0.02	0.04	0.01	0.02	0.05
2007	0.02	0.03	0.05	0.01	0.14	0.03	0.01	0.03	0.05
2008	0.02	0.04	0.11	0.01	0.03	0.02	0.02	0.05	0.07
2009	0.02	0.06	0.09	0.01	0.20	0.05	0.02	0.07	0.20
2010	0.01	0.05	0.08	0.02	0.12	0.03	0.02	0.05	0.09
2011	0.05	0.05	0.04	0.01	0.03	0.02	0.02	0.05	0.03
2012	0.07	0.05	0.05	0.02	0.40	0.03	0.02	0.02	0.07
2013	0.05	0.04	0.10	0.01	0.09	0.04	0.01	0.05	0.07
2014	0.04	0.03	0.11	0.01	0.08	0.04	0.02	0.04	0.07
2018	0.04	0.03	0.15	0.01	0.20	0.08	0.02	0.05	0.07

Analogamente, in figura 4.7, si descrive con un grafico l'andamento temporale di questa proprietà.

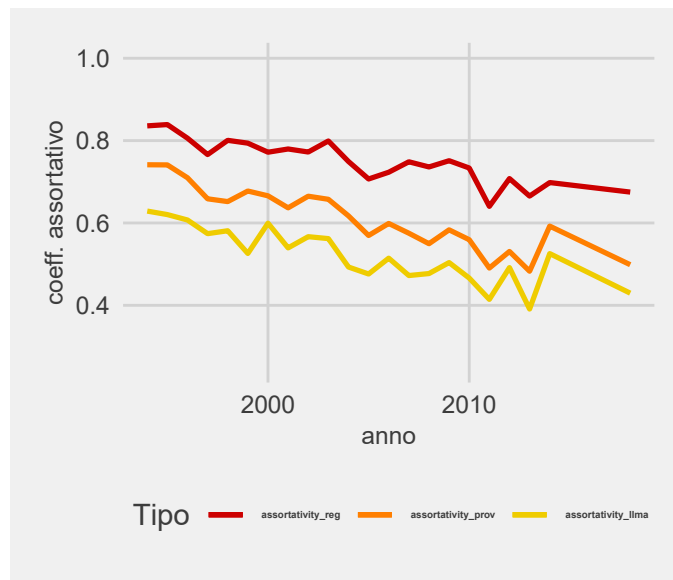
Figura 4.7: Analisi temporale della densità della rete dei principali sl



Ciò che emerge con chiarezza, confermando un trend osservabile anche nel campione completo, è che al crescere della dimensione del sistema locale (e quindi solitamente della sua rete), diminuisce la varianza del dato lungo gli anni presi in considerazione. Le reti che conservano nel tempo un valore di densità pressoché costante sono infatti quelle relative ai sll di Roma e Milano. Diversamente, in particolare i *network* di Bologna, Genova, Padova e Verona mostrano un andamento più variabile. Questa evidenza empirica trova spiegazione nell'idea che all'interno di reti più piccole, per costruzione (si veda 4.1), la presenza (o l'assenza) di un collegamento o di un nodo tra un anno e l'altro comporta una maggiore variazione della misura. Nonostante nel periodo considerato risulti uno dei sistemi locali con il maggior numero di collaborazioni brevettuali (4.2), il caso più emblematico riguarda il sistema locale di Bologna, che per l'anno 1995 presenta un valore di densità pari a 0.61: si tratta infatti di una rete con appena 7 nodi, rispetto ai 26 dell'anno precedente.

Un'altra misura spesso usata nello studio delle reti sociali è il coefficiente assortativo. Questa proprietà descrive la tendenza dei nodi a legare (o meno) con altri dalle proprietà simili (o differenti). È quindi una variabile versatile e che consente diverse applicazioni. Questo termine varia nell'intervallo  $[-1,1]$ : il limite inferiore rappresenta una rete in cui i nodi tendono a legarsi a vertici con caratteristiche diverse dalle proprie; accade il contrario quando il coefficiente assortativo è pari a 1. In questo elaborato si intende utilizzare questa variabile per comprendere la tendenza media dei nodi a formare collegamenti all'interno della stessa unità territoriale. Per raggiungere questo scopo, si è calcolato il coefficiente assortativo della rete fissando un periodo temporale di riferimento e utilizzando tre diversi ordini di grandezza: regionale, provinciale e a livello di sll. I risultati di questa analisi sono presentati in figura 4.8.

Figura 4.8: Analisi temporale del coefficiente assortativo a tre diversi livelli spaziali



Emerge con chiarezza (e in verità senza sorpresa) che il valore del coefficiente assortativo è

inversamente proporzionale al livello di dettaglio geografico preso in considerazione. Infatti, scegliendo come riferimento un anno centrale (i.e., 2006) rispetto al periodo di tempo analizzato, il coefficiente assortativo a livello regionale risulta 0.72, a livello provinciale 0.6 e a livello di sistema locale 0.51. Questo trend, oltre a sottolineare una prevedibile propensione dei nodi delle reti a legarsi ad altri che condividono la stessa area geografica, indica come sia più semplice che questo legame avvenga nella stessa regione, meno a livelli di dettaglio maggiori. In secondo luogo, questo grafico mostra una globale decrescita del coefficiente assortativo con il passare del tempo. A parità di altri fattori, questo andamento può essere giustificato da una maggiore inclinazione e/o facilità a condividere la brevettazione di un nuovo prodotto con realtà fisicamente più lontane, secondo un recente processo di *globalizzazione* delle reti di collaborazione brevettuale.

Proseguendo la descrizione delle reti, si presentano ora alcune proprietà in grado di descrivere il *network* ad un livello di dettaglio maggiore, concentrandosi sul comportamento e il ruolo dei singoli nodi. Una di queste è senza dubbio la misura di transitività dei nodi della rete, vale a dire la frequenza con cui le coppie di nodi  $j$  e  $k$  adiacenti ad uno specifico nodo  $i$  risultano anch'essi tra loro legati. Sfortunatamente, nonostante l'importanza che questa variabile ricopre nello studio delle reti sociali, come mostrato ad esempio in Breschi e Lenzi (2016) e Broekel e Boschma (2012), la scelta di studiare le reti secondo l'area funzionale dei sistemi locali conduce in questo caso a risultati poco interessanti. Infatti, più del 70 % dei SLL mostra un coefficiente di transitività pari a zero nell'intero periodo considerato. Le reti con un valore diverso da zero in almeno la metà degli anni considerati sono, in ordine esclusivamente alfabetico, Bologna, Firenze, Genova, Milano, Padova, Palermo, Roma e Torino. Rispetto alle analisi condotte sulla densità, ad eccezione del sistema locale di Palermo, che qui rappresenta una novità in contrasto con i risultati fino ad ora ottenuti, le poche realtà che offrono spunti interessanti sono i sistemi locali di dimensione maggiore e prevalentemente localizzati nel Nord del Paese. In tabella 4.3 si mostrano i risultati non nulli relativi ad alcuni di questi territori.

Tabella 4.3: Misura di transitività dei principali sll

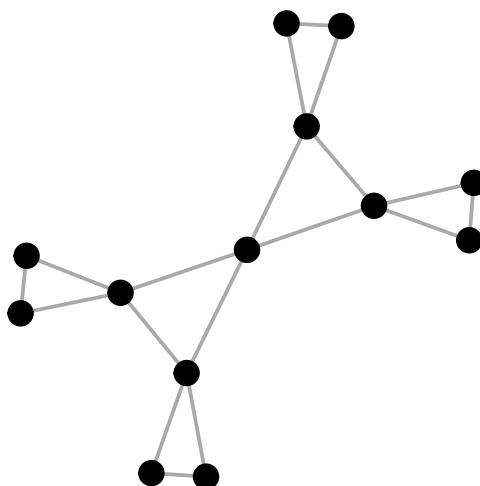
llma	1994	1995	1996	1998	2000	2001	2002	2003	2004	2006	2007	2008	2011	2012	2013	2018
Firenze	1.00	0.86	0.86	1.00	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60
Milano	0.42	0.63	0.77	0.60	0.97	0.94	0.95	0.75	0.75	0.96	0.56	0.46	0.50	0.98	0.92	0.27
Roma	0.90	0.95	1.00	0.99	1.00	0.82	0.95	0.86	0.50	0.93	1.00	1.00	0.69	0.89	0.77	0.81
Torino	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.82	1.00	0.82	1.00	0.92	0.87

Considerando che una misura così calcolata varia nell'intervallo  $[0,1]$ , i valori in tabella sono molto spesso uguali o prossimi al limite superiore, indicando che pur tra i pochi collegamenti della rete, fissato un nodo, il più delle volte i nodi adiacenti sono tra loro connessi.

Accanto alla scelta di un'area geografica non eccessivamente ampia, è possibile ipotizzare una seconda motivazione per motivare dei risultati così poco interessanti. In letteratura, infatti, il coefficiente di transitività di una rete è calcolato anche in modi alternativi. Definita una coppia di nodi  $i$  e  $j$ , una seconda misura valuta se esista almeno un nodo  $k$  connesso ad entrambi. Per quanto tra loro concettualmente simili, i due criteri possono condurre a risultati differenti. Pur rappresentando entrambi una misura di transitività, in Jackson et al. (2017) sono distinti con i termini "*clustering*" (la misura utilizzata in questo

elaborato) e "support" (criterio alternativo). In figura 4.9 si presenta una rete che pur avendo un coefficiente di transitività del secondo tipo pari a 1, mostra invece una misura ben inferiore a 1 secondo la definizione "clustering". Questa singolare applicazione può quindi ulteriormente giustificare i risultati poco incoraggianti ottenuti.

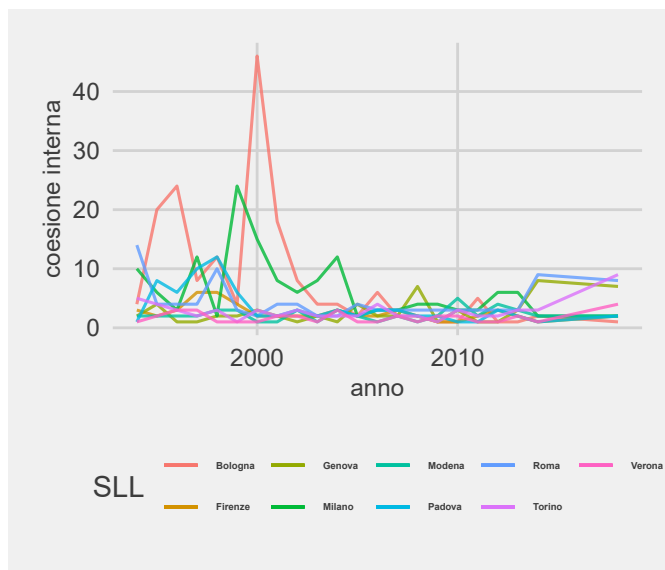
Figura 4.9: Rete con *support* pari a 1, ma coefficiente di *clustering* molto minore di 1



Un altro aspetto importante nello studio delle reti a livello *micro* consiste nel valutare la presenza di partizioni della rete con una frequenza di collegamenti superiore alla media. Questa proprietà è definita coesione interna e può essere misurata utilizzando il metodo *k-cores*. Questo procedimento prevede di raggruppare quei nodi che condividono lo stesso grado  $k^4$ . In particolare, è possibile conoscere il numero massimo di collegamenti ripetuti in una certa area geografica. Si tratta quindi di una misura che può assumere qualunque valore intero positivo. È verosimile che i valori più alti siano associati alle reti dei principali sistemi locali. Costruendo questa variabile a livello di sll, anche in questo caso sono presenti diversi valori nulli (più del 98 % del campione dei sistemi locali mostra almeno un valore nullo nel periodo di tempo considerato): è il caso delle reti con nessun collegamento o addirittura nessun nodo al loro interno. Anche in questa occasione, si provvede ad analizzare le reti che presentano i valori più interessanti: si tratta di Bologna, Firenze, Genova, Milano, Modena, Padova, Roma, Torino e Verona. Ancora una volta, sono elencati sistemi locali di grandi dimensioni e localizzati nel Centro-Nord. Le regioni che riportano i risultati migliori sono nuovamente l'Emilia Romagna e il Veneto. In figura 4.10 si mostra l'andamento nel tempo di questa variabile. In questo caso, di-

<sup>4</sup>Nella teoria dei grafi assume un ruolo importante il grado di un nodo della rete, pari al numero di archi che incidono sul nodo stesso. Si dice grado massimo di un grafo  $G$  il grado massimo dei suoi nodi, e, analogamente, grado minimo del grafo il grado minimo dei suoi nodi.

Figura 4.10: Analisi temporale del livello massimo di coesione interna dei principali sll



versamente da quanto mostrato in 4.7, i valori più alti di *coesione interna* sono relativi ai sistemi locali più grandi, come Milano e Bologna. Per costruzione, è infatti verosimile che il numero massimo di collegamenti ripetuti all'interno di una rete sia più alto per i *network* con un numero maggiore di collaborazioni brevettuali. Più regolare appare invece l'andamento degli altri sistemi locali, per i quali il valore di questa misura rimane confinato nell'intervallo [1,14].

L'ultima proprietà di rete che si sceglie qui di analizzare è una misura di centralità. L'idea è di verificare quanto ciascun nodo risulti centrale nel *network*. Nel caso in esame, ci si aspetta che gli inventori con una misura di centralità associata elevata siano in maggior misura responsabili del processo di collaborazione brevettuale. La misura che si intende qui studiare è definita *betweenness centrality*:

$$Bt = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (4.11)$$

dove il denominatore rappresenta il numero di percorsi minimi che legano il nodo  $s$  al nodo  $t$  e il numeratore quanti di questi percorsi attraversano il nodo  $v$ . Si sceglie di applicare questa definizione alla rete complessiva italiana in tre periodi temporali distinti (i.e., 1994, 2004, 2014), in modo che l'analisi non sia falsata da trend geografici locali o temporali. La prima rete considerata (i.e., 1994) presenta 698 nodi, di cui 694 hanno un valore di *betweenness* pari a 3 o inferiore. Spiccano invece in particolare due nodi, con  $Bt$  pari a 63.07 e 23.28: il primo fa riferimento a una S.p.a. operante nel campo dell'energia, con sede legale a San Donato Milanese e pertanto associata al sll del capoluogo lombardo; il secondo, è relativo a una S.p.a. leader nella progettazione e costruzione di grandi impianti onshore, anch'essa con sede principale a San Donato Milanese, ma anche presente nei comuni di Roma e Milano. Risulta quindi evidente l'importanza delle più grandi realtà

industriali del Paese nelle reti di collaborazioni brevettuali. Ancora una volta, il sistema locale di Milano si conferma come punto di riferimento per l'attività innovativa italiana. Il secondo *network* si riferisce invece all'anno 2004. In questo caso si registra una crescita del numero di inventori coinvolti in almeno una richiesta di brevettazione, qui pari a 910. Tuttavia, rispetto al caso precedente, i nodi principali presentano un livello di *betweenness* non eccessivamente elevato e rispettivamente pari a 15, 12 e 10. Il primo fa riferimento all'Università degli Studi di Milano: è infatti piuttosto frequente che la domanda di brevetto sia depositata da un ente pubblico come quello universitario, che investe nella ricerca buona parte delle proprie risorse umane e finanziarie. Allo stesso modo, il secondo nodo fa riferimento all'Università degli Studi di Cagliari e il terzo all'Università degli Studi di Trieste. Questa seconda analisi offre una nuova prospettiva, sottolineando l'importanza dei centri di ricerca universitari accanto alle più grandi realtà industriali del Paese. Infine, per l'anno 2014, il numero di richiedenti scende a 838, ma la misura di *betweenness centrality* acquista un'importanza maggiore: infatti, i primi 15 valori di questa speciale classifica sono superiori a quanto analizzato finora. Per costruzione della misura (4.11), questo cambiamento riflette non tanto un aumento del numero di collegamenti presenti nella rete, quanto invece il peso diverso che alcuni attori rivestono in questa circostanza. Confermando la tendenza osservata per l'anno 2004, i nodi con un valore di *Bt* più elevato si riferiscono a Università (Alma Mater di Bologna,  $Bt=1061$ , Università degli Studi di Genova,  $Bt=816$ ) o centri di ricerca, come il Consiglio Nazionale delle Ricerche, ente pubblico di ricerca con competenze multidisciplinari, o l'Istituto Italiano di Tecnologia. A conclusione di questa descrittiva delle principali proprietà della rete, si presentano due grafi esemplificativi dei ragionamenti proposti. In figura 4.11 si mostra la rete delle collaborazioni nel 2018. In particolare, si evidenziano in giallo i nodi con un *grado* superiore a 8. In 4.12 si distinguono i nodi appartenenti al sistema locale di Milano (i.e., in rosso), uno dei più attivi in termini di attività inventiva e collaborazioni, dal resto del *network* (i.e., in verde). A conferma dell'importanza di questo sll, alcuni dei nodi con maggior *grado* appartengono proprio al sistema locale del capoluogo lombardo.

Figura 4.11: Rete delle collaborazioni brevettuali nel 2018 (filtro sul grado)

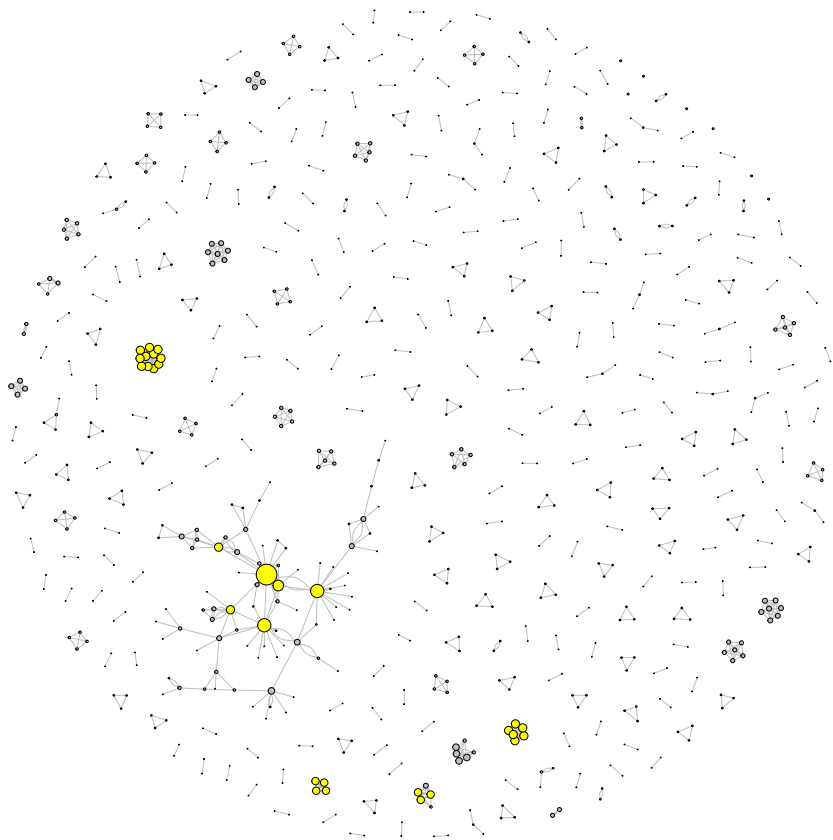
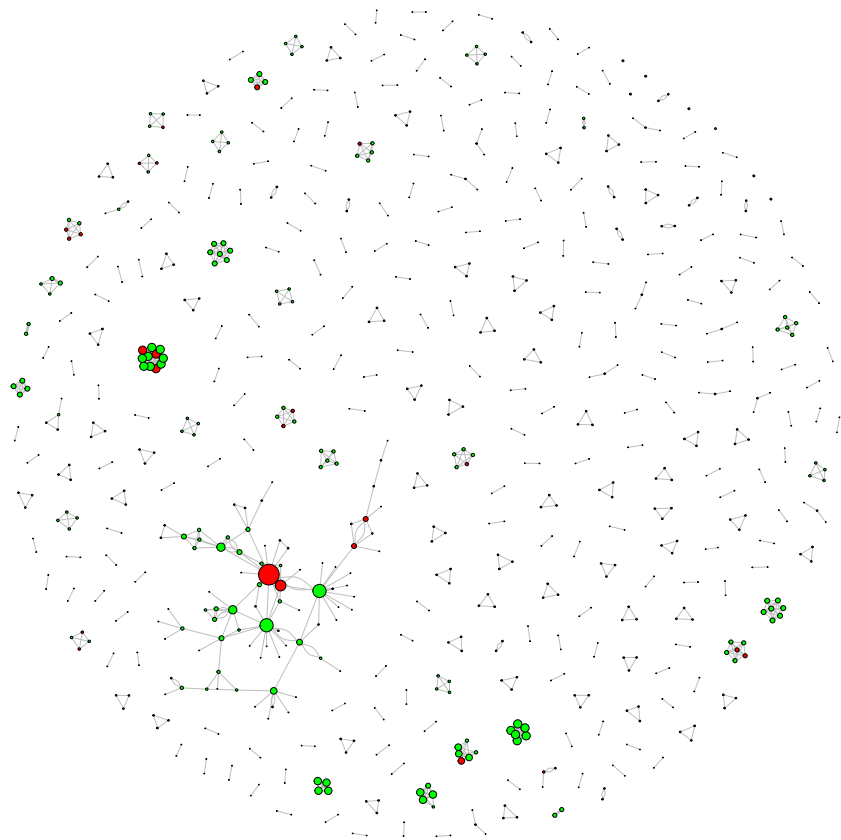




Figura 4.12: Rete delle collaborazioni brevettuali nel 2018 (filtro su grado e sll di Milano)



# Capitolo 5

## Analisi econometrica

### 5.1 Modello di regressione lineare

Nella più recente letteratura scientifica che si occupa di studiare l'attività inventiva di un territorio si fa un uso sempre più frequente dei modelli di regressione lineare. (Doring & Schnellenbach, 2006). Definito il generico modello:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad i = 1, \dots, n \quad (5.1)$$

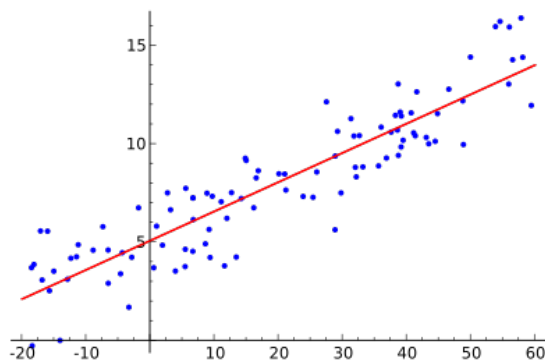
si descrivono le grandezze presenti:

- $X$ , variabile indipendente o regressore;
- $Y$ , variabile dipendente;
- $\beta_0$ , intercetta;
- $\beta_1$ , pendenza;
- $u_i$ , errore (o residuo) di regressione.

Un modello così costruito consente di stimare la pendenza della retta di regressione e, in ultima analisi, l'effetto causale su  $Y$  di una variazione unitaria di  $X$ . La pendenza della retta è determinata dal suo coefficiente angolare, che non è noto e deve quindi essere stimato utilizzando i dati che si ha a disposizione. In figura 5.1 si mostra schematicamente la retta di regressione, dove ciascun punto rappresenta un'osservazione del dataset e la differenza rispetto alla retta costituisce l'errore  $u_i$ . L'errore di regressione è costituito da fattori omessi. In generale, si tratta di fattori diversi dalla variabile  $X$ , che influenzano  $Y$ . I coefficienti  $\beta_i$  sono stimati secondo il metodo dei minimi quadrati ordinari (OLS, *ordinary least squares*). Lo stimatore OLS è dato da:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n [Y_i - (\beta_0 + \beta_1 X_i)]^2 \quad (5.2)$$

Figura 5.1: La generica retta di regressione lineare



Lo stimatore minimizza la somma dei quadrati delle differenze tra i valori reali (osservati) di  $Y_i$  e i valori predetti in base alla retta di regressione stimata ( $b_0 + b_1 X_i$ ).  
 Gli stimatori OLS della pendenza e dell'intercetta sono:

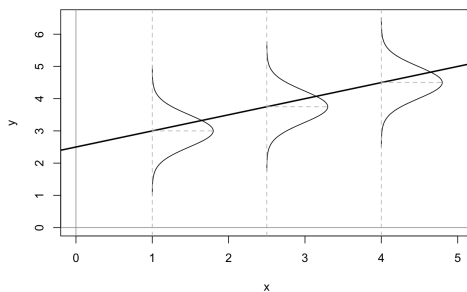
$$\hat{\beta}_1 = \frac{cov_{xy}}{var_x^2} \quad (5.3)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \quad (5.4)$$

Il metodo dei minimi quadrati si basa su alcune assunzioni, affinché si ottenga un elenco di stimatori corretti dei reali parametri della popolazione, che risultano invece ignoti:

1. la distribuzione di  $u$  condizionata a  $X$  ha media nulla, cioè  $E(u|X=x)=0$ . Il fatto che la distribuzione degli errori sia concentrata allo stesso modo attorno alla retta di regressione implica l'assenza di correlazione tra la variabile indipendente e l'errore  $u$   $i$ -esimo. In virtù di questo, lo stimatore  $\hat{\beta}_1$  risulta non distorto. In figura 5.2 si mostra la distribuzione degli errori rispetto alla retta di regressione. Questa configurazione si presenta con un esperimento controllato casualizzato;

Figura 5.2: Distribuzione degli errori attorno alla retta di regressione



2.  $(X_i, Y_i, i=1, \dots, n)$  sono indipendenti identicamente distribuite, ovvero sono scelte in modo casuale e a partire dalla stessa popolazione;
3. gli *outlier* in  $X$  e/o  $Y$  sono rari.

Definite queste assunzioni, ci si concentra sul calcolo degli stimatori; quando la numerosità  $n$  del campione è sufficientemente elevata, è possibile affermare:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\text{var}_v^2}{n(\text{var}_x^2)^2}\right), \quad v_i = (X_i - \mu)u_i \quad (5.5)$$

Minore è la differenza tra  $\beta_1$  e  $\hat{\beta}_1$ , tanto più lo stimatore calcolato risulterà consistente. Inoltre, l'equazione 5.5 mostra come la varianza dello stimatore sia inversamente proporzionale alla dimensione del campione e alla varianza dei regressori.

Si procede quindi a valutare come i regressori influenzino la variante dipendente e se questo contributo sia statisticamente significativo. Per fare questo, si conduce un test di ipotesi così strutturato:

$$H_0 : \beta_1 = \beta_{1,0} \quad (5.6)$$

$$H_0 : \beta_1 \neq \beta_{1,0} \quad (5.7)$$

In 5.6 si definisce l'ipotesi nulla del test, ovvero che lo stimatore sia uguale a zero: un coefficiente angolare nullo implica l'assenza di un effetto sulla variabile dipendente; in 5.7 è invece presentata l'ipotesi alternativa. Il test è condotto utilizzando la distribuzione  $t$  di Student, calcolando il *t-test* in questo modo:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\text{var}_{\hat{\beta}_1}}} \quad (5.8)$$

Il risultato di 5.8 consente di definire l'intervallo di confidenza tale per cui ci si attende che il coefficiente oggetto di analisi ricada con una data probabilità all'interno di un intervallo fissato. Per un'analisi più approfondita di questo concetto si rimanda ai risultati del modello econometrico, che saranno commentati nei capitoli successivi.

Giunti a questo livello di dettaglio, si aggiunge un ulteriore livello di complessità. Infatti, è necessario generalizzare il caso fino ad ora analizzato ad un modello di regressione lineare con regressori multipli  $(X_{1,\dots,n})$ :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} \quad (5.9)$$

La necessità di condurre un'analisi di questo tipo risponde fondamentalmente all'esigenza di evitare la distorsione da variabile omessa (i.e.,  $Y_i$  influenzata da variabili non esplicitate nel modello). In particolare, vale<sup>1</sup>:

$$\hat{\beta}_1 = \beta_1 + \frac{\sigma_u}{\sigma_x} \rho_{xu} \quad (5.10)$$

---

<sup>1</sup>Per una descrizione dettagliata di come i coefficienti siano in questo caso stimati si veda “Derivation of the least squares estimator for beta in matrix notation” (2015)

Poiché in  $u$  è presente la variabile omessa, il valore di  $\hat{\beta}_1$  risulterà sovrastimato. Alle assunzioni sul metodo OLS già citate, si aggiunge qui l'ipotesi di assenza di collinearità perfetta (i.e., si verifica quando uno dei regressori è funzione lineare esatta degli altri). In alternativa, è comunque possibile che si verifichi collinearità imperfetta quando due o più regressori sono altamente correlati: ciò significa che uno o più coefficienti saranno stimati in modo impreciso.

## 5.2 Specificazione del modello

Il modello di regressione lineare completo che si intende analizzare in questo elaborato è così strutturato:

$$\begin{aligned}
 Y_{it} = & \beta_0 + \beta_1 \text{density}_{it} + \beta_2 \text{maxcoreness}_{it} + \beta_3 \text{sproximity}_{it} \\
 & + \beta_4 \text{shapplicants}_{it} + \beta_5 \text{extbreadth}_{it} + \beta_6 \text{extdepth}_{it} \\
 & + \beta_7 \text{urvariety}_{it} + \beta_8 \text{srvariety}_{it} + \beta_9 \text{rvvariety}_{it} \\
 & + \text{gdp}_{it} + \text{education}_{it} + \text{rd}_{it} + \text{popdensity}_{it} + u_{it}
 \end{aligned} \tag{5.11}$$

In particolare, dando seguito a quanto già approfondito nel capitolo di costruzione dei regressori, si riporta una descrizione puntuale delle variabili coinvolte.

Variabile	Descrizione
Variabile dipendente	numero di brevetti pro capite
density	densità della rete di collaborazione
maxcoreness	misura di centralità di un nucleo della rete
sproximity	misura di transitività
shapplicants	quota di richiedenti per sistema locale
extbreadth	quota di collaborazioni con imprese esterne al sistema locale
extdepth	distribuzione delle collaborazioni con imprese esterne al sistema locale
urvariety	distribuzione delle collaborazioni con imprese esterne al sistema locale
srvariety	misura di entropia di ciascun settore industriale (semi-related)
rvvariety	misura di entropia di ciascun settore industriale (related)
gdp (per capita)	gross domestic product per capita
education	percentuale di residenti con un livello d'istruzione almeno universitario
rd	spese in ricerca e sviluppo
densitypop	densità di popolazione

Ciascuna variabile è costruita per ogni periodo  $t$  (anno) e per ogni  $i$ : il modello è stato costruito per sfruttare le potenzialità dei dati *panel*. Un dataset così realizzato consente di osservare il comportamento della variabile oggetto di interesse non solo attraverso le sezioni  $i$ -esime, ma anche lungo l'arco temporale selezionato. In particolare, l'ipotesi dell'elaborato è che una regressione lineare con effetti fissi sia la scelta migliore per il tipo di dato da trattare. Questo modello, in contrasto con quello a effetti casuali, consente di valutare il comportamento di tutte quelle variabili non esplicitamente specificate, ma potenzialmente caratteristiche distintive di ciascuna area geografica. È il caso di variabili come la cultura, la religione o il sesso, ragionevolmente immutate nel periodo temporale qui considerato, ma possibilmente erroneamente omesse dal modello. Utilizzare un modello a effetti fissi consente quindi di eliminare il problema della distorsione da variabile omessa quanto meno per quelle grandezze che non variano nel tempo nel periodo considerato. Tale scelta, al momento solo intuitivamente corretta, è quindi confermata empiricamente dall'*hausman test*: l'ipotesi nulla prevede che la differenza nei coefficienti tra i due modelli non sia sistematica. Rifiutare l'ipotesi nulla ( $p < 0.05$ ) equivale ad affermare che invece questa differenza è sistematica, confermando quindi l'intuizione iniziale secondo cui il modello a effetti fissi risulta più appropriato per studiare il campione a nostra disposizione.

## 5.3 Analisi descrittive

Si presentano ora le analisi preliminari svolte sul campione. In primo luogo, si calcolano le principali descrittive delle variabili utilizzate, che sono mostrate in tabella 5.1.

Tabella 5.1: Statistiche descrittive

Variabile	Count	Mean	Median	SD	Min	Max
Brevetti pro-capite	9165	9.594	1.000	38.983	0.000	933.333
Densità della rete (density)	3111	0.112	0.020	0.178	0.000	3.000
Misura di centralità (maxcoreness)	3111	0.818	1.000	1.222	0.000	18.000
Misura di transitività (sproximity)	9165	0.039	0.000	0.192	0.000	1.000
Quota di richiedenti (shapplicants)	9165	0.000	0.000	0.000	0.0000	0.001
Quota di collaborazioni con imprese esterne (extbreadth)	3078	0.715	0.889	0.359	0.000	1.000
Distribuzione delle collaborazioni con imprese esterne (extdepth)	2638	3.128	1.000	10.616	0.130	304.000
Misura di entropia di settore (unrelated variety)	5765	0.860	0.956	0.615	0.000	2.030
Misura di entropia di settore (semi-related variety)	5765	0.464	0.367	0.497	0.000	2.037
Misura di entropia di settore (related variety)	5765	1.034	1.055	0.748	0.000	3.194

Count = numero osservazioni; SD = deviazione standard

Ciò che appare subito evidente è che solo alcune delle variabili mostrano un valore diverso da zero per l'intero periodo considerato. Come già anticipato a valle della costruzione dei regressori, questo fenomeno non è determinato da un dataset non bilanciato, bensì dalla presenza di molte reti povere di nodi (e quindi di archi). Ne deriva che le proprietà del *network* di co-assegnazioni presentano spesso valore nullo o pari a zero, in base a come è costruito l'indicatore.

Come spiegato nella parte introduttiva del capitolo, l'affidabilità dell'analisi dipende anche dal tasso di correlazione che lega le variabili utilizzate nel modello. In quest'ottica, si calcola la matrice di correlazione, mostrata in figura 5.2.

Tabella 5.2: Matrice di correlazione

Variabile	1	2	3	4	5	6	7	8	9	10
1 Brevetti pro-capite	1.000									
2 Densità della rete	-0.028	1.000								
3 Misura di centralità	0.488	0.566	1.000							
4 Misura di transitività	0.357	0.263	0.620	1.000						
5 Quota di richiedenti	0.212	0.112	0.233	0.136	1.000					
6 Quota di collaborazioni con imprese esterne	-0.276	-0.574	-0.708	-0.541	-0.260	1.000				
7 Distribuzione delle collaborazioni con imprese esterne	-0.045	0.342	0.163	0.118	0.004	-0.121	1.000			
8 Misura di entropia di settore (unrelated)	0.343	0.123	0.357	0.261	0.359	-0.374	-0.003	1.000		
9 Misura di entropia (semi-related variety)	0.526	0.065	0.418	0.307	0.456	-0.404	-0.043	0.593	1.000	
10 Misura di entropia (related variety)	0.499	0.103	0.440	0.332	0.387	-0.417	-0.021	0.938	0.692	1.000

Tra i valori presenti nella matrice triangolare, alcuni sono da analizzare con cura: si tratta di quelli prossimi all'unità e, in particolare, superiori al valore soglia scelto di 0.65. Si ritiene infatti che un tasso di correlazione più alto del limite superiore proposto non risulti più accettabile, in quanto inserire in un modello di regressione lineare due variabili indipendenti altamente correlate comporta una stima dei coefficienti imprecisa e, in ultima analisi, delle conclusioni errate. Pertanto, nei modelli statistici che saranno presentati nei prossimi paragrafi non sarà valutato contemporaneamente, a coppie, l'impatto di *extdepth* e *maxcoreness*, *rvariety* e *srvariety*, *rvariety* e *uvariety*. In particolare, le ultime due variabili possono quasi essere definite *perfettamente correlate* ( $\rho = 0.938$ ).

Infine, prima di studiare i modelli di regressione, si effettua un controllo per verificare se

l'errore  $u$  sia o meno omoschedastico, ovvero se valga o no la relazione seguente:

$$\text{var}(u|X = x) = k, \quad (5.12)$$

con  $k =$  costante. Quando questo non avviene, poiché la varianza del termine d'errore dipende da  $x$ , si parla di eteroschedasticità. La valutazione di questo tipo di fenomeno<sup>2</sup>, in verità molto frequente, risulta fondamentale in quanto è necessario stimare in modo diverso lo *standard error* (SE) del coefficiente oggetto di analisi. Non correggere opportunamente questo fattore determina una sua sottostima e quindi un errore nel calcolo del *t-test*, che, come mostrato, guida l'inferenza e determina la stima dei coefficienti. In verità senza sorpresa, nel caso in esame si rifiuta l'ipotesi secondo cui  $u$  è omoschedastico, adattando di conseguenza i modelli di regressione.

---

<sup>2</sup>White Test



## 5.4 Risultati

Si presentano ora i risultati più interessanti dell'analisi condotta. In tabella 5.3 è possibile sintetizzare le informazioni fondamentali per interpretare un modello di regressione. Pur facendo sempre riferimento a quanto specificato in generale in 5.11, in particolare si mostrano tante colonne quanti sono i modelli scelti. Si tratta di minime ma necessarie variazioni affinché l'inferenza risulti robusta.

Tabella 5.3: Modelli di regressione

Modello	(1)	(2)	(3)	(4)	(5)	(6)	(7)
density	-17.699** (6.979)	-17.088** (6.719)	-10.297*** (3.008)	-29.610*** (10.278)	-29.221*** (10.253)	-29.453*** (10.209)	-19.905** (9.318)
maxcoreness	3.780** (1.878)	3.707** (1.830)		4.282** (1.990)	4.252** (1.990)	4.275** (1.986)	2.671 (1.950)
sproximity	-4.709* (2.433)	-3.959* (2.305)	0.737 (1.483)	-4.200 (2.683)	-4.195 (2.651)	-4.231 (2.697)	-3.312 (3.448)
shapplicants	141,253.293*** (15,333.395)	109,635.698*** (11,660.011)	131,319.850*** (14,388.461)	118,497.336*** (12,006.956)	123,588.903*** (16,147.626)	122,321.541*** (12,955.529)	110,849.242*** (12,432.501)
extbreadth			-0.888 (1.478)				
extdepth			0.019* (0.011)	0.030** (0.014)	0.029** (0.014)	0.029** (0.014)	0.020 (0.016)
uvariety					-3.459*** (1.077)		-1.880*** (0.653)
srvariety					1.132 (1.172)		
rvariety						-1.341 (0.914)	
gdp							0.007 (0.008)
education							-0.136 (0.454)
rd							-0.021 (0.142)
densitypop							-0.428* (0.222)
Constant	9.232*** (1.773)	8.482*** (1.725)	10.205*** (2.871)	9.533*** (2.012)	12.490*** (1.420)	11.223*** (1.727)	132.079* (72.436)
year dummies	NO	YES	YES	YES	YES	YES	YES
spatial dummies	YES	YES	YES	YES	YES	YES	YES
Observations	3,111	3,111	2,634	2,634	2,618	2,618	2,017
Number of groups	498	498	468	468	463	463	448
Average group size	6.247	6.247	5.628	5.628	5.654	5.654	4.502
Panel-level standard deviation	39.851	39.792	41.844	40.582	41.149	41.156	140.385
R-squared for within model	0.17527	0.20402	0.19178	0.21747	0.22078	0.21806	0.26551
R-squared for between model	0.08267	0.09763	0.04475	0.12825	0.10162	0.10116	0.19412
R-squared for overall model	0.11222	0.12157	0.04961	0.13746	0.12157	0.11542	0.26187
Root mean squared error	14.440	14.218	15.470	15.223	15.243	15.266	12.839

La variabile dipendente è calcolata utilizzando il numero di brevetti pro capite di ciascun sll (per 100k abitanti). Tutti i modelli di regressione includono il termine costante e nel modello (7) sono inserite le principali variabili di controllo utilizzate nei modelli di crescita economica (gdp, rd, education, densitypop). I modelli sono stimati con il metodo OLS e con un *vec* robusto (*variance estimator clustered* al livello dei controlli spaziali (i.e., sll)). Gli asterischi da uno a tre indicano un livello di significatività rispettivamente pari al 10%, 5% e 1%. Tra parentesi sono riportati i relativi errori standard.

Definita la variabile dipendente del modello, in ogni riga si mostra, leggendo la tabella da sinistra, il nome del regressore che si sta analizzando e, per ciascuna colonna, due valori (la cella vuota indica l'assenza di quella specifica variabile all'interno del modello). Il primo valore indica la stima del coefficiente  $\beta$ , il secondo, tra parentesi, l'errore standard. Gli asterischi indicano il livello di significatività della stima condotta.

Nella parte inferiore della tabella, dopo il termine costante, sono presenti una serie di indicatori statistici che completano l'analisi. Accanto a misure prevalentemente descrittive (e.g., il numero di osservazioni), si menziona uno dei parametri tipici di un *output* di questo tipo: il coefficiente di determinazione (i.e., *R-squared*). In generale, questo termine misura la varianza della variabile dipendente spiegata dai regressori; si tratta di un valore che varia nell'intervallo  $[0,1]$ . In particolare, poiché come detto si sceglie di utilizzare un modello *panel*, si distingue il valore *within* da quello *between*. Come si può intuitivamente capire, il primo indicatore misura quanto accade all'interno dello stesso gruppo (i.e., periodo), mentre il secondo riguarda l'analisi di gruppi diversi.

Si presentano ora i risultati dei modelli inseriti, uno per volta e commentati facendo opportuni riferimenti al lavoro che ispira questo elaborato di tesi (Innocenti et al., 2020).

Nel primo modello (1) si inseriscono le principali variabili di rete utilizzate in letteratura, insieme alla valutazione sulla quota di richiedenti: un primo contributo dell'elaborato consiste nel misurare quale di queste due categorie guidi maggiormente l'attività inventiva. In questo primo modello tutte le variabili risultano significative. Innanzitutto, è interessante notare come la densità della rete sia l'unica variabile con segno negativo: coerentemente con quanto riportato da Innocenti et al. (2020), una rete molto densa determina un flusso di conoscenza ridondante e quindi potenzialmente deleterio per l'attività inventiva del territorio (i.e., sll).

In secondo luogo, anche in questo modello risulta significativa (e con segno positivo) la misura di centralità di un nucleo all'interno della rete (*maxcoreness*). Questo risultato conferma l'ipotesi secondo cui una rete *coesa* a livello micro favorisce la produzione brevettuale.

Infine, si sottolinea il peso del processo di aggregazione (segno positivo), valutato tramite la quota dei richiedenti, sull'attività inventiva territoriale. Un *output* di questo tipo è giustificato dal fatto che un'elevata concentrazione di richiedenti facilita le *esternalità di conoscenza*.

Di più difficile interpretazione è invece la variabile *sproximity*, in parte per come l'indicatore è costruito (si veda 4.3); tra l'altro, anche in Innocenti et al. (2020) la misura di transitività non è significativa, come si vedrà nei modelli che seguono.

Nel secondo modello (2) si aggiunge un controllo temporale (*year dummies*), catturando in questo modo le variazioni nell'attività inventiva di un sll in precedenza erroneamente correlate alle variabili di rete. Pur essendoci una tendenza di questo tipo (i.e., ogni periodo risulta significativo e con segno quasi sempre positivo), questo non modifica i risultati del primo modello. Le *dummy* anno sono d'ora in avanti inserite in tutti i modelli, in aggiunta ai controlli spaziali per i sll.

Nel terzo (3) e quarto (4) modello si inseriscono le due variabili che rappresentano l'importanza dei collegamenti esterni alla rete del sll *i*-esimo per la sua produzione brevettuale. Tuttavia, in virtù dell'elevata correlazione tra *maxcoreness* e *extbreadth*, ciascun indicatore è valutato separatamente. Il primo risulta avere un impatto maggiore sulla variabile dipendente e si procede pertanto a privilegiare questa scelta anche nei successivi modelli. A conferma di ciò, la misura del coefficiente di determinazione complessivo diminuisce sensibilmente proprio in concomitanza dell'eliminazione della *maxcoreness* dal modello (3). La variabile *extdepth* risulta significativa (segno positivo), più di quanto verificato da Innocenti et al. (2020), indicando quindi una forte relazione tra la variabile dipendente e

la tipologia di collegamenti esterni (i.e., archi che legano un nodo del sll  $i$ -esimo oggetto di studio a un nodo appartenente a una rete  $j$ -esima, diversa da  $i$ ), con particolare attenzione alla loro distribuzione spaziale.

Nel quinto (5) e sesto (6) modello si introducono le variabili utilizzate per spiegare, tramite il codice IPC associato a ciascun brevetto, quanta *varietà* industriale sia presente in ciascun settore. Anche in questo caso, poiché la *rvarietà* risulta altamente correlata con le altre due variabili di entropia, si effettuano due test separatamente: come in Innocenti et al. (2020), la sola *uvarietà* risulta significativa (ma qui con segno negativo). La *srvarietà*, qui aggiunta seguendo la proposta di Castaldi et al. (2015), non sembra avere un ruolo sulla produzione inventiva. Si sottolinea che tutte le variabili precedentemente inserite nei modelli conservano la loro significatività. I risultati che emergono da questi modelli sembrano mostrare che un settore in cui il livello di diversificazione non sia elevato possa beneficiare in termini di attività inventiva prodotta.

Il settimo (7) e ultimo modello riprende quanto specificato in 5.11, aggiungendo i controlli solitamente utilizzati nei modelli di crescita economica. Dando seguito alle riflessioni sul tasso di correlazione  $\rho$  e circa il livello di significatività delle variabili coinvolte, *extbreadth*, *srvarietà* e *rvarietà* sono escluse dal modello. Nel modello completo anche *maxcoreness* e *extdepth* perdono significatività, evidenziando la prima importante differenza rispetto a Innocenti et al. (2020). Infatti, nel loro articolo l'importanza di un nucleo centrale all'interno della rete è conservata anche nel modello completo. Per il resto, *density*, *uvarietà* (segno negativo) e *shapplicants* (segno positivo) rimangono altamente significative. L'aggiunta dei controlli determina un aumento del coefficiente di determinazione complessivo. L'ampiezza dei dati a disposizione presso l'UIBM e l'ampia letteratura che negli ultimi anni sta emergendo sul tema hanno evidentemente imposto delle scelte metodologiche chiare in termini di pulizia del dataset, scelta dei regressori da inserire nei modelli e relativa costruzione. Con questa consapevolezza, sono state condotte delle analisi alternative, di cui si riporta un estratto in Appendice H. In particolare, si presentano gli stessi modelli qui mostrati, le cui variabili sono state però costruite a partire da un set di dati differente. Infatti, nel tentativo di ovviare alla presenza di reti eccessivamente "povere" (i.e., numero di nodi e/o di archi esiguo), ciascuna variabile è stata aggregata effettuando una media su tre anni, conservando quindi la struttura *panel*, ma con ciascun periodo che ora non è più rappresentato da un singolo anno, ma da un triennio. Rispetto al modello con dati annuali, si evidenziano le seguenti differenze:

1. la densità della rete non risulta sempre significativa;
2. la *maxcoreness* è invece qui significativa in tutti i modelli in cui è testata;
3. la misura di transitività (*sproximity*) è spesso altamente significativa;
4. l'importanza relativa della quota dei richiedenti rispetto alla popolazione residente è pressoché irrilevante;
5. le variabili che controllano per i collegamenti esterni al sll non sono mai significative;
6. non c'è nessuna evidenza della relazione tra misura di entropia del settore industriale e attività brevettuale.

Questo esempio è esemplificativo del fatto che un cambiamento anche minimo nelle ipotesi iniziali può generare dei risultati anche molto diversi o di difficile interpretazione (i.e., *shapplicants* mai significativo nei primi sei modelli, altamente significativo nell'ultimo). Al contrario, i modelli commentati in questo capitolo risultano in linea con la letteratura, in particolare con il lavoro pubblicato da Innocenti et al. (2020), nonché coerenti con le ipotesi propedeutiche all'analisi econometrica e precisate nei capitoli precedenti.



## Capitolo 6

# Conclusione e sviluppi futuri

All'interno della branca di ricerca che studia l'*economia della conoscenza*, l'elaborato di tesi si concentra sullo studio delle *esternalità di conoscenza* con lo scopo di applicare in modo costruttivo parte della metodologia che la letteratura ci ha offerto negli ultimi anni per studiare l'attività innovativa dei sistemi locali del lavoro italiani. In particolare, ispirato dal lavoro di Innocenti et al. (2020), la tesi si pone l'obiettivo di approfondire come i seguenti aspetti possano influenzare la produzione brevettuale italiana:

1. la struttura delle reti di conoscenza, qui approssimate dal *network* di co-assegnazioni brevettuali;
2. i processi di *agglomeration economies*;
3. la varietà industriale, misurata tramite i codici IPC associati a ciascun brevetto.

Dopo un attento processo di pulizia del dataset UIBM, si mostrano le analisi descrittive sul campione, costituito dai brevetti depositati dal 1994 al 2019, con esclusione di alcuni periodi temporali<sup>1</sup>.

Un primo contributo dell'elaborato consiste nella costruzione e analisi delle reti di co-assegnazioni brevettuali, focalizzando l'attenzione sulle principali proprietà proposte dalla letteratura (Jackson et al., 2017). Allo stesso modo e servendosi della classificazione IPC, si definisce il concetto di *varietà* di un settore industriale (Castaldi et al., 2015), valutando in questo modo quanto la produzione (i.e., brevettuale, in questo caso) sia trasversale rispetto al campo tecnologico.

A questo punto e una volta definito l'insieme delle variabili coinvolte, si costruisce un modello di regressione lineare per comprendere il ruolo di queste proprietà. Innanzitutto, la struttura della rete influenza l'attività inventiva dei sll analizzati: in particolare e coerentemente con il lavoro di Innocenti et al. (2020), una rete più densa ha un impatto negativo sulla produzione brevettuale; al contrario, la centralità del nucleo all'interno del *network* risulta essere un *driver* dell'innovazione.

---

<sup>1</sup>Appendice C

Inoltre, questo lavoro si inserisce all'interno del dibattito acceso su quale aspetto sia maggiormente rilevante tra le variabili di rete e i processi di *agglomeration economies*. Il secondo aspetto sembra prevalere, anche a seguito di un'analisi di robustezza dei modelli<sup>2</sup>.

Infine, i modelli studiati contribuiscono a studiare l'importanza della struttura industriale di un'area geografica, mostrando una correlazione negativa tra il numero di brevetti pro capite del sll e la misura di *varietà non correlata*: si tratta del risultato che maggiormente si discosta dal lavoro di Innocenti et al. (2020).

Occorre pertanto effettuare le opportune riflessioni circa il ruolo della struttura industriale del sll rispetto all'attività inventiva del territorio. Infatti, il segno negativo mostrato dalla variabile che misura l'entropia al più basso livello di dettaglio implica inequivocabilmente che è preferibile una struttura industriale le cui attività *non siano varie*. D'altronde, sarebbe intuitivamente atteso che questo risultato fosse complementare a un valore significativo, ma con segno opposto, della *varietà correlata*. Poiché ciò non accade, certamente questo aspetto rappresenta un primo spunto di riflessione per i lavori empirici futuri.

In secondo luogo, pur essendo un risultato meritevole di attenzione e con un senso ben preciso, affermare che la quota dei *richiedenti* risulti una variabile significativa nella produzione brevettuale italiana non implica necessariamente che la stessa relazione esista anche per la quota di *inventori*. Si tratta quindi di un altro spunto di riflessione meritevole di approfondimento, qualora i dati in sorgente fornissero anche l'anagrafica degli inventori, come accade presso l'EPO.

Inoltre, e in particolare nelle regioni italiane, l'attività inventiva di alcuni distretti industriali (e.g., attività manifatturiera) non è tracciabile tramite l'uso dei brevetti. Sarebbe utile poter migliorare le analisi svolte in questo elaborato integrando efficacemente misure alternative per misurare l'attività inventiva (e.g., spese in R&D, modelli di utilità, marchi). Come mostrato, i brevetti, pur garantendo una grande quantità di informazioni ad un costo relativamente basso, spiegano una minima parte dell'attività inventiva di un territorio. A questa criticità si aggiunge poi quella delle collaborazioni, che rappresentano una piccola parte dei brevetti totali fornendo spesso, come mostrato, un'informazione appena sufficiente per effettuare un'analisi puntuale. A tal proposito, in virtù della presenza di tanti valori nulli tra i regressori mostrati nell'elaborato, si potrebbe effettuare un'analisi di robustezza valutando modelli non basati sul metodo OLS e che possano sfruttare una distribuzione sottostante differente (e.g., *Count*, *Poisson*, *Negative Binomial*).

Certamente, è auspicabile che ciascuno di questi spunti di riflessione sia sviluppato nella direzione qui intrapresa, ovvero mediante l'uso di partizioni funzionali (i.e., sll) e non amministrative (e.g., regioni, province, comuni). Si ritiene infatti che la gestione problematica di reti con informazioni troppo scarse e la difficoltà nel reperire alcune misure a livello di sll, siano più che bilanciate dall'accuratezza di un'area geografica che ben descrive fenomeni di *esternalità di conoscenza* in quanto, per costruzione (Istat, 2015), non sono costituite da fenomeni storico-politici, ma dall'organizzazione (lavorativa) autonoma degli individui.

In conclusione, l'elaborato mostra risultati certamente incoraggianti, spesso in linea con le

---

<sup>2</sup>Appendice H

ipotesi formulate in fase di impostazione del modello e coerenti con la letteratura economica di riferimento. Non mancano nuovi spunti di riflessione, che rappresentano un punto di partenza per migliorare l'elaborato.

Nonostante alcuni autori si interrogino in merito al proliferare della letteratura che studia gli *spillover di conoscenza* (Breschi & Lissoni, 2001), in un contesto macroeconomico che evolve velocemente e con dinamiche sempre nuove, si ritiene che questo filone di ricerca debba essere ulteriormente esplorato per poter definire, in ultima analisi, delle linee guida in termini di *policy* per fare in modo che si possa beneficiare, a più livelli, di processi di *esternalità* certamente complessi, ma reali e tangibili.





# Riferimenti bibliografici

- Aarstad, J., Kvitastein, O. A. & Jakobsen, S.-E. (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. *Research Policy*, 45(4), 844–856.
- Archibugi, D. (1992). Patenting as an indicator of technological innovation: a review. *Science and public policy*, 19(6), 357–368.
- Basberg, B. L. (1987). Patents and the measurement of technological change: a survey of the literature. *Research policy*, 16(2-4), 131–141.
- Boschma, R. & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic geography*, 85(3), 289–311.
- Breschi, S. & Lenzi, C. (2016). Co-invention networks and inventive productivity in US cities. *Journal of Urban Economics*, 92, 66–75.
- Breschi, S. & Lissoni, F. (2001). Knowledge spillovers and local innovation systems: a critical survey. *Industrial and corporate change*, 10(4), 975–1005.
- Broekel, T. & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of economic geography*, 12(2), 409–433.
- 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.
- Doring, T. & Schnellenbach, J. (2006). What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature. *Regional Studies*, 40(3), 375–395.
- Ebersberger, B., Herstad, S. J. & Koller, C. (2014). Does the composition of regional knowledge bases influence extra-regional collaboration for innovation? *Applied Economics Letters*, 21(3), 201–204.
- Fleming, L., King III, C. & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, 18(6), 938–954.
- Fornahl, D., Broekel, T. & Boschma, R. (2011). What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. *Papers in regional science*, 90(2), 395–418.
- Frenken, K., Van Oort, F. & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies*, 41(5), 685–697.
- Gambardella, A., Harhoff, D. & Verspagen, B. (2008). The value of European patents. *European Management Review*, 5(2), 69–84.

- Griliches, Z., Hall, B. & Hausman, J. (1986). Patents and R&D: Is there a lag. *International Economic Review*, 27(2), 265–283.
- Hidalgo, C. A. (2022). Knowledge is non-fungible. *arXiv preprint arXiv:2205.02167*.
- Innocenti, N., Capone, F. & Lazzeretti, L. (2020). Knowledge networks and industrial structure for regional innovation: An analysis of patents collaborations in Italy. *Papers in Regional Science*, 99(1), 55–72.
- Istat. (2014). I sistemi locali del lavoro 2011.
- Istat. (2015). La nuova geografia dei sistemi locali.
- Jackson, M. O., Rogers, B. W. & Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1), 49–95.
- Jaffe, A. B., Trajtenberg, M. & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3), 577–598.
- Katila, R. (2000). Using patent data to measure innovation performance. *International Journal of Business Performance Management*, 2(1-3), 180–193.
- Laursen, K. & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic management journal*, 27(2), 131–150.
- Lobo, J. & Strumsky, D. (2008). Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics*, 63(3), 871–884.
- Marshall, A. (1890). Principles of Economics, London: Mac-Millan. *Originally published in.*
- Michael, D. K. & Battiston, S. (2009). From graph theory to models of economic networks. a tutorial. *Networks, topology and dynamics*, 23–63.
- OECD. (2009). OECD Patent Statistics Manual.
- Paci, R. & Usai, S. (2000). Technological enclaves and industrial districts: an analysis of the regional distribution of innovative activity in Europe. *Regional studies*, 34(2), 97–114.
- Pavitt, K. (1982). R&D, patenting and innovative activities: A statistical exploration. *research Policy*, 11(1), 33–51.
- Rigby, D. (2018). You have to (Br) enter to (Br) exit: The EU collaboration space. *Environment and Planning A: Economy and Space*, 50(3), 493–496.
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy*, 94(5), 1002–1037.
- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2), S71–S102.
- Tavassoli, S. & Carbonara, N. (2014). The role of knowledge variety and intensity for regional innovation. *Small Business Economics*, 43(2), 493–509.
- Theil, H. (1972). H. Theil. Statistical Decomposition Analysis. Amsterdam, North-Holland Publishing Co, 1972, XV p. 337 p., DFL 65.00. *Louvain Economic Review*, 42(2), 166–166.
- Van der Wouden, F. & Rigby, D. L. (2019). Co-inventor networks and knowledge production in specialized and diversified cities. *Papers in Regional Science*, 98(4), 1833–1853.

Wasserman, S., Faust, K. et al. (1994). Social network analysis: Methods and applications.



# Sitografia

Codice della proprietà industriale. (2022). <https://www.altalex.com/documents/codici-altalex/2014/10/30/codice-della-proprietà-industriale>

Derivation of the least squares estimator for beta in matrix notation. (2015). [https://economictheoryblog.com/2015/02/19/ols\\_estimator/](https://economictheoryblog.com/2015/02/19/ols_estimator/)









# Appendice A

# Esempio di brevetto

(19)   (11) **EP 2 634 128 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication: **04.09.2013 Bulletin 2013/36** (51) Int. Cl.: **B65H 75/24 (2006.01)**

(21) Application number: **12199120.2**

(22) Date of filing: **21.12.2012**

(84) Designated Contracting States: <b>AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR</b> Designated Extension States: <b>BA ME</b>	(71) Applicant: <b>TLK S.r.l.</b> <b>29121 Piacenza (IT)</b> (72) Inventor: <b>Scaravella, Carlo</b> <b>29121 Piacenza (IT)</b> (74) Representative: <b>Fisauli, Beatrice A. M.</b> <b>Con Lor S.p.A</b> <b>Via Renato Fucini, 5</b> <b>20133 Milano (IT)</b>
---	--

(30) Priority: **28.02.2012 EP 12157370**  
**19.04.2012 IT PC20120010**

(54) **Expansion mandrel for coils**

(57) The present invention relates to an expansion gripper comprising a hollow shaft (2) at one end of which there is produced a hub (3), a sleeve (5) mounted to slide along an axis (X) on said hub (3) with a first inner portion (6) and a second outer portion (7), mutually coaxial, a liner (8) connected integral with the hub (3), provided with a plurality of through openings (9) adapted to house inserts (10) that act as gripping elements of the gripper,

said second portion (7) being provided with through openings (14) adapted to house the same number of rolling means (13) interposed between the outer surface of said hub (3) and the lower surface of said inserts (10), said surfaces being provided with profiles (11, 12) adapted to cooperate with said rolling means (13) to control the movement of said inserts (10) from a closed position in which they retract inside the contour of the liner (8) to an open position in which they project from said liner (8).

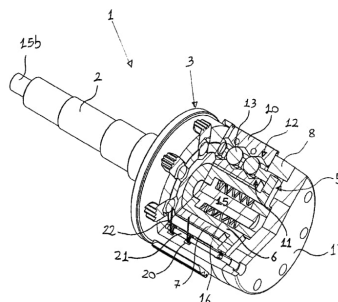


Fig. 1

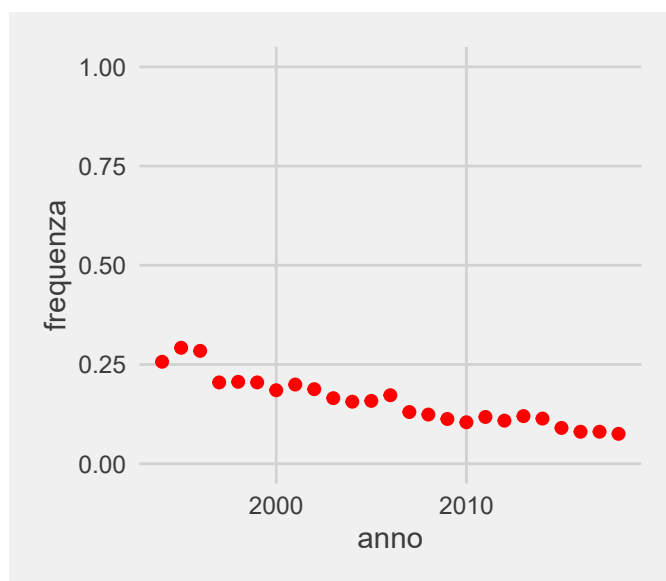
EP 2 634 128 A2

## Appendice B

# Pulizia dati (Parte I)

Come mostrato in figura B.1, la percentuale di record con valore nullo per il campo per il codice fiscale dei richiedenti, di cui si tiene traccia in ordinata (*frequenza*), rispetto al totale dei brevetti presenti nel periodo considerato (1994-2018), mostra un andamento lievemente decrescente nel tempo, probabilmente grazie ad una maggiore precisione nella raccolta dati da parte degli Uffici preposti.

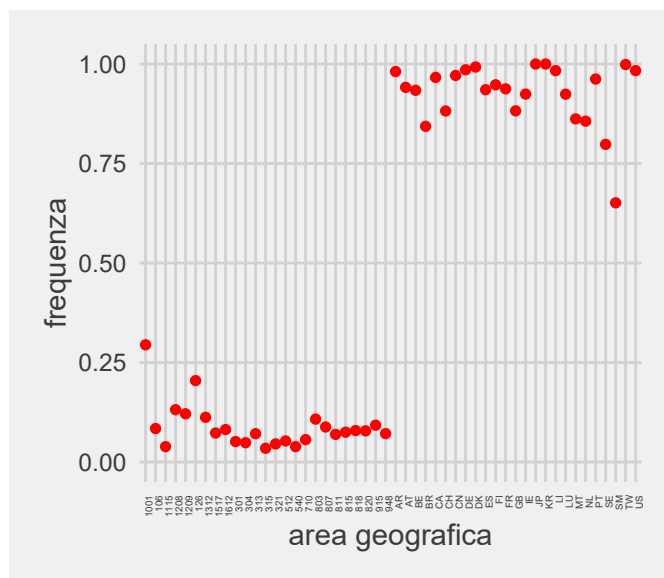
Figura B.1: Analisi temporale dei valori nulli per il codice fiscale dei richiedenti



Si integra quindi questa analisi aggiungendo una metrica simile, ma a livello spaziale. Con l'obiettivo di conservare, quanto meno in una prima fase del lavoro, tutte le informazioni relative ai brevetti stranieri, l'area geografica di riferimento per questa analisi preliminare sarà il sll a cui il brevetto è associato o, nel caso in cui si tratti di un brevetto straniero, la nazione relativa. Poichè la maggior parte dei valori nulli per il campo codice fiscale è associata a poche e specifiche aree e considerando che una statistica descrittiva ragionevole

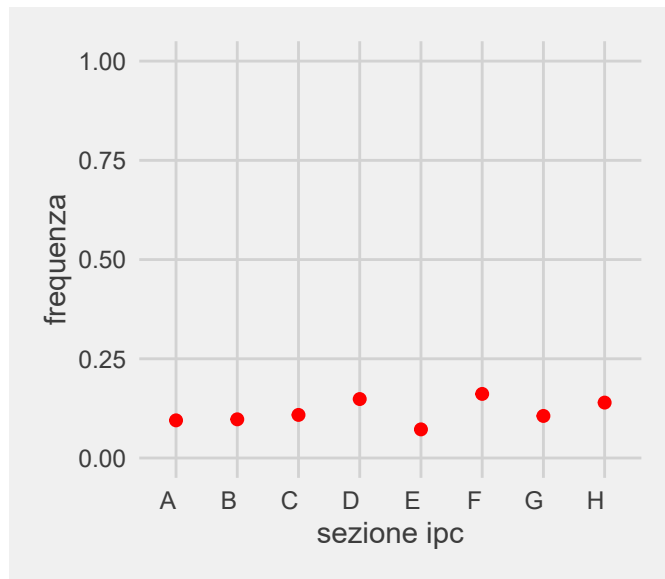
necessita di un numero minimo congruo di dati (non si considerano pertanto le regioni con pochi codici associati), si sceglie di concentrare l'attenzione sulle prime 50 zone in termini di valori nulli per il campo di interesse. La scelta è motivata dal fatto che l'82.7 % dei brevetti con valore di codice fiscale del richiedente nullo è associato a una di queste 50 aree geografiche. Si valuta a questo punto se esistono delle rilevanti differenze tra le zone selezionate. La figura B.2 mostra la frequenza relativa dei brevetti con valore nullo per il campo selezionato per ogni area.

Figura B.2: Analisi spaziale dei valori nulli per il codice fiscale dei richiedenti



Il grafico appare evidentemente sbilanciato: una prima metà mostra infatti una percentuale *frequenza* sostanzialmente ben inferiore al valore 0.25, una seconda parte presenta valori superiori a 0.75, indicando delle aree geografiche a cui sono associati dei brevetti che il più delle volte non possiedono l'informazione identificativa del richiedente. Questa dicotomia è giustificata dal fatto che nella parte destra del grafico ci si riferisce ai brevetti stranieri, i quali molto spesso sono privi di alcune delle informazioni più importanti, probabilmente facoltative in fase di deposito della domanda. Diversamente, i sistemi locali selezionati indicano una *frequenza* simile e che si attesta su valori molto più bassi. Infine, si effettua la stessa valutazione esaminando i codici IPC al livello delle 8 sezioni tecnologiche. La figura B.3 mostra come anche da questo punto di vista i record che contengono un valore nullo per il codice fiscale dei richiedenti sono distribuiti in maniera omogenea.

Figura B.3: Analisi per classi tecnologiche dei valori nulli per il codice fiscale dei richiedenti



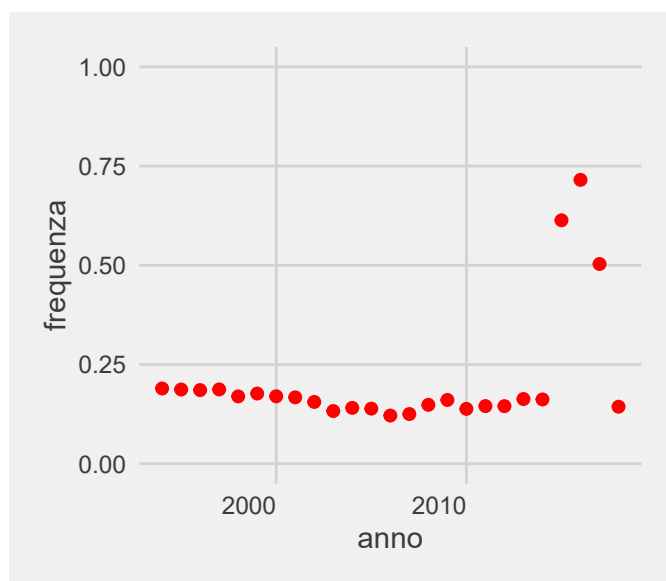


## Appendice C

# Pulizia dati (Parte II)

Se l'analisi per classi tecnologiche risulta in linea con quanto già verificato in figura B.3, non è possibile affermare lo stesso per l'analisi temporale: infatti, come mostrato in figura C.1, la frequenza relativa dei casi senza codice del sistema locale del lavoro è superiore al 50 % negli anni 2015, 2016 e 2017 e significativamente diversa rispetto agli analoghi valori degli altri anni del database considerato.

Figura C.1: Analisi temporale dei valori nulli per codice del sll



Questa notevole differenza richiede ulteriori approfondimenti per capire come trattare al meglio questo dato. Si sceglie come riferimento l'anno 2009, che si delinea come ottima approssimazione dei valori medi del dataset rispetto al numero delle domande di brevetto e alla loro distribuzione spaziale e tecnologica. Si procede quindi a valutare se i valori nulli per il codice del sll siano uniformemente distribuiti rispetto ai sistemi locali. Per fare ciò, si confrontano i tre periodi citati con il *benchmark* e si ripete la stessa operazione con

un periodo che non presenti le medesime carenze informative. Per quest'ultimo scopo si sceglie di considerare l'anno 2018, in quanto certamente temporalmente prossimo agli anni oggetto di analisi e non troppo vicino al 2009, con il rischio di incorporare possibili trend temporali che potrebbero condurre ad una valutazione distorta dei grafici. Le figure C.2, C.3, C.4 e C.5 mostrano i risultati ottenuti. Appare immediato come nei primi tre grafici la distribuzione dell'attività inventiva (misurata in ordinata dal contatore *frequenza*) sia significativamente diversa nella coppia di anni presa in considerazione.

Figura C.2: Confronto distribuzione brevettuale nel tempo.1

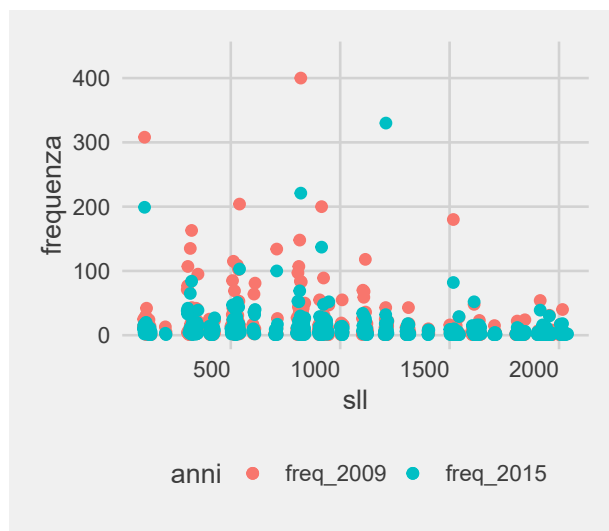


Figura C.3: Confronto distribuzione brevettuale nel tempo.2

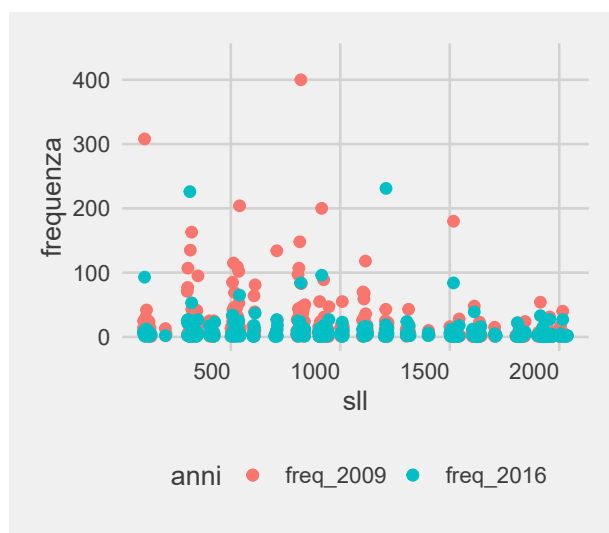


Figura C.4: Confronto distribuzione brevettuale nel tempo.3

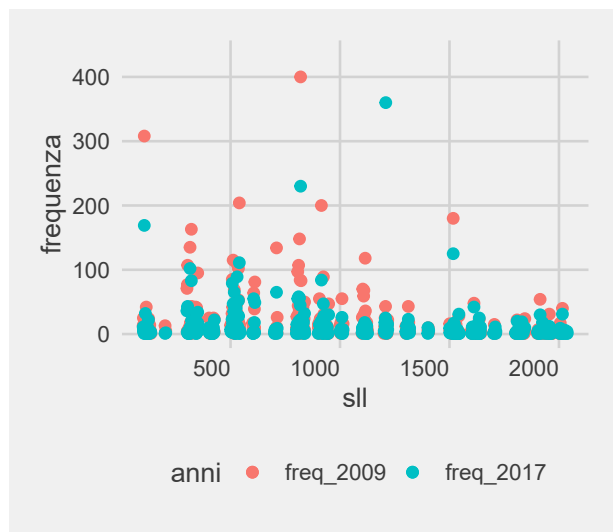
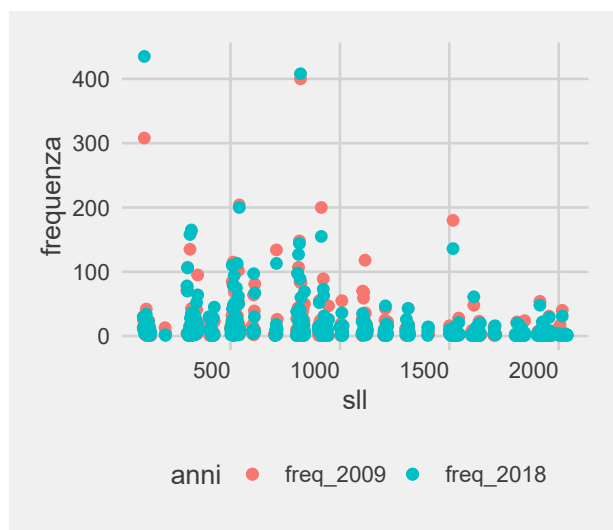


Figura C.5: Confronto distribuzione brevettuale nel tempo.4



Diversamente, la figura C.5 mostra differenze meno marcate e, in ultima analisi, un quadro accettabile, in quanto evidentemente le due distribuzioni non potranno mai essere perfettamente sovrapposte.





# Appendice D

## Classificazione IPC: sezione A

---

1	A01	Agriculture; forestry; animal husbandry; hunting; trapping; fishing
2	A21	Baking; equipment for making or processing doughs; doughs for baking
3	A22	Butchering; meat treatment; processing poultry or fish
4	A23	Foods or foodstuffs; their treatment, not covered by other classes
5	A24	Tobacco; cigars; cigarettes; smokers' requisites
6	A41	Wearing apparel
7	A42	Headwear
8	A43	Footwear
9	A44	Haberdashery; jewellery
10	A45	Hand or travelling articles
11	A46	Brushware
12	A47	Furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in general
13	A61	Medical or veterinary science; hygiene
14	A62	Life-saving; fire-fighting
15	A63	Sports; games; amusements
16	A99	Subject matter not otherwise provided for in this section

---



# Appendice E

## Classificazione IPC: sezione B

---

1	B01	Physical or chemical processes or apparatus in general
2	B02	Crushing, pulverising, or disintegrating; preparatory treatment of grain for milling
3	B03	Separation of solid materials using liquids or using pneumatic tables or JIGS; etc.
4	B04	Centrifugal apparatus or machines for carrying-out physical or chemical processes
5	B05	Spraying or atomising in general; applying liquids or other fluent materials to surfaces
6	B06	Generating or transmitting mechanical vibrations in general
7	B07	Separating solids from solids; sorting
8	B08	Cleaning
9	B09	Disposal of solid waste; reclamation of contaminated soil
10	B21	Mechanical metal-working without essentially removing material; punching metal
11	B22	Casting; powder metallurgy
12	B23	Machine tools; metal-working not otherwise provided for
13	B24	Grinding; polishing
14	B25	Hand tools; portable power-driven tools; handles for hand implements;
15	B26	Hand cutting tools; cutting; severing
16	B27	Working or preserving wood or similar material; nailing or stapling machines
17	B28	Working cement, clay, or stone
18	B29	Working of plastics; working of substances in a plastic state in general
19	B30	Presses
20	B31	Making articles of paper, cardboard or material worked in a manner analogous to paper
21	B32	Layered products
22	B33	Additive manufacturing technology
23	B41	Printing; lining machines; typewriters; stamps
24	B42	Bookbinding; albums; files; special printed matter
25	B43	Writing or drawing implements; bureau accessories
26	B44	Decorative arts
27	B60	Vehicles in general
28	B61	Railways
29	B62	Land vehicles for travelling otherwise than on rails
30	B63	Ships or other waterborne vessels; related equipment
31	B64	Aircraft; aviation; cosmonautics
32	B65	Conveying; packing; storing; handling thin or filamentary material
33	B66	Hoisting; lifting; hauling
34	B67	Opening or closing bottles, jars or similar containers; liquid handling
35	B68	Saddlery; upholstery
36	B81	Microstructural technology
37	B82	Nanotechnology
38	B99	Subject matter not otherwise provided for in this section

---



## Appendice F

# Algoritmo di creazione della *Edge list*

```
#CALCOLO VARIABILI DI RETE
```

```
#ordino dati
```

```
df_totale <- df_totale[order(df_totale$code, decreasing = FALSE)  
                        , ]
```

```
#nodi
```

```
nodes <- as.data.frame(cbind(df_totale$fiscal_code ,  
                             df_totale$llma ,  
                             df_totale$city_name ,  
                             df_totale$Year ,  
                             df_totale$cod_reg ,  
                             df_totale$identifier.y))
```

```
colnames(nodes) <- c("fiscal_code", "llma", "city_name", "year",  
                    "cod_reg", "cod_prov")
```

```
#distinct fatto per ogni anno (se lo stesso inventore Ã
```

```
#presente in due anni diversi ci deve essere due volte)
```

```
nodes <- distinct(nodes, fiscal_code, year, .keep_all = TRUE)
```

```
#edges
```

```
Edges <- data.frame()
```

```
#Ciclo for + while per costruire una edge_list
```

```

j <- 1
i <- 0

for(i in 1:(nrow(df_totale)-1)){

  j <- i

  while(df_totale[i,"code"] == df_totale[j+1, "code"]){

    #Estraggo fiscal_code (x2), code, year, llma (x2)

    result <- c(df_totale$fiscal_code[i],
                df_totale$fiscal_code[j+1],
                df_totale$code[i], df_totale$Year[i],
                df_totale$llma[i], df_totale$llma[j+1])

    Edges <- rbind(Edges, result)

    if(j < nrow(df_totale)-1) {

      j <- j + 1

    }
    else{
      break
    }
  }
}

colnames(Edges) <- c ("fiscal_code_from", "fiscal_code_to",
                    "code", "year", "llma_from", "llma_to")

```

## Appendice G

# Algoritmo di costruzione dei regressori

```
#density [edges/((N*(N-1)/2) + N)]  
  
density[z,k] <- edge_density(grafo_parziale , loops = TRUE)  
  
#ottengo '0' quando ho zero collegamenti con un certo  
numero di nodi presenti  
#ottengo NaN quando non ho nodi presenti all'interno  
di quel llma  
  
#max k-core  
  
coreness <- graph.coreness(grafo_parziale)  
maxCoreness[z,k] <- max(coreness)  
  
#warnings() #ho "-inf" come valore quando ho  
grafo_parziale fatto di nodi non collegati da archi,  
#i.e. Á un llma che ha nodi con collegamenti esterni  
a llma di riferimento  
  
#social_proximity  
  
social_proximity[z,k] <- transitivity(  
  grafo_parziale ,  
  type = "global",  
  vids = NULL,  
  weights = NULL,  
  isolates = "zero"  
)
```



```

#Share of applicants on population

N_applicants[z,k] <- count(distinct(y1_temp
[y1_temp$llma == j, ], fiscal_code, .keep_all = TRUE))
Share_applicants[z,k] <- N_applicants[z,k]
/Population_SLL[z,1]

#External links, BREADTH (numero di
collegamenti esterni a llma/collegamenti
totali llma)

edges_from <- count(Edges_temp
[Edges_temp$llma_from == j & Edges_temp$llma_to != j, ])
edges_to <- count(Edges_temp
[Edges_temp$llma_from != j & Edges_temp$llma_to == j, ])
Breadth[z,k] <- (edges_to + edges_from) /
(nrow(Edges_temp[Edges_temp$llma_from == j
| Edges_temp$llma_to == j, ]))

#External links, DEPTH (HHI[si calcola
concentrazione di ogni collegamento esterno]/ Breadth)

edges_depth_1 <- Edges_temp[Edges_temp$llma_from == j &
Edges_temp$llma_to != j, ]
edges_depth_2 <- Edges_temp[Edges_temp$llma_from != j &
Edges_temp$llma_to == j, ]
edges_depth <- c(edges_depth_2$llma_from,
edges_depth_1$llma_to)

shares_q <- c()

b <- table(edges_depth)
counts <- as.vector(b)
shares <- counts/length(counts)

v <- 1
for(w in 1:length(shares)){
  shares_q[w] <- shares[w]^2
  v <- v + 1
}

HHI <- sum(shares_q)

```

```

    Depth[z,k] <- HHI/Breadth[z,k]
  z <- z + 1
}

k <- k + 1
#Inizializzo le 3 variabili
df_reduced_ipc_temp <- data.frame()

UV <- data.frame(matrix(0, ncol = 22,
nrow = length(llmas_code)))
row.names(UV) <- LLMA$name
colnames(UV) <- anni

SRV <- data.frame(matrix(0, ncol = 22,
nrow = length(llmas_code)))
row.names(SRV) <- LLMA$name
colnames(SRV) <- anni

RV <- data.frame(matrix(0, ncol = 22,
nrow = length(llmas_code)))
row.names(RV) <- LLMA$name
colnames(RV) <- anni

z <- 1
k <- 1

for(i in anni){

  z <- 1
  df_reduced_ipc_temp <- df_reduced_ipc
  [df_reduced_ipc$year == i, ]

  for(j in llmas_code){

    df_ipc <- df_reduced_ipc_temp
    [df_reduced_ipc_temp$llma == j, ]
    if(nrow(df_ipc) > 0){

      #calcolo senza quote dei codici

      Pg <- df_ipc %>%
        group_by(section) %>%
        count(section)
      df_ipc$n <- as.numeric(df_ipc$n)
    }
  }
}

```

```

#correzione con quote

Adjust <- aggregate(df_ipc$n,
by=list(Section=df_ipc$section), FUN=sum)
Pg$n <- Adjust$x
Pg$shares <- c(Pg$n/sum(Pg$n))

#Unrelated

UV_sing <- c()
r <- 1

for(p in Pg$shares){

  UV_sing[r] <- p*log(1/p)
  r <- r + 1

}

UV[z,k] <- sum(UV_sing)

#Semi related

Ps <- df_ipc %>%
  group_by(digit3) %>%
  count(digit3)
df_ipc$n <- as.numeric(df_ipc$n)
Adjust2 <- aggregate(df_ipc$n,
by=list(Section=df_ipc$digit3), FUN=sum)
Ps$n <- Adjust2$x
Ps$shares <- c(Ps$n/sum(Ps$n))

SRV_sing <- c()
r <- 1

for(p in Ps$shares){

  SRV_sing[r] <- p*log(1/p)
  r <- r + 1

}

SRV_sum <- sum(SRV_sing)
SRV[z,k] <- SRV_sum - UV[z,k]

```

```

#Related

Pr <- df_ipc %>%
  group_by(digit4) %>%
  count(digit4)
df_ipc$n <- as.numeric(df_ipc$n)
Adjust3 <- aggregate(df_ipc$n,
by=list(Section=df_ipc$digit4), FUN=sum)
Pr$n <- Adjust3$x
Pr$shares <- c(Pr$n/sum(Pr$n))

RV_sing <- c()
r <- 1

for(p in Pr$shares){

  RV_sing[r] <- p*log(1/p)
  r <- r + 1

}

RV_sum <- sum(RV_sing)
RV[z,k] <- RV_sum - SRV[z,k]

z <- z + 1

} else{

UV[z,k] <- NA
SRV[z,k] <- NA
RV[z,k] <- NA

z <- z + 1

}

}

k <- k + 1
}
}

```



# Appendice H

## Analisi di robustezza dei modelli econometrici

Tabella H.1: Modelli di regressione

Modello	(1)	(2)	(3)	(4)	(5)	(6)	(7)
density	-7.183*** (2.291)	-7.292*** (2.274)	0.052 (2.753)	-8.277*** (3.149)	-8.309*** (3.178)	-8.289*** (3.159)	-2.649 (2.752)
maxcoreness	1.869*** (0.477)	1.808*** (0.464)		1.807*** (0.472)	1.804*** (0.473)	1.805*** (0.472)	0.805** (0.391)
sproximity	-2.867*** (0.746)	-2.375*** (0.663)	-0.691 (0.888)	-2.175*** (0.733)	-2.179*** (0.733)	-2.169*** (0.733)	-0.827 (0.821)
shapplicants	5,182.652* (3,114.709)	-1,962.258 (3,449.348)	3,397.231 (3,230.597)	-1983.084 (4,065.310)	-1,084.275 (4,768.331)	-971.080 (4,373.030)	-12,121.888*** (4,497.323)
extbreadth			-1.915 (1.890)				
extdepth			0.004 (0.010)	0.006 (0.009)	0.006 (0.009)	0.006 (0.009)	-0.006 (0.009)
uvariety					-0.679 (0.619)		0.967 (0.588)
srvariety					-0.546 (0.888)		
rvariety						-0.867 (0.578)	
gdp							-0.004 (0.005)
education							0.668* (0.374)
rd							0.000 (0.000)
densitypop							-0.086* (0.047)
Constant	13.720*** (0.672)	19.928*** (1.020)	24.156*** (2.410)	22.157*** (1.227)	23.425*** (1.396)	23.436*** (1.426)	39.942*** (19.631)
year dummies	NO	YES	YES	YES	YES	YES	YES
spatial dummies	YES	YES	YES	YES	YES	YES	YES
Observations	1,623	1,623	1,441	1,441	1,435	1,435	1,159
Number of groups	491	491	465	465	463	463	446
Average group size	3.305	3.305	3.099	3.099	3.099	3.099	2.599
Panel-level standard deviation	36.826	37.107	39.364	38.063	38.341	38.363	52.940
R-squared for within model	0.15207	0.22131	0.10002	0.23122	0.23200	0.23215	0.07609
R-squared for between model	0.51168	0.39095	0.00993	0.35683	0.29453	0.28740	0.22800
R-squared for overall model	0.26334	0.18843	0.01063	0.18348	0.15760	0.15325	0.23163
Root mean squared error	7.760	7.445	8.475	7.833	7.851	7.848	4.907

La variabile dipendente  $\dot{A}$  è calcolata utilizzando il numero di brevetti pro capite di ciascun sl (per 100k abitanti). Tutti i modelli di regressione includono il termine costante e nel modello (7) sono inserite le principali variabili di controllo utilizzate nei modelli di crescita economica (gdp, rd, education, densitypop). I modelli sono stimati con il metodo OLS e con un *variance estimator clustered* al livello dei controlli spaziali (i.e., sl). Gli asterischi da uno a tre indicano un livello di significatività rispettivamente pari al 10%, 5% e 1%. Tra parentesi sono riportati i relativi errori standard.