

Bari e Napoli prima dell'Alta Velocità un'analisi delle reti di conoscenza, dei flussi universitari e della mobilità lavorativa

Giulio Albano Francesco Prota

Bari e Napoli prima dell'Alta Velocità: un'analisi delle reti di conoscenza, dei flussi universitari e della mobilità lavorativa

Giulio Albano¹, Francesco Prota¹

¹ University of Bari, DIEF, Uniba (IT)

Abstract. Questo working paper costruisce una baseline empirica ex ante all'apertura della linea AV/HSR Napoli–Bari, con l'obiettivo di misurare lo “stato di integrazione” tra i due poli lungo tre canali complementari: collaborazione scientifica, mobilità studentesca e reti professionali. Sul piano cognitivo, un'analisi bibliometrica su Web of Science (1985–2024) documenta un rafforzamento persistente delle co-pubblicazioni tra istituzioni di Bari e Napoli, suggerendo la presenza di legami già consolidati che l'HSR potrebbe intensificare. Sul piano formativo, i microdati MUR mostrano che nel 2024/2025 Puglia e Campania presentano elevata ritenzione (73,7% e 87,2%), flussi bilaterali contenuti (Puglia→Campania 192, Campania→Puglia 76), e un ruolo centrale della Basilicata come bacino di uscita (self-flow 24,1%, con Basilicata→Puglia 611 e Basilicata→Campania 293). Sul piano professionale, si è sperimentato l'uso del webscraping su LinkedIn attraverso Proxy URL, enrichment e classificazioni LLM vincolate. Si è costruito un DF con 105 profili LinkedIn con connessioni effettive Bari–Napoli e 1.328 eventi: le traiettorie risultano raramente “a due nodi”, con il resto d'Italia come hub (destinazione finale 57% in stage4) e connessione spesso mediata dall'università (Bari 57,4%, Napoli 45,5). Nei passaggi lavorativi diretti emergono asimmetrie qualitative: Bari→Napoli più tecnico-accademico, Napoli→Bari più orientato a impresa e funzioni economico-manageriali. Nel complesso, i risultati indicano una relazione già connessa ma non pienamente integrata, e forniscono una base di riferimento multidimensionale per future valutazioni causali degli effetti post-AV/HSR, tenendo conto dei potenziali rischi di polarizzazione.

Keywords: High speed Rail, Linkedin, web scraping

1 Introduzione

Lo sviluppo della linea ferroviaria ad Alta Velocità (AV/HSR) Napoli–Bari rappresenta uno dei più significativi investimenti infrastrutturali per il Mezzogiorno, inserendosi nel quadro più ampio della rete transeuropea dei trasporti (TEN-T) lungo il Corridoio Scandinavo-Mediterraneo (Commissione Europea, 2021; Cascetta et al., 2020). La sua realizzazione porta con sé l'ambizione implicita di ridurre la distanza economica tra i sistemi urbani, migliorare l'accessibilità e favorire nuove forme di integrazione territoriale in una

regione storicamente frammentata (Pagliara et al., 2021). onnettere Adriatico e Tirreno significa quindi collocare l'infrastruttura all'interno di una strategia di sviluppo orientata alla crescita e alla coesione, che necessita di un coordinamento esplicito con le politiche territoriali e di specializzazione intelligente per evitare effetti polarizzanti e massimizzare le complementarità tra sistemi locali (Prota & Viesti, 2022).

In letteratura, l'HSR è interpretata come una tecnologia abilitante capace di ridurre i costi generalizzati di trasporto, rafforzare le interazioni *face-to-face* e, in determinate condizioni, sostenere processi di policentrismo e diffusione della conoscenza (Gutiérrez, 2001; Cheng, Loo, & Vickerman, 2015; Bosurgi et al., 2025). Tuttavia, l'evidenza empirica suggerisce risultati eterogenei e fortemente dipendenti dal contesto: è documentato il rischio di un "effetto sifone" (*siphon effect*), in cui i poli principali vengono rafforzati a scapito di quelli secondari, amplificando potenzialmente gli squilibri spaziali qualora le complementarità locali e la connettività dell'ultimo miglio risultino deboli (Chen & Haynes, 2015; Zhang, Wan, & Yang, 2019; Albano & Pagliara, 2025).

In questo quadro, un limite ricorrente negli studi *ex-ante* e nelle valutazioni di impatto territoriale è la difficoltà di misurare l'effettivo "stato di integrazione" tra le città prima che l'infrastruttura diventi operativa. Ciò è particolarmente vero per le dimensioni immateriali dell'integrazione, legate alla circolazione della conoscenza e del capitale umano. Questo lavoro muove dalla premessa che, per comprendere in modo credibile i futuri impatti della Napoli–Bari, sia essenziale costruire una *baseline* empirica multidimensionale. Tale base di riferimento deve osservare, attraverso fonti di dati diverse ma coerenti, tre canali chiave attraverso i quali l'HSR influenza lo sviluppo: (i) la cooperazione scientifica e la capacità innovativa; (ii) la mobilità studentesca e la formazione del capitale umano; (iii) le reti professionali e i segnali di mobilità del lavoro qualificato.

In primo luogo, la letteratura suggerisce che l'aumento della velocità di trasporto faciliti la collaborazione tra lavoratori altamente qualificati distribuiti nello spazio, specialmente nelle attività *knowledge-intensive* dove la qualità e la frequenza delle interazioni influenzano direttamente la produttività (Dong, Zheng, & Kahn, 2020). Evidenze recenti indicano che i collegamenti HSR possono stimolare la performance innovativa degli attori accademici e rafforzare i meccanismi di trasferimento tecnologico, migliorando il processo di *matching* tra università e imprese (Li et al., 2024). In linea con questo filone, e adottando un approccio bibliometrico per misurare l'integrazione scientifica (Katz & Martin, 1997), lo studio integra un'analisi basata su *Web of Science*. Ciò permette di ricostruire l'evoluzione di lungo periodo della produzione scientifica e delle co-pubblicazioni tra gli atenei delle due aree (1985–2024), evidenziando un marcato incremento delle collaborazioni nell'ultimo

decennio e una crescente densità di rete che potrebbe costituire un terreno fertile per ulteriori sinergie dopo l'apertura della linea.

In secondo luogo, il capitale umano si forma e si redistribuisce attraverso la mobilità studentesca. I collegamenti HSR possono influenzare le scelte di iscrizione — incidendo sull'attrattività, sulla ritenzione e sul pendolarismo a lunga distanza — interagendo al contempo con le differenze strutturali tra i sistemi urbani e universitari (Delaplace et al., 2014). Nel caso Bari–Napoli, il sistema accademico appare eterogeneo, con differenze significative nella dimensione degli atenei, nella loro composizione e nel peso delle università telematiche. Questi fattori sono potenzialmente decisivi per interpretare i flussi "reali" di mobilità rispetto ai semplici dati di iscrizione. Di conseguenza, il paper include un'analisi dei flussi di immatricolazione dai dataset del MUR (Ministero dell'Università e della Ricerca), ricostruendo matrici origine-destinazione e indicatori sintetici di scambio formativo, con un focus specifico sulle traiettorie interregionali e sui pattern di concentrazione/dispersione.

In terzo luogo, la misurazione dell'integrazione territoriale richiede l'osservazione della dimensione "post-formazione", ovvero la mobilità e l'interazione professionale. Parallelamente alla letteratura emergente che utilizza tracce digitali e reti spaziali — come i segnali derivanti dai social media — per interpretare le interazioni tra i luoghi (Wu et al., 2023), proponiamo una strategia basata sui dati di LinkedIn. Attraverso tecniche di *scraping* e arricchimento dei dati (*data enrichment*) per estrarre segnali sulle traiettorie educative e di carriera collegate ai nodi di Bari e Napoli, ci allineiamo alle recenti tendenze metodologiche nell'uso dei "Big Data" per l'analisi del mercato del lavoro (Zherdev, 2020; Stateva et al., 2021). L'obiettivo non è sostituire le fonti amministrative sul lavoro, ma costruire indicatori esplorativi e replicabili sulla struttura delle reti professionali e sulle direzioni della mobilità.

Il contributo di questo *working paper* è dunque duplice. Sul piano sostantivo, fornisce una *baseline* integrata della "capacità di connessione" tra Bari e Napoli lungo tre dimensioni (scienza, formazione e lavoro qualificato) che vengono spesso analizzate separatamente. Come evidenziato da Prota e Calò (2021), i sistemi di monitoraggio e valutazione delle politiche territoriali risultano decisivi per orientare gli investimenti verso traiettorie effettivamente trasformatrici e per evitare che le risorse si disperdano in interventi non coordinati.

Sul piano metodologico, propone un disegno multi-sorgente che combina bibliometria, dati amministrativi e tracce digitali. Questo approccio mira a rendere osservabili i meccanismi di integrazione e i potenziali *trade-off* distributivi, in coerenza con l'evidenza che gli effetti dell'HSR possono essere allo stesso tempo abilitanti e polarizzanti (Albano & Pagliara, 2025; Zhang, Wan, & Yang, 2019; Chen & Haynes, 2015).

2 Integrazione della produzione scientifica tra Bari e Napoli

La collaborazione scientifica rappresenta una dimensione centrale dell'integrazione cognitiva tra territori, in quanto riflette la capacità di sistemi accademici distinti di condividere conoscenza, risorse e competenze specializzate (Katz & Martin, 1997). La letteratura sui sistemi di innovazione evidenzia come la prossimità geografica, pur non essendo una condizione necessaria, continui a svolgere un ruolo rilevante nel facilitare le interazioni conoscitive e la circolazione di conoscenza tacita (Boschma, 2005; Ponds et al., 2007). In questo senso, la densità delle collaborazioni accademiche è comunemente utilizzata come indicatore dell'integrazione spaziale dei sistemi di ricerca (Balland, 2012).

Nel contesto delle infrastrutture di trasporto, numerosi studi mostrano come la riduzione dei tempi di viaggio possa favorire la collaborazione tra ricercatori localizzati in città diverse, riducendo i costi degli incontri faccia-a-faccia e facilitando scambi ripetuti e coordinamento progettuale (Gutiérrez, 2001; Dong, Zheng, & Kahn, 2020). Evidenze empiriche indicano inoltre che l'alta velocità ferroviaria è associata a un aumento delle co-pubblicazioni tra università collegate, soprattutto in presenza di reti scientifiche già attive (Wang et al., 2020; Li et al., 2024; Garmendia et al., 2012; He et al., 2020).

Tuttavia, l'evidenza rimane eterogenea e fortemente dipendente dal contesto istituzionale. Come discusso in Albano e Pagliara (2025), l'HSR tende prevalentemente a rafforzare reti di collaborazione preesistenti, mentre i suoi effetti redistributivi tra poli accademici di diversa dimensione risultano ambigui. In questo quadro, l'analisi delle co-pubblicazioni viene qui utilizzata in chiave descrittiva per mappare il grado di integrazione cognitiva tra sistemi universitari che saranno interconnessi dal collegamento Napoli–Bari, senza finalità di identificazione causale.

Alla luce di queste considerazioni, la sezione analizza la struttura e l'evoluzione della collaborazione scientifica tra Bari e Napoli attraverso dati bibliometrici estratti da Web of Science, osservando la dinamica delle pubblicazioni congiunte nel periodo 1985–2024 e la configurazione delle reti accademiche interurbane (Glänzel & Schubert, 2004).

2.1 Dati e Metodologia

L'analisi si basa su un dataset bibliometrico estratto da Web of Science, comprendente le pubblicazioni scientifiche indicizzate nel periodo 1985–2024. Sono stati selezionati tutti gli articoli che includono almeno un'affiliazione riconducibile a istituzioni universitarie localizzate nelle aree urbane di Bari e Napoli. L'utilizzo delle co-pubblicazioni come unità di osservazione è consolidato in letteratura quale proxy delle interazioni conoscitive tra istituzioni e territori (Katz & Martin, 1997; Ponds, Van Oort, & Frenken, 2007).

A partire da questo insieme di record, sono stati costruiti tre indicatori principali:

- (i) produzione scientifica totale per anno e per area urbana;
- (ii) co-pubblicazioni interurbane, definite come articoli con almeno un'affiliazione a Bari e una a Napoli;
- (iii) densità della rete di collaborazione, calcolata a partire dalle relazioni di co-authorship tra istituzioni.

L'approccio adottato è descrittivo e longitudinale, in linea con la letteratura che utilizza dati bibliometrici per analizzare l'evoluzione delle reti scientifiche nel tempo in assenza di disegni controfattuali (Hoekman, Frenken, & Van Oort, 2009; Balland et al., 2020). L'obiettivo è documentare l'evoluzione storica delle interazioni scientifiche tra i due sistemi universitari, fornendo una baseline informativa per l'interpretazione di potenziali effetti futuri legati al miglioramento dell'accessibilità.

2.2 Risultati

Il sistema universitario lungo l'asse Napoli–Bari, sintetizzato nelle [Tabella 1](#) e [Figura 2.1](#), mostra una marcata eterogeneità istituzionale e dimensionale. Napoli presenta un ecosistema variegato, caratterizzato dalla coesistenza di grandi poli statali storici, come l'Università Federico II (oltre 71.000 iscritti), e un forte comparto non statale come l'Università Telematica "Pegaso", che con quasi 100.000 iscritti e un'elevata incidenza di docenti a contratto (68,8%), delinea un modello formativo dematerializzato che impone una riflessione specifica sulle dinamiche di mobilità fisica.

Al contrario, il polo di Bari appare più concentrato e legato a modelli accademici tradizionali. L'Università "Aldo Moro" (circa 40.000 iscritti) funge da principale hub generalista della regione, affiancata dal Politecnico di Bari, istituzione d'eccellenza con una forte specializzazione tecnica e una bassa incidenza di docenti a contratto (15,1%).

Tabella 1

Atenei tra Bari e Napoli

Ateneo	Tipo di istituto	Iscritti 2023/2024	QS Rankin g 2025	Personale Docente e ricercatore	Di cui prof a contratto	Personal e non docente
Università Telematica "Pegaso"	Non statale	99.556	N.C.	369	68,8%	300

U.d.S. di Napoli "Federico II"	Statale	71.170	347	4.080	13,3%	2.376
U.d.S. della Campania "Luigi Vanvitelli"	Statale	22.948	651-700	1.381	13,5%	1.315
U.d.S. di Napoli "Parthenope"	Statale	13.387	1001-1200	448	12,1%	300
U.d.S. di Napoli "L'Orientale"	Statale	9.328	51-100	357	18,2%	271
U.d.S. "Suor Orsola Benincasa"	Non statale	8.800	N.C.	442	75,4%	201
Totale Napoli		225.189		7.077	0,34	4.763
Totale non telematiche Na.		125.633		6.708	0,27	4.463
U.d.S. di Bari Aldo Moro	Statale	39.870	801-850	2.158	15,1%	1.415
Politecnico di Bari	Statale	10.036	580	679	21,6%	276
Libera Università Mediterranea LUM	Non statale	2.016	N.C.	369	79,1%	75
Totale Bari		51.922		3.206	38,6%	1.766
Totale		277.111		10.283	36%	6.529
Totale non telematiche		177.555		9.914	33%	6.229

Tabella 1 Elaborazione dell'autore su dati Ministero dell'Università e della Ricerca (MUR), Portale USTAT (2025)

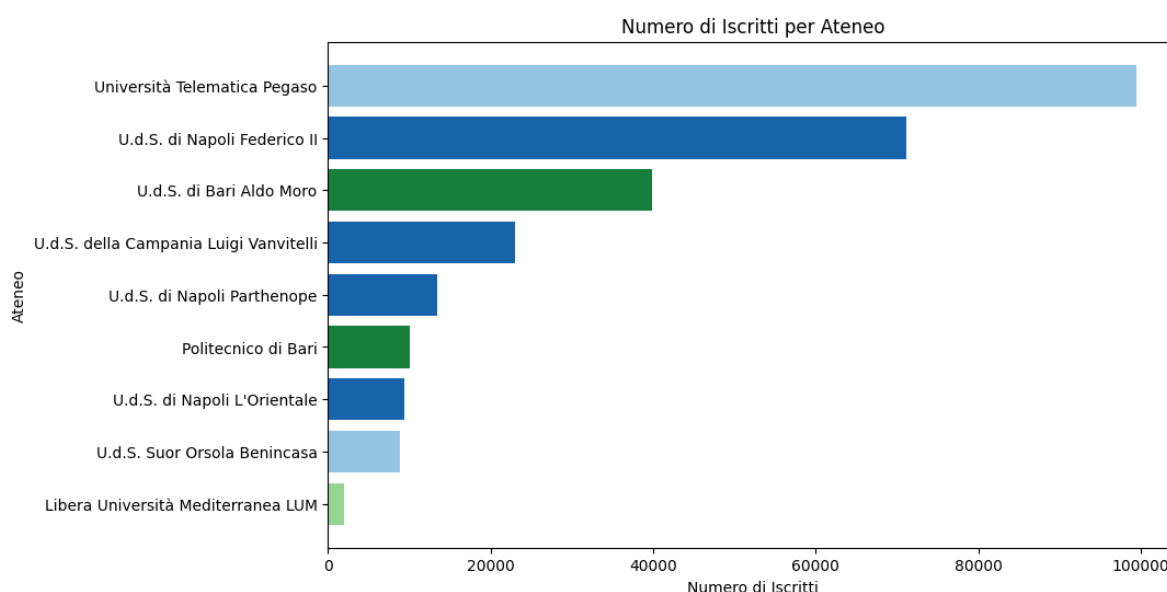


Figura 2.1 Elaborazione dell'autore su dati Ministero dell'Università e della Ricerca (MUR), Portale USTAT (2025)

La produzione scientifica degli atenei localizzati nelle aree urbane di Bari e Napoli evidenzia una crescita sostenuta lungo l'intero periodo di osservazione (1985–2024), come riportato in [Tabella A.1](#) e

illustrato in [Figura 2.2](#). Entrambi i sistemi universitari mostrano un aumento progressivo del numero di pubblicazioni indicizzate in Web of Science, in linea con le dinamiche di lungo periodo del sistema accademico.

Nel confronto tra le due aree, Napoli presenta livelli di produzione scientifica sistematicamente superiori rispetto a Bari, riflettendo la diversa scala e articolazione dei rispettivi sistemi universitari. In particolare, l'Università degli Studi di Napoli "Federico II" rappresenta il principale contributore in termini di volume di pubblicazioni, mentre altri atenei statali dell'area napoletana mostrano una crescita rilevante nel tempo. A Bari, l'Università degli Studi "Aldo Moro" concentra la quota prevalente della produzione scientifica, affiancata da una dinamica di crescita progressiva del Politecnico di Bari; le istituzioni di minori dimensioni risultano invece marginali.

Dal punto di vista temporale, entrambe le aree urbane registrano un'accelerazione della crescita della produzione scientifica a partire dalla metà degli anni 2010. Tale andamento è coerente con processi di rafforzamento delle politiche di ricerca, internazionalizzazione e integrazione nelle reti scientifiche, ampiamente documentati nella letteratura sulle traiettorie cumulative della conoscenza (Hoekman, Frenken, & Van Oort, 2009; Balland et al., 2020).

Nel complesso, questi risultati delineano due sistemi universitari caratterizzati da scale differenti ma accomunati da una dinamica di crescita nel lungo periodo.

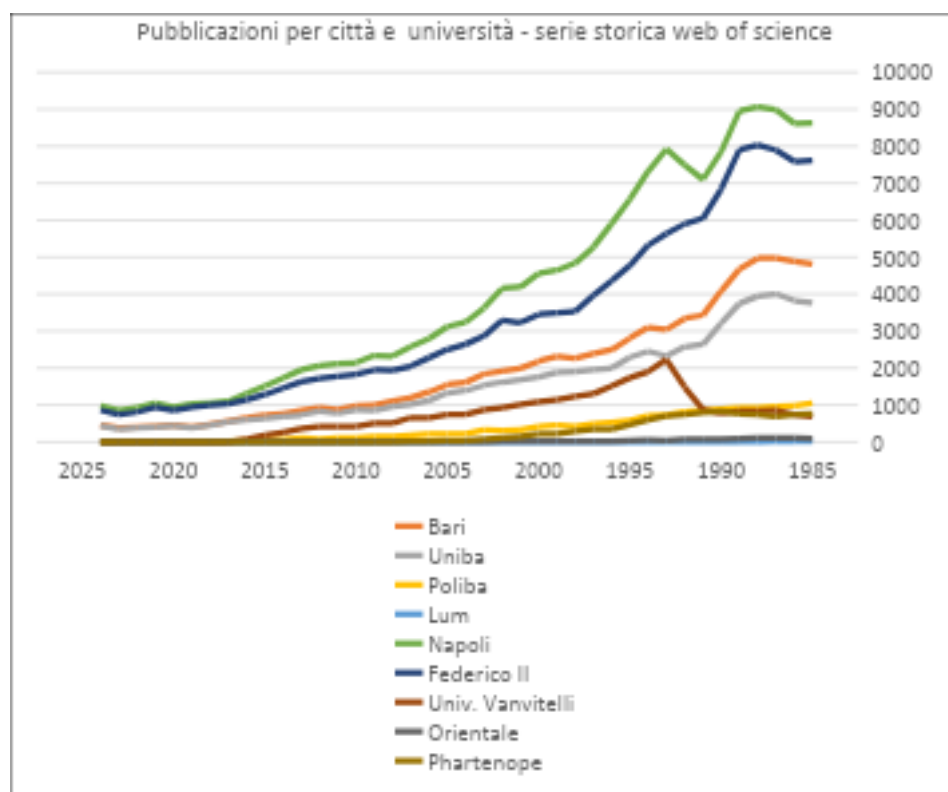


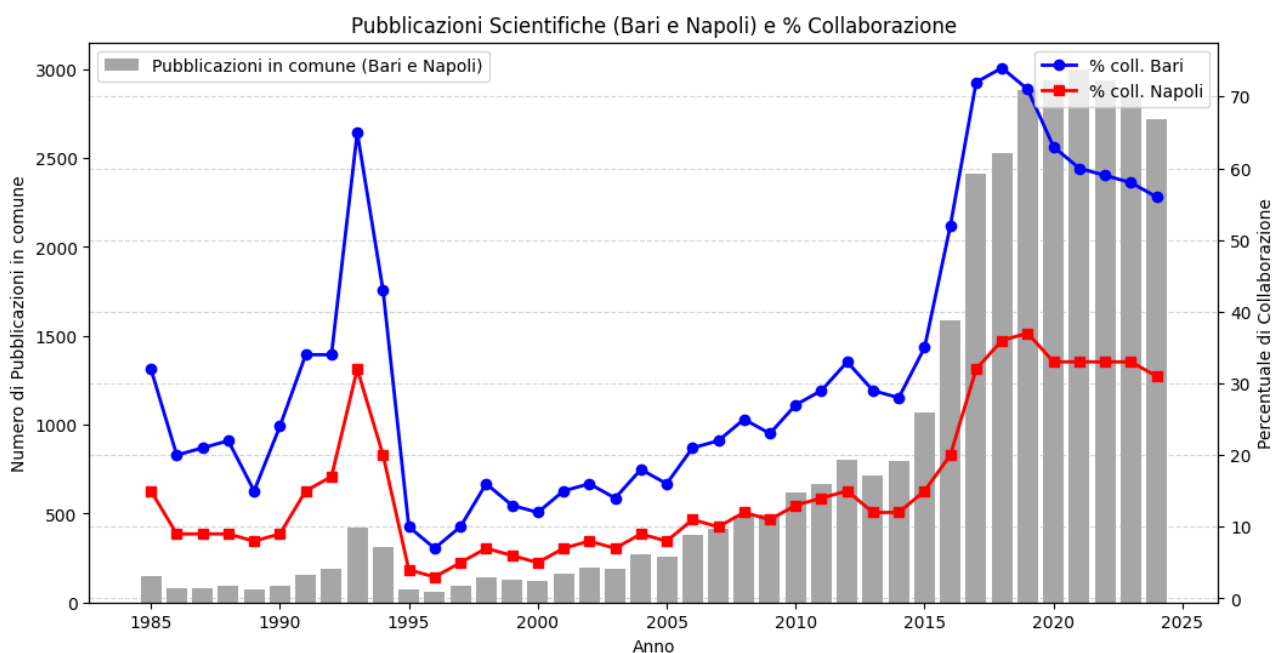
Figura 2.2 Elaborazione dell'autore su dati Web of Science

Le collaborazioni scientifiche tra gli atenei di Bari e Napoli mostrano una crescita marcata e persistente lungo l'intero periodo di osservazione (1985–2024), come documentato in [Tabella A.2](#) e [Figura 2.3](#). Il numero di pubblicazioni congiunte aumenta progressivamente nel tempo, con un'accelerazione evidente a partire dalla metà degli anni 2000 e una dinamica particolarmente intensa negli anni più recenti.

Sebbene Napoli presenti livelli assoluti di produzione scientifica superiori, la quota di pubblicazioni realizzate in collaborazione con istituzioni baresi risulta sistematicamente più elevata per Bari. Questa asimmetria è coerente con quanto osservato in letteratura per sistemi universitari di diversa scala, in cui gli atenei di dimensioni relativamente minori tendono a mostrare una maggiore apertura verso collaborazioni esterne come strategia di accesso a risorse scientifiche e reputazionali complementari (Ponds, Van Oort, & Frenken, 2007; Hoekman, Frenken, & Van Oort, 2009).

Dal punto di vista temporale, l'incremento della quota di co-pubblicazioni suggerisce il progressivo consolidamento di relazioni scientifiche interurbane, in linea con dinamiche cumulative tipiche delle reti di conoscenza accademica (Balland et al., 2020). Tali evidenze indicano l'esistenza di una base di integrazione cognitiva già strutturata tra i due sistemi universitari, che precede l'entrata in esercizio della linea ferroviaria ad alta velocità.

In questo contesto, il collegamento HSR Napoli–Bari può essere interpretato come un potenziale fattore abilitante piuttosto che come un generatore ex novo di collaborazioni, agendo sulla riduzione dei costi di interazione e sulla frequenza degli scambi tra ricercatori già inseriti in reti scientifiche condivise (Dong, Zheng, & Kahn, 2020;



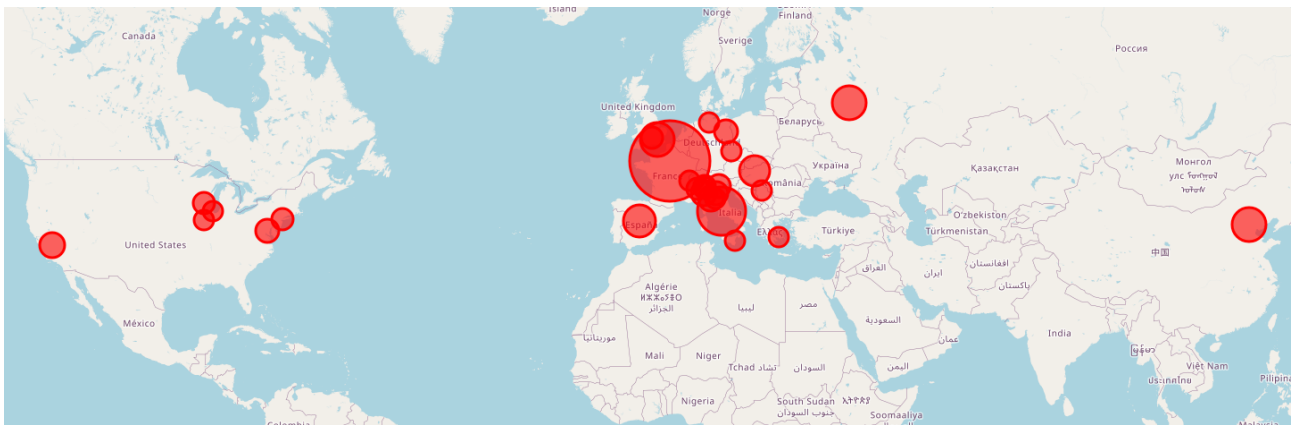
Albano & Pagliara, 2025).

L'analisi a livello dipartimentale ([Tabella A.4](#)) evidenzia una forte concentrazione delle collaborazioni in specifici ambiti scientifici, in particolare nelle scienze applicate, nella fisica e nelle discipline tecnologiche. Alcune strutture accademiche, come i dipartimenti e le scuole di area scientifico-tecnologica dell'Università Federico II, emergono come nodi centrali della rete, riflettendo dinamiche tipiche delle collaborazioni scientifiche ad alta intensità di capitale umano e



The map displays the distribution of the 100 largest cities in Europe. The cities are represented by red circles, where the size of the circle corresponds to the city's population. The map covers a wide area of Europe, from the British Isles in the west to the borders of Russia and Turkey in the east, and from Scandinavia in the north to North Africa in the south. Major urban centers are clearly visible, including London, Paris, Moscow, Berlin, and Rome. The map also shows numerous smaller cities and towns across the continent, providing a comprehensive overview of European urban distribution.

grado di integrazione dei due sistemi universitari in reti globali della conoscenza.



2.3 Conclusioni della Sezione 2 – Produzione scientifica e integrazione cognitiva

L'analisi bibliometrica presentata in questa sezione fornisce una ricostruzione sistematica della struttura e dell'evoluzione delle collaborazioni scientifiche tra i sistemi universitari di Bari e Napoli nel periodo precedente all'entrata in esercizio della linea ferroviaria ad alta velocità. In quanto indagine ex ante, l'obiettivo non è valutare l'impatto causale dell'infrastruttura, ma identificare il grado di integrazione cognitiva già esistente e le condizioni strutturali entro cui eventuali effetti futuri potrebbero manifestarsi.

I risultati mostrano che entrambi i sistemi universitari sono caratterizzati da una crescita sostenuta della produzione scientifica nel lungo periodo e da un progressivo rafforzamento delle collaborazioni interurbane. In particolare, l'aumento della quota di co-pubblicazioni tra Bari e Napoli e il loro inserimento in reti scientifiche ampie e internazionalizzate indicano l'esistenza di legami conoscitivi già consolidati, sviluppatisi indipendentemente dalla nuova infrastruttura di trasporto. Tali evidenze sono coerenti con la letteratura che interpreta le collaborazioni scientifiche come processi cumulativi, fortemente dipendenti da traiettorie storiche e dalla presenza di nodi accademici centrali (Hoekman, Frenken, & Van Oort, 2009; Balland et al., 2020).

Alla luce di questi risultati, l'alta velocità ferroviaria Napoli–Bari può essere interpretata, in prospettiva, come un potenziale fattore abilitante piuttosto che come un determinante autonomo della collaborazione scientifica. Studi empirici condotti in altri contesti mostrano infatti che l'HSR tende a rafforzare la frequenza e l'intensità delle interazioni tra università già collegate da relazioni di ricerca, riducendo i costi di coordinamento e facilitando incontri ripetuti, piuttosto che generare ex novo reti scientifiche tra attori precedentemente disconnessi (Gutiérrez, 2001; Dong, Zheng, & Kahn, 2020; Wang et al., 2020; Li et al., 2024). In linea con la letteratura sulle reti scientifiche globali, ci si aspetta che i benefici di un miglioramento dell'accessibilità fisica tendano a manifestarsi soprattutto come rafforzamento e intensificazione di relazioni preesistenti, piuttosto che come creazione ex novo di legami scientifici (Wagner, 2008; Adams, 2013; Frenken, Hardeman, & Hoekman, 2010).

Al tempo stesso, la letteratura segnala che tali effetti possono risultare asimmetrici e dipendenti dalla struttura dei sistemi accademici coinvolti. In assenza di politiche di accompagnamento, il miglioramento dell'accessibilità può tradursi in un rafforzamento selettivo dei poli scientifici più grandi e meglio posizionati, con il rischio di accentuare dinamiche di polarizzazione (“siphon effect”) nella produzione e nella visibilità scientifica (Chen & Haynes, 2015; Albano & Pagliara, 2025). In questo senso, l'evidenza ex ante presentata in questa sezione suggerisce che l'impatto dell'HSR sulla collaborazione scientifica dipenderà in larga misura dalla capacità delle istituzioni locali di valorizzare reti già esistenti e di inserirle in strategie coordinate di sviluppo della ricerca.

Nel complesso, l'analisi contribuisce a definire una baseline empirica della “capacità di integrazione scientifica” tra Bari e Napoli, utile per interpretare in modo più rigoroso eventuali cambiamenti osservabili nel periodo post-apertura della linea. Questa baseline costituisce inoltre un riferimento fondamentale per confrontare gli effetti dell'HSR sulle altre dimensioni dell'integrazione territoriale analizzate nel paper, in particolare la mobilità studentesca e le reti professionali.

3 Mobilità studentesca tra regioni

La mobilità studentesca rappresenta una componente fondamentale dei processi di integrazione territoriale e di formazione del capitale umano, in quanto riflette le scelte individuali di investimento in istruzione e la capacità dei sistemi universitari di attrarre, trattenere e redistribuire talenti nello spazio. In letteratura, i flussi di immatricolazione sono ampiamente utilizzati come *proxy* della circolazione del capitale umano in fase formativa e come indicatori anticipatori di dinamiche migratorie più strutturali (Faggian, McCann, & Sheppard, 2007; Venhorst, Van Dijk, & Van Wissen, 2011).

Le scelte di mobilità degli studenti risultano influenzate da un insieme eterogeneo di fattori, tra cui la qualità percepita degli atenei e

l'accessibilità spaziale (Sá, Florax, & Rietveld, 2004). In particolare, il "costo della distanza" rimane una barriera significativa: l'evidenza empirica mostra come la probabilità di iscrizione decresca con l'aumentare dei tempi di viaggio, rendendo le infrastrutture di trasporto variabili cruciali per l'equità nell'accesso all'istruzione (Frenette, 2006; Dotti et al., 2013). In questo contesto, l'Alta Velocità Ferroviaria (AV/HSR) agisce riducendo i costi generalizzati della mobilità, rendendo sostenibili forme di pendolarismo di lungo raggio e modificando i bacini di utenza degli atenei (Garmendia, Ureña, & Coronado, 2011; Delaplace et al., 2014).

Diversi contributi evidenziano come l'HSR possa innescare meccanismi di competizione spaziale tra università, favorendo potenzialmente la concentrazione degli studenti verso i poli accademici più rinomati o meglio connessi (Ureña, Menerault, & Garmendia, 2009; Moyano, Coronado, & Garmendia, 2018). Tuttavia, l'impatto non è uniforme: mentre alcuni atenei di medie dimensioni possono beneficiare di una maggiore attrattività, esiste il rischio di un deflusso netto verso i nodi principali della rete, consolidando modelli di sviluppo polarizzati (Chen & Haynes, 2017; Pagliara et al., 2019).

Evidenza recente per l'Italia conferma questa ambivalenza: l'apertura di una fermata AV aumenta in media dell'8% i deflussi studenteschi dalle province trattate, senza generare afflussi compensativi, con effetti concentrati nel Mezzogiorno e nei movimenti di lunga distanza (Frattola et al., 2026). L'accessibilità può quindi facilitare dinamiche di "brain drain" dalle aree periferiche verso nodi già forti.

Nel caso italiano, la mobilità studentesca riflette profonde asimmetrie tra Nord e Sud, con flussi storicamente orientati verso le regioni settentrionali per ragioni occupazionali e di qualità percepita (Di Pietro, 2012; Bratti & Verzillo, 2019). Tuttavia, lungo l'asse Napoli-Bari, la mobilità assume una dimensione interregionale complessa, mediata dal ruolo della Basilicata. Questa regione, posizionata strategicamente lungo il corridoio ma caratterizzata da un sistema universitario di dimensioni ridotte, rappresenta un bacino fondamentale per interpretare la capacità di attrazione reciproca tra i due poli metropolitani e l'efficacia delle reti di trasporto nel trattenere capitale umano all'interno del Mezzogiorno (Carteni et al., 2017).

Alla luce di questi elementi, la sezione analizza i flussi di immatricolazione tra Bari e Napoli, estendendo l'osservazione alla Basilicata per mappare la geografia degli scambi formativi nell'area del Sud-Est. Utilizzando i dati amministrativi del Ministero dell'Università e della Ricerca (MUR), l'obiettivo è ricostruire una *baseline* empirica dell'integrazione formativa pre-HSR.

3.1 Dati e Metodologia

L'analisi utilizza i microdati sugli immatricolati per regione di residenza e sede del corso, forniti dal Ministero dell'Università e della Ricerca. Il dataset contiene, per ciascun anno accademico, informazioni sul numero di immatricolati (Imm), la regione e provincia di residenza dello studente e l'ateneo di iscrizione.

L'analisi è limitata agli studenti residenti in Puglia, Campania e Basilicata, selezionando esclusivamente le osservazioni in cui la variabile RegioneRES appartiene a questo insieme.

Gli atenei sono stati ricondotti a una regione di localizzazione tramite una mappatura manuale (AteneoNOME → Regione_ATENEO), che consente di aggregare coerentemente le sedi universitarie a livello regionale.

3.2 Risultati

La [Figura 3.1](#) rappresenta i flussi di immatricolazione in valore assoluto dagli studenti residenti in Puglia, Campania e Basilicata verso tutte le regioni italiane nell'anno accademico 2024/2025. I corrispondenti valori numerici sono riportati in Appendice, [Tabella A.5](#).

Per la Puglia, su un totale di 21.528 immatricolati, 15.861 studenti (73,7%) si iscrivono ad atenei localizzati nella stessa regione (self-flow). I flussi in uscita più rilevanti sono diretti verso Emilia-Romagna (912 studenti, 4,2%), Lombardia (838, 3,9%) e Lazio (706, 3,3%). I flussi verso la Campania ammontano a 192 studenti (0,9%), mentre quelli verso la Basilicata sono pari a 117 studenti (0,5%). Nel complesso, la mobilità extra-regionale pugliese si concentra prevalentemente verso regioni del Centro-Nord.

Per la Campania, il self-flow risulta particolarmente elevato: 28.714 studenti su 32.924 rimangono nella regione, pari a 87,2% del totale. I principali flussi in uscita sono diretti verso il Lazio (1.210 studenti, 3,7%) e la Lombardia (566, 1,7%), seguiti da Abruzzo (361, 1,1%) ed Emilia-Romagna (262, 0,8%). I flussi verso la Puglia risultano quantitativamente contenuti (76 studenti, 0,2%), mentre quelli verso la Basilicata sono pari a 55 studenti (0,17%).

La Basilicata presenta un profilo nettamente distinto. A fronte di 2.719 immatricolati complessivi, solo 656 studenti (24,1%) rimangono nella regione, evidenziando una bassa capacità di ritenzione. Il principale flusso in uscita è diretto verso la Puglia (611 studenti, 22,5%), seguito dalla Campania (293, 10,8%). Tra le regioni del Centro-Nord, i flussi più consistenti sono verso Emilia-Romagna (241, 8,9%), Lazio (222, 8,2%) e Lombardia (175, 6,4%).

Nel complesso, i dati mostrano una forte eterogeneità nei livelli di self-retention tra le tre regioni: elevata per Campania e Puglia, molto più contenuta per la Basilicata. I flussi interregionali tra le tre regioni considerate risultano quantitativamente limitati in valore assoluto, ad eccezione del collegamento Basilicata → Puglia, mentre una quota rilevante della mobilità in uscita, soprattutto per Puglia e Basilicata, è assorbita da regioni del Centro-Nord.

I risultati sono anche riassunti nella matrice in [Figura 3.2](#).

Flussi di immatricolazione – Puglia, Campania e Basilicata (2024/2025)
Flussi assoluti verso tutte le regioni italiane

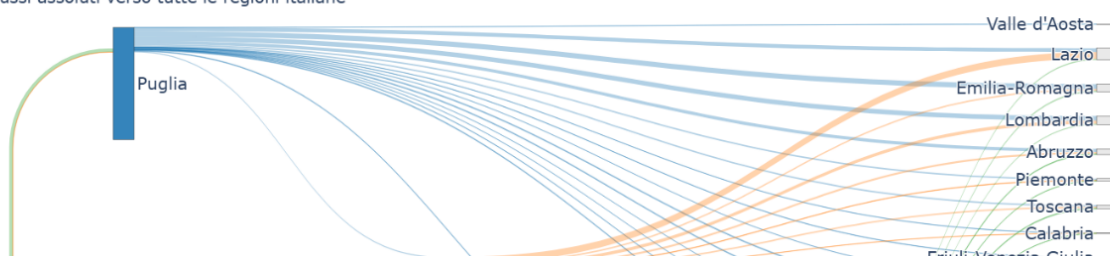


Figura 3.1 Elaborazione dell'autore su dati MIUR

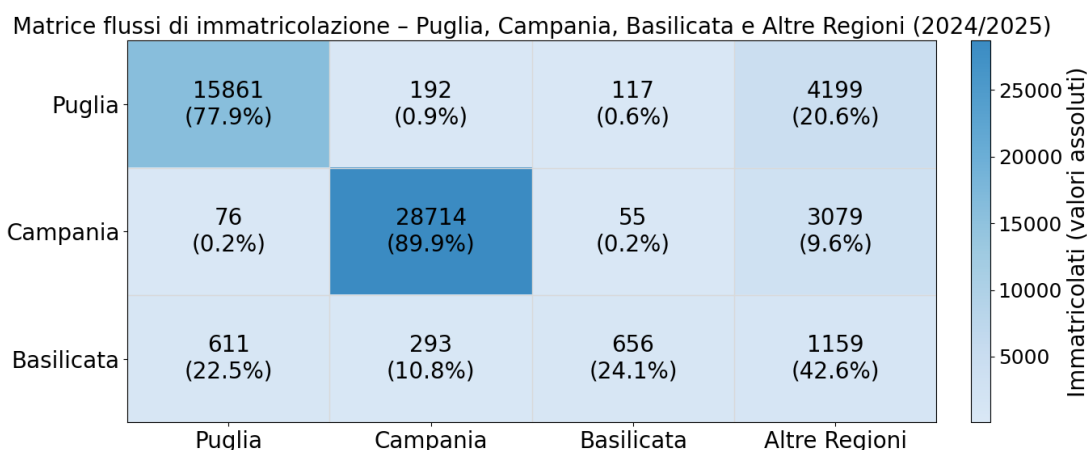


Figura 3.2 Elaborazione dell'autore sui dati MIUR

Osservano il fenomeno nel tempo, La [Tabella A.6](#) riporta la serie storica dei principali flussi interregionali di immatricolazione tra Puglia, Campania e Basilicata nel periodo 2009/2010–2024/2025.

Il grafico in [Figura 3.3](#) mostra questi dati in valore assoluto, mentre la [Figura 3.4](#) in valori percentuali, con una scala separata per la Basilicata che valori estremamente superiori a quelli di Puglia e Campania.

I flussi Puglia → Campania mostrano un andamento complessivamente crescente nel periodo considerato, passando da 79 immatricolati nel 2009/2010 a 192 nel 2024/2025, con un massimo locale di 277 studenti nel 2020/2021. I flussi Puglia → Basilicata risultano più contenuti e relativamente stabili, oscillando tra 72 e 141 studenti, senza una chiara tendenza di lungo periodo.

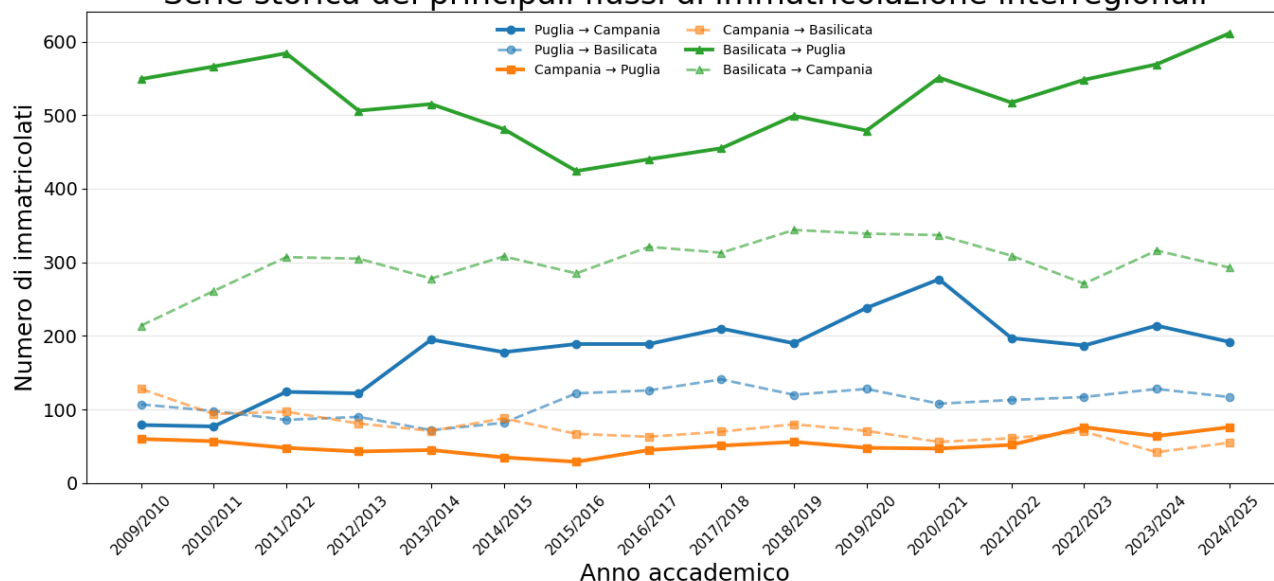
I flussi Campania → Puglia sono sistematicamente inferiori rispetto alla direzione opposta, con valori compresi tra 29 e 76 immatricolati nel periodo osservato. Analogamente, i flussi Campania → Basilicata si collocano su livelli contenuti, mostrando una lieve riduzione negli ultimi anni, da 128 studenti nel 2009/2010 a 55 nel 2024/2025.

La Basilicata presenta flussi di dimensione nettamente superiore verso entrambe le regioni considerate. In particolare, i flussi Basilicata → Puglia risultano stabilmente elevati lungo tutta la serie storica, variando tra 424 e 611 studenti, con un valore massimo nell'ultimo anno disponibile (611 nel 2024/2025). I flussi Basilicata → Campania mostrano livelli anch'essi consistenti, compresi tra 214 e 344 immatricolati, con oscillazioni contenute nel tempo.

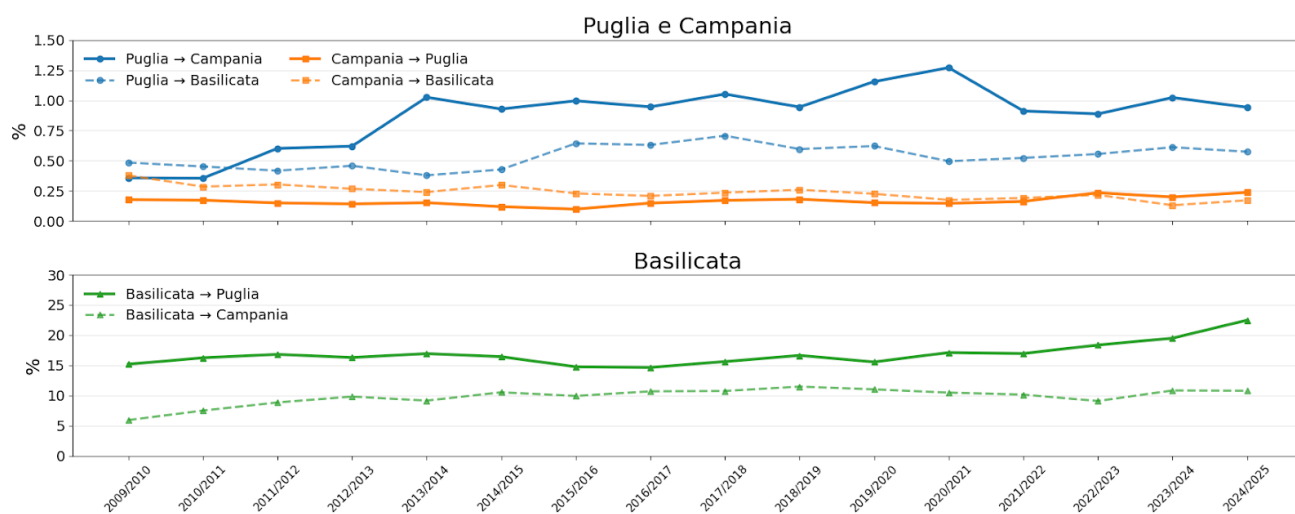
Nel complesso, la tabella evidenzia una persistente asimmetria nei flussi bilaterali, con la Basilicata caratterizzata da una mobilità in uscita di ordine di grandezza superiore rispetto a Puglia e Campania,

mentre i flussi tra Puglia e Campania rimangono quantitativamente limitati e relativamente stabili nel tempo.

Serie storica dei principali flussi di immatricolazione interregionali



Serie storica dei flussi interregionali di immatricolazione (percentuale sul totale della regione di origine)



Portando l'analisi a livello provinciale, emergono rilevanti eterogeneità territoriali all'interno delle tre regioni considerate. I risultati sono sintetizzati nella Tabella A.7, mentre le Figure successive mostrano la distribuzione percentuale delle destinazioni universitarie per provincia di residenza nell'anno accademico 2024/2025. I grafici di Puglia e Campania sono stati tagliati sull'asse

Y al 60% per valorizzare i flussi tra le due regioni, altrimenti difficili da visualizzare.

Per la Puglia ([Figura 3.5](#)), la quota di studenti che rimane nella regione è generalmente elevata, ma presenta una variabilità non trascurabile tra le province. Bari e Barletta-Andria-Trani mostrano livelli di ritenzione superiori all'85% (rispettivamente 87,6% e 85,4%), mentre Taranto (67,6%), Brindisi (71,9%), Lecce (72,1%) e Foggia (72,9%) evidenziano una maggiore propensione alla mobilità extra-regionale. I flussi diretti verso la Campania risultano complessivamente contenuti, ma sono relativamente più elevati nelle province di Foggia (1,9%) e Taranto (1,4%), che rappresentano i principali punti di connessione pugliesi verso il sistema universitario campano. I flussi verso la Basilicata sono invece marginali in tutte le province pugliesi (valori inferiori o prossimi all'1%), mentre una quota rilevante della mobilità in uscita è assorbita da altre regioni del Paese, con percentuali che raggiungono il 29,8% a Taranto e superano il 25% in Foggia, Lecce e Brindisi.

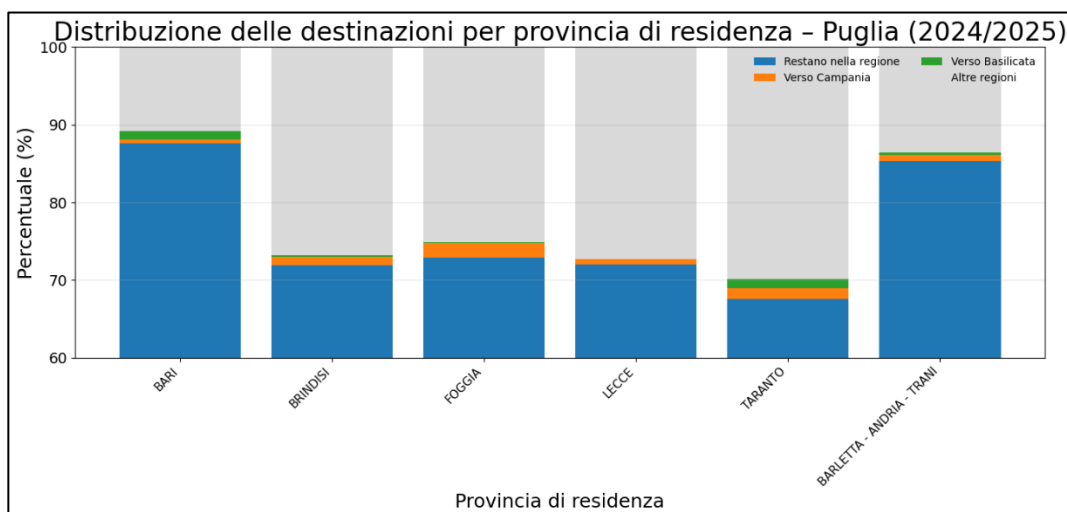


Figura 3.5 Elaborazione dell'autore su dati MIUR

La Campania ([Figura 3.6](#)) presenta un profilo provinciale caratterizzato da una capacità di ritenzione molto elevata e più omogenea. Napoli costituisce un caso particolarmente accentuato, con

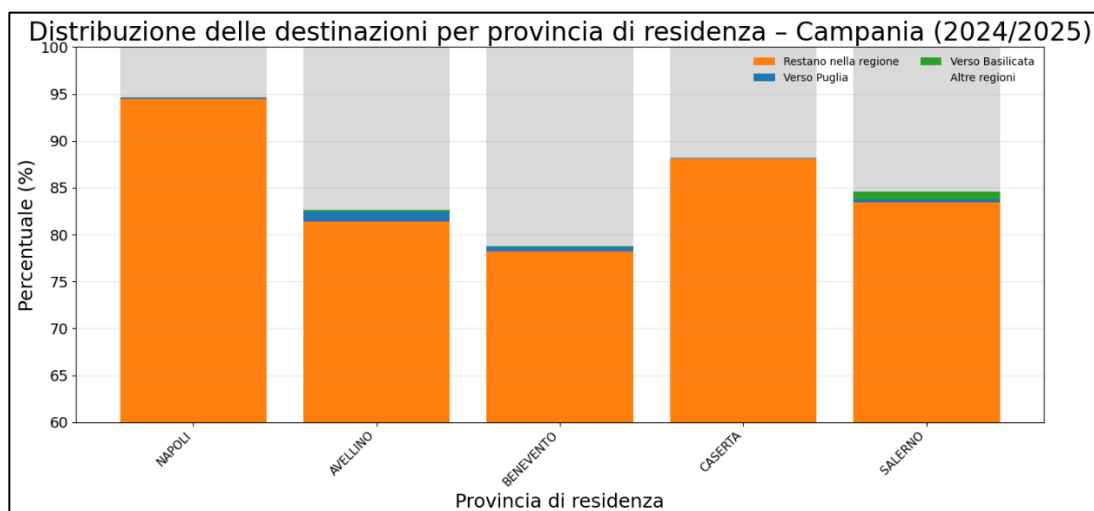


Figura 3.6 Elaborazione dell'Autore su Dati MIUR

il 94,5% degli immatricolati che rimane nella regione, seguita da Caserta (88,1%). Anche nelle province interne, come Avellino e Benevento, la quota di studenti che resta in Campania supera stabilmente il 78%. I flussi verso la Puglia e la Basilicata risultano trascurabili in tutte le province (generalmente inferiori all'1%), confermando una debole integrazione interregionale in uscita. La mobilità extra-regionale campana è principalmente rivolta verso altre regioni diverse da Puglia e Basilicata, con valori che oscillano tra il 5,3% di Napoli e il 21,2% di Benevento.

Il quadro cambia radicalmente per la Basilicata ([Figura 3.7](#)), che mostra una bassa capacità di ritenzione in entrambe le province. A Matera, solo l'11,3% degli studenti rimane nella regione, mentre a Potenza la quota sale al 30,7%, rimanendo comunque nettamente inferiore rispetto alle province pugliesi e campane. I flussi verso la Puglia sono particolarmente rilevanti, soprattutto per Matera, dove raggiungono il 42,7%, configurando un asse di mobilità molto intenso. Anche i flussi verso la Campania risultano significativi, in particolare per Potenza (14,8%), mentre sono più contenuti per Matera (2,8%). In entrambe le province lucane, una quota consistente degli studenti si dirige verso altre regioni del Paese, con percentuali superiori al 42%, evidenziando una forte dispersione geografica della mobilità in uscita.

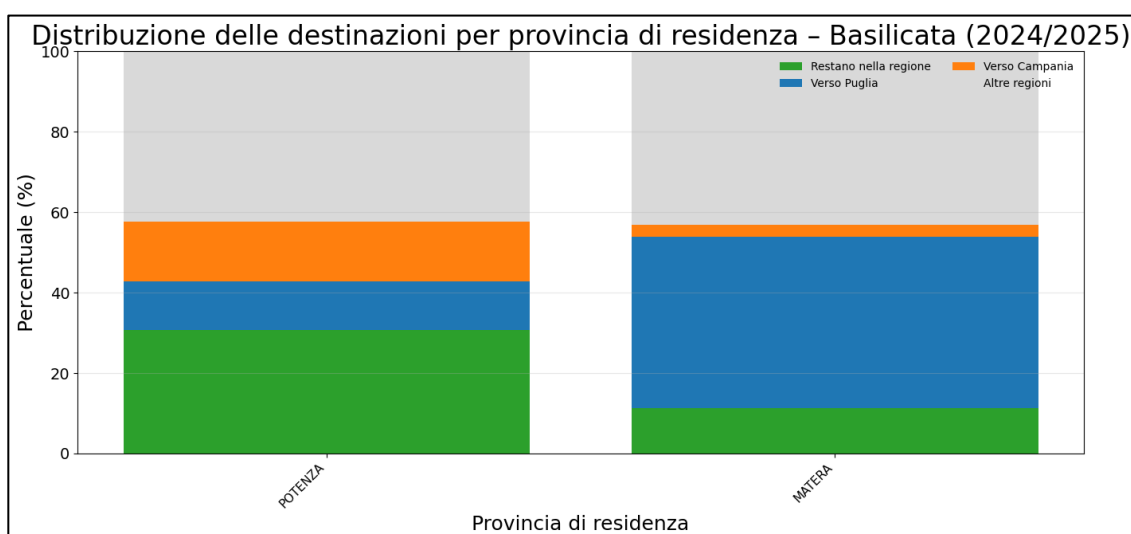


Figura 3.7 Elaborazione dell'Autore su Dati MIUR

Nel complesso, l'analisi provinciale conferma e rafforza i risultati osservati a livello regionale, mettendo in luce una marcata asimmetria nei flussi tra le tre regioni considerate. In particolare, la Basilicata si distingue per un'elevata mobilità in uscita concentrata verso la Puglia e, in misura minore, verso la Campania, mentre i flussi tra Puglia e Campania restano quantitativamente limitati e circoscritti ad alcune province specifiche. Allo stesso tempo, la mobilità verso il resto del

Paese assume un ruolo centrale soprattutto nelle province pugliesi e lucane con minore capacità di ritenzione.

3.3 Conclusioni dell'analisi di mobilità tra studenti

Questo capitolo ha analizzato la mobilità studentesca tra Puglia, Campania e Basilicata su scala regionale, temporale e provinciale, documentando un sistema fortemente asimmetrico. I risultati mostrano un'elevata capacità di ritenzione per Puglia e Campania, a fronte di una mobilità in uscita strutturalmente elevata per la Basilicata, che emerge come principale bacino di origine dei flussi interregionali lungo il corridoio Napoli–Bari. I flussi bilaterali tra Puglia e Campania risultano invece quantitativamente limitati e stabili nel tempo.

L'analisi provinciale evidenzia una marcata eterogeneità territoriale, con flussi concentrati in specifiche aree di confine. In Puglia, Foggia e Taranto rappresentano i principali punti di connessione verso il sistema universitario campano, mentre in Basilicata il flusso verso la Puglia assume dimensioni dominanti, soprattutto nella provincia di Matera. Questi risultati indicano che la mobilità formativa nell'area è già oggi mediata da una struttura territoriale complessa, nella quale le province interne svolgono un ruolo cruciale.

L'entrata in esercizio della linea AV/AC Napoli–Bari è pertanto destinata a intervenire su relazioni formative preesistenti, riducendo i costi generalizzati di accesso agli atenei lungo il corridoio. In linea con la letteratura, è plausibile attendersi un rafforzamento dei flussi interregionali di breve e medio raggio e un ampliamento dei bacini di utenza universitari (Garmendia, Ureña, & Coronado, 2011; Moyano, Coronado, & Garmendia, 2018). Tuttavia, l'evidenza internazionale suggerisce che tali effetti possano risultare asimmetrici, accentuando processi di concentrazione verso i poli più attrattivi e rafforzando gerarchie territoriali esistenti (Ureña, Menerault, & Garmendia, 2009; Chen & Haynes, 2017), con potenziali effetti avversi per le regioni a bassa capacità di ritenzione come la Basilicata.

I risultati costituiscono quindi una baseline empirica pre-HSR, funzionale allo sviluppo di valutazioni causali future. Analisi basate su approcci quasi-sperimentali, quali Difference-in-Differences o Synthetic Control Method, potranno identificare l'impatto dell'HSR sui flussi di immatricolazione e la presenza di effetti eterogenei lungo il corridoio (Gibbons & Machin, 2005; Abadie, Diamond, & Hainmueller, 2010; Cascetta et al., 2020). Estensioni successive potranno integrare informazioni sugli esiti occupazionali e sulla mobilità post-laurea, valutando se l'aumento della mobilità formativa si traduca in una maggiore circolazione o in una perdita netta di capitale umano (Faggian, McCann, & Sheppard, 2007; Venhorst, Van Dijk, & Van Wissen, 2011).

Nel complesso, lo studio mostra come la mobilità studentesca rappresenti un indicatore sensibile dei potenziali effetti redistributivi delle grandi infrastrutture di trasporto. La linea Napoli–Bari offre un'opportunità empirica unica per valutare se l'HSR possa favorire una maggiore integrazione formativa nel Mezzogiorno o, al contrario, rafforzare dinamiche di polarizzazione già in atto.

4 Mobilità professionale: data scraping da LinkedIn con interpretazione con linguaggio LLM

Nel contesto italiano, l'esperienza decennale dell'alta velocità ha evidenziato effetti rilevanti su crescita, mobilità e integrazione territoriale (Cascetta et al., 2020). La disponibilità di “digital traces” generate dalle piattaforme online ha ampliato in modo sostanziale la frontiera empirica per lo studio dei mercati del lavoro, della mobilità e delle reti professionali. In particolare, i dati ricavati da profili professionali online consentono di osservare traiettorie occupazionali, transizioni tra imprese e connessioni tra territori con una granularità preclusa dalle statistiche tradizionali (Arribas-Bel, 2014). In questo quadro, LinkedIn è trattato come una fonte informativa utile per costruire proxy di capitale umano e “pipeline” di mobilità, coerentemente con la letteratura che interpreta i network sociali come determinanti dell'allocazione lavorativa (Granovetter, 1973; Stateva et al., 2021).

L'interesse per i dati di rete è ulteriormente motivato dall'evidenza che la struttura delle connessioni, inclusi i legami “deboli”, influenza la probabilità di transizione occupazionale e la diffusione dell'informazione, con implicazioni dirette per la geografia della mobilità (Gee et al., 2017; Macy, 2023). Sul piano territoriale, i profili LinkedIn costituiscono un osservatorio complementare ai dati amministrativi per descrivere configurazioni relazionali lungo corridoi infrastrutturali come l'asse Napoli–Bari.

L'analisi, pur riconoscendo i limiti di selezione e copertura tipici dei dati di piattaforma (World Bank, 2018), è ritenuta informativa per mappare le connessioni professionali potenzialmente rilevanti per il corridoio e costituire una baseline per successive verifiche. Dal punto di vista metodologico, l'impiego di modelli linguistici per la standardizzazione e classificazione di campi testuali non strutturati (ruoli, imprese, località) si inserisce nel dibattito sui “foundation models” come infrastruttura general-purpose per la codifica di dati economici (Bommasani et al., 2021; Korinek, 2023), garantendo scalabilità pur richiedendo una rigorosa validazione procedurale per mitigare rischi di opacità e bias sistematici.

4.1 Dati e Metodologia

Ad oggi non esiste un accesso strutturato, pubblico e riproducibile ai microdati dei profili LinkedIn che consenta di identificare in modo esaustivo insiemi di utenti sulla base di connessioni territoriali multiple. In assenza di un frame campionario ufficiale, la costruzione di un dataset di profili professionalmente connessi sia alla città di Bari sia alla città di Napoli richiede pertanto l'adozione di strategie indirette di identificazione e raccolta.

La metodologia adottata in questo lavoro si configura come una *search-based data collection pipeline* costruita in modo iterativo e orientata alla riproducibilità. Essa privilegia la tracciabilità

delle unità di analisi (profili individuali) e la separazione concettuale tra identificazione dei profili, arricchimento informativo e costruzione delle variabili analitiche. Il processo si articola in cinque fasi principali.

(a) Identificazione dei profili tramite Proxy URL.

In una prima fase, gli URL di profili LinkedIn pubblici ([linkedin.com/in/...](https://www.linkedin.com/in/...)) sono stati individuati tramite tre differenti famiglie di query interrogando il motore di ricerca *Proxy URL* Brave Search accessibile via API. Questo approccio consente di recuperare, a partire da query testuali strutturate, gli URL dei profili indicizzati pubblicamente, senza accedere direttamente ai contenuti completi delle pagine della piattaforma.

Le tre famiglie di query sono state progettate per intercettare profili potenzialmente connessi alle città di Bari e Napoli secondo logiche metodologicamente complementari:

- (i) query semantiche dirette basate su riferimenti territoriali espliciti (Bari/Napoli) e formulazioni di mobilità studio–lavoro;
- (ii) query generate sistematicamente tramite combinazioni di località, istituzioni accademiche e verbi di traiettoria, finalizzate ad ampliare la copertura;
- (iii) query string-based con operatori di ricerca e filtri semantici ex post sugli snippet, come analisi di robustezza.

Questa articolazione mira ad ampliare la copertura complessiva e a ridurre il rischio di selezione indotta da una singola formulazione di ricerca. Il dettaglio operativo delle query è riportato in Appendice ([Tabella A.8](#)).

Dal punto di vista tecnico, la ricerca via Proxy URL consente l'accesso esclusivamente a informazioni di anteprima indicizzate (ad es. URL del profilo, titolo sintetico, snippet testuale), senza visibilità diretta sul contenuto completo del profilo. Tale caratteristica comporta un duplice trade-off metodologico. Da un lato, limita la profondità informativa disponibile nella fase di identificazione iniziale, rendendo impossibile una selezione esaustiva o completamente rappresentativa dell'universo dei profili LinkedIn. Dall'altro lato, garantisce una procedura replicabile, tracciabile e conforme a un utilizzo indiretto di dati pubblicamente indicizzati, separando nettamente la fase di identificazione dei profili da quella di arricchimento informativo successivo.

Gli URL raccolti sono stati successivamente normalizzati e deduplicati, eliminando le ripetizioni generate dalla sovrapposizione tra query. Questo passaggio è cruciale per assicurare che ciascun profilo costituisca un'unica unità statistica e per evitare sovra-rappresentazioni meccaniche dovute alla strategia multi-query adottata.

(b) Arricchimento dei profili (*enrichment*).

Gli URL dei profili LinkedIn, precedentemente deduplicati, sono stati utilizzati come chiavi di accesso per il servizio di *profile enrichment Enrich Layer* accessibile via API, progettato per estrarre informazioni strutturate a partire da profili pubblicamente visibili. Il servizio restituisce, ove disponibili, campi standardizzati relativi al percorso professionale e formativo dell'utente, tra cui esperienze lavorative, istruzione, *headline* professionale, località associate allo studio e al lavoro, e ulteriori metadati testuali.

Dal punto di vista metodologico, l'arricchimento consente di trasformare un insieme di identificatori (URL) in un dataset non tabulare in formato Json, in cui le informazioni sono state rese comparabili tra profili eterogenei attraverso parsing logico. Questo passaggio è cruciale per separare la fase di identificazione dei profili dalla fase di osservazione delle loro caratteristiche, riducendo la dipendenza da parsing diretto di pagine HTML e aumentando la stabilità della pipeline rispetto a cambiamenti nell'interfaccia della piattaforma.

È importante sottolineare che l'arricchimento si basa esclusivamente su informazioni pubblicamente accessibili e indicizzabili, e non garantisce una copertura completa e uniforme per tutti i campi e per tutti i profili. La disponibilità e la qualità delle informazioni estratte dipendono infatti sia dalle impostazioni di visibilità dell'utente sia dalla struttura del profilo stesso, introducendo una potenziale eterogeneità informativa. In particolare, alcune sezioni possono risultare parziali o assenti, e le informazioni testuali non seguono uno standard semantico univoco.

Questi limiti sono affrontati in due modi. In primo luogo, le variabili direttamente osservabili vengono costruite in modo conservativo, evitando imputazioni arbitrarie. In secondo luogo, l'arricchimento costituisce la base informativa per una successiva fase di classificazione assistita da modelli linguistici, che consente di normalizzare e sintetizzare informazioni testuali eterogenee mantenendo una chiara distinzione tra dati osservati e dati inferiti.

(d) Costruzione delle variabili osservabili e inferite

La costruzione delle variabili analitiche avviene lungo una pipeline sequenziale che separa (i) trasformazioni deterministiche e pienamente replicabili dei campi strutturati, (ii) indicatori derivati da regole di ricodifica territoriale e temporale, (iii) attributi inferiti tramite classificazione semantica vincolata con modelli linguistici, impiegati esclusivamente quando l'informazione non è disponibile in forma strutturata o comparabile tra profili.

(d.1) Parsing strutturato e normalizzazione dei campi osservabili

A partire dal JSON restituito dal servizio di enrichment, le sezioni **education** ed **experience** vengono trasformate in liste di record strutturati (una per profilo), normalizzando i campi chiave (istituzione, titolo di studio, ruolo, località, anni di inizio/fine). Questo passaggio produce variabili osservabili direttamente riconducibili alla fonte (ad esempio `education_parsed`, `experience_parsed`) e consente di trattare in modo omogeneo profili eterogenei, preservando la granularità evento-specifica (più esperienze formative e lavorative per la stessa unità individuale).

Per rendere confrontabili istituzioni e località, vengono applicate procedure di pulizia e normalizzazione testuale (lowercasing, rimozione di caratteri non informativi, gestione di varianti linguistiche), riducendo l'eterogeneità superficiale delle stringhe senza introdurre inferenze sul contenuto.

(d.2) Costruzione del dataset evento-livello e ricodifica territoriale

Per l'analisi della mobilità territoriale è cruciale disporre di una rappresentazione “long” che renda osservabili, per ciascun profilo, gli eventi in ordine temporale. Si costruisce quindi un dataset evento-livello (`df_long`) in cui ogni riga corrisponde a un episodio di **formazione** o **lavoro**, con campi standardizzati (tipologia evento, istituzione, ruolo, anni, località).

La componente territoriale è ricodificata in categorie discrete con logiche conservative, distinguendo tra **core cities** (Bari, Napoli) e categorie residuali informative (Italia non-core, Estero, Unknown). In particolare:

1. **Eventi di studio:** la localizzazione è ottenuta tramite un dizionario di mapping tra istituzioni (normalizzate) e macro-aree (`STUDY_LOOKUP`), costruito e aggiornato iterativamente. I casi non riconosciuti rimangono “Unknown” o vengono trattati con procedure di recupero mirate solo quando giustificato da evidenza testuale.
2. **Eventi di lavoro:** la localizzazione è derivata dalla stringa di località associata all'esperienza, tramite regole basate su toponimi e pattern ricorrenti, con priorità alle aree Bari/Campania e fallback a Italia/Estero quando l'informazione è generica o non discriminante.

Questa ricodifica non mira a ricostruire una residenza “vera”, ma a costruire una variabile territoriale **comparabile e auditabile** tra profili, adatta a definire connessioni e transizioni lungo l'asse Bari–Napoli.

(d.3) Variabili di connessione territoriale e costruzione delle traiettorie

A partire dal dataset evento-livello si costruiscono indicatori di connessione territoriale e sequenze temporali. In primo luogo, per ciascun profilo si definiscono flag binari che sintetizzano la presenza di eventi in aree core lungo due dimensioni:

- `studied_in_bari`, `studied_in_napoli`, calcolati sulle sole esperienze formative;
- `worked_in_bari`, `worked_in_napoli`, calcolati sulle sole esperienze lavorative.

In secondo luogo, si ricostruiscono le traiettorie territoriali ordinando gli eventi per anno di inizio e introducendo una priorità (educazione prima del lavoro a parità di tempo). Le sequenze territoriali vengono poi (i) depurate dai segmenti non informativi (“Unknown”, intesi come gap informativi e non come località), (ii) compresse eliminando ripetizioni contigue della stessa categoria. Ne derivano tre rappresentazioni:

- `trajectory_total` (sequenza complessiva studio+lavoro),
- `trajectory_edu` (solo studio),
- `trajectory_work` (solo lavoro),

insieme alle corrispondenti versioni in lista (`*_list`) e a misure di intensità descrittiva (numero di eventi e numero di “step” dopo pulizia e compressione). Questa costruzione consente di distinguere profili con esposizione territoriale limitata da profili effettivamente “multi-nodo”, evitando che la molteplicità di esperienze nella stessa area generi artificialmente mobilità.

Infine, le traiettorie complessive sono riclassificate in tipologie sintetiche (`trajectory_type`), separando almeno: profili senza transizioni (stayer), transizioni one-way tra Bari e Napoli, ritorni (BA↔NA con stesso nodo iniziale e finale), combinazioni con Estero, e casi residuali.

(d.4) Variabile di origine: `origin_final` e razionale analitico

Una variabile chiave per l’analisi dei flussi direzionali è l’**origine** del profilo, intesa come il **primo ancoraggio territoriale informativo** osservabile nella biografia (studio o lavoro). Si definisce quindi `origin_final` come la prima categoria territoriale non-missing/non-unknown nella sequenza ordinata degli eventi, applicando una ricodifica “larga” che, quando possibile, mappa province e regioni alla corrispondente città core (ad esempio Provincia di Bari/Puglia → Bari, Provincia di Napoli/Campania → Napoli). Tutto ciò che è informazione nazionale non-core rimane Italia, mentre Estero viene mantenuto come categoria separata.

Questa scelta risponde a un'esigenza metodologica precisa: l'origine non è interpretata come luogo di nascita o residenza, ma come baseline territoriale osservabile da cui misurare la direzione delle prime transizioni rilevanti. In modo coerente, viene costruita anche la variabile `first_work_paper`, definita come la prima localizzazione lavorativa informativa (Bari, Napoli, Italia) lungo la storia occupazionale, utile per sintetizzare la rotta principale "origine → primo lavoro".

(d.5) Variabili di profilo rule-based: genere e proxy di età

Accanto alle variabili territoriali, il dataset profilo-livello include attributi demografici costruiti con criteri conservativi:

- **Genere:** assegnato tramite un mapping name-based a partire dal nome proprio normalizzato (`first_name_clean`), con categoria mancante quando l'informazione non è disponibile o non classificabile.
- **Età (proxy):** stimata a partire dal primo anno osservabile di istruzione (`first_edu_year`), con fallback al primo anno di esperienza lavorativa (`first_job_year`) quando l'istruzione non è temporizzata. La procedura produce una stima (`age_proxy`, o versione estesa quando integrata) e un indicatore della fonte ("education" vs "job"), così da rendere trasparente l'eterogeneità informativa.

L'obiettivo di queste variabili non è la misurazione demografica puntuale, ma l'introduzione di controlli descrittivi sulla composizione del campione e sulla selettività dei flussi.

(d.6) Variabili inferite tramite LLM: classificazione semantica vincolata

Alcune dimensioni, pur centrali per l'interpretazione economica della mobilità, non sono disponibili in forma strutturata e omogenea nei profili (o risultano altamente eterogenee nelle stringhe). In questi casi si utilizza un modello linguistico (OpenAI, `gpt-4.1-mini`) come **classificatore semantico vincolato**, con tre principi operativi:

1. **Set di etichette chiuso:** il modello può scegliere solo tra categorie predefinite.
2. **Temperatura zero** e output controllato (solo etichetta, nessun testo libero).
3. **Auditabilità:** salvataggio degli output grezzi (file `.jsonl`) e merge deterministico nel dataset finale.

In particolare, sono state inferite:

- `highest_education`: livello massimo di istruzione in classi discrete, costruito su input testuale/strutturato delle esperienze formative quando il parsing non consente un ordinamento univoco o quando i titoli sono non standard.

- `last_employment_type`: tipologia dell'ultimo impiego (ad esempio pubblico, privato, autonomo, non classificabile), basata sull'ultimo episodio lavorativo identificato (ruolo e istituzione).
- `work_area`: macro-area professionale corrente (ad esempio STEM/tech, economia-finanza, diritto-policy, accademia, sanità, cultura-media, pubblica amministrazione, imprenditoria, altro), derivata combinando segnali da ruolo, headline/occupation e istituzione.

La distinzione tra variabili osservate e variabili inferite viene mantenuta esplicitamente nel dataset e nelle analisi descrittive successive, in modo da permettere verifiche di robustezza e un'interpretazione “a strati” dei risultati: traiettorie e connessioni territoriali sono rule-based, mentre le dimensioni qualitative del capitale umano sono introdotte come classificazioni semantiche sotto vincoli.

4.2 Risultati

Come mostrato nella [Tabella 2](#), le tre famiglie di query basate su Proxy URL hanno prodotto complessivamente 2.972 URL, ridotti a 2.604 dopo normalizzazione e deduplicazione. Dopo il parsing è stato possibile individuare i profili che effettivamente contenessero passaggi sia su Bari che su Napoli, **il campione analitico risulta composto da 105 profili**, che costituiscono il nucleo empirico su cui sono calcolate tutte le statistiche e le figure riportate nelle pagine seguenti.

I risultati sono organizzati per livelli informativi, coerentemente con l'architettura del dataset. In primo luogo utilizziamo un dataset profilo-livello (`df_v5`, variabili in [Tabella A. 9](#)) che integra attributi demografici e di capitale umano, incluse variabili osservabili (ad esempio, prima esperienza lavorativa, anno di primo episodio formativo, proxy di età) e variabili inferite tramite classificazione semantica vincolata (ad esempio, livello di istruzione, area professionale, tipo di ultimo impiego). In secondo luogo, , utilizziamo un dataset evento-livello (`df_long`, variabili in [Tabella A. 10](#)), che contiene 1.328 episodi complessivi di studio e lavoro associati ai 105 profili e consente di descrivere la composizione e la distribuzione delle esperienze nel tempo e nello spazio. In terzo luogo, utilizziamo un dataset traiettoria-livello (`df_traj`, variabili in [Tabella A. 11](#)) che ricostruisce, per ciascun individuo, la sequenza territoriale complessiva e le componenti separate di studio e lavoro, sintetizzandole in indicatori di intensità della mobilità (numero di eventi, numero di “step” territoriali) e in una tipologia discreta di traiettoria.

Totale profili analizzati	2604
Profili con connessione a Bari	644
Profili con connessione a Napoli	1559
Profili con connessioni a Bari e Napoli	105
Profili con nessuna connessione con entrambe le città	685

La [Figura 4.1](#) descrive la composizione demografica del sottocampione di profili LinkedIn con origine formativa a Bari o a Napoli, utilizzato per le analisi successive. La distribuzione per origine evidenzia una composizione sbilanciata verso l'area napoletana. In particolare, **49 profili su 105 hanno origine Napoli (46,7%), 37 hanno origine Bari (35,2%) e 19 risultano classificati come Italia (18,1%)**. Questa asimmetria è coerente con due elementi: (i) la maggiore densità di ecosistemi formativi e lavorativi nell'area napoletana, che può aumentare la probabilità di osservare “agganci” territoriali multipli nei profili, e (ii) la natura search-based della raccolta, che tende a riflettere l'indicizzazione e la visibilità differenziale dei profili oltre che i fenomeni reali. Di conseguenza, la figura va letta come composizione del campione informativo e non come stima di prevalenza nella popolazione.

La distribuzione per età (proxy, costruita per l'intero campione) evidenzia profili concentrati soprattutto nelle classi **30–49**, con

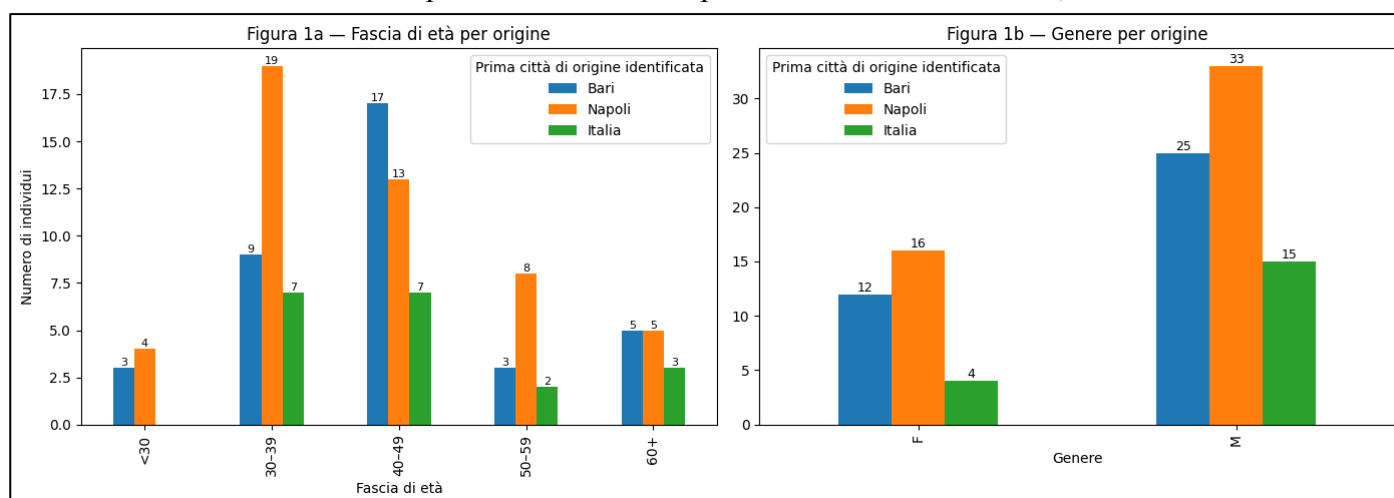


Figura 4.1 Elaborazione dell'autore sul campione

differenze nella composizione tra origini.

Per i profili con origine Bari, la distribuzione è spostata verso classi più mature: la fascia 40–49 è dominante (45,9%), seguita da 30–39 (24,3%). Le classi 50–59 e <30 incidono in modo simmetrico e contenuto (8,1% ciascuna), mentre 60+ rappresenta 13,5%. Nel complesso, l'origine Bari appare associata a traiettorie con maggiore anzianità anagrafica osservabile, coerente con percorsi professionali più lunghi o con mobilità interurbana che si manifesta in fasi relativamente avanzate della carriera.

Per i profili con origine Napoli, la distribuzione è invece più concentrata nella fascia 30–39 (38,8%), seguita da 40–49 (26,5%) e 50–59 (16,3%). Le fasce estreme rimangono minoritarie (<30: 8,2%, 60+: 10,2%). Rispetto a Bari, Napoli mostra quindi una maggiore incidenza di profili in età tipica di consolidamento iniziale della carriera, suggerendo una mobilità Bari–Napoli osservabile più frequentemente in una fase relativamente precoce del ciclo lavorativo.

La categoria Italia (che include anche i profili con passaggi internazionali accorpati) presenta una distribuzione più “bimodale” tra 30–39 e 40–49 (entrambe 36,8%), con una quota residuale nelle fasce più alte (60+: 15,8%) e assenza di profili <30. Dato il numero più ridotto di osservazioni in questa origine, la lettura va mantenuta prudente, ma la figura suggerisce che i profili non ancorati inizialmente a una delle due aree core si collocano prevalentemente nel segmento centrale dell’età lavorativa.

Nel complesso, la Figura 4.1 evidenzia che, a parità di selezione su profili con doppia connessione territoriale Bari–Napoli, l’origine si associa a una diversa struttura per età: più “matura” per Bari e più concentrata nei 30–39 per Napoli.

La [figura 4.2](#) sul titolo di studio evidenzia un campione fortemente selezionato verso profili ad alto capitale umano formale, il che è spiegabile con un principio di selezione avversa: i profili più “modesti” usano meno LinkedIn. In tutte le origini prevalgono Laurea, Master/Post-laurea e PhD/Abilitazione, mentre la quota di Diploma è residuale e le categorie non classificabili sono marginali.

La Figura 4.2 conferma che il campione LinkedIn intercetta soprattutto profili ad alto capitale umano, coerentemente con un bias di selezione della piattaforma: le carriere meno qualificate o meno formalizzate risultano meno presenti e meno “tracciabili”. In tutte le origini dominano Laurea, Master/Post-laurea e PhD/Abilitazione, mentre il Diploma è residuale. Tra i profili con origine Bari (n=37) la distribuzione è bilanciata tra Laurea e Master/Post-laurea (13 casi ciascuno), con una quota rilevante di PhD/Abilitazione (9); tra quelli con origine Napoli (n=49) il profilo è analogo ma più spostato verso credenziali avanzate, con 16 PhD/Abilitazione, oltre a 17 Laurea e 15 Master/Post-laurea; la categoria Italia (n=19) resta anch’essa “alta” (8 Master/Post-laurea e 5 PhD/Abilitazione), coerentemente con traiettorie più articolate e spesso extra-locali.

La componente settoriale rafforza la stessa lettura: la mobilità osservabile Bari–Napoli si colloca prevalentemente in ambiti knowledge-intensive, con una presenza robusta di Accademia/Ricerca, STEM/Engineering/Tech e Impresa/Imprenditoria. Per origine Bari i due cluster principali sono ricerca e STEM (9 casi ciascuno), mentre per origine

Napoli emerge una maggiore incidenza di funzioni economico-manageriali e imprenditoriali, con 9 profili in Economia/Finanza/Management e 8 in Impresa/Imprenditoria, oltre a 11 in ricerca e 8 in STEM; la categoria Italia, più piccola, concentra soprattutto impresa (5) e mantiene quote comparabili in ricerca e sanità (3 e 3). Nel complesso, la figura non descrive la struttura “vera” del mercato del lavoro, ma il profilo del sottocampione più visibile su LinkedIn, e va quindi letta come evidenza comparativa interna tra origini e come indicazione della selettività del capitale umano che alimenta le connessioni professionali lungo l’asse Bari–Napoli.

La [Figura 4.3](#) sintetizza due aspetti complementari delle traiettorie territoriali dei profili con origine Bari e Napoli (n=81 complessivi, 33 Bari e 48 Napoli): (i) quanto le carriere osservate si sviluppino esclusivamente tra i due poli oppure includano ulteriori nodi nazionali/internazionali, (ii) la presenza di un ritorno verso l’area di origine lungo la sequenza degli eventi osservabili.

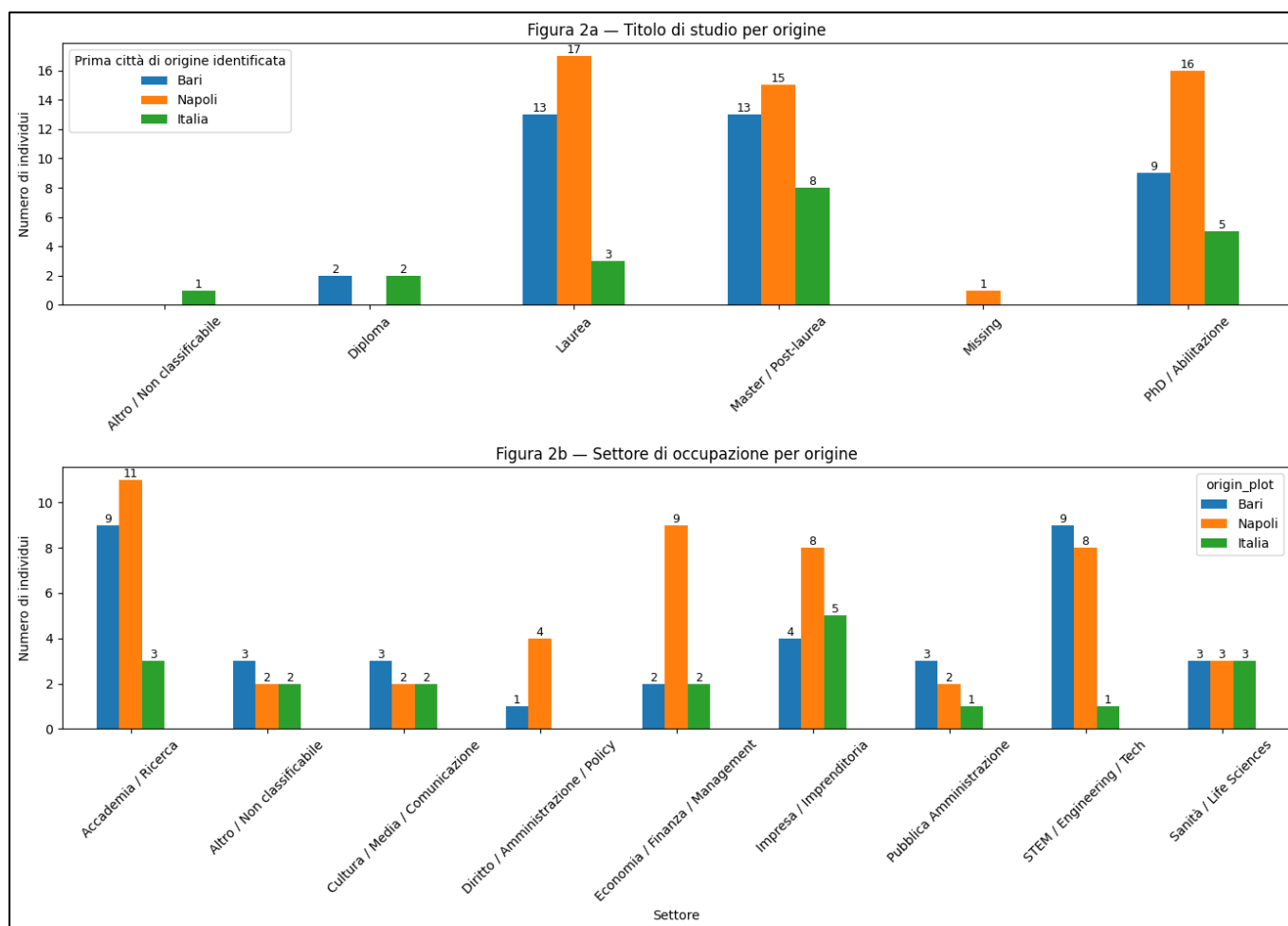


Figura 4.2 Elaborazione dell'autore su dati

Sul primo punto, emerge che le traiettorie “a due nodi” (solo Bari–Napoli) sono rare e praticamente assenti tra i profili con origine Napoli. In particolare, tra i profili con origine Bari una quota limitata (4 su 33, 12,1%) presenta una traiettoria circoscritta ai due poli, mentre la maggioranza include almeno un passaggio in altri territori italiani (18, 54,5%) o in Estero (11, 33,3%). Tra i profili con origine Napoli nessuna traiettoria risulta “solo Bari–Napoli”, e la mobilità osservata è quasi interamente inserita in percorsi più ampi: Italia aggiuntiva (27 su 48, 56,2%) e soprattutto Estero (21, 43,8%). In termini interpretativi, questo pattern suggerisce che l’asse Bari–Napoli, almeno nel sottocampione LinkedIn selezionato, funzioni più come segmento di traiettorie multi-nodo (nazionali e/o internazionali) che come canale bilaterale “chiuso”.

Sul secondo punto, la componente di ritorno risulta minoritaria in entrambi i gruppi, ma leggermente più frequente per origine Napoli: tornano 3 su 33 profili con origine Bari (9,1%) e 7 su 48 con origine Napoli (14,6%). Anche senza inferenze causali, la differenza è coerente con l’idea che una parte delle connessioni interurbane osservate sia compatibile con traiettorie di andata (verso un nodo percepito come opportunità) più che con percorsi circolari stabili. Va osservato che la bassa quota di ritorni può essere dovuta anche alla consistente presenza di persone ancora nel pieno della propria carriera.

Infine, come per le figure precedenti, queste evidenze vanno lette alla luce del bias di piattaforma: LinkedIn tende a rendere più osservabili carriere lunghe, qualificate e internazionalizzate, e quindi può amplificare l’incidenza di traiettorie multi-nodo e con passaggi esteri.

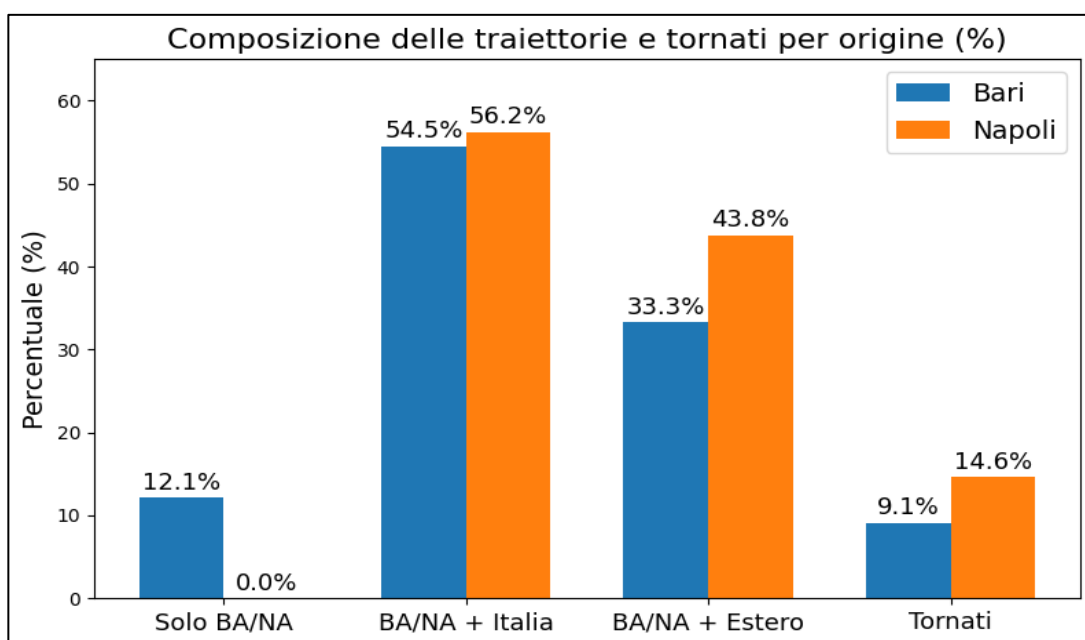


Figura 4.3 Elaborazione dell'autore sul campione

La [Figura 4.4](#) rappresenta, tramite chord diagram, i passaggi territoriali evento-per-evento ricostruiti dalle traiettorie individuali. Ogni arco misura il numero di transizioni tra un'area e la successiva nella sequenza temporale (dopo la compressione delle ripetizioni consecutive, così da contare solo i cambi di area effettivi). È importante notare che l'unità di conteggio non è l'individuo ma la transizione, quindi il totale dei flussi può superare il numero di profili.

L'evidenza principale è che la mobilità osservata lungo l'asse Bari–Napoli è inserita in traiettorie più ampie, in cui “Italia” (intesa come resto del territorio nazionale non ricondotto ai due poli) svolge un ruolo di nodo di snodo. I flussi più intensi sono infatti quelli che coinvolgono Italia: Italia→Bari (68), Napoli→Italia (68), Bari→Italia (60) e Italia→Napoli (51). Accanto a questi, si osserva anche una componente bilaterale non trascurabile tra i due poli: Bari→Napoli (56) e Napoli→Bari (49), che segnala l'esistenza di connessioni dirette tra le due aree, ma non in forma “chiusa” o isolata dal resto del sistema.

La componente internazionale è presente e asimmetrica: i passaggi verso Estero sono consistenti sia da Bari (27) sia da Napoli (29), ma soprattutto dal resto d'Italia (35). Simmetricamente, i rientri da Estero avvengono più spesso verso Italia (36) e, in misura minore, verso Bari (29) e Napoli (19), suggerendo che l'esperienza internazionale, nel campione osservato, si reinserisca frequentemente in traiettorie multi-nodo.

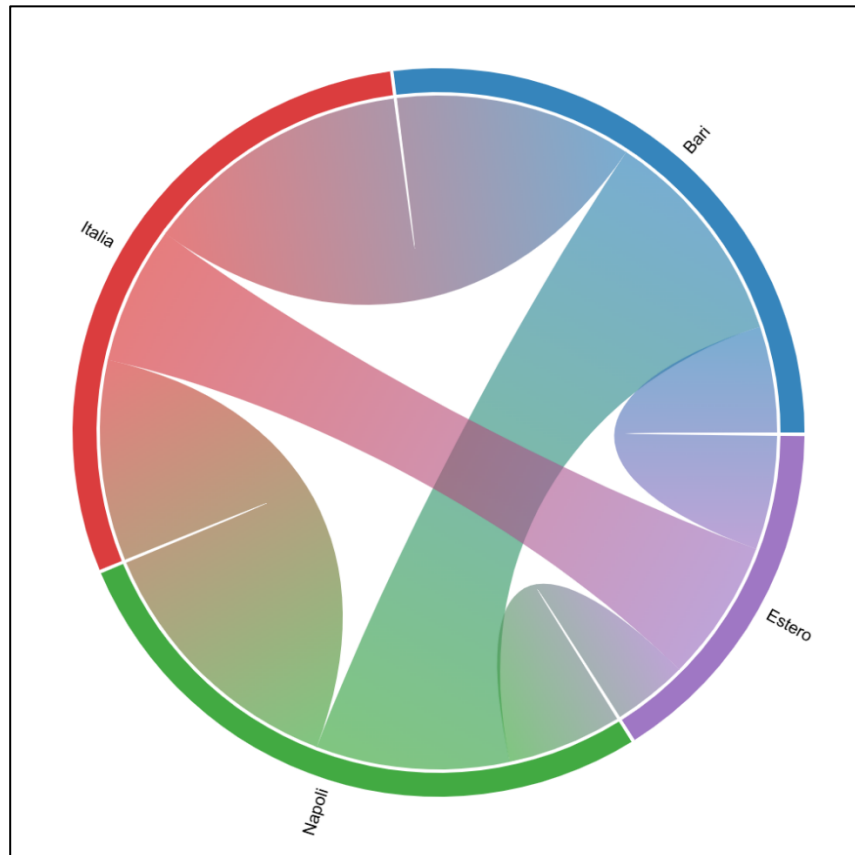


Figura 4.4 – Grafico Chord dei flussi - Elaborazione dell'autore sul campione

La matrice in [Figura 4.5](#), che normalizza per origine, conferma la stessa struttura in termini relativi. Da Bari, il passo successivo è quasi equamente ripartito tra Italia (42,0%) e Napoli (39,2%), con Estero al 18,9%. Da Napoli prevale Italia (46,6%), poi Bari (33,6%) ed Estero (19,9%). Da Italia i flussi si dirigono più spesso verso Bari (44,2%) che verso Napoli (33,1%), con una quota verso Estero pari al 22,7%. Da Estero il rientro è soprattutto verso Italia (42,9%), poi Bari (34,5%) e Napoli (22,6%). Nel complesso, l'asse Bari–Napoli emerge come segmento importante ma inserito in traiettorie multi-nodo, con Italia e, in misura minore, Estero come componenti strutturali. Queste evidenze vanno lette tenendo presente il bias di LinkedIn: la piattaforma sovrarappresenta carriere qualificate, mobili e ben documentate, quindi tende ad amplificare la visibilità di traiettorie articolate e internazionalizzate. Il valore delle due figure è soprattutto comparativo, perché rende trasparente l'architettura relativa dei passaggi e il ruolo di “hub” esercitato dal resto d'Italia.

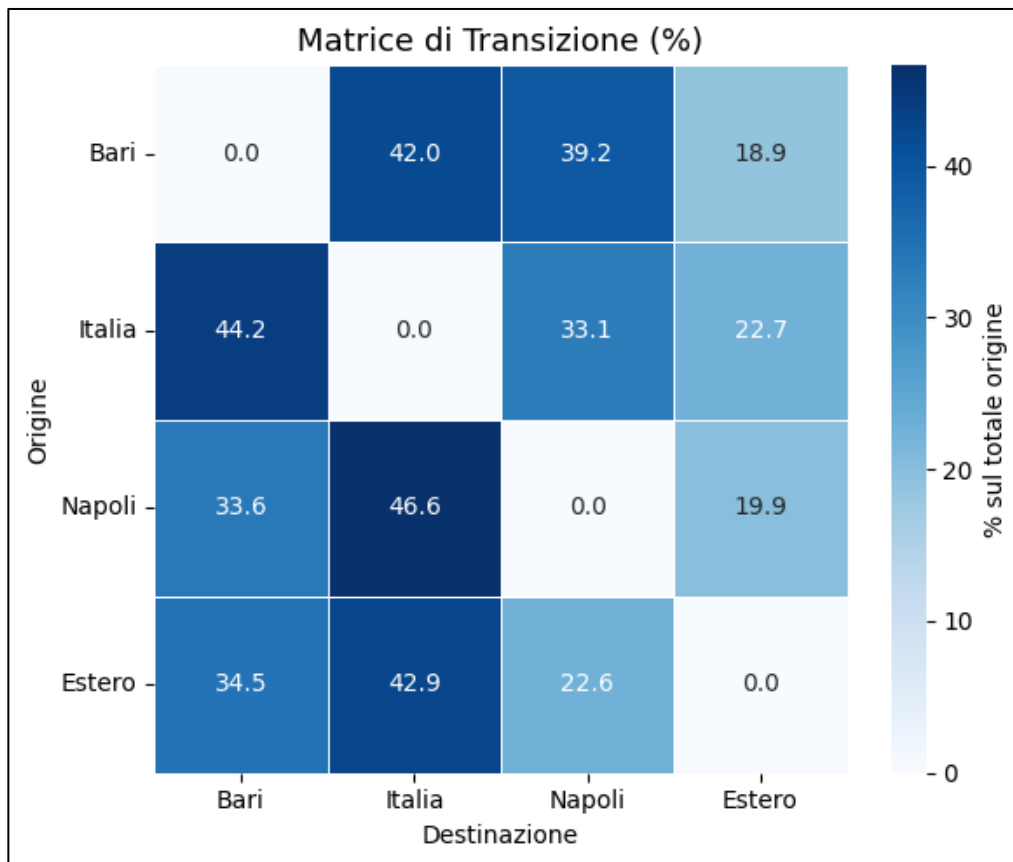
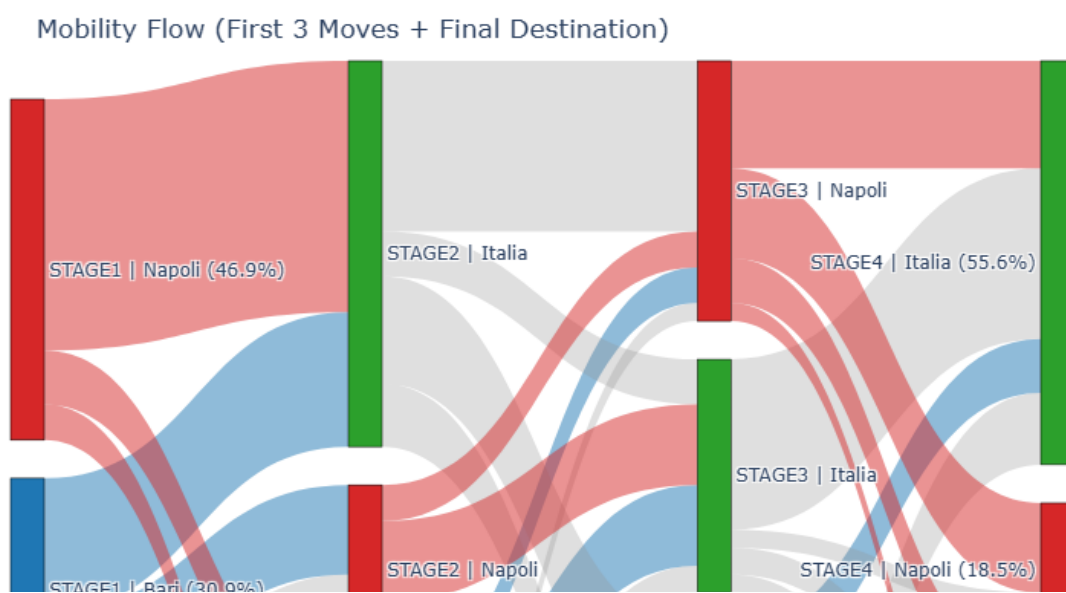


Figura 4.5 Elaborazione dell'autore sul campione

La [Figura 4.6](#) ricostruisce le traiettorie in forma dinamica su quattro stadi, riportando i primi tre passaggi territoriali osservati (dopo compressione delle ripetizioni consecutive) e, come quarto stadio, la destinazione finale effettiva (l'ultimo nodo della sequenza). La lettura conferma in modo molto netto quanto già suggerito dalla chord e dalla matrice: la mobilità Bari–Napoli, nel sottocampione informativo, si inserisce spesso in percorsi multi-nodo in cui “Italia” (resto del Paese) ha un ruolo centrale e, soprattutto, tende a diventare l'esito finale più frequente.

Già allo stadio iniziale, la sequenza parte più spesso da Napoli (38 casi, 44,2%) e Bari (25, 29,1%), ma una quota rilevante



entra nel percorso con un primo ancoraggio in Italia (23, 26,7%), segnale che per una parte dei profili la componente “extra-polo” è presente sin dalle prime tappe osservabili. Lo stadio 2 è dominato da Italia (43 casi), e allo stadio 3 Italia e Napoli risultano perfettamente bilanciate (29 e 29), mentre Bari rimane più contenuta (16) ed Estero cresce (12), indicando che l’articolazione della traiettoria si intensifica proprio nei passaggi intermedi. L’esito finale (stadio 4) è chiaramente concentrato su Italia con 49 casi, pari al 57,0%, mentre i finali su Napoli (15, 17,4%) e Bari (14, 16,3%) sono minoritari e Estero rappresenta il 9,3% (8 casi). In altre parole, il nodo “Italia” non è solo un passaggio, ma appare come una destinazione conclusiva ricorrente nella biografia osservabile.

Ma quando avviene il passaggio tra Bari e Napoli? La [Figura 4.7](#) scompone i passaggi Bari↔Napoli in base al “trigger” osservabile nel profilo, distinguendo tra università e momenti della carriera lavorativa. Per chi “parte” da Bari il collegamento con Napoli emerge soprattutto come mobilità per università (27 casi, 57,4%), mentre i passaggi riconducibili al lavoro sono più spesso “intermedi” (lavoro successivo: 16, 34,0%) e raramente coincidono con il primo o ultimo lavoro (2 e 2, 4,3% ciascuno). Per chi “parte” da Napoli la componente universitaria resta dominante (20, 45,5%), ma la mobilità appare più “work-oriented” rispetto ai baresi: oltre ai passaggi per lavoro successivo (10, 22,7%), cresce soprattutto la quota associata all’ultimo lavoro (9, 20,5%) e, in misura minore, al primo lavoro (5, 11,4%). In sintesi, nel campione LinkedIn la connessione tra i due poli è molto spesso mediata dall’università, ma i profili con origine Napoli mostrano una traccia più frequente di aggancio “forte” sul lato occupazionale (inizio/fine carriera).

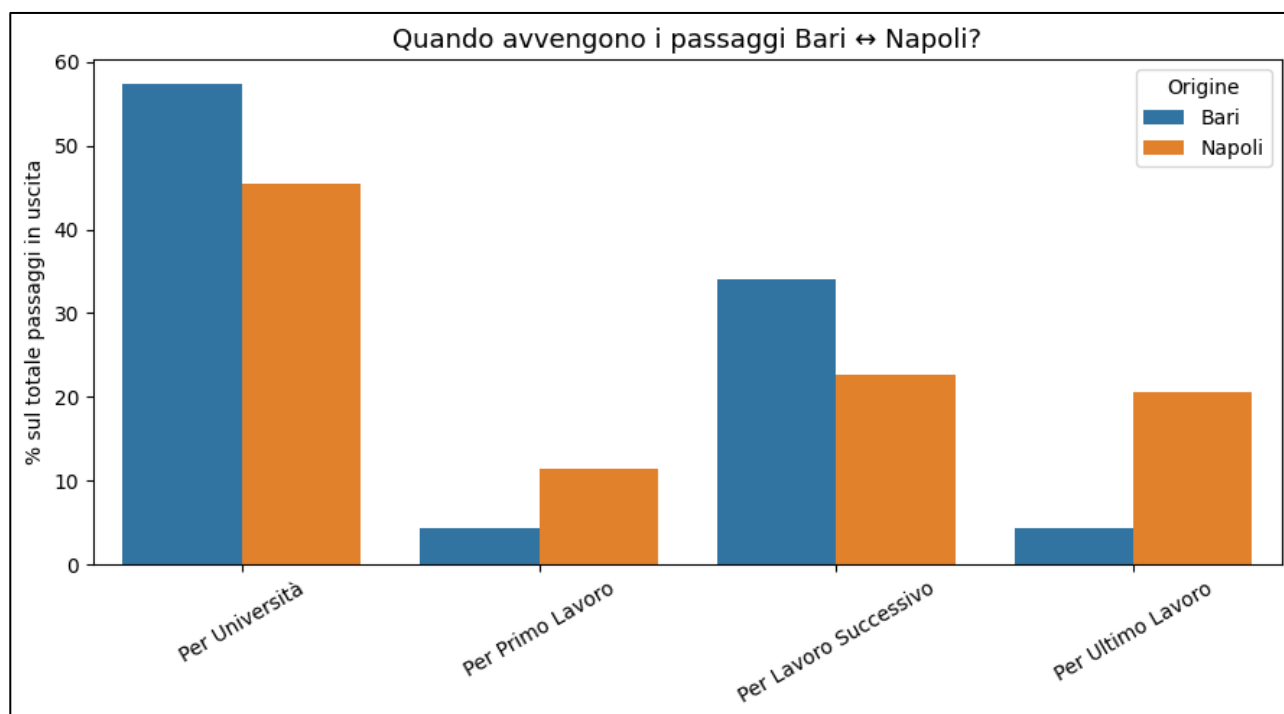


Figura 4.7 Elaborazione dell'autore sul Campione

Le [Figure 4.8](#) e [4.9](#) qualificano i passaggi lavorativi diretti tra i due poli, mostrando la composizione dei profili che passano a un'occupazione a Napoli partendo da Bari e, simmetricamente, a un'occupazione a Bari partendo da Napoli.

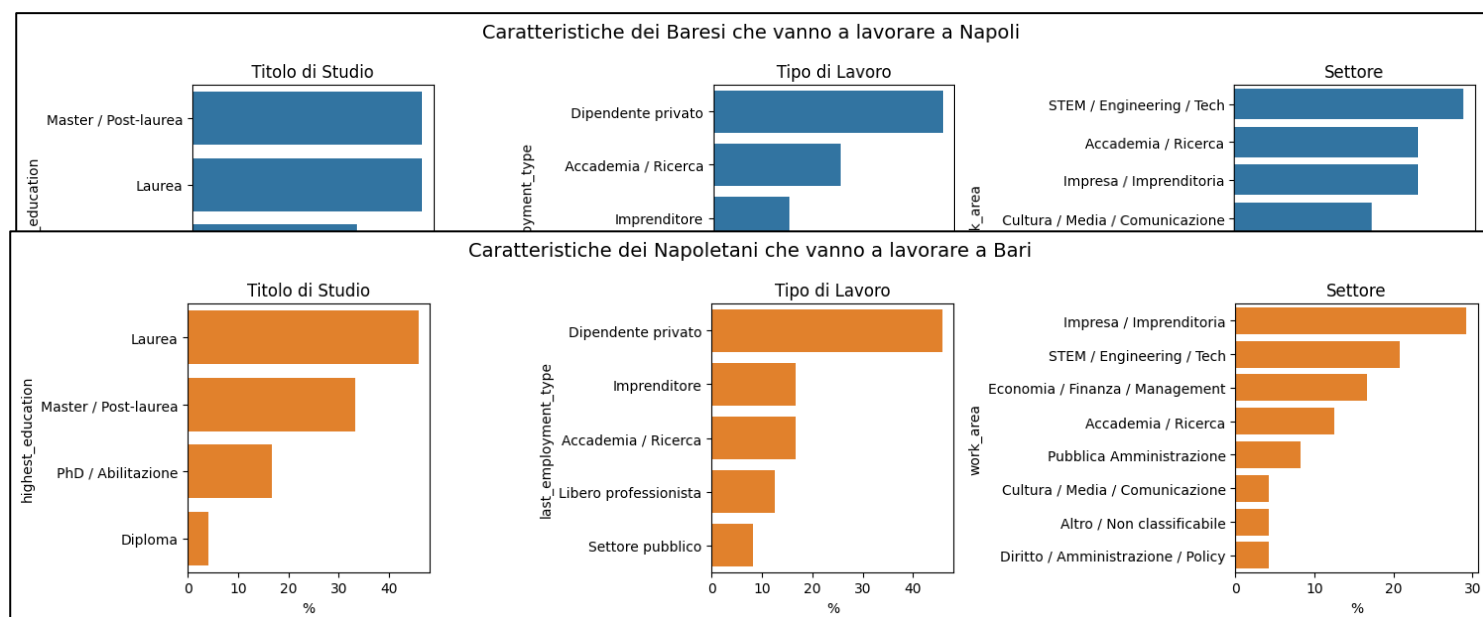


Figura 4.9 Elaborazione dell'autore sul campione

Nel flusso Bari→Napoli ([Figura 4.8](#)) il capitale umano è molto elevato: Laurea e Master/Post-laurea pesano entrambi 35%, mentre PhD/Abilitazione raggiunge 25% e il Diploma resta marginale (5%). Anche il tipo di impiego è nettamente orientato

al segmento qualificato e formale: prevale il dipendente privato (45%), ma una quota rilevante è in Accademia/Ricerca (25%), seguita da imprenditori (15%) e liberi professionisti (10%). La struttura settoriale è coerente: dominano STEM/Engineering/Tech (25%), Accademia/Ricerca (20%) e Impresa/Imprenditoria (20%), mentre Cultura/Media pesa 15% e Economia/Finanza/Management 10%; Sanità e Diritto/Policy sono residuali (5% ciascuno). In sintesi, il collegamento occupazionale Bari→Napoli che emerge dai profili LinkedIn è concentrato su traiettorie tecnico-scientifiche, accademiche e imprenditoriali, con una base di lavoro dipendente privato ampia ma affiancata da componenti “knowledge-intensive” non marginali.

Nel flusso opposto Napoli→Bari ([Figura 4.9](#)) l’istruzione resta alta ma con un profilo leggermente diverso: la Laurea è più frequente (45,8%) rispetto al Master/Post (33,3%), mentre PhD/Abilitazione scende al 16,7%; il Diploma è ancora residuale (4,2%). Anche qui la modalità prevalente è dipendente privato (45,8%), ma si osserva una maggiore articolazione verso funzioni “di sistema” e istituzionali: imprenditori e Accademia/Ricerca sono entrambi al 16,7%, i liberi professionisti al 12,5%, e compare una quota di settore pubblico (8,3%) che nel flusso Bari→Napoli non emerge. Sul piano settoriale, Bari attrae soprattutto profili legati a Impresa/Imprenditoria (29,2%), seguiti da STEM/Engineering/Tech (20,8%) e Economia/Finanza/Management (16,7%); Accademia/Ricerca è più contenuta (12,5%) e compaiono Pubblica Amministrazione (8,3%) e una coda di categorie minori (ognuna 4,2%). In sintesi, il flusso Napoli→Bari appare relativamente più orientato verso impresa e funzioni economico-manageriali, e include una componente pubblica, mentre il flusso Bari→Napoli mostra una quota più elevata di profili in ricerca/accademia e una struttura più centrata su STEM.

Queste differenze vanno lette come selettività interna del campione LinkedIn: non descrivono la composizione “vera” dei pendolari del lavoro tra le due aree, ma la composizione dei profili per cui il passaggio è osservabile e ben documentato. Proprio per questo, il confronto Bari→Napoli vs Napoli→Bari è informativo: segnala che, all’interno del segmento tracciabile, le due direzioni non sono simmetriche né per titoli né per settori.

4.3 Conclusioni dell’analisi di mobilità professionale

In questa sezione abbiamo costruito, con una pipeline search-based riproducibile (Proxy URL → enrichment →

parsing rule-based → classificazioni LLM vincolate), un dataset di 105 profili LinkedIn con connessioni effettive tra Bari e Napoli, organizzato su tre livelli (evento, profilo, traiettoria). I risultati descrittivi indicano che l'asse Bari–Napoli emerge raramente come canale “chiuso”: le traiettorie sono tipicamente multi-nodo e il resto d'Italia opera come hub, sia in valori assoluti (Italia→Bari 68, Napoli→Italia 68) sia come destinazione finale prevalente (57% in stage4). I profili osservabili sono fortemente selezionati verso alto capitale umano e settori knowledge-intensive, coerentemente con il bias di LinkedIn, che sovrarappresenta carriere qualificate, mobili e documentate. Il timing della connessione tra i due poli è spesso mediato dall'università (Bari 57,4%, Napoli 45,5%), ma per i profili con origine Napoli si osserva più spesso un aggancio occupazionale “forte” (ultimo lavoro 20,5% vs 4,3% per Bari). Nei passaggi lavorativi diretti, Bari→Napoli appare più tecnico-accademico (STEM 25%, Accademia 20%), mentre Napoli→Bari è relativamente più orientato a impresa e funzioni economico-manageriali e include una quota pubblica (PA 8,3%). Nel complesso, i digital traces consentono di mappare una “pipeline” di mobilità professionalmente rilevante lungo il corridoio, ma le evidenze vanno interpretate come struttura del sottocampione visibile e informativamente ricco, utile soprattutto per confronti interni e come baseline per integrazioni con fonti amministrative.

5 Conclusioni sulle relazioni tra Bari e Napoli

Le tre evidenze del paper convergono su un punto: tra Bari e Napoli esiste già una relazione “connessa”, ma non ancora un sistema integrato, perché i legami osservabili sono fortemente mediati da reti più ampie, con il resto d'Italia (e in parte l'estero) che opera da hub strutturale. Sul versante cognitivo, le co-pubblicazioni 1985–2024 indicano una base di collaborazione già consolidata e in accelerazione nell'ultimo decennio, coerente con dinamiche cumulative, l'AV/HSR può quindi agire soprattutto da moltiplicatore di relazioni preesistenti, non da generatore ex novo. Sul versante formativo, i flussi amministrativi mostrano che Puglia e Campania sono altamente “auto-ritentive” e che l'integrazione bilaterale è quantitativamente piccola e stabile nel tempo, mentre la Basilicata è il vero snodo del corridoio, con uscita strutturalmente alta e direzione dominante verso la Puglia. Sul versante professionale, i digital traces (105 profili con doppia connessione) confermano che l'asse Bari–Napoli raramente è un canale chiuso, e che la connessione è spesso universitaria, con differenze direzionali non simmetriche nei passaggi lavorativi (Bari→Napoli più STEM/accademia, Napoli→Bari più

impresa/manageriale e con una quota pubblica). Nel complesso, la Napoli–Bari si inserisce in un contesto in cui i meccanismi di integrazione esistono, ma sono selettivi e potenzialmente polarizzanti: senza politiche di accompagnamento su ultimo miglio, attrattività universitaria e capacità di trattenere capitale umano, il rischio è che la riduzione dei tempi rafforzi soprattutto i poli già forti (siphon effect) invece di produrre complementarità. Il contributo principale del working paper è quindi una baseline multidimensionale pre-apertura, replicabile e comparabile, utile per leggere gli effetti post-AV/HSR non solo in termini di accessibilità, ma di evoluzione delle reti di conoscenza, dei bacini formativi e delle pipeline di lavoro qualificato lungo il corridoio.

References

1. **Adams, J. (2013).** The fourth age of research. *Nature*, 497(7451), 557–560. <https://doi.org/10.1038/497557a>
2. **Albano, G., & Pagliara, F. (2025).** High-speed rail and socioeconomic inequality: a systematic bibliometric analysis of research trends, methodologies and thematic structures. *Railway Sciences*, 4(6), 783–814. <https://doi.org/10.1108/RS-09-2025-0032>
3. **Balland, P. A. (2012).** Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies*, 46(6), 741–756. <https://doi.org/10.1080/00343404.2010.529121>
4. **Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. (2020).** Smart specialization policy in the European Union: A geographical, economic and political perspective. *Journal of Economic Geography*, 20(2), 325–354. <https://doi.org/10.1093/jeg/lbz021>
5. **Boschma, R. (2005).** Proximity and innovation: a critical assessment. *Regional Studies*, 39(1), 61–74. <https://doi.org/10.1080/0034340052000320887>
6. **Bratti, M., & Verzillo, S. (2019).** Geography, university quality and students' mobility: Evidence from Italy. *Regional Studies*, 53(11), 1545–1558. <https://doi.org/10.1080/00343404.2019.1585523>
7. **Bosurgi, G., Coropulis, S., Giacomello, G., Giunta, M., Intignano, M., La Placa, A., Leonardi, G., Loprencipe, G., Pasquini, E., Roberti, R., Rondinella, F., & Vignali, V. (2025).** Infrastrutture ferroviarie. Didattica, ricerca, normativa e nuove sfide: Rapporto tecnico della S.I.I.V. Napoli: FedOA – Federico II University Press. <https://doi.org/10.6093/978-88-6887-372-1>
8. **Cascetta, E., Carteni, A., & Henke, I. (2020).** High speed rail systems: Characteristics, impacts and planning criteria. *Transportation Research Part A: Policy and Practice*, 132, 11–21. <https://doi.org/10.1016/j.tra.2019.10.005>
9. **Cascetta, E., Carteni, A., Henke, I., & Pagliara, F. (2020).** Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post

evaluation and future perspectives. *Transportation Research Part A: Policy and Practice*, 139, 412–428. <https://doi.org/10.1016/j.tra.2020.07.008>

10. **Chen, Z., & Haynes, K. E. (2015).** High-speed rail and regional economic disparity in China: A decomposition analysis. *Transportation Research Record*, 2517(1), 65–72. <https://doi.org/10.3141/2517-08>
11. **Chen, Z., & Silva, E. A. (2013).** The regional-level economic outcome of high-speed rail investment in China: A multivariate time series analysis. *Transport Policy*, 28, 34–45. <https://doi.org/10.1016/j.tranpol.2013.01.001>
12. **Cheng, Y., Loo, B. P. Y., & Vickerman, R. (2015).** High-speed rail networks, inter-regional accessibility and regional inequality in China. *Journal of Transport Geography*, 42, 125–136. <https://doi.org/10.1016/j.jtrangeo.2014.12.001>
13. **Commissione Europea (2021).** *The Scandinavian-Mediterranean Corridor: Fourth Work Plan of the European Coordinator*. Brussels: Directorate-General for Mobility and Transport.
14. **Delaplace, M., Pagliara, F., Perrin, J., & Mermet, S. (2014).** Can High-Speed Rail foster the mobility of students? The case of the University of Naples Federico II. *Transport Policy*, 32, 170–178. <https://doi.org/10.1016/j.tranpol.2014.01.008>
15. **Di Pietro, G. (2012).** Does the distance to the nearest university affect the probability of moving to study? Evidence from Italy. *Economics Bulletin*, 32(1), 849–859.
16. **Dong, X., Zheng, S., & Kahn, M. E. (2020).** The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115, 103212. <https://doi.org/10.1016/j.jue.2019.103212>
17. **Dotti, N. F., Fratesi, U., Lenzi, C., & Percoco, M. (2013).** Local nodes of global networks: The geography of graduate migrations in Italy. *Regional Studies*, 47(9), 1470–1484. <https://doi.org/10.1080/00343404.2012.671451>
18. **Faggian, A., McCann, P., & Sheppard, S. (2007).** Some evidence that women are more mobile than men: Graduate migration in Great Britain. *Journal of Regional Science*, 47(3), 517–539. <https://doi.org/10.1111/j.1467-9787.2007.00517.x>
19. **Frattola, E., Lazzaro, E., Lopresti, I., & Tartaglia, M. (2026).** Brain drain on the fast train: High-speed connections and student mobility. *Temi di discussione (Working Papers)*, No. 1519, Banca d'Italia. <https://doi.org/10.32057/0.TD.2026.1519>
20. **Frenette, M. (2006).** Too far to go on? Distance to school and university participation. *Education Economics*, 14(1), 31–58. <https://doi.org/10.1080/09645290500481865>
21. **Frenken, K., Hardeman, S., & Hoekman, J. (2010).** Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, 3(3), 222–232. <https://doi.org/10.1016/j.joi.2009.03.005>
22. **Garmendia, M., Ureña, J. M., & Coronado, J. M. (2012).** Long-distance trips on high-speed trains: Can they be considered as part of the daily commute? *Transport Policy*, 23, 71–82. <https://doi.org/10.1016/j.tranpol.2012.06.014>
23. **Glänzel, W., & Schubert, A. (2004).** Analysing scientific networks through co-authorship. In *Handbook of quantitative science and*

technology research (pp. 257–276). Springer.
https://doi.org/10.1007/1-4020-2755-9_12

24. **Gutiérrez, J. (2001).** Location, synergy and accessibility: the case of the Madrid–Seville high-speed line. *Journal of Transport Geography*, 9(4), 229–239. [https://doi.org/10.1016/S0966-6923\(01\)00016-X](https://doi.org/10.1016/S0966-6923(01)00016-X)
25. **He, C., Zhou, Y., & Huang, Z. (2020).** High-speed rail and regional innovation: Evidence from China. *Regional Studies*, 54(4), 548–560. <https://doi.org/10.1080/00343404.2019.1583023>
26. **Hoekman, J., Frenken, K., & Van Oort, F. (2009).** The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, 43(3), 721–738. <https://doi.org/10.1007/s00168-008-0252-9>
27. **Katz, J. S., & Martin, B. R. (1997).** What is research collaboration? *Research Policy*, 26(1), 1–18. [https://doi.org/10.1016/S0495-9149\(96\)00017-1](https://doi.org/10.1016/S0495-9149(96)00017-1)
28. **Li, J., Li, Z., & Li, S. (2024).** High-speed rail and the spatial flow of university innovation capital: Evidence from China. *Transport Policy*, 146, 11–25. <https://doi.org/10.1016/j.tranpol.2023.11.004>
29. **Ministero dell'Università e della Ricerca (MUR). (2025).** *Portale dei dati dell'istruzione superiore - USTAT*. [Dataset]. <https://ustat.mur.gov.it>
30. **Moyano, A., Coronado, J. M., & Garmendia, M. (2018).** High-speed rail and student mobility: a demand and supply approach. *Transportation*, 45(6), 1641–1663. <https://doi.org/10.1007/s11116-017-9774-y>
31. **Pagliara, F., Mauriello, F., & Garofalo, A. (2021).** Exploring the impact of High Speed Rail on tourism in Italy. *Transportation Research Interdisciplinary Perspectives*, 10, 100343. <https://doi.org/10.1016/j.trip.2021.100343>
32. **Ponds, R., Van Oort, F., & Frenken, K. (2007).** The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3), 423–443. <https://doi.org/10.1111/j.1435-5957.2007.00126.x>
33. **Prota, F., & Viesti, G. (2022).** Linking the ‘Recovery and Resilience Plan’ and Smart Specialisation. The Italian Case. JRC Working Papers on Territorial Modelling and Analysis, No. 10/2022, European Commission, Seville. <https://publications.jrc.ec.europa.eu/repository/handle/JRC130071>
34. **Prota, F., & Calò, E. (2024).** *How to Evaluate the Smart Specialisation Strategy? A Methodological Proposal*. SSRN Working Paper
35. **Sá, C., Florax, R. J., & Rietveld, P. (2004).** Determinants of the regional mobility of students in the Netherlands: A discrete choice model. *The Annals of Regional Science*, 38(3), 467–492. <https://doi.org/10.1007/s00168-003-0145-8>
36. **Stateva, M., et al. (2021).** *Using LinkedIn data to measure the international mobility of professionals*. Eurostat Statistical Reports.
37. **Ureña, J. M., Menerault, P., & Garmendia, M. (2009).** The high-speed rail challenge for intermediate cities: A national, regional and local perspective. *Cities*, 26(5), 266–279. <https://doi.org/10.1016/j.cities.2009.07.001>
38. **Venhorst, V., Van Dijk, J., & Van Wissen, L. (2011).** An analysis of the migration patterns of Dutch higher education graduates. *Spatial Economic Analysis*, 6(1), 109–132. <https://doi.org/10.1080/17421772.2010.540032>

39. **Wagner, C. S. (2008).** *The New Invisible College: Science for Development*. Brookings Institution Press.
40. **Wang, Y., Ni, L., & Liu, J. (2020).** Does high-speed rail facilitate academic collaboration? Evidence from China. *Journal of Transport Geography*, 82, 102573. <https://doi.org/10.1016/j.jtrangeo.2019.102573>
41. **Wu, J., Wang, S., & Liu, X. (2023).** High-speed rail, spatial networks, and urban economic resilience: A perspective of social media data and input-output analysis. *Cities*, 134, 104184. <https://doi.org/10.1016/j.cities.2022.104184>
42. **Zhang, X., Wan, G., & Yang, J. (2019).** Technical efficiency, high-speed rail and regional inequality in China. *China & World Economy*, 27(2), 64–91. <https://doi.org/10.1111/cwe.12273>
43. **Zherdev, N. (2020).** LinkedIn as a data source for studying labor mobility and professional networks. *Journal of Economic Geographies*, 20(4).
44. **Abadie, A., Diamond, A., & Hainmueller, J. (2010).** Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505. <https://doi.org/10.1198/jasa.2009.ap08746>
45. **Cascetta, E., Pagliara, F., Papola, A., & Vassallo, J. M. (2020).** High-speed rail and regional development: Evidence from Italy. *Transportation Research Part A: Policy and Practice*, 132, 12–28. <https://doi.org/10.1016/j.tra.2018.12.022>
46. **Chen, Z., & Haynes, K. E. (2017).** Impact of high-speed rail on regional economic disparity in China. *Journal of Transport Geography*, 65, 80–91. <https://doi.org/10.1016/j.jtrangeo.2017.10.009>
47. **Faggian, A., McCann, P., & Sheppard, S. (2007).** Human capital, higher education and graduate migration: An analysis of Scottish and Welsh students. *Urban Studies*, 44(13), 2511–2528. <https://doi.org/10.1080/00420980701666667>
48. **Garmendia, M., Ureña, J. M., & Coronado, J. M. (2011).** Long-distance trips in a sparsely populated region: The impact of high-speed rail. *Transport Policy*, 18(1), 187–196. <https://doi.org/10.1016/j.tranpol.2010.07.006>
49. **Gibbons, S., & Machin, S. (2005).** Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), 148–169. <https://doi.org/10.1016/j.jue.2004.10.002>
50. **Moyano, A., Coronado, J. M., & Garmendia, M. (2018).** High-speed rail and urban transformation: Spatial and economic impacts. *Cities*, 74, 89–98. <https://doi.org/10.1016/j.cities.2017.11.006>
51. **Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., & Fowler, J. H. (2017).** The strength of weak ties in a large-scale networks. *Nature Human Behaviour*, 1(1), 0005. <https://doi.org/10.1038/s41562-016-0005>
52. **Granovetter, M. S. (1973).** The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.1086/225469>
53. **Korinek, A. (2023).** Generative AI for Economic Research: Use Cases and Implications for Economists. *Journal of Economic Literature* (forthcoming). <https://doi.org/10.3386/w31174>
54. **Macy, M. (2023).** Social networks and the strength of weak ties: A review. *Annual Review of Sociology*, 49, 1–20. <https://doi.org/10.1146/annurev-soc-031021-114402>
55. **Stateva, M., et al. (2021).** Using LinkedIn data to measure the international mobility of professionals. *Eurostat Statistical Reports*. <https://doi.org/10.2785/29166>

56. World Bank. (2018). Digital Traces: Using New Data to Measure Development. World Bank Group. <http://hdl.handle.net/10986/30374>

57.

Appendice

Tabella A.1

Pubblicazioni scientifiche - Web of Science									
Anno	Bari	Uniba	Poliba	Lum	Napoli	Feder. II	Vanvitelli	Oriental e	Phartenop e
2024	4815	3758	1057	26	8635	7624	698	89	757
2023	4889	3826	987	15	8620	7585	745	105	740
2022	4974	4004	970	15	8996	7896	849	103	685
2021	4968	3943	930	7	9066	8030	838	113	739
2020	4683	3746	925	7	8953	7907	843	101	760
2019	4085	3205	895	1	7850	6842	777	78	838
2018	3431	2639	852	2	7101	6049	874	83	797
2017	3345	2576	822	1	7487	5898	1481	74	751
2016	3040	2328	720	0	7925	5633	2234	38	713
2015	3102	2452	706	0	7321	5322	1905	50	597
2014	2809	2288	583	0	6564	4776	1740	36	478
2013	2494	1998	528	0	5906	4359	1514	29	347
2012	2390	1951	488	0	5270	3956	1307	21	343
2011	2267	1906	411	0	4836	3535	1226	25	287
2010	2313	1888	452	0	4644	3493	1148	16	216
2009	2186	1754	417	0	4565	3451	1086	22	220
2008	1991	1678	321	0	4190	3226	1011	14	147
2007	1927	1620	307	0	4160	3290	923	13	108
2006	1839	1533	333	0	3626	2864	873	7	68
2005	1611	1387	225	0	3246	2642	740	6	31
2004	1551	1331	227	0	3120	2504	741	4	31
2003	1354	1125	223	0	2798	2281	660	3	17
2002	1200	1020	179	0	2579	2052	659	2	22
2001	1107	965	150	0	2326	1932	505	1	4
2000	998	854	141	0	2343	1948	507	3	0
1999	981	872	108	0	2145	1827	414	1	0
1998	879	771	110	0	2110	1777	424	1	0
1997	918	837	84	0	2071	1722	413	4	0
1996	847	734	110	0	1956	1634	366	1	0
1995	756	692	59	0	1728	1461	262	1	0
1994	727	644	80	0	1518	1281	184	2	0
1993	646	599	41	0	1309	1141	83	3	0
1992	560	546	14	0	1102	1032	3	2	0
1991	470	462	3	0	1057	990	1	1	0
1990	400	387	5	0	1028	940	1	2	0
1989	459	428	8	0	939	852	2	2	0
1988	423	405	12	0	1054	940	1	0	0
1987	401	389	5	0	922	808	2	0	0
1986	376	354	4	0	852	738	2	0	0
1985	451	436	5	0	973	863	0	0	0
Tot	79213	64619	14590	98	162891	133101	28042	1056	9696

Tabella A.2

Pubblicazioni scientifiche - Web of Science					
Anno	Bari e Napoli	Bari	% coll. Bari	Napoli	% coll. Napoli
2024	2718	4815	56%	8635	31%
2023	2840	4889	58%	8620	33%
2022	2936	4974	59%	8996	33%
2021	3001	4968	60%	9066	33%
2020	2940	4683	63%	8953	33%
2019	2883	4085	71%	7850	37%
2018	2530	3431	74%	7101	36%
2017	2414	3345	72%	7487	32%
2016	1585	3040	52%	7925	20%
2015	1072	3102	35%	7321	15%
2014	795	2809	28%	6564	12%
2013	717	2494	29%	5906	12%
2012	800	2390	33%	5270	15%
2011	666	2267	29%	4836	14%
2010	621	2313	27%	4644	13%
2009	496	2186	23%	4565	11%
2008	492	1991	25%	4190	12%
2007	416	1927	22%	4160	10%
2006	383	1839	21%	3626	11%
2005	256	1611	16%	3246	8%
2004	274	1551	18%	3120	9%
2003	189	1354	14%	2798	7%
2002	196	1200	16%	2579	8%
2001	161	1107	15%	2326	7%
2000	123	998	12%	2343	5%
1999	127	981	13%	2145	6%
1998	140	879	16%	2110	7%
1997	95	918	10%	2071	5%
1996	62	847	7%	1956	3%
1995	74	756	10%	1728	4%
1994	309	727	43%	1518	20%
1993	419	646	65%	1309	32%
1992	191	560	34%	1102	17%
1991	158	470	34%	1057	15%
1990	95	400	24%	1028	9%
1989	71	459	15%	939	8%
1988	93	423	22%	1054	9%
1987	83	401	21%	922	9%
1986	77	376	20%	852	9%
1985	146	451	32%	973	15%
Tot	33644	79213		162891	

**Tabella
A.3**

Istituti affiliati in Pubblicazioni scientifiche con Ba e Na – top 50- Source: Web of Science		
Istituzioni	Numero pubblicazioni	% pubblicazioni
UNIVERSITY OF NAPLES FEDERICO II INSTITUTO NAZIONALE DI FISICA NUCLEARE INFN *	26,707	78,7
SAPIENZA UNIVERSITY OF ROME	6,256	18,4
UNIVERSITA' DEGLI STUDI DI BARI ALDO MORO CENTRE	5,967	17,6
NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS CONSIGLIO NAZIONALE DELLE RICERCHE CNR	5,02	14,8
UNIVERSITY OF PISA	4,93	14,5
UNIVERSITA' DELLA CAMPANIA	4,628	13,6
	4,617	13,6

VANVITE		
LLI		
UNIVE		
RSITY	4,466	
OF		
GENOA		13,2
PARTH		
ENOPE		
UNIVERS	4,397	
ITY		
NAPLES		13,0
UNIVE		
RSITY		
OF	4,337	
CALIFOR		
NIA		
SYSTEM		12,8
UNIVE		
RSITY		
OF	4,229	
BOLOGN		
A		12,5
CEA	4,209	12,4
UNITE		
D		
STATES		
DEPART		
MENT	4,176	
OF		
ENERGY		
DOE		12,3
UNIVE		
RSITE	4,123	
PARIS		
SACLAY		12,1
HELM		
HOLTZ	3,954	
ASSOCIA		
TION		11,6
CNRS		
NATION		
AL		
INSTITU		
TE OF		
NUCLEA	3,887	
R AND		
PARTICL		
E		
PHYSICS		
IN2P3		11,5
CONSE		
JO		
SUPERIO		
R DE		
INVESTI		
GACION	3,861	
ES		
CIENTIFI		
CAS		
CSIC		11,4

RUSSEAN ACADEMY OF SCIENCES	3,843	11,3
CHINESE ACADEMY OF SCIENCES	3,809	11,2
UNIVERSITY OF PADUA	3,771	11,1
UNIVERSITY SYSTEM OF OHIO	3,65	10,8
UNIVERSITY OF WISCONSIN	3,51	10,3
SYSTEM UNIVERSITY OF TURIN	3,483	10,3
UK RESEARCH INNOVATION	3,463	10,2
UKRI UNIVERSITY OF MILAN	3,462	10,2
STATE UNIVERSITY OF NEW YORK	3,436	10,1
SUNY SYSTEM SCIENCE TECHNOLOGY FACILITIES	3,428	10,1
COUNCIL STFC STFC RUTHERFORD APPLETON	3,425	10,1

LABORATORY UNIVERSITY OF WISCONSIN MADISON	3,393		10,0
UNIVERSITY OF PAVIA OHIO STATE UNIVERSITY	3,344		9,9
UNIVERSITY OF IOWA INSTITUTE OF HIGH ENERGY PHYSICS	3,307		9,7
CAS MASSACHUSETTS INSTITUTE OF TECHNOLOGY	3,301		9,7
MIT UNIVERSITY OF CHICAGO	3,159		9,3
UNIVERSITY OF FLORENCE HUNTER EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH CERN	3,068		9,0
NATIONAL RESEARCH CENTRE KURCHA	3,048		9,0
	3,033		8,9
	3,03		8,9
	2,976		
			8,8
	2,925		
			8,6

TOV INSTITU TE CHAR LES UNIVERS ITY	2,9	
PRAGUE BOSTO N UNIVERS ITY	2,876	8,5
HUN REN WIGNER RESEAR CH	2,864	8,5
CENTRE FOR PHYSICS NATIO NAL		8,4
KAPODIS TRIAN UNIVERS ITY OF	2,849	
ATHENS UNIVE RSITY OF	2,837	8,4
CATANIA UNIVE RSITE		8,4
PARIS CITE	2,817	
AUTO NOMOUS UNIVERS ITY OF	2,814	8,3
MADRID UNIVE RSITY OF	2,81	
ILLINOIS SYSTEM DEUTS CHES ELEKTR ONEN	2,77	8,3
SYNCHR OTRON DESY UNIVE RSITY		8,2
OF BELGRA DE	2,756	
		8,1

* INFN ha più sedi sul territorio
Nazionale, di cui una a Bari e una a
Napoli

Tabella A.4

**Dipartimenti affiliati in Pubblicazioni scientifiche con Base
Na – top 50-**

Source: Web of Science

Istituzione	Numero pubblicazioni	% pubblicazioni
University Of Naples Federico Ii Polytechnic And Applied Sciences School	8,586	25,279
University Of Naples Federico Ii Department Of Physics Ettore Pancini	4,038	11,889
University Of Naples Federico Ii School Of Medicine And Surgery	2,691	7,923
University Of Rome La Sapienza Faculty Of Mathematical Physical And Natural Sciences	2,37	6,978
University Of Rome La Sapienza Department Of Physics	2,228	6,56
Lomonosov Moscow State University Faculty Of Physics	2,195	6,463
University Of Warwick Faculty Of Science Engineering And Medicine	2,175	6,404
University Of Warwick Department Of Physics	2,155	6,345
University Of Milan Faculty Of Science And Technology	2,125	6,256
Lomonosov Moscow State University Skobeltsyn Institute Of Nuclear Physics	2,103	6,192

University Of Genoa School Of Mathematical Physical And Natural Sciences	2,093	6,162
University Of Genoa Department Of Physics	2,057	6,056
University Of Milan Department Of Physics Aldo Pontremoli	2,052	6,042
University Of Milan Sciences And Technologies Of Physics	2,052	6,042
University Of Pisa Department Of Physics	2,018	5,941
University Of California Santa Cruz Department Of Physics	1,936	5,7
University Of California Santa Cruz Division Of Physical Biological Sciences	1,936	5,7
University Of California Santa Cruz Santa Cruz Institute For Particle Physics	1,936	5,7
University Of Oxford Mathematical Physical And Life Sciences Division	1,907	5,615
University Of Oxford Department Of Physics	1,881	5,538
Heidelberg University Department Of Physics And Astronomy	1,852	5,453
Tu Dresden Faculty Of Mathematics And Natural Sciences	1,797	5,291
Tu Dresden Department Of Physics	1,793	5,279

Tu Dresden Institute Of Nuclear And Particle Physics	1,789	5,267
Heidelberg University Institute Of Physics	1,765	5,197
Uppsala University Faculty Of Science And Technology	1,763	5,191
Uppsala University Department Of Physics And Astronomy	1,755	5,167
Comenius University In Bratislava Faculty Of Mathematics Physics And Informatics	1,705	5,02
University Of Naples Federico Ii School Of Agricultural Science And Veterinary Medicine	1,692	4,982
Johannes Gutenberg University Mainz Faculty Of Physics Mathematics And Computer Science	1,66	4,887
University Of Rome Tor Vergata Macro Area Of Mathematical Physical And Natural Sciences	1,65	4,858
University Of Rome Tor Vergata Department Of Physics	1,607	4,731
Humboldt University Of Berlin Faculty Of Mathematics And Natural Sciences	1,589	4,678
Humboldt University Of	1,586	4,67

Berlin Institute Of Physics		
The University Of Tokyo	1,585	4,667
Graduate School Of Science		
Faculty Of Science		
Tel Aviv University	1,577	4,643
Raymond And Beverly Sackler Faculty Of Exact Sciences		
Tel Aviv University	1,571	4,625
School Of Physics And Astronomy		
The University Of Texas At Austin College Of Natural Sciences	1,563	4,602
The University Of Texas At Austin	1,546	4,552
Department Of Physics		
Charles University	1,54	4,534
Faculty Of Mathematics And Physics		
The University Of Tokyo	1,528	4,499
Graduate School Of Science		
Faculty Of Science		
Department Of Physics		
University Of Toronto Faculty Of Arts Science	1,513	4,455
Roma Tre University	1,507	4,437
Department Of Mathematics And Physics		
Middle East Technical University	1,479	4,354
Faculty Of Arts And Sciences		
Middle East Technical University	1,477	4,349
Department Of Physics		

The University Of Edinburgh College Of Science And Engineering	1,477	4,349
Nagoya University School Of Science Graduate School Of Science	1,473	4,337
Boston University College Of Arts And Sciences	1,469	4,325
University College London Faculty Of Mathematical And Physical Sciences	1,467	4,319
Boston University Department Of Physics	1,459	4,296

Tabella A.5

Flussi studenti in valore assoluto da Puglia, Campania, Basilicata

Origine	Destinazione	Immatricolati	Anno
Puglia	Abruzzo	551	2024/2025
Puglia	Basilicata	117	2024/2025
Puglia	Calabria	32	2024/2025
Puglia	Campania	192	2024/2025
Puglia	Emilia-Romagna	912	2024/2025
Puglia	Friuli-Venezia Giulia	49	2024/2025
Puglia	Lazio	706	2024/2025
Puglia	Liguria	15	2024/2025
Puglia	Lombardia	838	2024/2025
Puglia	Marche	148	2024/2025
Puglia	Piemonte	278	2024/2025
Puglia	Puglia	15861	2024/2025
Puglia	Sardegna	6	2024/2025
Puglia	Sicilia	26	2024/2025
Puglia	Toscana	282	2024/2025
Puglia	Trentino-Alto Adige	73	2024/2025
Puglia	Umbria	143	2024/2025
Puglia	Valle d'Aosta	1	2024/2025
Puglia	Veneto	139	2024/2025
Campania	Abruzzo	361	2024/2025
Campania	Basilicata	55	2024/2025
Campania	Calabria	38	2024/2025
Campania	Campania	28714	2024/2025

		Campania		Emilia-Romagna		262		2024/2025
		Campania		Friuli-Venezia Giulia		9		2024/2025
		Campania		Lazio		1210		2024/2025
		Campania		Liguria		4		2024/2025
		Campania		Lombardia		566		2024/2025
		Campania		Marche		30		2024/2025
		Campania		Piemonte		82		2024/2025
		Campania		Puglia		76		2024/2025
		Campania		Sardegna		6		2024/2025
		Campania		Sicilia		28		2024/2025
		Campania		Toscana		256		2024/2025
Tabella A.6		Campania		Trentino-Alto Adige		16		2024/2025
Serie storica flussi interregionali (immatricolati)		Campania		Umbria		149		2024/2025
Anno	Puglia → Campania	Campania → Basilicata	Campania → Puglia	Veneto → Campania	Campania → Basilicata	Basilicata → Campania		
2009/2010	79	107	60	128	549	214		2024/2025
2010/2011	77	98	57	94	566	261		2024/2025
2011/2012	124	86	48	97	584	307		2024/2025
2012/2013	122	90	43	81	506	305		2024/2025
2013/2014	195	72	45	71	515	278		2024/2025
2014/2015	178	82	35	88	481	308		2024/2025
2015/2016	189	122	29	Puglia		611		2024/2025
2016/2017	189	126	45	Sardegna	424	285		2024/2025
2017/2018	210	141	51	Sicilia	440	321		2024/2025
2018/2019	190	120	56	Toscana		133		2024/2025
2019/2020	238	118	48	Trentino-Alto Adige	455	313		2024/2025
2020/2021	277	108	47	Umbria	499	344		2024/2025
2021/2022	197	113	52	Veneto		37		2024/2025
Tabella A.7								
Flussi A.A.	2022/2023	187	117	76	70	548	271	Provinciali 2024/2025
Regione di residenza	2023/2024	214	128	64	42	569	316	Altre regioni (%)
	2024/2025	192	117	76 (%)	55	611	293 (%)	
Elaborazione dell'autore su dati MIUR								
Basilicata	MATERA	11.3		42.7		2.8	0.0	43.2
Basilicata	POTENZA	30.7		12.2		14.8	0.0	42.3
Campania	AVELLINO	81.4		1.0		0.0	0.2	17.4

Campania	BENEVENTO	78.2	0.5	0.0	0.1	21.2
Campania	CASERTA	88.1	0.1	0.0	0.0	11.8
Campania	NAPOLI	94.5	0.1	0.0	0.0	5.3
Campania	SALERNO	83.4	0.3	0.0	0.8	15.5
Puglia	BARI	87.6	0.0	0.5	1.1	10.8
Puglia	BARLETTA - ANDRIA - TRANI	85.4	0.0	0.7	0.3	13.6
Puglia	BRINDISI	71.9	0.0	1.0	0.3	26.8
Puglia	FOGGIA	72.9	0.0	1.9	0.1	25.1
Puglia	LECCE	72.1	0.0	0.6	0.0	27.3
Puglia	TARANTO	67.6	0.0	1.4	1.2	29.8

Elaborazione dell'autore su dati MIUR

Tabella A.8

Query per l'Individuazione dei profili LinkedIn connessi sia a Bari che a Napoli con Proxy Url		
Query_Block	Query_ID	Query_Text
BLOCCO 1 – Query manuali mirate (seed queries)		
Q1	Q1.1	LinkedIn Bari Napoli
Q1	Q1.2	LinkedIn Napoli Bari
Q1	Q1.3	LinkedIn Bari to Naples
Q1	Q1.4	LinkedIn Naples to Bari
Q1	Q1.5	LinkedIn studied in Bari works in Naples
Q1	Q1.6	LinkedIn studied in Naples works in Bari
Q1	Q1.7	LinkedIn graduated Bari working Naples
Q1	Q1.8	LinkedIn graduated Naples working Bari
Q1	Q1.9	LinkedIn Università di Bari Napoli
Q1	Q1.10	LinkedIn Università degli Studi di Bari Napoli
Q1	Q1.11	LinkedIn Federico II Bari
Q1	Q1.12	LinkedIn Politecnico di Bari Napoli
Q1	Q1.13	LinkedIn Parthenope Bari
Q1	Q1.14	LinkedIn Provincia di Bari Napoli
Q1	Q1.15	LinkedIn Provincia di Napoli Bari
Q1	Q1.16	LinkedIn Software Engineer Bari Napoli
Q1	Q1.17	LinkedIn Ingegnere Bari Napoli
Q1	Q1.18	LinkedIn Consultant Bari Napoli
Q1	Q1.19	LinkedIn Manager Bari Napoli
Q1	Q1.20	LinkedIn Analyst Bari Napoli
Q1	Q1.21	LinkedIn Bari Naples Italy
Q1	Q1.22	LinkedIn Naples Bari Italy

Q1	Q1.23	LinkedIn relocation Bari Naples
Q1	Q1.24	LinkedIn relocation Naples Bari
BLOCCO 2 – Query generate automaticamente (alta copertura)		
Q2	Q2.1	LinkedIn {city_1} {city_2}
Q2	Q2.2	LinkedIn from {city_1} working in {city_2}
Q2	Q2.3	LinkedIn {verb} {city_1} {city_2}
Q2	Q2.4	{verb} {city_1} moved to {city_2} LinkedIn
Q2	Q2.5	LinkedIn {university} {city}
Q2	Q2.6	{university} alumni {city} LinkedIn
Q2	Q2.7	{university} {city} profilo LinkedIn
Q2	Q2.8	LinkedIn {movement_pattern}
BLOCCO 3 – Query string-based con filtro semantico ex post		
Q3	Q3.1	"LinkedIn" "Bari" "Napoli"
Q3	Q3.2	"LinkedIn" "Bari" "Naples"
Q3	Q3.3	"Bari" "Napoli" site:linkedin.com/in
Q3	Q3.4	"Bari" "Naples" site:linkedin.com/in
Q3	Q3.5	"Napoli" "Bari" site:linkedin.com/in
Q3	Q3.6	"Naples" "Bari" site:linkedin.com/in
Q3	Q3.7	"LinkedIn profile" "Bari" "Napoli"
Q3	Q3.8	"LinkedIn career" "Bari" "Napoli"
Q3	Q3.9	"works in Naples" "Bari" "LinkedIn"
Q3	Q3.10	"works in Bari" "Napoli" "LinkedIn"
Q3	Q3.11	"ha studiato a Bari" "Napoli" "LinkedIn"
Q3	Q3.12	"ha studiato a Napoli" "Bari" "LinkedIn"
Q3	Q3.13	"studied in Bari" "Naples" "LinkedIn"
Q3	Q3.14	"studied in Naples" "Bari" "LinkedIn"
Q3	Q3.15	"Bari Napoli" "LinkedIn"
Q3	Q3.16	"Napoli Bari" "LinkedIn"

Tabella A.9

Dataset finale – Variabili utilizzate nell’analisi LinkedIn

Variable	Type	Source	Description
linkedin_url	Text	Raw (Proxy URL / Enrichment)	Unique profile identifier (LinkedIn public URL)
raw_json	Text	Raw (Enrichment)	Raw enrichment payload (original JSON content)
status	Categorical	Raw (Enrichment)	Status flag of the enrichment / parsing pipeline for the profile
json	Text	Processed (Storage)	Serialized/cleaned JSON representation used in the pipeline
education_parsed	Text	Parsed (Rule-based)	Structured education records parsed from raw_json
experience_parsed	Text	Parsed (Rule-based)	Structured work experience records parsed from raw_json
first_edu_year	Continuous	Derived (Rule-based)	First observed education start year (timeline anchor)

first_edu_city	Categorical	Derived (Rule-based)	City/area of first observed education (mapped from institution/location)
first_job_year	Continuous	Derived (Rule-based)	First observed work start year (from experience timeline)
first_job_city	Categorical	Derived (Rule-based)	City/area of first observed job (mapped from work location)
city_timeline	Text	Derived (Rule-based)	Ordered sequence of city/area states across education+work events (compressed)
first_migration_year	Continuous	Derived (Rule-based)	Year of first detected inter-area transition in the timeline
years_to_first_migration	Continuous	Derived (Rule-based)	Time (years) from first anchor event to first migration event
first_migration_trigger	Categorical	Derived (Rule-based)	Trigger of first Bari↔Napoli connection (e.g., University / first job / subsequent job / last job)
public_identifier	Text	Raw (Enrichment)	LinkedIn public identifier/slug (when available)
first_name	Text	Raw (Enrichment)	First name extracted from profile
last_name	Text	Raw (Enrichment)	Last name extracted from profile
full_name	Text	Raw (Enrichment)	Full name string extracted from profile
city	Categorical	Raw (Enrichment)	Profile location city field (as provided by enrichment)
country	Categorical	Raw (Enrichment)	Profile location country field (as provided by enrichment)
location_str	Text	Raw (Enrichment)	Free-text location string from profile header
headline	Text	Raw (Enrichment)	Headline / tagline text from profile
summary	Text	Raw (Enrichment)	About/summary free-text field (when available)
occupation	Text	Raw (Enrichment)	Current occupation/title field (when available)
first_name_clean	Text	Derived (Rule-based)	Normalized first name (cleaning for matching/mapping)
gender	Categorical	Derived (Rule-based)	Gender proxy inferred via name-based mapping (conservative; missing if unclear)
origin_city	Categorical	Derived (Rule-based)	First informative territorial anchor (origin_final: Bari / Napoli / Italia / Estero)
highest_education	Categorical	LLM-based (Constrained)	Highest education category inferred with closed-label LLM classifier
last_employment_type	Categorical	LLM-based (Constrained)	Employment type of last job (private/public/self-employed/academia/other) inferred by LLM
work_area	Categorical	LLM-based (Constrained)	Macro professional sector (STEM, academia, finance, etc.) inferred by LLM
age_proxy	Continuous	Derived (Rule-based)	Age proxy inferred from first_edu_year (fallback: first_job_year) with fixed assumptions

Tabella A.10

Dataset evento-livello (df_long) — Variabili utilizzate nell'analisi LinkedIn

Variable	Type	Source	Description
linkedin_url	Text	Raw (Proxy URL / Enrichment)	Profile identifier linking events to individuals
event_type	Categorical	Parsed (Rule-based)	Event type flag (education/work) from parsed sections

institution	Text	Raw (Enrichment)	Institution/employer name as extracted from profile
institution_norm	Text	Derived (Rule-based)	Normalized institution string (lowercase, punctuation/variants cleaned)
area	Categorical	Derived (Rule-based)	Territorial category assigned to event (Bari/Napoli/Italia/Estero/Unknown) before collapsing
degree	Text	Raw (Enrichment)	Degree title (education events only)
field	Text	Raw (Enrichment)	Field of study (education events only)
city	Text	Raw/Parsed (Enrichment)	City/location string associated with the event (when available)
country	Text	Raw/Parsed (Enrichment)	Country string associated with the event (when available)
start_year	Continuous	Parsed (Rule-based)	Start year extracted/converted from date fields
end_year	Continuous	Parsed (Rule-based)	End year extracted/converted from date fields (can be missing for ongoing)
role	Text	Raw (Enrichment)	Role/title text for work events
area_core	Categorical	Derived (Rule-based)	Collapsed core area for analysis (Bari/Napoli/None) used to detect BA↔NA transitions

Tabella A.11

Dataset traiettoria-livello (df_traj) — Variabili utilizzate nell'analisi LinkedIn

Variable	Type	Source	Description
linkedin_url	Text	Derived (Key)	Profile identifier (one row per individual trajectory)
n_events_total	Continuous	Derived (Rule-based)	Total number of events (education+work) in df_long for the profile
n_events_edu	Continuous	Derived (Rule-based)	Number of education events for the profile
n_events_work	Continuous	Derived (Rule-based)	Number of work events for the profile
n_steps_total	Continuous	Derived (Rule-based)	Number of territorial steps after cleaning/compression (education+work)
n_steps_edu	Continuous	Derived (Rule-based)	Number of territorial steps after cleaning/compression (education only)
n_steps_work	Continuous	Derived (Rule-based)	Number of territorial steps after cleaning/compression (work only)
trajectory_total	Text	Derived (Rule-based)	Territorial sequence (education+work) after removing Unknown and compressing repeats
trajectory_edu	Text	Derived (Rule-based)	Territorial sequence for education events only (cleaned/compressed)
trajectory_work	Text	Derived (Rule-based)	Territorial sequence for work events only (cleaned/compressed)
trajectory_total_list	Text	Derived (Rule-based)	List representation of trajectory_total (ordered categories)
trajectory_edu_list	Text	Derived (Rule-based)	List representation of trajectory_edu
trajectory_work_list	Text	Derived (Rule-based)	List representation of trajectory_work
studied_in_bari	Binary	Derived (Rule-based)	Indicator: at least one education event mapped to Bari
studied_in_napoli	Binary	Derived (Rule-based)	Indicator: at least one education event mapped to Napoli

worked_in_bari	Binary	Derived (Rule-based)	Indicator: at least one work event mapped to Bari
worked_in_napoli	Binary	Derived (Rule-based)	Indicator: at least one work event mapped to Napoli
trajectory_type	Categorical	Derived (Rule-based)	Synthetic trajectory class (stayer/one-way/return/multi-node/estero-combo/residual)