



UNIVERSITÀ DI PISA

**Dipartimento di Filologia, Letteratura, Linguistica
Laurea Magistrale in Informatica Umanistica**

Tesi di Laurea

LinkedIn Data:

**uno sguardo attraverso LinkedIn alle carriere
degli alumni e delle alumnae degli atenei toscani**

**Relatore:
Prof. Giulio Rossetti**

**Candidato:
Gianmarco Di Mauro**

**Correlatrice:
Dott.ssa Laura Pollacci**

Anno Accademico 2020/2021

Abstract

Argomento cardine di questo studio è l'utilizzo di dati non convenzionali per l'analisi delle carriere e della mobilità di soggetti altamente formati. In particolare, la fonte di dati utilizzata è il professional network LinkedIn, mentre soggetto dello studio sono gli alumni e le alumnae di alcuni atenei toscani.

Questa tesi descrive l'approccio elaborato per la raccolta e la pulizia dei dati al fine di ottenere un'analisi esplorativa.

L'analisi è stata condotta su un campione di 9405 utenti di LinkedIn, di cui 2372 riportanti evidenze di spostamenti internazionali per motivi di studio e/o lavoro.

Sommario

1	Introduzione	5
2	Stato dell'Arte	10
2.1	Mobilità in Italia	13
2.2	Studi con dati non convenzionali	15
3	LinkedIn	19
4	Metodologia	23
4.1	Raccolta dati	24
4.2	Preparazione dati	26
4.2.1	Dati sensibili: il nome utente e la pseudonimizzazione	26
4.2.2	Trasformazione dei dati geografici	28
4.2.3	Trasformazioni dei dati sull'istruzione	31
4.2.4	Trasformazione dati professionali	38
4.2.5	Mobilità internazionale	40
5	Caso di studio	42
5.1	Dataset	44
5.2	Analisi Statistica	47
5.2.1	Località geografica	47
5.2.2	Percorsi di studio	49
5.2.3	Esperienze lavorative	56
5.3	Mobilità Internazionale	61

5.3.1	Differenze a livello accademico	66
5.3.2	Differenze percorsi professionali	69
5.4	Discussione	73
6	Problematiche e Sviluppi futuri	75
7	Conclusioni	78
A	Tabelle	82
B	Tabella Mobilità	104

1 || Introduzione

Un argomento caldo all'interno del dibattito pubblico è quello del **brain drain**, più comunemente detto in Italia “fuga dei cervelli”. Con questo termine si indica, in senso ampio, il trasferimento fuori dai confini nazionali di soggetti altamente qualificati, cioè di persone che hanno raggiunto traguardi molto alti nel settore accademico¹, in quello professionale², sportivo o artistico [Milio et al., 2012].

A suscitare tanto interesse sono le conseguenze politiche, economiche e sociali che tali spostamenti provocano: il brain drain è infatti la principale causa di drenaggio di persone, dato che in quasi tutti i paesi del mondo si registra un percentuale maggiore di espatri tra i lavoratori qualificati rispetto a tutte le altre categorie [Docquier and Marfouk, 2004]. Nella visione tradizionale, il termine “brain drain” ha quindi una connotazione negativa, in quanto inserito nella dinamica tra Nord e Sud del mondo [Meyer, 2001] o più in generale in quella tra paesi sviluppati e paesi in via di sviluppo. In questo quadro, i paesi più sviluppati godrebbero del capitale umano formato a spese degli altri stati, senza che questi ne ricavano alcun ritorno economico.

Studi recenti hanno tuttavia portato nuova luce sul dibattito, evidenziandone la complessità e l'impatto su scala globale. Al fenomeno del brain drain è stato accostato la sua controparte, cioè il **brain gain**, indicante invece l'afflusso verso una nazione di persone altamente qualificate. Il saldo di uno stato andrebbe dunque considerato sul totale dei flussi bidirezionali, cioè su quello che viene definito in letteratura come **brain exchange**. A ciò si aggiunge il fenomeno del rientro in patria o a ciò che viene definito come **brain circulation**, ovvero un percorso di perfezionamento tenuto all'estero, a livello accademico o professionale, con successivo ritorno nel paese di origine [Milio et al., 2012]. La complessità del fenomeno si rifletterebbe soprattutto nel mondo accademico. L'approccio marcatamente internazionale delle più recenti ricerche e collaborazioni mostra una sostanziale ininfluenza della localizzazione dell'attività di ricerca, spostandone invece l'attenzione soprattutto sui risultati, con influenze spesso a livello globale come visto relativamente alla gestione della pandemia di Covid-19 [Giroud and Ivarsson, 2020].

¹Ricercatori quindi, ma più in generale tutti i possessori di un titolo di educazione terziaria, cioè di una laurea o di un master.

²Come professionisti di progetto, consulenti specializzati e business man.

Come però evidenzia [Milio et al., 2012], gli interscambi di flussi migratori tra soggetti qualificati avvengono principalmente tra quelle regioni del Nord del mondo già sviluppate, mentre i fenomeni migratori riguardanti le nazioni in via di sviluppo si caratterizzano generalmente come unidirezionali. Guardando ad esempio al caso dell'Italia, le analisi più recenti hanno evidenziato come negli ultimi anni si sia registrata una crescita considerevole del fenomeno del brain drain³, senza che a questa corrispondesse una crescita nel numero di ingressi di soggetti altamente formati nel paese (brain gain), definendo quindi i movimenti in entrata e in un uscita dal paese come non equiparabili [Brandi, 2017] e soprattutto come dispendiosi, con un costo medio per lo stato di circa 110.000 euro per emigrato [Boffo and Gagliardi, 2017].

Il dibattito tra chi sostiene una visione globalizzata del fenomeno e quelli che invece ne sottolineano le criticità è quindi ancora aperto. Complice di questa frattura è la mancanza di una fonte univoca di dati, che rende difficoltosa la quantificazione del fenomeno e ne limita gli studi. Per affrontare l'argomento è quindi generalmente necessario incrociare più fonti al fine di restituire una stima quanto più valida possibile.

A livello internazionale, le due fonti più autorevoli sono la Banca Mondiale⁴ e l'OECD⁵, ovvero l'Organizzazione per la Cooperazione e lo Sviluppo Economico. In entrambi questi istituti vengono studiati i flussi migratori che intercorrono tra le nazioni e le conseguenze economiche e politiche che ne derivano. I dati raccolti da questi enti sono utili per quantificare le dimensioni dei fenomeni migratori, ma fanno riferimento esclusivamente ai cosiddetti "paesi sviluppati" e mancano di informazioni relative al livello accademico e al settore professionale dei migranti, che generalmente devono essere estrapolate successivamente dai dati raccolti dalle singole nazioni. In Italia una fonte consolidata di dati è l'AIRE⁶, ovvero l'Anagrafe degli Italiani Residenti all'Estero, i cui dati, manchevoli di in-

³Rispetto al 2013 si è registrato un aumento del 41,8% nel numero di laureati che lasciano il paese [dei Conti, 2021].

⁴<https://www.worldbank.org/en/home>

⁵<https://www.oecd.org/>

⁶<https://www.esteri.it/it/servizi-consolari-e-visti/italiani-all-estero/aire0/>

formazioni sul settore professionale, possono essere implementati con quelli raccolti dall'ISTAT⁷, cioè l'Istituto Nazionale di Statistica, o con quelli del consorzio interuniversitario AlmaLaurea⁸.

Per superare questa carenza nelle fonti tradizionali, una nuova prospettiva sull'argomento è stata apportata dall'utilizzo di **dati non convenzionali**, come quelli ricavabili da fonti online. Un tale approccio è riscontrabile in alcuni studi innovativi dell'ultimo decennio, come [Bogdan et al., 2014] e [Zagheni and Weber, 2012], che utilizzano dati ricavati dai social network per studiare, analizzare e predire i fenomeni di mobilità internazionale. I vantaggi nell'utilizzo di questi dati rispetto a quelli tradizionali sono diversi:

- la reperibilità, essendo dati facilmente accessibili tramite internet e generalmente resi disponibili dagli stessi utenti che utilizzano le piattaforme social;
- l'utilizzo di tecnologie innovative, come le tecniche di geolocalizzazione, che permettono di definire con un elevato grado di accuratezza la posizione geografica dei vari soggetti;
- la copertura, data dall'ingente numero di utenti che ogni giorno accedono ai social network da ogni angolo del globo.

Questa tesi propone una metodologia di lavoro (d'ora in avanti *framework*) per il monitoraggio dello scambio culturale tra istituzioni. Gli obiettivi di questo framework sono principalmente due. Il primo è la creazione di un dataset strutturato e standardizzato di dati relativi alle carriere professionali degli alumni di una o più istituzioni fornite dall'utente. Il secondo, l'analisi del dataset ottenuto al fine di ottenere informazioni utili per lo studio dello scambio culturale tra istituzioni e paesi, come le dimensioni del fenomeno e i settori professionali più interessati dagli scambi.

In generale, il framework sfrutta le caratteristiche intrinseche del social network LinkedIn per la collezione di informazioni relative agli alumni. I dati così ottenuti hanno copertura globale e descrivono un panorama eterogeneo di

⁷<https://www.istat.it>

⁸<https://www.almalaurea.it/>

studenti e professionisti. Attraverso operazioni di preprocessing supervisionato, i dati collezionati vengono anonimizzati, riordinati cronologicamente e standardizzati grazie alla creazione di metodi di Elaborazione del Linguaggio Naturale (NLP). Sul dataset ottenuto vengono, infine, effettuate analisi su basi temporali, geografico-spaziali, professionali ed educative. Grazie all'ampia copertura dei dati, è possibile compiere studi su enti molto diversi e distanti tra loro. Inoltre, data la costante disponibilità di dati continuamente aggiornati, è possibile effettuare un monitoraggio nel tempo delle aree e degli istituti studiati.

Il framework proposto è stato applicato allo studio di circa 10.000 alumni e alumnae⁹ di cinque atenei toscani: Università degli Studi di Firenze, Università degli Studi di Siena, Università di Pisa, Scuola Normale Superiore e Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna. Questa analisi permette di osservare le connessioni degli atenei toscani nel mondo, sia da un punto di vista geografico che istituzionale, individuando gli enti (e relative nazioni) da cui provengono gli studenti iscritti e quelli verso cui vanno gli studenti una volta completato il percorso di studi. Al contempo permette di studiare possibili attrattività o punti critici delle offerte accademiche attraverso le analisi dei curriculum degli studenti per settore disciplinare.

I capitoli della tesi affrontano ciascuno un aspetto diverso di quanto riassunto fino ad ora. Nel Capitolo 2 viene introdotto lo stato dell'arte, che riassume gli studi più significativi sulla mobilità internazionale e sull'utilizzo di dati non convenzionali. Il Capitolo 3 presenta la risorsa LinkedIn, i dati che fornisce, i vantaggi e le sue limitazioni. Nel Capitolo 4 vengono illustrate le operazioni di raccolta dati e di preparazione effettuate, mentre nel Capitolo 5 vengono esposti i risultati ottenuti dall'analisi esplorativa dei dati e dallo studio della mobilità. Il Capitolo 6 espone le considerazioni finali e suggerisce eventuali sviluppi futuri e il Capitolo 7 riassume brevemente i risultati ottenuti. Infine, le Appendici A e B contengono alcune tabelle contenente i dati grezzi ottenuti dal nostro dataset.

⁹Con i termini "alumnus" e "alumna" ci si riferisce in questo lavoro all'insieme di studenti ed ex-studenti e di studentesse ed ex-studentesse di una data università.

2 || Stato dell'Arte

La prima definizione del fenomeno del brain drain avviene in Inghilterra, negli anni del dopoguerra. Ad inaugurare gli studi è la Royal Society che, sollecitata dalla corona inglese e dalla crescente preoccupazione sui risvolti politici del fenomeno, nel 1963 pubblica un report di analisi sulle partenze degli scienziati britannici che, con un trend in crescita dalla fine degli anni '50, lasciavano il paese in favore degli Stati Uniti d'America [Balmer et al., 2009]. Negli anni a seguire, diversi studi hanno contribuito a inquadrare il fenomeno globale del brain drain, cioè del trasferimento fuori dai confini nazionali dei soggetti altamente qualificati. Volendo schematizzare tutta la letteratura prodotta sull'argomento, è possibile individuare due principali filoni di pensiero.

Nei primi studi sull'argomento, anch'essi prodotti intorno agli anni '60, la prospettiva adottata è quella *nazionalista*. Questa prospettiva, presente ancora oggi in diversi studi, sarà a lungo quella dominante e per questo definita anche come "*standard view*" [Milio et al., 2012]. Secondo questa prospettiva il fenomeno del brain drain va inquadrato nell'ottica geopolitica della divisione tra Nord e Sud del mondo [Meyer, 2001], per cui questi movimenti partirebbero dai paesi meno sviluppati verso quelli più sviluppati in cerca di opportunità lavorative ed economiche migliori. L'attenzione è quindi rivolta principalmente alla perdita di risorse dei paesi di partenza, che formerebbero a proprie spese il capitale umano successivamente sfruttato da nazioni già sviluppate, e agli effetti socioeconomici di questa perdita [Bhagwati and Hamada, 1974]. Tra gli studiosi appartenenti a questa corrente di pensiero non vi è comune accordo sugli effetti del brain drain. Secondo la teoria neoclassica dello sviluppo economico adottata negli studi di Grubel e Scott, l'emigrazione alleggerirebbe la pressione sociale causata dalla disoccupazione e arricchirebbe indirettamente il paese di partenza attraverso le rimesse di denaro effettuate dall'emigrato [Grubel and Scott, 1966], come visto, ad esempio, nel caso della grande migrazione italiana del XIX secolo [Lutz, 1961]. Contrariamente, diversi studi empirici, tra cui [Bhagwati and Hamada, 1974] e [Rodriguez, 1975], evidenziano invece i costi per le nazioni di partenza, che soffrono del mancato ritorno degli investimenti fatti per la formazione di queste persone.

Alla fine degli anni '90, da alcuni studi basati sulle osservazioni di [Boulier, 1999] e [Rudolph and Hillmann, 1998], emerge una nuova prospettiva sull'argomento, una prospettiva *internazionale*. Questo filone, definito successivamente come "circolazionista" [Gaillard and Gaillard, 1997] [Johnson and Regets, 1998], porterà nuova luce sul fenomeno, evidenziandone specificità fino ad ora trascurate. Se nella "standard view" si aveva una focalizzazione sugli effetti delle migrazioni nei paesi di partenza, in questa prospettiva si evidenziano i benefici in termini di sviluppo globale: ad esempio, in [Milio et al., 2012] si sottolinea come la produzione di conoscenza nel settore scientifico produca valore in tutte le nazioni, a prescindere da quale sia quella in cui è stata prodotta. Il punto di partenza alla base di questa nuova prospettiva è che i movimenti delle persone altamente qualificate non sono unidirezionali, ma "policentrici, circolatori, temporanei e soggetti a fenomeni di scambio tra Paesi" [Milio et al., 2012]. Si inizia quindi a parlare di "brain circulation" e di "brain exchange", sostituendo alla connotazione geopolitica della standard view una di matrice sociologica, dove ad essere messe in risalto sono le connessioni internazionali. Sempre [Milio et al., 2012] suggerisce come i movimenti internazionali cambiano a seconda del settore in analisi: ogni nazione "esporta" e "importa" capitale umano in diversi settori, contribuendo in qualche modo ad un bilanciamento globale.

Diverse obiezioni sono state avanzate nel corso degli anni a questa visione. Se è vero che in diversi studi sono stati evidenziati interscambi di personale altamente formato tra paesi in via di sviluppo, si veda [Boulier, 1999], [Rudolph and Hillmann, 1998] o anche [Logan,], i dati dell'OECD riportano che più dell'85% dei flussi mondiali hanno come meta i paesi del mondo più sviluppati [Docquier and Marfouk, 2004]. I paesi con maggiore forza attrattiva sono Stati Uniti, Canada, Germania, Inghilterra e Australia. Tuttavia, è difficile fare una stima per tutti i paesi del mondo, dato che per molti non si hanno statistiche accurate sugli ingressi e le uscite dal paese [Milio et al., 2012].

Studi prodotti nell'ultimo decennio hanno criticato il carattere fortemente ideologico e "ottimista" di questa prospettiva [Bönisch-Brednich, 2016], che non tiene in considerazione le cause politiche, sociali ed economiche dei movimenti

internazionali, sia in relazione alle ragioni della partenza, sia in relazione alle ragioni della scelta dei paesi di approdo. Su queste premesse è stata recentemente proposta una nuova metodologia di studio del fenomeno delle migrazioni, chiamato “*aspiration-(cap)ability*”. In questo framework è prevista una fase preliminare di analisi delle motivazioni sottostanti gli spostamenti attraverso l’utilizzo di sondaggi, e, solo successivamente, una quantificazione del fenomeno migratorio.

2.1 | *Mobilità in Italia*

Il fenomeno del brain drain in Italia è strettamente correlato alla precaria situazione economica del paese. Fino al 2001, anno dello scoppio della crisi finanziaria che ha investito i paesi mediterranei, la situazione economica italiana è caratterizzata da una certa stazionarietà nel mercato del lavoro e da un aumento della disoccupazione giovanile [Minnecci, 2015], che si traduce con una crescita nel numero di partenze dei laureati già dalla fine degli anni ‘90, con una media di circa 3.000 partenze all’anno [Brandi and Avveduto, 2004]. All’inizio degli anni 2000, quindi alle porte della crisi finanziaria, circa il 7% dei laureati italiani lascia il paese. Il livello di drenaggio in Italia in questi anni si configura come medio-basso, in quanto al di sotto sia della media europea che di quella mondiale, influenzata da nazioni come i Caraibi, Haiti e Giamaica aventi un tasso di drenaggio dell’80% [Docquier and Marfouk, 2004].

Con lo scoppio della crisi finanziaria le partenze dall’Italia non possono che aumentare. Minnecci [Minnecci, 2015] stima che tra il 2000 e il 2010 siano partiti circa 30.000 giovani l’anno, ma il picco nella crescita del fenomeno si registra nel 2013, con un incremento del 21% del numero di partenze rispetto all’anno precedente. Nell’arco 2010-2015 si contano un totale di 430.813 partenze, di cui il 30% riguardano i possessori di un titolo di studio terziario [Boffo and Gagliardi, 2017].

Relativamente al fenomeno negli ultimi 6 anni non si hanno statistiche aggiornate, ma secondo quanto dichiarato dalla corte dei conti nel report di Marzo 2021, rispetto al 2013 si è registrato un aumento del 41,8% nel numero di laureati che lasciano il paese [dei Conti, 2021].

In generale, quindi, dagli anni 90 in poi si continua a registrare un aumento in percentuale del numero di persone altamente formate che lasciano il paese. In [Becker et al., 2004], Becker suggerisce che in media chi parte è generalmente più istruito di chi rimane. Dalle analisi fatte sui dati OECD relativi all'anno 2015 in [Brandi, 2017] si evince che le destinazioni principali degli spostamenti italiani sono Regno Unito, Francia, Germania e Stati Uniti d'America. Diversi studi concordano, tra cui la stessa Brandi e [Minnecci, 2015], nel ritenere che le ragioni dietro questi spostamenti vadano ricercate nelle migliori opportunità di ricerca offerte in queste nazioni, oltre che da un numero maggiore di investimenti fatti dallo stato al settore della ricerca. Specialmente nel Regno Unito, il settore accademico è in grande espansione e ciò comporta più possibilità lavorative e stipendi più alti rispetto alla media italiana. Un'altra meta italiana individuata in [Minnecci, 2015] è la Spagna, cui forza d'attrazione nei confronti dell'Italia è data principalmente dalla vicinanza culturale e dal recente sviluppo di Madrid e Barcellona.

Nonostante le mete siano generalmente ricche di opportunità lavorative e di situazioni economiche favorevoli, la situazione lavorativa degli emigrati italiani, ma non solo, è spesso precaria, con diseguaglianze di reddito e di opportunità [Pugliese, 2017].

La quasi totalità degli studi qui citati¹ convengono nel definire più problematica per l'Italia non la fuga annuale di giovani laureati, ma la mancanza di attrazione nei confronti di persone altamente qualificate provenienti dall'estero e quindi nell'incapacità di bilanciare i fenomeni di brain drain e brain gain. Secondo i dati OECD del 2005, solo il 12,2% degli stranieri in Italia ha una laurea o un titolo equivalente, contro la media del 18,6% degli altri paesi europei. Lo studio [Milio et al., 2012] suggerisce che ciò sia dovuto alla mancanza di politiche nazionali mirate a facilitare l'entrata di persone qualificate in Italia. Nella classifica stilata dal Business Intelligence Unit, l'Italia è la ventitreesima nazione per capacità di "attrarre e produrre talenti", al pari merito con la Grecia e al di sotto di tutte le altre nazioni dell'Europa Occidentale [Milio et al., 2012].

¹Tra cui [Balduzzi and Rosina, 2011], [Milio et al., 2012], [Boffo and Gagliardi, 2017], [Minnecci, 2015].

2.2 | *Studi con dati non convenzionali*

Una delle più grandi limitazioni allo studio della mobilità internazionale è la scarsa disponibilità di dati. Al giorno d'oggi non esiste una fonte univoca ed esaustiva di informazioni sui fenomeni migratori, ma diverse fonti, spesso collegate ai sistemi di censimento nazionali, i cui dati necessitano di essere integrati tra loro per studiare il fenomeno. Questi dati sono generalmente soggetti a politiche e restrizioni dei vari paesi, che possono decidere di renderli disponibili integralmente, parzialmente, dietro compenso o non renderli disponibili affatto. In questo contesto, uno studio in scala globale dei fenomeni migratori risulta particolarmente complicato, se non impossibile, mentre invece, secondo [Hui et al., 2010], sarebbe addirittura necessario per aiutarci a capire e prevenire la diffusione di epidemie.

Dagli anni 90, le fonti più autorevoli sul fenomeno sono la Banca Mondiale e l'Organizzazione per la Cooperazione e lo Sviluppo Economico [Boffo and Gagliardi, 2017], due istituzioni di carattere internazionale attive nel settore degli studi finanziari ed economici. Entrambi gli enti hanno costruito dei database per lo studio dei fenomeni migratori incrociando i dati dei censimenti di diversi paesi, utili a quantificare le dimensioni degli spostamenti, e dati provenienti da diverse fonti utili a stimare le percentuali di espatri di persone altamente qualificate [Beltrame, 2007]. Le informazioni presenti in questi database non fanno però riferimento a tutto il mondo: i dati dell'OECD si riferiscono esclusivamente ai 38 paesi OECD, escludendo di fatto le nazioni dell'Africa, dell'Asia (ad eccezione di Giappone e Turchia) e di parte del Sud America, comprendendo quindi solo l'85% circa di tutte le tratte mondiali [Docquier and Marfouk, 2004]; i dati della banca mondiale, invece, fanno riferimento a quelle sole nazioni che hanno sviluppato un proprio sistema di censimento, escludendo le nazioni in via di sviluppo che non hanno ancora strumenti di monitoraggio demografico. Inoltre, i dati raccolti dagli enti nazionali (come l'ISTAT e l'AIRE italiani) e ripresi in questi dataset fanno riferimento esclusivamente ad una piccola parte dei migranti, ovvero a quelli che decidono di spostare la propria residenza fuori dal loro paese d'origine, registrandosi quindi nelle liste dello stato ospitante.

I dati raccolti da questi enti sono puramente quantitativi e non sono disaggregati per settore professionale. Al momento, le informazioni sul settore professionale sono disponibili solo per gli Stati Uniti d'America, grazie all'operato della National Science Foundation, che attraverso il sistema SESTAT, Scientists and Engineers Statistical Data System, raccoglie informazioni dettagliate sugli emigrati altamente qualificati negli USA [Beltrame, 2007].

Nell'ultimo decennio, molti ricercatori hanno visto l'opportunità di superare questa tradizionale mancanza di dati nella diffusione, sempre più pervasiva, di dispositivi mobili sempre connessi ad internet e nella nascita del cosiddetto "web 2.0". Internet è infatti ormai diffuso in tutto il mondo. Secondo le stime di [We Are Social, 2021], a Marzo 2021 il 59,5% della popolazione mondiale è connessa ad internet e il 53,6% è attivo su almeno un social network. I dati prodotti dall'attività di queste persone su internet hanno un grande potenziale nello studio demografico grazie alla loro portata globale, all'essere generati e resi disponibili in tempo reale e alla precisione della geolocalizzazione [Bogdan et al., 2014].

Lo studio [Brockmann and Theis, 2008], ad esempio, ha utilizzato i dati di diversi siti di tracciamento di denaro, wheresgeorge.com e geocaching.com, per inferire pattern di mobilità internazionale. I dati del social network Twitter sono stati utilizzati in [Ferrari et al., 2011] per misurare gli spostamenti interni all'area metropolitana di New York e in [Zagheni et al., 2014] per analizzare i movimenti internazionali, evidenziando un calo degli spostamenti verso gli Stati Uniti, specialmente dal vicino Messico. In [Pultar and Raubal, 2009], invece, i dati del servizio online CouchSurfing sono stati utilizzati per cercare di ricostruire gli itinerari di un campione di turisti.

Uno studio sugli stock dei migranti è stato effettuato in [Zagheni and Weber, 2012] attraverso gli indirizzi IP di un grosso campione di email inviate tramite il servizio offerto da Yahoo!. Lo studio ha evidenziato come il crollo delle migrazioni verso gli Stati Uniti si concentrasse principalmente nei primi anni del nuovo millennio, probabilmente a causa del collasso della "dot-com bubble" nel 2000 e alla crisi economica del 2008, e come invece il tasso di migrazioni sia ripreso a salire a partire dal 2010, soprattutto relativamente alla fascia d'età compresa tra i 25 e i 35 anni e

più tra le persone di sesso femminile che in quelle di sesso maschile.

Una fonte particolarmente interessante per lo studio della mobilità internazionale e, in particolar modo, del fenomeno della brain circulation è LinkedIn². La peculiarità di questo professional network è quella di contenere al suo interno dati professionali e accademici di milioni di utenti, caratterizzati sia da un punto di vista cronologico che geografico, cioè riportando per ogni informazione la data e il luogo di riferimento. I dati ottenibili risultano quindi disgregati dal punto di vista settoriale, geografico e temporale, hanno una copertura globale, sono relativamente poco costosi e sono sempre disponibili [Johnson et al.,].

Una particolarità di LinkedIn, evidenziata sia in [Mas-Bleda et al., 2014] che in [Baruffaldi et al., 2017], è quella di riuscire ad attirare come utenza molte persone altamente qualificate, fascia generalmente poco attiva nei social network. Il lavoro [Mas-Bleda et al., 2014] ha evidenziato come LinkedIn sia il social network più utilizzato dagli “European Highly Cited Scientist”, in sigla “EHC”, ovvero gli scienziati europei più citati nelle pubblicazioni. Questa categoria è generalmente caratterizzata da un basso utilizzo delle diverse piattaforme social, ma tra queste la più utilizzata risulta appunto essere LinkedIn, superando piattaforme di fatto più specializzate nel settore, come *Academia.edu* e *Google Scholar*. Nella ricerca del 2017 [Baruffaldi et al., 2017] viene invece indagato l’utilizzo della piattaforma da parte dei possessori di un titolo di dottorato. L’analisi ha evidenziato come ad usare maggiormente il professional network siano quei ricercatori con molte collaborazioni internazionali o quelli impiegati presso enti privati. Secondo gli autori ciò suggerirebbe l’importanza di LinkedIn come strumento per la ricerca di opportunità di carriera all’infuori dall’ambiente universitario.

Lo studio [Bogdan et al., 2014] utilizza i dati di LinkedIn per quantificare gli spostamenti di persone altamente qualificate verso gli Stati Uniti negli anni compresi tra il 1990 e il 2012. Attraverso la suddivisione dei dati ottenuti per livello accademico, l’analisi ha evidenziato come nella prima decade degli anni 2000 si registri una diminuzione degli ingressi di persone altamente formate nel

²Informazioni più specifiche su LinkedIn e sul suo funzionamento sono riportati nel Capitolo 3.

paese: la percentuale di possessori di un titolo di istruzione terziaria di primo livello che entrano negli USA crolla dal 33% dell'anno 2000 al 17% del 2012; per i possessori di un titolo di istruzione terziaria di secondo livello, la percentuale scende dal 27% al 12%; per i possessori di un dottorato, dal 29% al 18%. Il settore più colpito è quello STEM³, i cui migranti passano da essere il 37% sul campione complessivo al 15%, con un calo di 22 punti percentuali, contro un calo di solo 12 punti per il settore non-STEM. Lo stesso studio evidenzia come in quegli anni si registri un calo del fenomeno del brain gain anche in Europa e Canada, mentre aumentano i flussi verso l'Australia, l'Africa, il Sud America e, soprattutto, l'Asia, che nel 2012 ha ricevuto il 25% dei flussi migratori di persone altamente formate di tutto il mondo.

Nello studio [Johnson et al.,] vengono analizzati i profili di LinkedIn contrassegnati come “aperti a opportunità all'estero”, cioè i profili di quegli utenti potenzialmente interessati a spostarsi fuori dal paese d'origine. I risultati ottenuti hanno evidenziato come molti utenti europei siano disponibili a spostarsi all'interno del continente e come molte nazioni esercitino tra loro una attrazione mutuale, motivata probabilmente da ragioni culturali, come per Spagna e Italia o Danimarca e Svezia. Relativamente ai dati collezionati, ottenuti tramite web-scraping delle informazioni mostrate dalla piattaforma di LinkedIn ai possessori di un abbonamento Recruiter, gli autori suggeriscono come questi non siano rappresentativi di tutta la popolazione, ma siano condizionati dal “Digital Divide”.

Come descritto in precedenza, analogamente a quanto fatto in [Bogdan et al., 2014] e [Johnson et al.,], in questo lavoro utilizzeremo i dati di LinkedIn per collezionare un dataset sugli alumni e le alumnae di alcuni atenei toscani, caratterizzato dal punto di vista del livello accademico e del settore professionale. Attraverso le indicazioni riportate dagli alumni sul proprio profilo su esperienze accademiche e professionali svolte e alla località in cui queste sono avvenute, sono state studiate le loro caratteristiche e descritte le loro carriere, con particolare attenzione agli spostamenti internazionali.

³“Science, Technology, Engineering and Mathematics”, ovvero Scienza, Tecnologica, Ingegneria e Matematica.

3 || LinkedIn

Ideato nel 2002 da Reid Hofferman e ufficialmente lanciato nel mercato il 5 Maggio del 2003, Linkiedin, con 774 milioni di utenti sparsi in più 200 nazioni, è attualmente il più grande *professional network* del mondo[[Linkedin, 2021a](#)]¹. Con “professional network” si intende una tipologia di social network incentrata su relazioni di tipo lavorativo, differentemente da social network come Facebook, ad esempio, che si basano invece sulle relazioni di tipo personale. Altri esempi di professional network sono il francese *Viadeo* e il tedesco *Xing*, diffusi principalmente nelle rispettive aree nazionali, o *Academia.edu*, molto utilizzato in ambito accademico.

La filosofia dietro questo genere di piattaforme rimanda alla **teoria dei sei gradi di separazione** del sociologo Granovetter [[Granovetter, 1973](#)], secondo la quale i cosiddetti “*legami deboli*”, cioè i contatti che abbiamo al di fuori della cerchia di parenti e di amici intimi, sarebbero fondamentali nella diffusione delle informazioni. È attraverso questi contatti che noi riusciamo a collegarci con realtà e gruppi sociali diversi dai nostri e ottenere informazioni che altrimenti impiegherebbero molto più tempo per raggiungerci. Secondo questa stessa teoria, l’attività sociale favorisce il mercato del lavoro, permettendoci di ottenere più facilmente informazioni e aggiornamenti relativi al nostro settore professionale. Secondo lo studio [[Granovetter, 2018](#)] le metodologie informali di ricerca di lavoro, come l’utilizzo di contatti personali, sono generalmente più fruttuose rispetto a quelle formali, come ad esempio l’affidarsi ad agenzie per l’impiego. Le piattaforme di professional network si presentano quindi agli utenti come uno spazio in cui è possibile recuperare e condividere informazioni lavorative, nonché dove ampliare la propria rete professionale.

Dato questo stretto collegamento con il mondo del lavoro, i profili degli utenti sono generalmente strutturati alla stregua di un curriculum vitae, riportando dati sia sui traguardi accademici raggiunti, che sulle esperienze lavorative pregresse e attuali. Una volta aggregati, questi dati possono essere utilizzati per disegnare una immagine rappresentative di diversi gruppi sociali, come potrebbero essere gli abitanti di una determinata area geografica o, come nel nostro caso,

¹I dati fanno riferimento a ottobre 2021.

gli alumni e le alumnae di una università.

La scelta di utilizzare LinkedIn come piattaforma per la raccolta dei dati si motiva, dunque, nella presenza all'interno del social network di tali informazioni professionali e nell'ampia diffusione a livello globale raggiunta dalla piattaforma. Secondo i dati pubblicati in *"LinkedIn's SelfService Advertising Tool"* [LinkedIn, 2021b] e ripresi in *"Digital 2021"* [We Are Social, 2021], a livello geografico la piattaforma è utilizzata in più di 200 nazioni, con principali poli di affluenza provenienti dalle Americhe e dall'Europa Occidentale. Relativamente all'Asia, vi è una considerevole diffusione del social network in India e in Cina, mentre risulta meno presente in Russia e Giappone. I 15 milioni di italiani iscritti fanno della nostra nazione la decima per presenza sul social network: LinkedIn è infatti l'ottavo social media più utilizzato nello stivale tra gli internauti compresi tra i 16 e i 64 anni di età.

La diffusione di LinkedIn sembrerebbe essere tuttora in crescita: dall'inizio della emergenza sanitaria data dalla pandemia di Covid-19, LinkedIn è passato dai 675 milioni di utenti globali nel gennaio 2020, ai 774 milioni di settembre 2021, evidenziando dunque una crescita degli utenti globali del circa 15% [We Are Social, 2021].

L'utenza media della piattaforma è composta prevalentemente da giovani nella fascia di età compresa tra i 25 e i 34 anni, principalmente di sesso maschile. L'analisi condotta da *We Are Social* riporta che il 22,6% degli intervistati ha dichiarato di usare il social network principalmente per ragioni professionali e che circa il 20% degli account seguiti sono contatti di lavoro o aziende rilevanti nel proprio settore [We Are Social, 2021].

Per quanto riguarda la gestione dei dati, le politiche dell'azienda sono molto stringenti. In generale, i dati relativi ai singoli utenti sono visionabili semplicemente con la creazione di un profilo personale. Tuttavia, in alcuni casi si potrebbe incorrere in alcune limitazioni dovute alla "distanza" in termini di rete tra il nostro profilo e quello dell'utente che si desidera visionare. La visibilità dei vari profili e dei contenuti di questi è regolamentata secondo il concetto di **"grado di connessione"**, ovvero della distanza, espressa in "amicizie in comune", tra due profili. Fanno parte della rete personale di un utente i profili fino ad una distanza mas-

sima di 3 gradi². I profili al di fuori di questa rete vengono nascosti, parzialmente o integralmente, dalla piattaforma. Questo limite è oltrepassabile attraverso la sottoscrizione ad uno degli abbonamenti a pagamento.

I dati aggregati sono inoltre disponibili attraverso un servizio di *API*³ che LinkedIn mette a disposizione a sviluppatori e ricercatori, ma soltanto previa autorizzazione.

Le informazioni personali visionabili direttamente dalla pagina profilo di un utente di LinkedIn sono diversi e raggruppabili in diverse categorie. Sebbene questa lista non sia totalmente esaustiva, in quanto ogni pagina profilo è altamente personalizzabile, le principali informazioni recuperabili sono:

- **Dati biografici**, come il nome e la posizione geografica.
- **Dati accademici**, relativi ai corsi di studio intrapresi dall'utente. Informazioni, quindi, su cosa ha studiato, quando e dove.
- **Dati professionali**, cioè dove ha lavorato, per quanto tempo e presso chi.
- **Dati sulle competenze**, cioè le abilità personali dell'utente spendibili nel mondo del lavoro (ad esempio saper programmare, saper utilizzare un certo software o parlare più lingue).
- **Dati relazionali**, gli utenti con cui interagisce sul social network.
- **Dati sui contenuti**, ovvero i contenuti che pubblica, le pagine che segue e gli articoli con cui interagisce.

²Ovvero gli amici degli amici di amici.

³Application Programming Interface. Maggiori informazioni al sito <https://developer.linkedin.com/product-catalog>

4 || Metodologia

Questo capitolo descrive la metodologia sviluppata in questa tesi per l'analisi statistica delle carriere degli studenti e delle studentesse che gravitano intorno ad un determinato ateneo. Ogni sezione illustrerà una delle fasi costituenti del framework e ne spiegherà brevemente il funzionamento, fornendo alcuni esempi laddove necessario.

Le fasi principali del nostro framework sono tre:

- la fase di **raccolta dei dati**;
- la fase di *preprocessing*, quindi di **preparazione dei dati**;
- la fase di **analisi**.

Le prime due fasi del framework servono alla costruzione di un dataset sulle esperienze accademiche e lavorative delle *alumnae* e degli *alumni* di una determinata università e alla preparazione di questo in vista delle successive operazioni di analisi.

L'algoritmo utilizzato e i dati ottenuti sono disponibili su GitHub all'indirizzo <https://github.com/GianmarcoDiMauro/Framework-LinkedInData>.

4.1 | *Raccolta dati*

L'operazione di raccolta dei dati si articola in due fasi. In un primo momento vi è la ricerca e la selezione all'interno di LinkedIn dei profili utente utili ai fini dello studio. Questa selezione è fatta indicando al motore di ricerca implementato nella piattaforma le istituzioni universitarie scelte, ottenendo così i nominativi degli utenti che vi hanno studiato. Da questo processo si ottiene una lista dei nominativi e dei rispettivi link ai profili utente, dalla quale in un secondo momento verranno estratti i dati di nostro interesse che saranno poi memorizzati in locale. Un esempio dei dati estratti dalla piattaforma è mostrato in Figura 4.1. Le informazioni recuperate da questo procedimento riguardano:

- Il nominativo dello studente, oscurato per motivi legati alla privacy.¹

¹Informazioni aggiuntive rispetto alle tecniche di anonimizzazione adottate sono disponibili al paragrafo 4.2.1

- La località di riferimento per il profilo.
- I percorsi d'istruzione.
- Le esperienze professionali.
- L'URL al profilo.

LinkedIn fornisce quattro tipi di valori relativi ai percorsi d'istruzione: la descrizione generale del percorso di studi, l'istituto che l'ha ospitato, la data di inizio e la data di fine del percorso. I dati sulle esperienze professionali comprendono: la mansione svolta, la compagnia presso cui si è lavorato, la durata dell'esperienza, la località dove si è lavorato, la data di inizio e la data di fine. Tutti questi dati vengono collezionati come stringhe di testo. Unica eccezione è la descrizione del percorso di studi, che invece viene memorizzata come una lista, contenente all'interno stringhe di testo.

```
{
  "Name": "          ",
  "Location": "Siena, Tuscany, Italy",
  "Experiences": [
    {
      "position_title": "Local exchange officer",
      "company": "AISO - Italian Association for Dental Students Full-time",
      "duration": "2 yrs 2 mos",
      "location": "Siena, Tuscany, Italy",
      "from_date": "2019",
      "to_date": "",
      "url": null
    }
  ],
  "Education": [
    {
      "from_date": "2017",
      "to_date": "2023",
      "university": "Universit\u00e0 degli Studi di Siena",
      "degree": [
        "Master's degree",
        "Dentistry"
      ],
      "url": "https://www.linkedin.com/school/13890/?legacySchoolId=13890"
    }
  ],
  "Linkedin_url": "https://www.linkedin.com/in/abdalrauf-kaddour-485aa914b?miniProfileUrn=urn%3Ali%3An",
  "Contact_url": []
}
```

FIGURE 4.1: Esempio di dati relativi ad un profilo utente

4.2 | *Preparazione dati*

Una volta aggregati i dati si presentano come in Figura 4.2 e in Figura 4.3, dove ad ogni riga corrisponde un **record**, cioè un determinato studente, mentre per ogni colonna abbiamo un **attributo** diverso, cioè una specifica tipologia di informazione.

Così ordinati in attributi, i valori vengono “ripuliti” in modo tale da facilitare il processo di analisi statistica. Le linee guida dietro questi passaggi di preparazione dei dati sono:

- **Pseudonimizzazione**, per tutelare la privacy degli studenti.
- **Standardizzazione**, per ricondurre i valori di ogni attributo ad un modello univoco.
- **Organizzazione dei dati**, cioè ordinamento dei dati in macro-categorie.
- **Aggiunta di informazioni** altrimenti assenti nel dataset.

Mentre le operazioni di pseudonimizzazione, organizzazione e di aggiunta di informazioni sono eseguite in maniera automatica, le operazioni di standardizzazione sono tutte eseguite in maniera semi-supervisionata e la scelta di eseguirle o meno è lasciata all’utente, a causa dell’ingente quantitativo di tempo richiesto. Come approfondito in 4.2.2, in 4.2.3 e in 7, si è scelta questa implementazione per evitare che vengano fatte associazioni erranee a causa della somiglianza lessicale di alcuni valori.

Questo paragrafo spiega la logica dietro le trasformazioni effettuate sui vari attributi. Per facilitare la comprensione, l’argomento verrà suddiviso in cinque sotto-paragrafi: la pseudonimizzazione; gli attributi geografici; gli attributi relativi ai percorsi d’istruzione; gli attributi professionali; la mobilità internazionale.

4.2.1 - *Dati sensibili: il nome utente e la pseudonimizzazione*

L’attributo “Name” rappresenta il nominativo scelto dallo studente sulla piattaforma. Nella maggior parte dei casi il nominativo scelto sembrerebbe poter coincidere

Name	Location	Dg1	Dg1University	Dg1FromDate	Dg1ToDate	Dg2	Dg2University	Dg2FromDate	
0	0	Puchong, Selangor, Malaysia	[]	Asia Pacific University of Technology and Inno...	None	None	['MBA', 'International business', 'portrait an...]	Edith Cowan University	2006
1	1	Milan, Lombardy, Italy	['Modern Literature', 'Modern Literature and C...]	Università di Bologna	1990	1996	['Master', 'Media Science and Technology']	Istituto Universitario di Studi Superiori, Pavia	1999
2	2	Pavia, Lombardy, Italy	['Bachelor's Degree', 'Civil Engineering']	Kocaeli Üniversitesi	2009	2014	['Master's Degree', 'Earthquake Engineering']	Boğaziçi University	2014

FIGURE 4.2: Screenshot dati aggregati pt.1

PT1	PT1Company	PT1Duration	PT1Location	PT1FromDate	PT1ToDate	PT1Url	PT2	PT2Company	PT2Duration
VP IM Architect	AmBank Berhad	6 yrs 10 mos	kuala lumpur	2009	2016	None	Ent. Data Architect / VP	CIMB	2 yrs 5 mos
Freelance Contributor	Forbes Italia	2 yrs	Milan Area, Italy	2018	2020	None	Freelance Contributor	MF Milano Finanza	2 yrs
Junior Structural Engineer	OTS Engineering and Consultancy Ltd. Co.	4 mos	Istanbul, Turkey	2013	2013	None	Structural Engineering Intern	Emay Engineering and Consulting	2 mos

FIGURE 4.3: Screenshot dati aggregati pt.2

con il nome reale dello studente, sebbene accanto a questo, secondo la nostra esperienza, possano comparire titoli professionali, come “Dott.re” o “Avv.” o anche *emoticon* e simboli testuali.

Al fine di assicurare la privacy degli studenti e a causa della presenza di informazioni potenzialmente sensibili, all’interno del nostro lavoro tutti i record sono stati anonimizzati, sostituendo il nome dello studente con un numero identificativo univoco. La corrispondenza tra l’identificativo e il nome dello studente è stata quindi registrata separatamente dal dataset.

In termini strettamente tecnici, l’operazione eseguita rientra nel campo della **pseudonimizzazione**. Secondo l’articolo 4 della **GDPR**², ovvero la regolamentazione generale sulla protezione dei dati approvata dal parlamento europeo

²General Data Protection Regulation

nel 2016, per pseudonimizzazione si intendono quei processi di trattamento dei dati finalizzati a impedire il collegamento tra i dati posseduti e la persona fisica a cui essi appartengono senza l'utilizzo di informazioni aggiuntive, si distingue dall'anonimizzazione in quanto quest'ultima prevede la completa impossibilità di risalire alla persona fisica [Parlamento europeo e Consiglio europeo, 2016].

4.2.2 - Trasformazione dei dati geografici

Per le informazioni geografiche ottenute relativamente ad un singolo profilo è necessario fare una distinzione: la località che nel dataset è contenuta all'interno dell'attributo "Location" è "generica", è cioè una indicazione da parte dello studente sulla posizione geografica di riferimento per il profilo; le altre località presenti sono invece "specifiche" e si riferiscono invece ai luoghi in cui sono state svolte le varie esperienze riportate dallo studente.

I possibili valori che otteniamo da LinkedIn relativamente alla località generica del profilo possono presentarsi seguendo due modelli ricorrenti. Ciò è dato dal fatto che all'interno del social network la scelta della località di riferimento del profilo è *controllata*, cioè l'utente sceglie la propria località a partire da una serie di opzioni già standardizzate presentate da LinkedIn.

Il primo modello presenta le informazioni geografiche in maniera esaustiva e completa, seguendo lo schema "**Città, Stato, Nazione**", dove con "stato" intendiamo il primo livello di suddivisione del territorio di una nazione, cioè quello che in Italia definiremmo come "regione". Alcuni esempi sono mostrati in Figura 4.2, dove abbiamo come valori di *Location*:

- "Puchong, Selangor, Malaysia";
- "Milan, Lombardy, Italy";
- "Pavia, Lombardy, Italy".

Alternativamente, questo modello può presentarsi in maniera parziale, riportando solo alcune delle informazioni. Nel dettaglio, sono state individuate le forme "**Nazione**" e "**Stato, Nazione**". Alcuni esempi sono:

- “South Korea”³;
- “Lombardy, Italy”⁴;

Il secondo modello si presenta come “**Nome area metropolitana**” ed è quindi costituito solo dal nominativo dell’area metropolitana e non fornisce indicazioni relativi alla nazione o allo stato, risultando quindi meno esaustivo. Eventualmente, il nominativo può riferirsi non tanto ad una effettiva città metropolitana, ma più in generale all’*hinterland* di una città. Due esempi relativi a questo secondo modello sono:

- “Greater Milan Metropolitan Area”⁵;
- “Greater Viareggio Metropolitan Area”⁶;

dove nel primo caso viene indicata una reale città metropolitana, mentre nel secondo ci si riferisce più probabilmente all’*hinterland* del comune toscano di Viareggio, non rientrando questo nella regolamentazione per le città metropolitane.

Diversamente da quanto visto fino ad ora, i dati geografici relativi alle esperienze professionali non seguono degli standard. Questo perché all’interno di LinkedIn il campo relativo alle esperienze professionali non è *controllato* come il precedente, ma è *semi-controllato*: cioè l’utente può sia scegliere la località da una lista di valori standardizzati sia inserirla manualmente in un campo libero. Questa tipologia di dati è quindi molto sporca ed è possibile trovare all’interno sia valori che seguono i due modelli indicati precedentemente, sia valori che non seguono alcuno schema. Un esempio è mostrato in Figura 4.3, dove il terzo record ha per valore di “PT1Location”:

- “Istanbul, Turkey”.

³Valore dell’attributo *Location* nel record con identificativo 16.

⁴Valore dell’attributo *Location* nel record con identificativo 500.

⁵Valore dell’attributo *Location* nel record con identificativo 174.

⁶Valore dell’attributo *Location* nel record con identificativo 4876.

Questo valore segue una schema “città, nazione”, che non è stato riscontrato tra i valori dell’attributo Location. Un altro esempio di informazioni non standardizzate è:

- “Concordia Sulla Secchia, San Prospero, San Possidonio”⁷;

in cui appaiono tre diversi comuni italiani elencati consecutivamente.

Per tutte le informazioni geografiche raccolte si è deciso di mantenere esclusivamente le informazioni relative alla nazione, sostituendo e talora rimuovendo le altre informazioni aggiuntive su stato e città. Lo standard che si è deciso di adottare relativamente a questa tipologia di dati è l’utilizzo della lingua inglese per i nomi delle nazioni e l’utilizzo della maiuscola per il primo carattere del nome.

Per apportare queste modifiche, l’algoritmo sviluppato estrae dalla stringa iniziale l’ultima porzione di testo, che spesso coincide con il riferimento alla nazione. Questa viene controllata attraverso un confronto con una lista contenente i nomi standardizzati di tutte le nazioni. Il confronto avviene tramite l’utilizzo della libreria “Fuzzywuzzy”⁸, che utilizza la “distanza di Levenshtein” per individuare il valore presente nella lista più simile alla porzione di testo ottenuta dalla stringa di partenza. Se il grado di similitudine tra le due supera il 96%, si assume che la porzione di testo ottenuta indichi effettivamente il nominativo di una nazione e che la forma standardizzata del nominativo di questa nazione sia quello individuato da Fuzzywuzzy nella lista.

Qualora il grado di similitudine non superasse la suddetta soglia, si suppone che la porzione di testo estratta non contenga il riferimento alla nazione. In questo caso, l’algoritmo confronta l’intera stringa di partenza con un *dizionario* appositamente creato, cui *chiavi* sono i nominativi di alcune città ed aree metropolitane di tutto il mondo, mentre i *valori* associati a queste sono i nomi standardizzati delle relative nazioni. Se la stringa di partenza corrisponde ad una delle chiavi del dizionario, questa viene sostituita nel dataset con il valore corrispondente. Nel caso in cui anche questo confronto fosse negativo, l’algoritmo non

⁷Valore dell’attributo *PT1Location* nel record con identificativo 108.

⁸Fuzzywuzzy: <https://github.com/seatgeek/fuzzywuzzy>

è capace di risalire alla nazione di riferimento in maniera automatica. Allora, se il dato fa riferimento alla località generica del profilo, l'algoritmo stampa a schermo la stringa per cui non è riuscito a risalire alla nazione e permette la modifica manuale del dato. Se il testo inserito corrisponde al nome standardizzato di una nazione, l'algoritmo sostituisce il valore ed aggiorna il dizionario in modo che l'associazione fatta resti memorizzata⁹.

Se l'algoritmo non riesce a risalire alla nazione di riferimento né in maniera automatica né, se previsto, in maniera manuale, l'algoritmo sostituisce alla stringa di partenza un valore vuoto, al fine di evitare errori nel momento dell'analisi dei fenomeni di mobilità internazionale.

Un'ulteriore modifica relativa ai dati geografici riguarda l'aggiunta della nazione in cui si è seguito un determinato corso di studi. L'algoritmo assegna ad ogni percorso d'istruzione indicato una nazione attraverso l'utilizzo di un dizionario appositamente creato, che associa ai nomi dei vari istituti la nazione in cui questi sono situati. Se il nome dell'istituto in cui si è svolto il percorso è presente all'interno del dizionario, l'informazione relativa alla nazione viene aggiunta ai dati sul corso di studi.

4.2.3 - Trasformazioni dei dati sull'istruzione

Come anticipato, la descrizione del percorso di studi si presenta come una lista contenente un numero variabile di stringhe. Le informazioni contenute in essa riguardano principalmente l'argomento del corso di studi e la tipologia di percorso seguito, ma a queste se ne aggiungono molte altre che possono variare sensibilmente da dato a dato. In linea generale, queste descrizioni seguono il modello “[**Tipologia di percorso**’, **Nome corso di studi**’, **Altre informazioni**]’”. Tuttavia questo modello è soltanto una generalizzazione di quanto presente concretamente nei dati, in quanto la struttura di queste descrizioni varia molto sia

⁹Questa implementazione è stata effettuata soltanto sui dati relativi alla località generale del profilo perché, come scritto precedentemente, già standardizzati secondo delle norme interne a LinkedIn. Essendo gli altri dati geografici non standardizzati, e quindi molto sporchi, la gestione in manuale di questi dati richiederebbe da parte di colui che esegue il codice uno sforzo e una quantità di tempo maggiore.

nell'ordine con cui presenta le informazioni, sia nelle tipologie delle informazioni stesse. Alcuni esempi sono:

- “[‘Bachelor’s degree’, ‘Computer engineering’, ‘100/110’]”¹⁰;
- “[‘Modern Literature’, ‘Modern Literature and Critical Theory Media Science and Technology’]”¹¹;
- “[‘Laurea breve in Scienze della Comunicazione’, ‘-’]”¹².

Come si evince, i tre esempi proposti sono infatti molto diversi tra loro. Se il primo ricalca fedelmente il modello proposto, non è possibile dire lo stesso per il secondo e per il terzo. Nel secondo, manca la tipologia di corso seguito ed è presente soltanto il nome del percorso accademico e una descrizione degli argomenti studiati. Nel terzo, invece, vengono presentate entrambe le informazioni principali, ma sono inserite all'interno della stessa stringa e ne è poi presente una seconda vuota.

Le informazioni non sono quindi ordinate secondo uno schema preciso e dagli esempi è possibile notare come non siano neanche uniformi dal punto di vista della terminologia: ne è un esempio la presenza all'interno del dataset di descrizioni in lingue diverse, ma anche la presenza di nomenclature differenti come “laurea breve” e “laurea triennale”.

Per avere una rappresentazione più strutturata su questi dati, si è creato un sistema di categorie e parole chiave: cioè si sono create delle categorie, rispettivamente per la tipologia del corso di studi e per la materia di studio, e si sono associate a queste delle parole chiave che le contraddistinguessero.

Il nostro algoritmo ricerca per ogni stringa all'interno della lista queste parole chiave e deduce le categorie più rappresentative per il percorso di studi dalla presenza o meno di queste. La ricerca viene effettuata una stringa alla volta e viene interrotta nel caso venisse trovata una corrispondenza. Ciò perché, come visto nel modello generale, le informazioni principali sono generalmente contenute nella

¹⁰Valore dell'attributo Dg1 nel record con identificativo 96.

¹¹Valore dell'attributo Dg1 nel record con identificativo 1.

¹²Valore dell'attributo Dg1 nel record con identificativo 12768.

Livello Accademico	Parole Chiave
High School	liceo, high school, istituto tecnico, itis, maturità
Bachelor	bachelor, laurea breve, triennale, primo livello, primo ciclo
Master	master, laurea magistrale, laurea ciclo unico, secondo livello, secondo ciclo
Ph.D	ph.d, phd, dottorato, doctor of philosophy

TABLE 4.1: Esempi di parole chiave utilizzate per individuare il livello accademico

prima e nella seconda stringa e le informazioni successive potrebbero potenzialmente portare a degli errori nel caso non si stabilisse questo genere di gerarchia.

È stato effettuato un controllo su un campione di 100 esperienze accademiche, facendo valutare la classificazione fatta dall’algoritmo da alcuni lettori umani, per osservare in che percentuali le categorie scelte dal nostro algoritmo potessero considerarsi appropriate. Relativamente al livello accademico, nell’ 88% dei casi, la classificazione effettuata dall’algoritmo è stata valutata come adeguata: nel 56% i tester hanno indicato la scelta fatta dall’algoritmo come corretta; mentre nel 32% i tester hanno confermato l’impossibilità di risalire alla tipologia del percorso di studi dalla descrizione fatta dallo studente. Nel restante 12% l’algoritmo non ha indicato alcun livello accademico, sebbene secondo i tester le informazioni presenti nella descrizione fossero sufficienti per risalire alla tipologia. Per il campo di studi, nel 52% dei casi la classificazione effettuata è stata valutata come coerente con quanto indicato nel record, nel 16% come erronea e nel 10% non erano invece presenti indicazioni relative al campo di studi. Nel restante 22% la classificazione effettuata è stata ritenuta “non valutabile” in quanto riferita a percorsi accademici interdisciplinari.

La Tabella 4.1 mostra le quattro categorie scelte per rappresentare i percorsi di studi all’interno dei dati raccolti e alcune delle parole chiave ad esse collegate. Le quattro categorie sono:

- **High School:** la scuola secondaria di secondo grado, la maturità.
- **Bachelor:** la laurea triennale, altrimenti detta laurea breve.
- **Master:** ovvero il secondo livello di istruzione universitaria, quindi laurea

Campo Studi	Esempi Parole Chiave
Agrarian Studies	natura, ambiente, enologia
Architecture	architettura
Economics	economia, finanza, business, marketing
Engineering	ingegneria, robotica, elettronica
Health Profession	infermieristica, fisioterapia, exercise science
Humanities	comunicazione, giornalismo, art, moda
Informatics	informatica, computer science, data science
Law	law, diritto, giurisprudenza,
Literatures, Languages and Cultures	letteratura, filologia, linguistica, lingue
Math and Physic	matematica, math, fisica
Medicine	medicina, chirurgia, cardiologia, surgery
Natural Science	biologia, chimica, geologia
Pharmacology	farmacia, pharma
Psicology	psicologia
Science Education	formazione, educazione, pedagogia
Social Science	sociologia, antropologia, pace
Story and Philosophy	story, filosofia, archeologia
Veterinary Studies	veterinaria, fauna

TABLE 4.2: Esempi parole chiave utilizzate per individuare l'argomento degli studi

specialistica e lauree magistrali o a ciclo unico. A queste si aggiungono i master di I e II livello.

- **Ph.D:** il dottorato

La Tabella 4.2 mostra le 18 categorie individuate relativamente al campo di studi e alcuni esempi di parole chiave utilizzate per capire se il percorso di studi fatto dall'utente rientri in quello specifico settore.

I campi di studio individuati sono¹³:

- **“Agrarian Studies”**, Agraria.
- **“Architecture”**, Architettura;

¹³Vogliamo sottolineare come questa classificazione sia inevitabilmente esemplificativa, in quanto mirata a proporre uno schema riassuntivo dei diversi ambiti di studio presenti all'interno del mondo accademico. In alcuni casi si è scelto di raggruppare alcuni campi di studio che, a nostro parere, possono essere considerati affini, sebbene poi indiscutibilmente distinti. Inoltre la classificazione non tiene conto dell'esistenza di interi settori dal carattere interdisciplinare, che si collocano quindi a metà tra due, o più, delle categorie proposte.

- **“Economics”**, Economia;
- **“Engineering”**, Ingegneria;
- **“Health Profession”**, Infermieristica e le altre professioni sanitarie;
- **“Humanities”**, che racchiude diverse realtà accademiche, come Editoria, Giornalismo, Scienze dello spettacolo e la stessa Informatica Umanistica;
- **“Informatics”**, Informatica e le sue branche;
- **“Law”**, ovvero Giurisprudenza;
- **“Literatures, Languages and Culture”**, Lingue e Lettere, con le rispettive branche;
- **“Math and Physic”**, Matematica e Fisica;
- **“Medicine”**, Medicina e le sue specializzazioni;
- **“Natural Science”**, le scienze naturali, quindi Biologia, Chimica, Geologia;
- **“Pharmacology”**, Farmacia;
- **“Psicology”**, Psicologia;
- **“Science Education”**, Scienza della Formazione;
- **“Social Science”**, che racchiude corsi di studi come Sociologia, Scienze per la Pace, ecc. ;
- **“Story and Philosophy”**, Storia e Filosofia;
- **“Veterinary Studies”**, Veterinaria;

Le categorie proposte si riferiscono volutamente ai titoli e ai campi di studio di carattere prettamente accademico. In realtà, all'interno di LinkedIn è possibile trovare informazioni relative anche a esperienze di studio esterne al mondo universitario. Ai fini del nostro progetto, che si basa sull'analisi delle carriere delle persone altamente formate, si è deciso di concentrarsi esclusivamente sul

settore accademico. Per questa ragione, le operazioni di trasformazione dei nomi degli istituti ospitanti i corsi di studio si sono concentrate quasi esclusivamente sull'uniformare la nomenclatura delle università e non quelle degli altri enti.

All'interno di LinkedIn, il campo in cui indicare il nome dell'istituzione è libero. Nei dati, questo comporta la presenza di molteplici valori che fanno riferimento ad un unico istituto. Prendendo ad esempio il caso dell'Università di Pisa, alcuni valori riscontrati sono:

- “Università di Pisa”¹⁴
- “Università degli Studi di Pisa”¹⁵
- “University of Pisa”¹⁶
- “Università di Pisa / University of Pisa”¹⁷
- “Unipi”¹⁸

Per ricondurre tutte le nomenclature ad un'unica forma standardizzata, sono stati creati due strumenti di controllo che permettono all'algoritmo di “riconoscere” l'istituto a cui la nomenclatura si riferisce. Il primo è una lista che elenca le nomenclature più comuni per ogni università. Il secondo, un dizionario che associa le varie nomenclature “alternative” al nome ufficiale dell'università. Ad esempio, i riferimenti all'Università di Pisa presenti nella lista sono: “Università di Pisa”, “Università degli Studi di Pisa” e “University of Pisa”. Invece, le chiavi presenti nel dizionario sono soltanto “Università degli Studi di Pisa” e “University of Pisa”, mentre il valore di questi è appunto il nome ufficiale dell'istituto, ovvero “Università di Pisa”.

Il procedimento consiste nell'estrazione dal dataset di tutti i *token* relativi ai nomi delle istituzioni, cioè tutti i valori utilizzati almeno una volta per indicare l'ente presso cui si è studiato, e nel calcolo della frequenza di questi, cioè

¹⁴Valore dell'attributo *Dg2Location* nel record con identificativo 33.

¹⁵Valore dell'attributo *Dg1Location* nel record con identificativo 897.

¹⁶Valore dell'attributo *Dg1Location* nel record con identificativo 167.

¹⁷Valore dell'attributo *Dg1Location* nel record con identificativo 1750.

¹⁸Valore dell'attributo *Dg1Location* nel record con identificativo 2674.

quante volte un dato token appare nel dataset. Attraverso l'utilizzo della libreria FuzzyWuzzy (Paragrafo 4.2.2) ogni token viene confrontato con la lista delle nomenclature comuni e viene individuata la nomenclatura più simile ad esso. Se il livello di similitudine è pari o superiore al 93%, si assume che la nomenclatura individuata nella lista e il token si riferiscano al medesimo ente. La nomenclatura viene allora confrontata con il dizionario. Se la nomenclatura individuata corrisponde ad una delle chiavi del dizionario, l'algoritmo sostituisce alle occorrenze del token nel dataset il valore che nel dizionario è associato alla chiave. Altrimenti, il token viene sostituito con la nomenclatura così come individuata nella lista. Qualora il confronto con la lista non restituisse alcuna nomenclatura avente con il token un valore di similitudine pari o superiore al 93%, opzionalmente è possibile avviare una procedura semi-controllata di modifica manuale del valore. Se il token analizzato raggiunge un numero minimo di 10 occorrenze, l'algoritmo abbassa la soglia similitudine richiesta per quel token e cerca all'interno della lista le 5 nomenclature con valore di similitudine con il token pari o superiore al 65%. Alle sei stringhe così ottenute, ovvero il token di partenze e le 5 nomenclature, vengono rimossi alcuni vocaboli che, essendo presenti nella maggioranza dei nomi degli istituti universitari, potrebbero comportare dei bias nel momento del calcolo della similitudine. In particolare, i termini che vengono rimossi dalle stringhe sono "università", "university", "universität", "université", "universidad", "studi", "libera", "free".

L'algoritmo procede confrontando le nomenclature modificate con il token modificato. Le nomenclature con un valore di similitudine superiore al 50% vengono stampate a schermo insieme al token originale, ed è richiesto a colui che esegue il codice di indicare quale delle cinque nomenclature dovrebbe sostituire il token all'interno del dataset. Qualora nessuna delle opzioni venisse scelta, i valori non vengono modificati nel dataset ed è possibile scegliere di indicare il token come nomenclatura standard per quell'ente, inserendolo nell'apposita lista e indicando la nazione che ospita l'ente.

Un'ultima modifica effettuata sui dati relativi all'istruzione riguarda la struttura con cui questi vengono memorizzati in locale. Le informazioni relative ad un

unico percorso sono state organizzate dentro un dizionario le cui chiavi sono:

- “**University**”, cioè il nome dell’ente ospitante.
- “**UniNation**”, cioè la nazione in cui è situato l’ente.
- “**Study Field**”, cioè l’argomento di studio.
- “**From Date**”, cioè la data di inizio.
- “**To Date**”, cioè la data di fine.
- “**RawInfo**”, cioè la descrizione del percorso così come descritta dallo studente.

I dati così strutturati sono stati ordinati in 5 attributi, ovvero: “High School”, “Bachelor’s Degree”, “Master’s Degree”, “Ph.D” e “Others”. I valori di questi attributi sono delle liste. Man mano che l’algoritmo individua il livello accademico a cui fanno riferimento le varie esperienze di studio descritte dall’utente, queste vengono inserite nelle apposite liste, a seconda che siano “High School”, “Bachelor’s Degree”, ecc.

I percorsi di studio non categorizzati dall’algoritmo, vengono inseriti all’interno della lista dell’attributo “Others”. Un esempio è mostrato in Figura 4.4,

4.2.4 - Trasformazione dati professionali

Anche i dati relativi alle mansioni professionali sono stati organizzati in alcune macro-categorie, principalmente per sovvenire all’alto numero di valori unici presenti. Si consideri che all’interno dell’attributo “PT1” del dataset è possibile contare più di 8.000 professioni diverse.

Come fattori discriminanti per la categorizzazione si è scelto il prestigio che una determinata mansione comporta e il ruolo che questa ha all’interno della propria azienda in termini di responsabilità. Per effettuare questa scrematura delle mansioni si è fatto riferimento ai titoli professionali associati ad esse, come “*junior*” e “*senior*” o “*associate*” e “*responsabile*”.

I quattro livelli individuati sono:

```

1 df.loc[4099]['Others']

[{'University': 'enea',
  'UniNation': '',
  'Study Field': 'Economics',
  'From Date': '2014',
  'To Date': '2014',
  'RawInfo': "['Project Management']"},
 {'University': 'datacamp',
  'UniNation': '',
  'Study Field': 'Informatics',
  'From Date': '2016',
  'To Date': '2016',
  'RawInfo': "['certification', 'intro python for data science']"},
 {'University': 'life learning',
  'UniNation': '',
  'Study Field': None,
  'From Date': '2021',
  'To Date': '2021',
  'RawInfo': "['Certification', 'Programmazione in Python', '98/100']"}]

```

FIGURE 4.4: Esempio valori relativi ai percorsi di studio

Livello	Parole Chiave
Internship	volontario, stagista, tirocinante, trainee
Prima Fascia	contractor, associate, member, executiver, junior
Seconda Fascia	manager, coordinatore, responsabile, senior
Terza Fascia	founder, owner, capo, presidente, CEO, rector, direttore

TABLE 4.3: Esempi parole chiave usate per classificare le professioni

- **Internship**
- **Prima Fascia**
- **Seconda Fascia**
- **Terza Fascia**

Dove con *Internship* si intendono stage, tirocini ed esperienze di volontariato; *Prima* e *Seconda Fascia* indicano lavori più o meno di responsabilità e con *Terza Fascia*, i vertici d'azienda e gli imprenditori.

Le operazioni effettuate sono del tutto speculari a quelle eseguite sulla descrizione del percorso di studi, descritte nel Paragrafo 4.2.3. Se la stringa contiene una delle parole chiave indicate, la stringa viene sostituita con la categoria associata a quella parola chiave.

Nella Tabella 4.3 sono riportati alcuni esempi delle parole chiave associate ai vari livelli professionali.

Nel caso in cui la stringa non contenesse nessuna delle parole chiave scelte, si suppone che il ruolo svolto dall'utente non preveda alcun titolo professionale e che quindi sia una mansione a basso livello di responsabilità e quindi appartenente a "Prima Fascia".

A questi quattro livelli selezionati se ne aggiunge un quinto:

- **Ph.D**

Durante il lavoro sui dati si è infatti notato che in diversi record le informazioni relative ad un eventuale dottorato erano presentate tra le esperienze lavorative piuttosto che in quelle accademiche. Per ragione di coerenza interna, si è deciso di considerare questi dati come esperienze di studio e non come esperienze professionali. Per ciò, in linea con quanto descritto nel Paragrafo 4.2.3, in questa fase di preparazione dei dati, i riferimenti ai dottorati vengono "etichettati" come "*Ph.D*" e successivamente inseriti nell'apposito attributo.

Un'altra operazione eseguita sui dati professionali consiste nel controllo per ogni record della data di inizio di tutte le esperienze lavorative, in modo tale da riorganizzarlo così che questo riporti le esperienze professionali in ordine cronologico, cioè dalla più antica alla più recente.

4.2.5 - Mobilità internazionale

L'ultima modifica effettuata alla struttura del dataset consiste nell'aggiunta dell'attributo, "*Mobility*". Con questo attributo viene evidenziato se lo studente a cui il record fa riferimento si è mai spostato da una nazione all'altra per motivi di studio o di lavoro, cioè se rappresenta un caso di **mobilità internazionale**.

Per dedurre questa informazioni, l'algoritmo controlla le varie località indicate negli attributi "*Location*", "*Dg Nation*" e "*PT Location*" e assegna un valore all'attributo "*Mobility*" a seconda se da questi si evince uno spostamento, banalmente perché diverse tra loro, o no. L'attributo potrà quindi assumere due valori diversi:

- “**y**”, nel caso in cui vi sia un fenomeno di mobilità internazionale
- “**n**”, nel caso in cui non ci fosse.

Alla fine della fase di normalizzazione dei dati si ottiene un dataset strutturato contenente per ogni studente:

- informazioni geografiche, standardizzate in maniera da mostrare solo il riferimento alla nazione ed esclusivamente in lingua inglese.
- informazioni sui percorsi formativi seguiti, suddivisi per titolo ottenuto e caratterizzati con informazioni geografiche e temporali;
- informazioni sulle esperienze lavorative, categorizzate in 4 macro-settori, con relative informazioni geografiche e temporali;
- un valore che indica se il soggetto si è mai spostato per studio o lavoro fuori dall'Italia.

5 || Caso di studio

L'approccio presentato nel Capitolo 4 è stato applicato a un caso di studio al fine di testarne le reali possibilità. Come anticipato nel Capitolo 1, sono stati presi in esame i dati relativi a circa 10.000 alumnae e alumni di alcune università toscane.

Questa analisi permette di osservare le connessioni degli atenei toscani nel mondo, evidenziando le criticità e i punti di forza delle offerte formative disponibili.

I cinque atenei che si è deciso di considerare per questo lavoro sono:

- **Scuola Normale Superiore**
- **Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna**
- **Università degli Studi di Firenze**
- **Università degli Studi di Siena**
- **Università di Pisa**

Questa scelta viene motivata principalmente dal funzionamento stesso di LinkedIn, che, come scritto nel Capitolo 3, limita la visibilità dei dati degli utenti a seconda della distanza che questi hanno all'interno della propria rete¹, limitandoci quindi nella scelta agli atenei più vicini alla nostra cerchia professionale, che nel nostro caso sono, appunto, l'Università di Pisa e gli altri atenei toscani.

A questa ragione puramente pratica, si aggiungono le specificità proprie di questa regione e delle sue Università. La Toscana è infatti la seconda regione italiana per numero di università pubbliche, contandone ben sette. Sommando gli iscritti di ogni ateneo, ogni anno si contano circa 110.000 studenti che gravitano intorno a questi poli culturali, tra cui spiccano per grandezza l'Università degli Studi di Firenze e l'Università di Pisa, entrambe tra le 10 università più grandi

¹In realtà il problema si pone più ad un livello teorico che pratico. Secondo lo studio [Backstrom et al., 2012] la distanza media tra due utenti di Facebook è di circa 3,57 gradi, cioè di poco più di tre profili. Supponendo che anche all'interno di LinkedIn la distanza tra due utenti sia pressapoco la stessa, il problema della visibilità dei dati diventa marginale in quanto si applicherebbe solo a quei pochi profili con una distanza superiore ai 3 gradi.

d'Italia², superando annualmente i 40.000 iscritti. Un'altra peculiarità è la presenza nella regione di due delle tre scuole superiori universitarie autonome istituite dallo stato, ovvero la Scuola Normale Superiore e la Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna.

Dall'analisi si è invece deciso di escludere l'Università per Stranieri di Siena e la Scuola IMT Alti Studi Lucca, a ragione della formazione strettamente settoriale che questi istituti offrono.

5.1 | *Dataset*

Il dataset collezionato è composto da 9405 record e da 92 attributi, che sono:

- **Name**, cui valore è un numero intero, rappresentante l'identificativo dello studente.
- **Location**, cui valore è una stringa di testo. Il contenuto di questa è il nome in lingua inglese della nazione indicata dallo studente come località di riferimento.
- **High School**, il cui valore è una lista contenente le informazioni relative ai titoli di istruzione secondaria indicati dallo studente. Le informazioni sono organizzate all'interno di dizionari.
- **Bachelor's Degree**, il cui valore è una lista contenente le informazioni relative ai titoli di laurea indicate dallo studente. Le informazioni sono organizzate all'interno di dizionari.
- **Master's Degree**, il cui valore è una lista contenente le informazioni relative ai titoli di laurea indicate dall'utente. Le informazioni sono organizzate all'interno di dizionari.

²I dati fanno riferimento a quanto pubblicato dal MIUR al link in bibliografia [[Ministero dell'Istruzione, dell'Università, della Ricerca, 2021](#)]

- **Others**, il cui valore è una lista contenente le informazioni relative ai vari corsi di studio non riconosciuti dal nostro algoritmo. Le informazioni sono organizzate all'interno di dizionari.
- **PT**, che indica il livello a cui appartiene la mansione svolta dallo studente. L'attributo può assumere i valori "Internship", "Prima Fascia", "Seconda Fascia", "Terza Fascia" a seconda del livello di responsabilità associato alla mansione.
- **PTCompany**, il cui valore è una stringa di testo indicante la compagnia presso cui si è svolta la mansione indicata in PT.
- **PTDuration**, per quanto tempo si è svolta la mansione indicata in PT. Il valore è indicato in una stringa di testo.
- **PTLocation**, la località in cui si è svolta la mansione indicata in PT. Il valore è indicato in una stringa di testo.
- **PTFromDate**, il cui valore è una stringa di testo indicante l'anno in cui ha avuto inizio l'esperienza lavorativa indicata in PT.
- **PTToDate**, il cui valore è una stringa di testo indicante l'anno in cui ha avuto fine l'esperienza lavorativa indicata in PT.
- **PTUrl**, il cui valore è una stringa di testo contenente l'URL al sito della compagnia indicata in PTCompany.
- **Mobility**, ovvero un valore binario che indica se, a partire dagli altri dati, è possibile risalire ad un fenomeno di mobilità internazionale o no. I valori che l'attributo può assumere sono "y", se vi è un cambio di nazione, "n" altrimenti.

Il dataset include diversi gruppi di attributi relativi alle esperienze professionali - PT, PTCompany, PTDuration, PTLocation, PTFromDate, PTToDate e PTUrl. In particolare, sono presenti undici gruppi di attributi professionali numerati, cioè:

- PT₁, PT₁Company, PT₁Duration...

- PT₂, PT₂Company, PT₂Duration...
- PT₃...
- ...
- PT₁₁, PT₁₁Company, PT₁₁Duration...

I diversi percorsi di studio sono invece organizzati dentro gli appositi dizionari come segue:

- **University**: a questa chiave corrisponde una stringa di testo indicante l'ente ospitante il percorso di studi.
- **UniNation**: a cui corrisponde una stringa di testo indicante la nazione in cui è situato l'ente.
- **Study Field**: a cui corrisponde una stringa di testo indicante l'argomento di studio.
- **From Date**: a cui corrisponde una stringa di testo indicante l'anno di inizio del percorso di studi.
- **To Date**: a cui corrisponde una stringa di testo indicante l'anno di fine del percorso di studi.
- **RawInfo**: a cui corrisponde una lista contenente le informazioni date dallo studente sul percorso di studio.

Relativamente all'attributo Location è importante sottolineare l'ambiguità del dato. Al momento dell'inserimento su LinkedIn dell'informazione, non viene specificato se l'utente deve indicare il luogo di nascita, il proprio domicilio, la residenza o il luogo in cui si preferisce cercare lavoro. Ad esempio, il terzo record nell'immagine 4.2 ha come valore di Location "*Pavia, Lombardy, Italia*", tuttavia dagli altri dati presenti nel record è possibile individuare come data di trasferimento in Italia l'anno 2016, essendo tutte le altre attività precedenti localizzate in Turchia, come osservabile nei valori di PT₁Location e PT₂Location dello stesso

Name	Location	High School	Bachelor's Degree	Master's Degree	Ph.D	Others	PT1	PT1Company	PT1Duration	PT1Location	PT1FromDate	PT1ToDate
33	Italy	None	[[{"University": "Università degli Studi di Tre..."}	[[{"University": "Università di Pisa", "UniNati..."}	None	[[{"University": "scuola secondaria di secondo ..."}]	Prima Fascia	Fondazione Bruno Kessler - FBK Internship	3 mos	Italy	2016	2016
68	Pakistan	None	None	[[{"University": "international islamic univers..."}	None	None	Prima Fascia	Baizai CNG Filling Station	4 yrs 5 mos	Pakistan	2017	None
72	Italy	None	None	[[{"University": "Università di Pisa", "UniNati..."}	[[{"University": "Scuola Normale Superiore di P..."}	[[{"University": "international summer scho..."}	Internship	Fermilab	3 mos	United States	2016	2016

FIGURE 5.1: Screenshot dati puliti aggregati

record mostrati in 4.3. Un esempio opposto ha come valore di Location “*La Spezia, Liguria, Italy*”, luogo in cui lo studente ha conseguito il diploma di maturità. Tuttavia, i corsi universitari e le attività successive hanno come località “*Pisa, Toscana, Italy*”.

La Figura 5.1 mostra un esempio dei dati contenuti nel dataset.

5.2 | *Analisi Statistica*

In questo paragrafo è presentata l’analisi esplorativa del dataset ottenuto mediante la procedura descritta nel Capitolo 4. Lo studio mira a descrivere, attraverso le statistiche ottenute, le carriere accademiche e professionali degli alumni e delle alumne dei cinque atenei toscani. Il capitolo è suddiviso in tre sezioni: una relativa alla posizione geografica degli studenti e delle studentesse, una al livello di istruzione e un’ultima alle esperienze professionali.

5.2.1 - *Località geografica*

Come specificato nel Paragrafo 5.1, i valori indicati nell’attributo Location sono ambigui, in quanto indicano una località generica di riferimento per il profilo. Tuttavia, il dato permette di identificare i luoghi geografici interessati dai movimenti degli studenti e delle studentesse.

Nel dataset, il 99,8% dei record ha un valore di Location: il dato è quindi presente nella quasi totalità dei record, con eccezione di appena 19 record su 9405. A

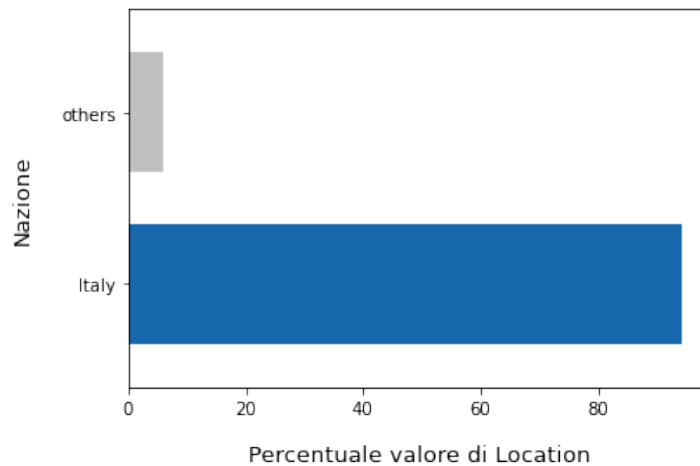


FIGURE 5.2: Confronto tra il valore di Location "Italy" e l'insieme degli altri.

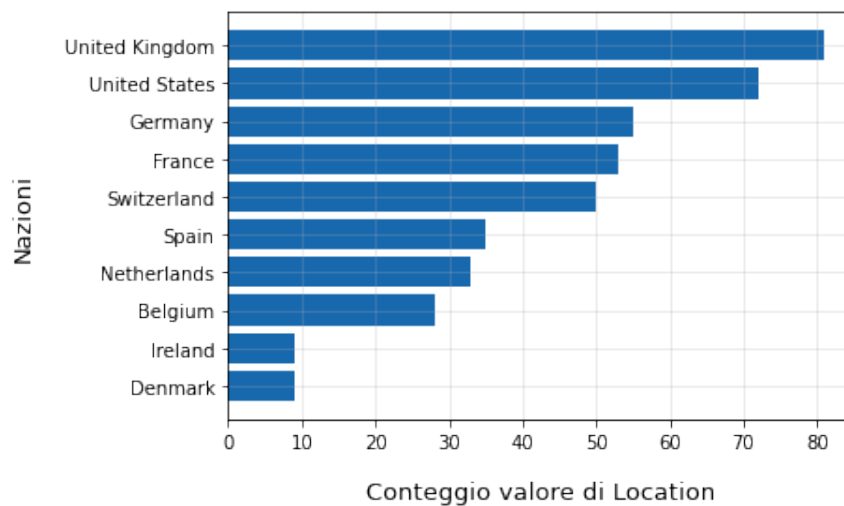


FIGURE 5.3: Conteggio dei 10 valori di Location più comuni dopo "Italy".

dodici di questi è stato attribuito il valore di Location "Italy", in quanto all'interno del record non erano indicate altre nazionalità, mentre per i restanti sette è stato mantenuto il valore nullo, in quanto erano presenti nel record più nazioni.

Come mostrato in Figura 5.2, il 94,17% degli utenti ha indicato l'Italia come località generica del profilo, mentre il restante 5,83% ha indicato una nazione estera³. Tra queste primeggiano le nazioni dell'Europa, e particolarmente dell'Europa Occidentale, ma nessuna di queste compare in maniera significativa: la nazione estera più presente è il Regno Unito, che però viene indicata come valore di Lo-

³I dati per esteso sui valori dell'attributo Location sono riportati in Appendice A, nella Tabella A.1.

Tipologia percorso di studio	Conteggio
High School	900
Bachelor's Degree	3883
Master's Degree	6732
Ph.D	2961
Others	9520

TABLE 5.1: Conteggio dei percorsi di studio per tipologia.

cation solo nello 0,81% dei record.

La Figura 5.3 propone un elenco delle 10 nazioni estere maggiormente indicate come località del profilo e riporta per ognuna di esse il numero di record che ne fanno riferimento all'interno dell'attributo Location. Nelle prime posizioni ci sono i due paesi anglofoni, cioè il Regno Unito e gli Stati Uniti, unica nazione extra europea presente tra le prime dieci. Entrambi questi valori contano più di 70 occorrenze. Le nazioni successive sono invece più vicine geograficamente all'Italia, ovvero Germania, Francia e Svizzera, ciascuna indicata come località di riferimento da circa cinquanta utenti. A seguire, Spagna, Paesi Bassi e Belgio, con circa 30 utenti l'una e infine Danimarca e Irlanda, che vengono indicate da meno di 10 utenti.

5.2.2 - Percorsi di studio

Nel dataset sono presenti un totale di 23.996 riferimenti a percorsi di studio. In media, per ogni profilo si hanno 2,5 indicazioni a esperienze di studio. Il numero massimo di esperienze riscontrate in un profilo è tre, mentre il minimo è uno, avendo filtrato i dati secondo l'aver studiato o meno in uno degli atenei toscani.

La Tabella 5.1 riporta il numero di percorsi di studio presenti nel dataset per ogni tipologia rilevata dall'algorithm, mentre la Figura 5.4 riporta il numero di studenti che hanno indicato almeno un percorso di studio in una determinata categoria ⁴.

⁴Per la lettura dei dati in Figura 5.4, è importante sottolineare che le informazioni ottenute dai profili LinkedIn non sono sempre complete: ad esempio, alcuni utenti inseriscono esclusivamente il livello accademico conseguito più alto, non riportando informazioni sui percorsi precedenti. Il numero di persone che hanno conseguito una laurea di secondo livello o un dottorato va quindi considerato singolarmente, e non in relazione ai dati relativi ai livelli accademici inferiori.

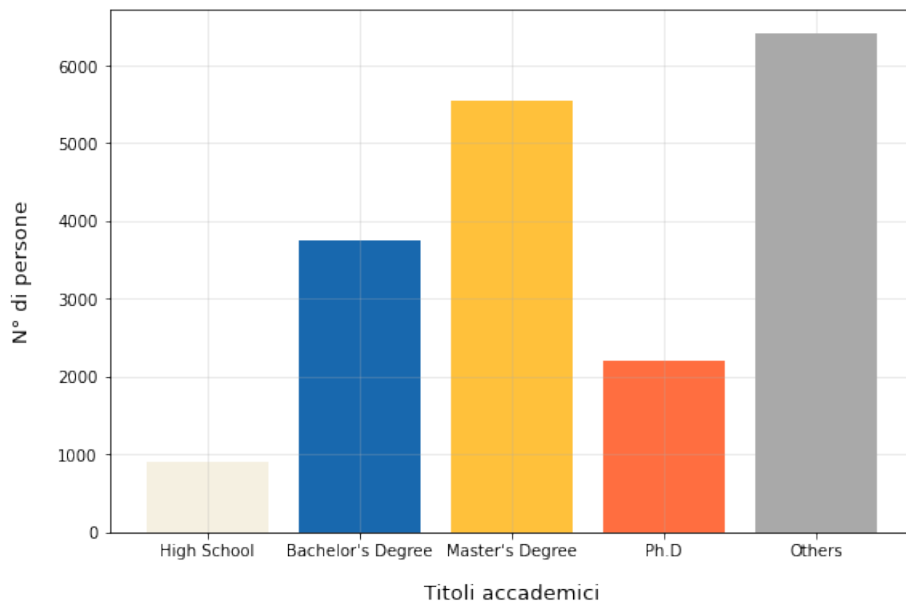


FIGURE 5.4: Distribuzione esperienze accademiche per studenti.

Il 74% degli studenti che hanno indicato di aver conseguito una laurea triennale ha proseguito gli studi ottenendo un titolo di II livello. Di questi, il 13% ha successivamente ottenuto un dottorato. Tra gli studenti che hanno indicato di possedere un titolo di II livello, laurea magistrale o specialistica, il 23% ha ottenuto un dottorato. Inoltre, secondo i dati in nostro possesso relativi al Ph.D, possiamo affermare che mediamente un dottore su tre consegue il dottorato più di una volta. Questo è dimostrato dalla differenza tra il numero di studenti che hanno indicato di possedere un dottorato, ovvero 2202, e il numero totale di dottorati, 2961.

In Figura 5.5 sono riportati in ordine i nominativi degli istituti frequentati dalla maggior parte degli studenti. Ad apparire insieme ai cinque atenei toscani ci sono esclusivamente università italiane.

Più nel dettaglio, hanno frequentato l'Università di Pisa 4428 studenti (47%), 2612 (28%) hanno studiato nell'Università degli Studi di Firenze, 2049 (22%) all'Università degli Studi di Siena, 948 (10%) alla Scuola superiore di studi universitari e di perfezionamento Sant'Anna, 853 (9%) alla Scuola Normale Superiore. A seguire, l'Università degli Studi di Bologna, l'Università degli Studi di Roma "La Sapienza", l'Università commerciale Luigi Bocconi, l'Università degli Studi di Trento e l'Università degli Studi di Catania, frequentate solo dall'1% degli studenti. Infine, il 74% ha poi

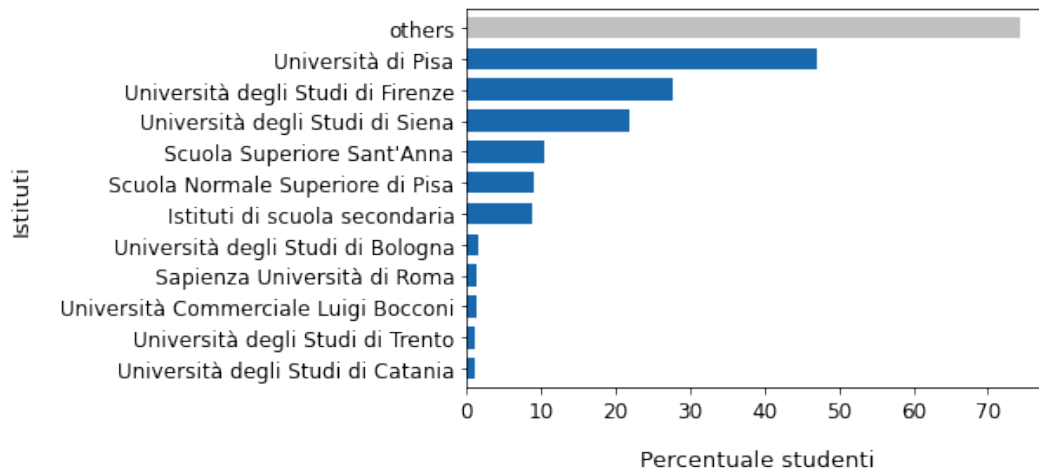


FIGURE 5.5: Percentuali degli studenti che hanno studiato in un istituto.

studiato anche in altre università, che se prese singolarmente sono minoritarie nel nostro dataset, in quanto frequentate da meno di 100 studenti.

La London School of Economics and Political Science è l'università estera con più studenti (34), seguita dal King's College London (27) e dall'Università Complutense di Madrid (25).

In generale, la nazione estera in cui hanno studiato più alumni e alumne degli atenei toscani è il Regno Unito, dove hanno studiato 205 persone. La Spagna è la seconda, con 125 e gli Stati Uniti terzi, con 105. A seguire Francia, Germania e Belgio con circa 50 studenti, Paesi Bassi con 37 e Irlanda, Svizzera e Danimarca con circa 25⁵. Eccezione fatta per la Spagna, i dati sembrerebbero coincidere con quanto evidenziato relativamente alla località del profilo nel Paragrafo 5.2.1: le prime posizioni sono occupate dalle due nazioni anglofone, a cui seguono le nazioni dell'Europa Occidentale, con particolare rilievo di Francia e Germania.

Per i dati relativi a Bachelor's Degree, Master's Degree e Ph.D, si hanno informazioni relativi alla nazione di riferimento nell'85% dei casi, mentre per i dati riferiti ad Others la percentuale scende al 65%, probabilmente per la presenza di percorsi di studio non istituzionali.

Di tutti i 23.996 percorsi di studio collezionati, il 59,6% si sono svolti nei cinque atenei toscani analizzati. Questo valore sale al 68,7% se si escludono i

⁵I conteggi degli utenti che hanno studiato nelle varie nazioni sono riportati nella tabella A.2 in AppendiceA

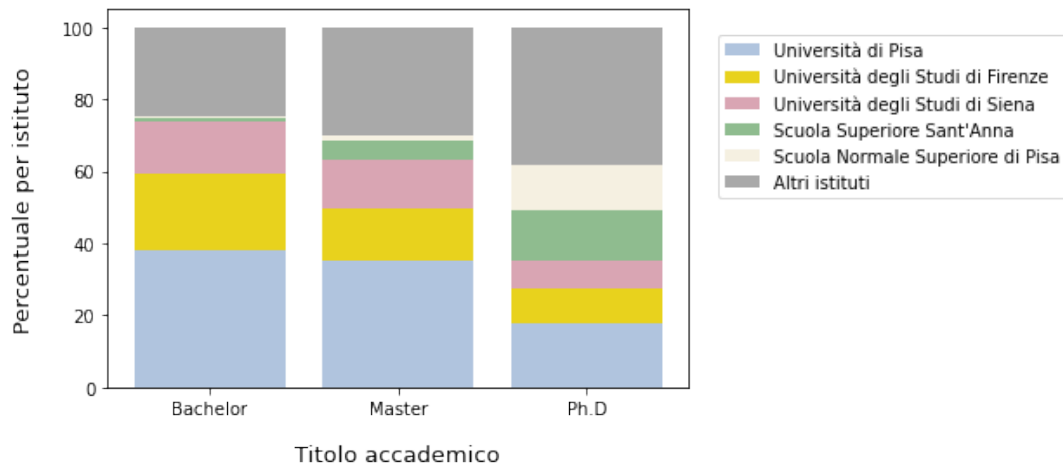


FIGURE 5.6: Confronto tra tipologia di percorso di studi e istituto.

percorsi categorizzati come High School e come Others.

All'interno degli atenei toscani ci sono delle differenze nel numero di studenti che affrontano i vari livelli accademici. Come mostrato nel grafico 5.6, le proporzioni degli utenti che studiano nelle cinque università non rimangono costanti all'aumentare del livello accademico.

In particolare si può notare che la percentuale degli studenti delle università di Firenze, Pisa e Siena è maggiore per le lauree di primo livello, aggirandosi intorno al 74%, scende al 64% per le lauree di secondo livello, fino ad arrivare al 35% per i dottorati. Una tendenza inversa si riscontra invece per le due scuole superiori, la Normale e la Sant'Anna, sul 2% relativamente ai Bachelor's Degree, al 6% per Master's Degree, fino ad un 27% per i Ph.D.

Relativamente al livello temporale, abbiamo informazioni sulla data di inizio dei percorsi accademici nel 93% dei dati, percentuale che cresce rimuovendo le informazione categorizzate come Others.

Il periodo coperto è di 80 anni: la data del primo percorso di studi è il 1941, mentre quella dell'ultimo è il 2021. Il numero di informazioni possedute per ogni anno non è però costante. Come raffigurato in Figura 5.7, i dati relativi agli studi tra gli anni 40 e gli anni 60 sono meno di 10, mentre la maggior parte fanno riferimento agli anni che vanno dal 2000 ad oggi⁶.

⁶Il conteggio dei percorsi di studio anno per anno è riportato in Tabella A.3 in Appendice A.

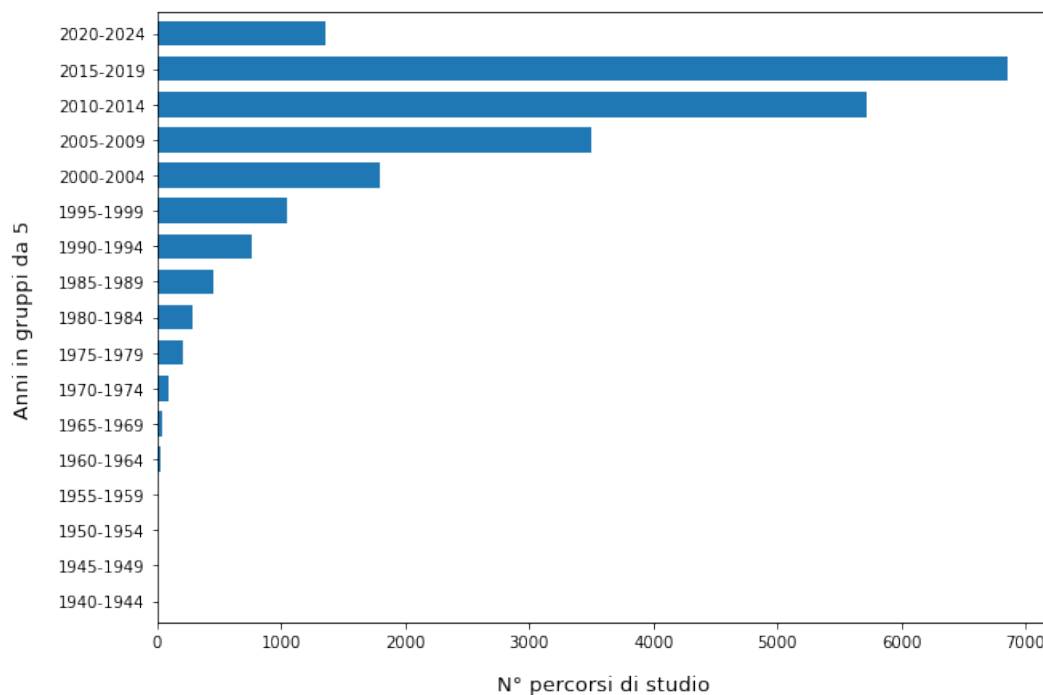


FIGURE 5.7: Distribuzione di percorsi di studio per gruppi di 5 anni.

La Figura 5.8 mostra i dati temporali dei soli percorsi di studio toscani. Da un confronto con la Figura 5.7 è possibile notare una sovrapposibilità dei grafici. Ciò significa che, da un punto di vista temporale, i dati sugli atenei Toscani sono coerenti con quelli generali del dataset⁷.

Relativamente al campo di studio, gli argomenti più studiati sono Economia (17,3% dei percorsi analizzati), Scienze Naturali (8,0%), Discipline Umanistiche (7,8%) e Ingegneria(6,8%). Economia è il corso più rappresentato, avente più del doppio dei dati rispetto al secondo, Scienze Naturali. Professioni Sanitarie e Veterinaria sono invece i corsi meno rappresentati: meno dell' 1% dei nostri dati riguarda questi campi.

La Figura 5.9 mostra i rapporti di proporzione che ci sono tra tutti i campi di studio presenti all'interno del dataset. Ogni rettangolo rappresenta una delle 18 macro categorie utilizzate e la sua dimensione è proporzionale al numero di percorsi di studio presenti nel dataset che fanno riferimento ad essa. Similmente, la Figura 5.10 propone una visione d'insieme degli argomenti più studiati per ogni livello accademico. Le proporzioni sembrano coerenti con quanto riscon-

⁷Il conteggio dei percorsi toscani anno per anno è riportato in Tabella A.4 in Appendice A.

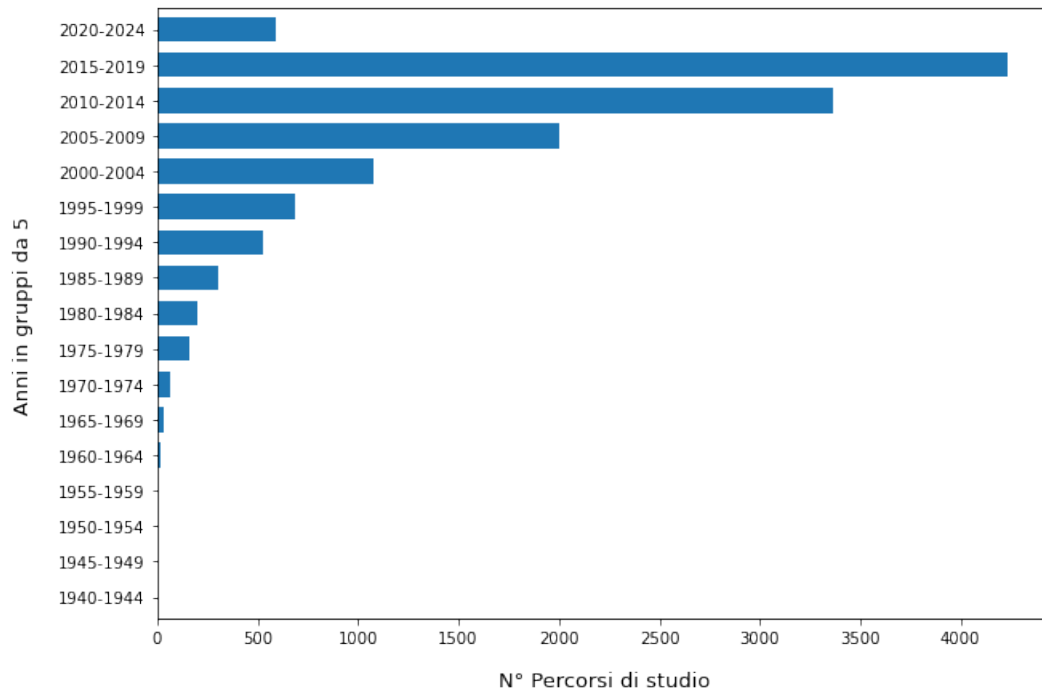


FIGURE 5.8: Distribuzione percorsi di studio fatti in Toscana per gruppi di 5 anni.

trato precedentemente. Nei percorsi di studio di primo e secondo livello e in quelli categorizzati come Others vi è una netta prevalenza di studi di Economia, seguita da quelli sulle scienze naturali e sulle materie umanistiche. Una lieve differenza è presente nei Ph.D, in cui il campo più studiato è invece Scienze Naturali, con Economia a seguire⁸.

Facendo un focus sulle materie più studiate in Toscana, queste rimangono Economia, Scienze Naturali e Ingegneria.⁹ Per quest'ultima materia, il 62% dei percorsi di studio totali presenti nel dataset sono stati effettuati in uno degli atenei toscani. La media per le altre materie si aggira intorno al 45%, con una deviazione standard di 10 punti percentuali. Anche Veterinaria (60%) e Scienze Naturali (58%) si pongono sopra questa media. Al contrario, solo il 22% degli studi di Scienze della Formazione sono stati fatti in Toscana.

La Figura 5.11 mostra i flussi di studenti che cambiano argomento di studi nel passaggio da un livello accademico all'altro¹⁰. I rettangoli a sinistra, al centro

⁸Il conteggio dei percorsi di studio per argomento è riportato in Tabella A.5 in Appendice A.

⁹Il conteggio dei percorsi di studio fatti in Toscana per argomento è riportato in Tabella A.6 in Appendice A.

¹⁰Una versione interattiva del grafico è disponibile al link <https://github.com/>

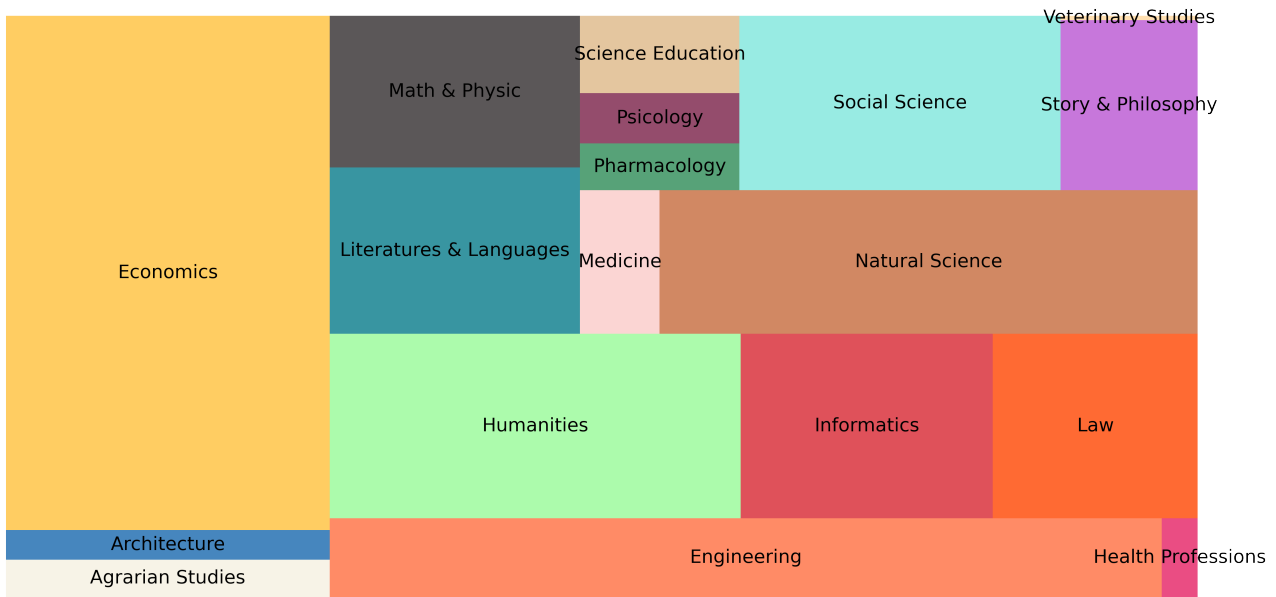


FIGURE 5.9: Rapporti di proporzione tra gli argomenti di studio.

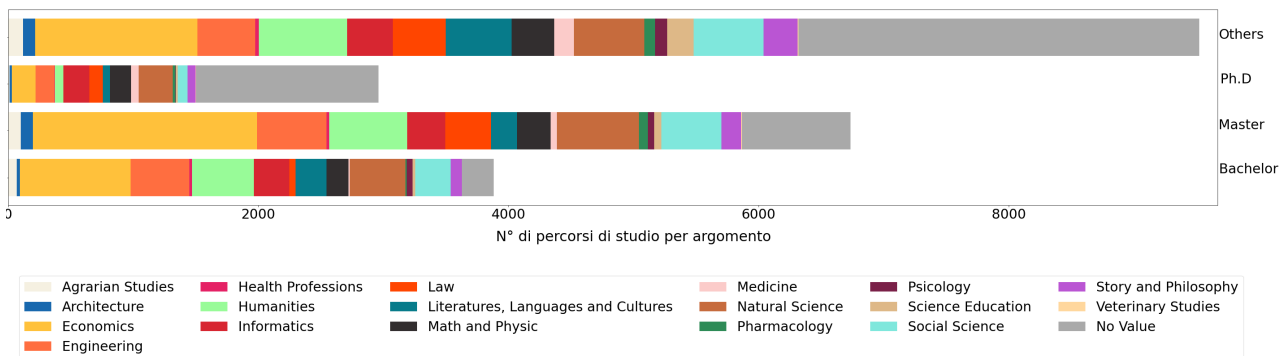


FIGURE 5.10: Argomenti di studio per tipologia di percorso.

e a destra nell'immagine rappresentano i vari argomenti di studi e la loro altezza è proporzionale al numero di persone che hanno studiato quell'argomento. A sinistra ci sono i percorsi classificati come Bachelor's Degree, al centro Master's Degree, a destra Ph.D. Lo spessore delle linee che li collegano indicano il numero di persone che passano da un livello accademico all'altro.

In generale si nota la tendenza a confermare il proprio ambito di studi man mano che ci si specializza. Tuttavia, tutti gli argomenti registrano una piccola "perdita" di studenti, che si spostano verso quei settori che abbiamo visto essere i più comuni. In particolare, Economia è la materia che attira più studenti proveni-

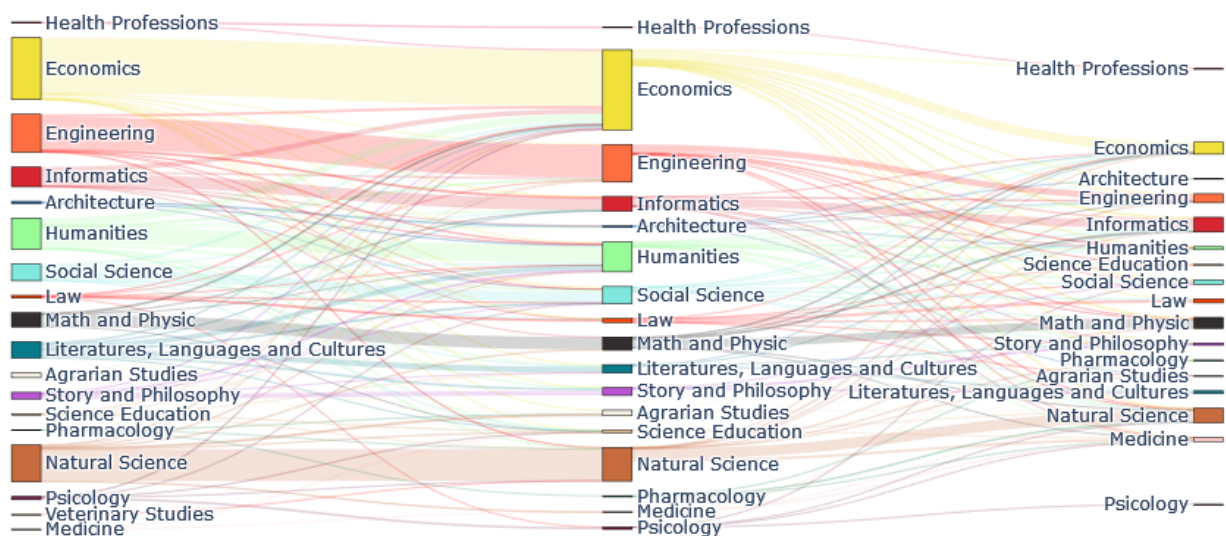


FIGURE 5.11: Flussi di persone che cambiano argomento di studio da un livello accademico all'altro.

enti da altre discipline, specialmente come laurea di secondo livello, ma attirano molti utenti anche Informatica, Scienze Sociali e gli Studi Umanistici.

Dalla Figura 5.11 si evince anche come il numero di studenti che decidono di continuare gli studi prendendo un dottorato dopo aver ottenuto un titolo di secondo livello in Economia o in Discipline Umanistiche sia in percentuale molto basso (13%). Al contrario è molto alto per Informatica e per Matematica e Fisica (80% circa)¹¹.

5.2.3 - Esperienze lavorative

Nel dataset si contano un totale di 33.441 riferimenti ad esperienze professionali. In media, per ogni studente sono indicate 3 o 4 mansioni svolte.

La Figura 5.12 mostra quanti studenti hanno indicato nel proprio profilo un certo numero di professioni svolte. Il 96,3% di questi ha riportato almeno una

¹¹Percentuali molto alte sono registrate anche in Architettura, Farmacia, Legge e Medicina. Tuttavia, data l'insufficienza dei dati, si è deciso di tralasciarle.

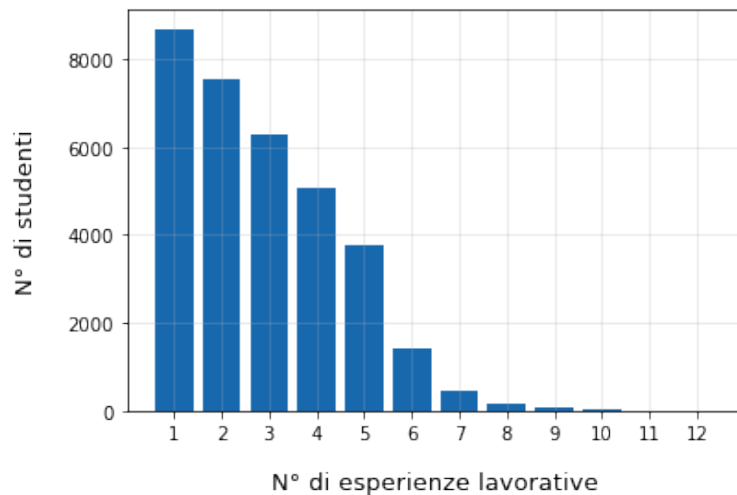


FIGURE 5.12: Distribuzione esperienze lavorative per studenti.

esperienza professionale, mentre il 7,7% (724 studenti) nessuna. Il numero di studenti che hanno indicato N esperienze professionali decresce gradualmente fino a 6, dove invece si registra un calo più netto: dei 3.774 studenti che hanno indicato almeno 5 esperienze professionali, circa un terzo, cioè 1.400 studenti, ne hanno indicato 6 o più. Infine, soltanto un record presenta 12 esperienze professionali, il numero più alto registrato nel dataset.

Nella Figura 5.13 sono riportate le percentuali degli studenti che hanno riportato esattamente N esperienze lavorative. La maggioranza, cioè il 25%, ne ha 5.

In generale, l'87% degli studenti ha svolto da 1 a 6 professioni diverse e solo il 5% ne ha svolte 7 o più.

Il 54% dei dati sulle esperienze professionali contengono informazioni sulla nazione in cui si è lavorato. Di queste, il 78% (il 42% sul totale) si sono svolte in Italia. Oltre all'Italia, le nazioni più comuni sono ancora una volta quelle dell'Europa Occidentale, tra cui primeggiano Regno Unito, Francia e Germania. L'unica nazione extra europea ad aver dato lavoro ad un numero relativamente alto di utenti è gli Stati Uniti d'America, che come mostrato in Figura 5.14 si posizionano al secondo posto, sotto il Regno Unito¹².

¹²Il conteggio degli utenti che hanno lavorato per ogni nazione è riportato in Tabella A.7 in Appendice A.

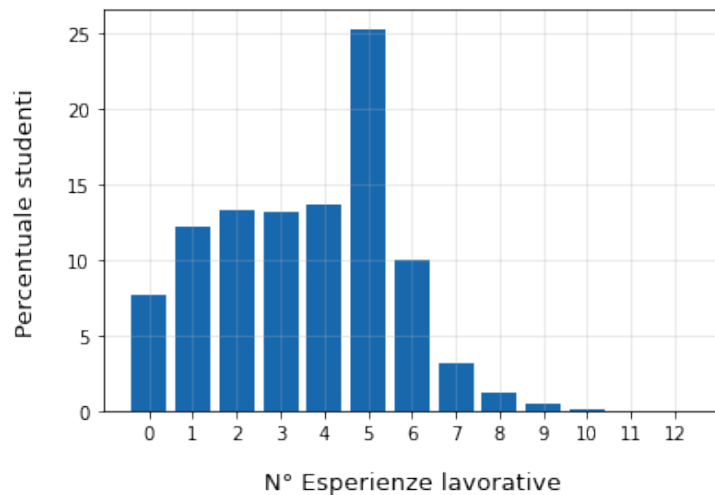


FIGURE 5.13: Percentuali studenti per numero di esperienze lavorative.

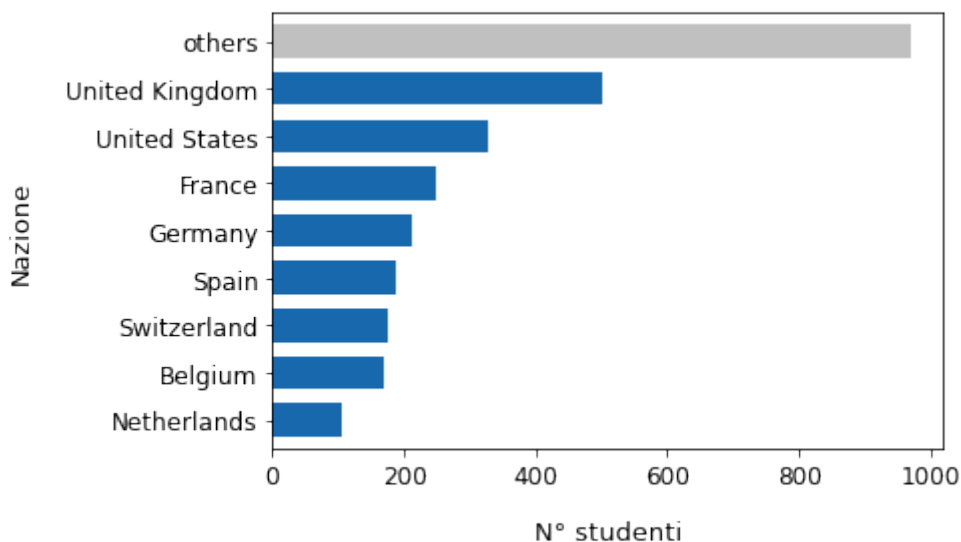


FIGURE 5.14: Numero di studenti che hanno lavorato nelle nazioni estere principali.

La data di inizio è riportata nel 98% delle esperienze lavorative. Quelle più recenti si riferiscono al 2021, mentre le più datate risalgono al 1929. Tuttavia, le esperienze antecedenti al 1955 sono soltanto 3, mentre il 96% delle altre esperienze si sono svolte nel ventennio che va dagli anni 2000 ai giorni odierni¹³ (Figura 5.15).

Come mostrato in Figura 5.16, tra le prime esperienze lavorative sono molto frequenti figure di Internship, cioè di tirocinio, e di Primo Livello, cioè figure junior o senza titolo professionale. Una prevalenza di mansioni di Secondo e Terzo livello

¹³Il conteggio delle esperienze lavorative anno per anno è riportato in Tabella A.8 in Appendice A.

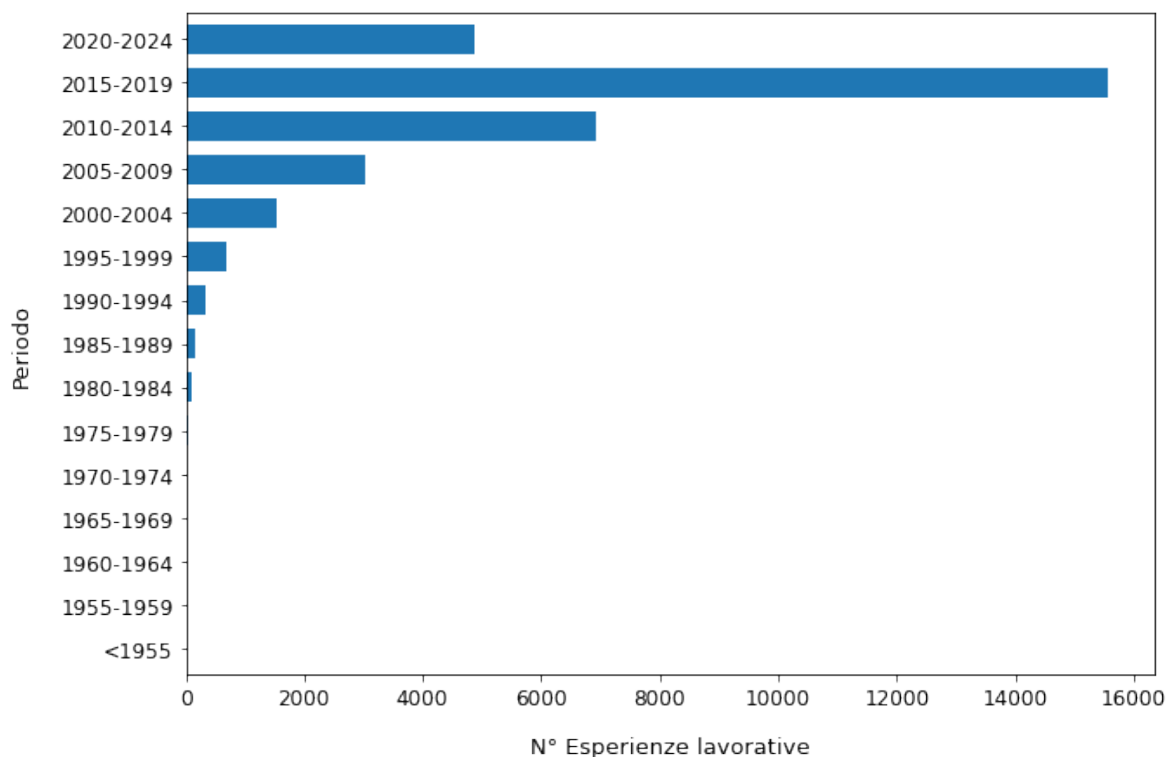


FIGURE 5.15: Conteggio esperienze lavorative in gruppi di 5 anni.

rispetto a quelle di primo livello non viene mai registrata, ma si può notare una progressiva parificazione dalla sesta esperienza lavorativa in poi¹⁴. Come si evince dal grafico (Figura 5.16), l’Internship è la seconda tipologia di esperienza più comune come primo approccio al mondo del lavoro. Il numero di Internship tende però a diminuire rapidamente: già come seconda esperienza lavorativa, questo è pari al numero di lavori di Seconda Fascia. Superata la sesta esperienza lavorativa, il numero di Internship è inferiore alla decina.

I lavori di Prima Fascia sono quelli più diffusi e primeggiano sia come prima, seconda, terza, quarta e quinta esperienza lavorativa. Nella sesta, invece, vi è un crollo netto nel numero di esperienze che diventa comparabile con quello delle mansioni di Secondo e di Terzo livello.

Il numero di queste ultime rimane costante per le prime cinque esperienze lavorative degli utenti e tende a diminuire dalla sesta in poi. Tuttavia, come mostrato in Figura 5.17, in percentuale si registra un aumento delle mansioni

¹⁴Il conteggio è riportato in Tabella A.9 in Appendice A.

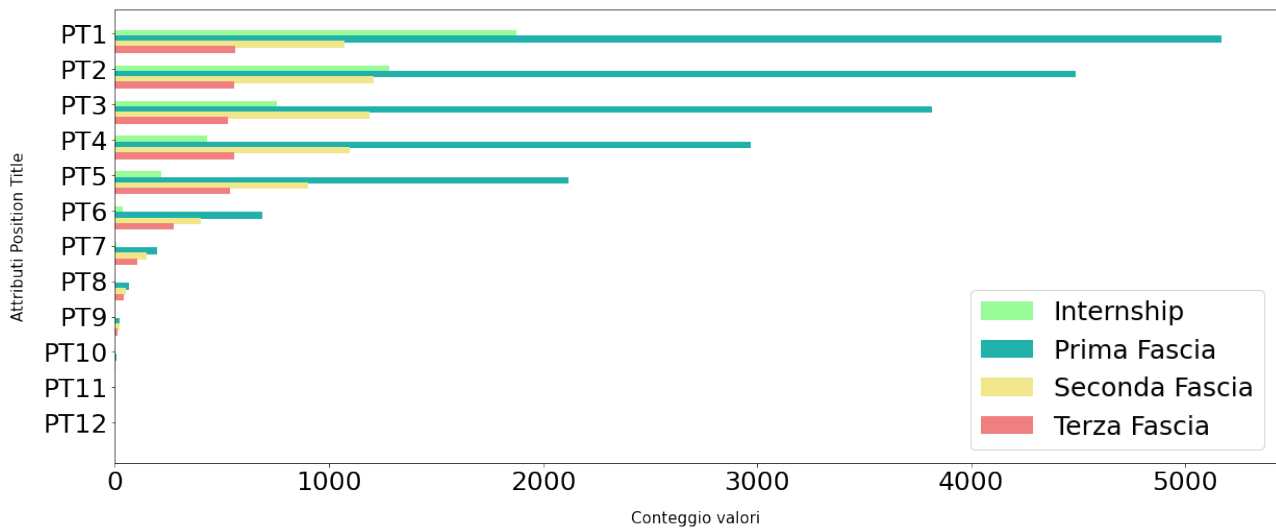


FIGURE 5.16: Suddivisione per fasce delle professioni svolte.

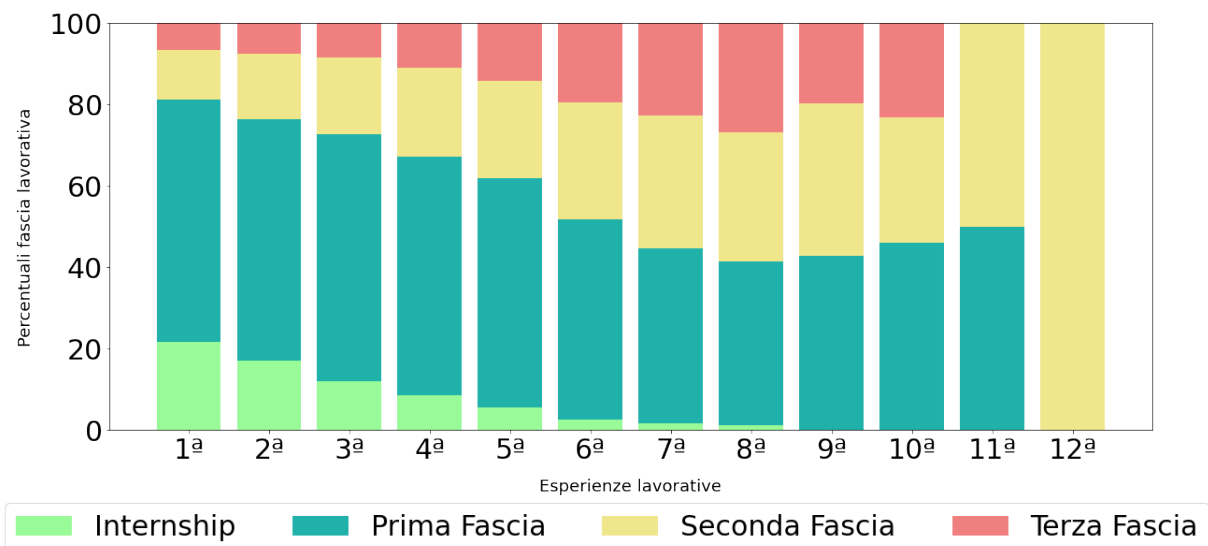


FIGURE 5.17: Percentuali livello professionale.

di livello più alto man mano che si fanno esperienze lavorative. In particolare, le mansioni di Secondo livello sono in percentuale maggiore già dalla seconda esperienza lavorativa ed aumentano fino all'ottava. Le mansioni di Terza Fascia invece, cioè i ruoli dirigenziali, mantengono una percentuale inferiore al 10% per le prime tre esperienze lavorative ed aumentano in percentuale solo dalla quarta esperienza in poi¹⁵.

¹⁵Per quanto riguarda le undicesime e dodicesime esperienze lavorative, i dati a nostra disposizione sono insufficienti per una analisi statistica. Per questa ragione sono stati tralasciati nella trattazione.

5.3 | *Mobilità Internazionale*

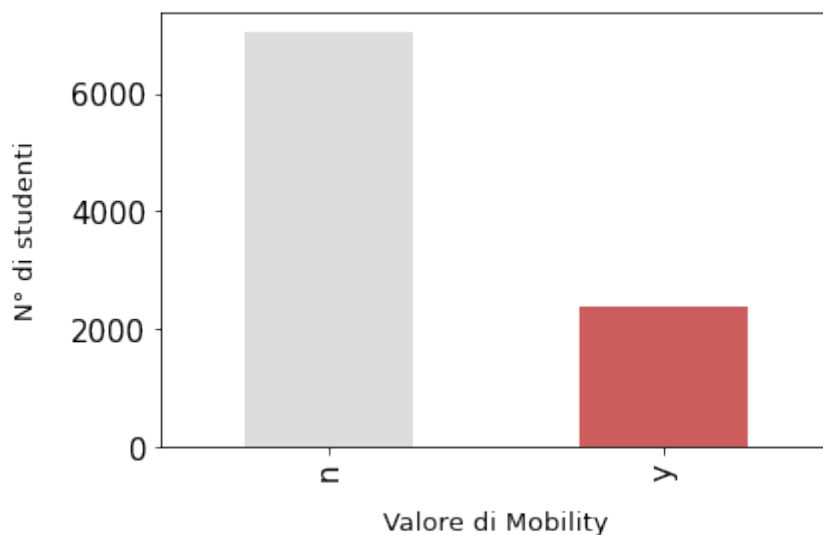


FIGURE 5.18: Conteggio utenti che si spostano all'estero.

Gli studenti che per ragioni di studio o lavoro sono stati almeno una volta fuori dall'Italia sono 2.372, ovvero il 25,2% del totale (Figura 5.18). Il 75,3% di chi parte fa esperienza solo in un'altra nazione, che come si è già sottolineato è spesso una nazione dell'Europa Occidentale, e solo il restante 24,7% si sposta più volte in nazioni diverse.

In totale le nazioni raggiunte dagli studenti sono 125¹⁶, cioè la quasi totalità degli stati riconosciuti sovrani a livello internazionale, ma solo poche di queste contano più di una decina di trasferimenti.

¹⁶Le 125 nazioni sono: Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Benin, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic, Chile, China, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, El Salvador, Eritrea, Estonia, Ethiopia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Haiti, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Ivory Coast, Japan, Jordan, Kazakhstan, Kenya, Kosovo, Latvia, Lebanon, Libya, Liechtenstein, Lithuania, Luxembourg, Malaysia, Maldives, Malta, Mauritania, Mexico, Moldova, Montenegro, Morocco, Myanmar, Netherlands, New Zealand, Niger, Nigeria, Norway, Oman, Pakistan, Palestine, Panama, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russia, Samoa, San Marino, Saudi Arabia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, South Africa, South Korea, South Sudan, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syria, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States, Uzbekistan, Vatican City, Venezuela, Vietnam, Zambia.

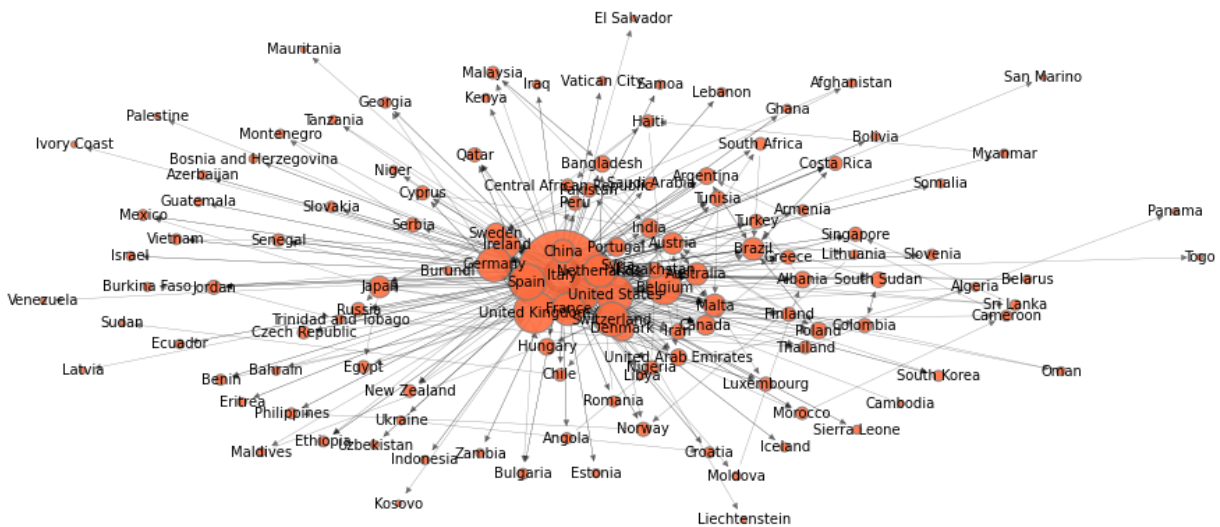


FIGURE 5.19: Rete degli spostamenti degli utenti.

La Figura 5.19 rappresenta la rete direzionata dei movimenti degli studenti, in cui:

- i nodi rappresentano le nazioni. Un nodo è tanto grande quanto è grande il numero di nodi a cui è collegato.
- gli archi rappresentano gruppi di utenti che si spostano da una nazione all'altra.

Il nodo più grande, situato al centro della rete, rappresenta l'Italia, che come si nota è collegato con quasi tutte le altre nazioni. Disposti intorno, ci sono gli altri nodi principali: Regno Unito, Stati Uniti e le altre nazioni già evidenziate nel Paragrafo 5.2.

Oltre ad essere i nodi con il più alto numero di studenti trasferiti, queste nazioni sono anche quelle che contano il maggior numero di collegamenti con terze nazioni¹⁷:

- Sono stati nel Regno Unito un totale di 565 studenti, della quale 502 (88,8%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 49 (8,7%).

¹⁷I dati quantitativi sugli spostamenti tra nazioni degli studenti sono riportati nell'Appendice B.

- Sono stati negli Stati Uniti d'America un totale di 372 studenti, della quale 341 (91,6%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 38 (10,2%).
- Sono stati in Spagna un totale di 267 studenti, della quale 227 provenienti dall'Italia (85,0%). Gli studenti che si sono poi spostati verso altre nazioni estere sono 23 (8,6%).
- Sono stati in Francia un totale di 251 studenti, della quale 227 (90,4%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 21(8,3%).
- Sono stati in Germania 232 studenti, di cui 214 (92,2%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 18 (7,7%).
- Sono stati in Svizzera 185 studenti, di cui 173 (93,5%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 20 (10,8%).
- Sono stati in Belgio 180 studenti, di cui 169 (93,8%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 15 (8,3%).
- Sono stati nei Paesi Bassi un totale di 128 studenti, di cui 120 (93,8%) provenienti dall'Italia. Gli studenti che si sono poi spostati verso altre nazioni estere sono 13 (10,2%).

Delle 125 nazioni estere presenti, sono solo 8 quelle con cui non si registrano connessioni dirette con l'Italia, ovvero: Afghanistan, Camerun, Haiti, Liechtenstein, Lituania, Panama, Sudan e Trinidad e Tobago. Gli scambi di persone tra queste nazioni e l'Italia avvengono attraverso il passaggio degli studenti in altre nazioni, presso cui questi studiano o lavorano prima di trasferirsi nuovamente.

Oltre a queste, ci sono altre 28 nazioni i cui collegamenti diretti con l'Italia sono unidirezionali.

Le nazioni da cui si registrano movimenti diretti verso l'Italia ma verso cui non si registrano movimenti dall'Italia, sono 4: Cambogia, Egitto, Montenegro e Tanzania. La Figura 5.20 riporta in arancione i nodi rappresentanti queste quattro

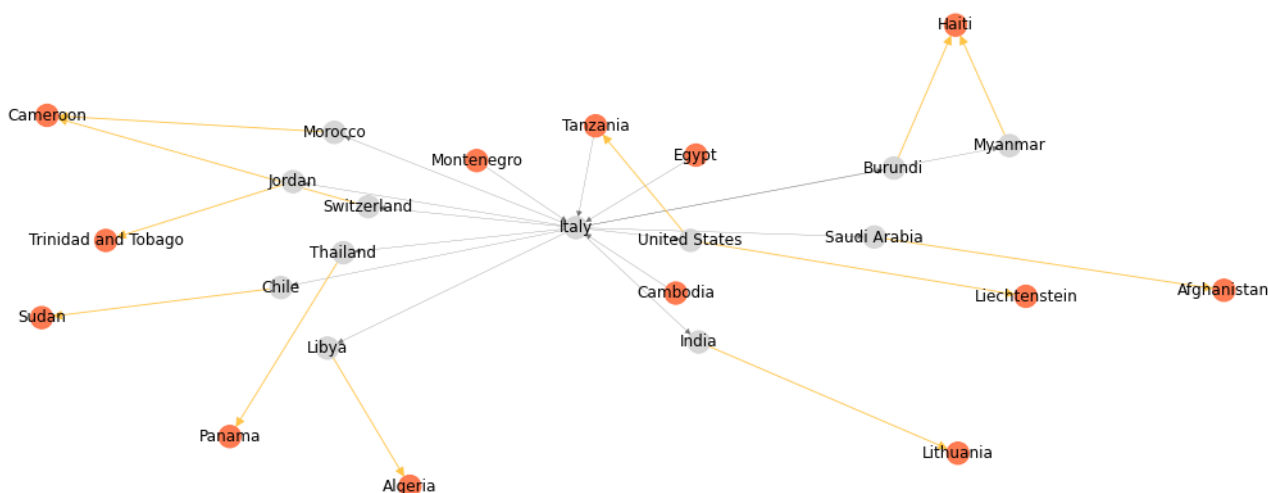


FIGURE 5.20: Nazioni non collegate direttamente con l'Italia (1) .

nazioni, a cui si aggiungono gli otto nodi delle nazioni non collegate direttamente con l'Italia. I collegamenti in giallo rappresentano i movimenti verso queste dodici nazioni, mentre quelli in nero rappresentano sia i movimenti dall'Italia che i movimenti dalle 12 nazioni all'Italia.

Le nazioni da cui non partono movimenti diretti verso l'Italia ma che invece ricevono spostamenti dall'Italia sono 24: Algeria, Bahrein, Burundi, Costa d'Avorio, El Salvador, Kosovo, Lettonia, Libia, Malaysia, Maldive, Mauritania, Moldavia, Myanmar, Nigeria, Nuova Zelanda, Oman, Palestina, San Marino, Siria, Sri Lanka, Thailandia, Togo, Ucraina, Venezuela.

La Figura 5.21 riporta in arancione i nodi di queste nazioni, a cui si aggiungono gli otto nodi delle nazioni non collegate direttamente con l'Italia. I collegamenti in azzurro rappresentano i movimenti dall'Italia, mentre quelli in giallo sono verso l'Italia. I collegamenti grigio chiaro sono i movimenti da e verso i nodi in arancione.

Ricapitolando, dai dati si evince che, oltre che in Italia, mediamente chi parte fa esperienza soltanto in una seconda nazione. Le mete predilette sono quelle dell'Europa Centro-Occidentale, a cui si aggiunge gli Stati Uniti d'America. L'Italia invia e riceve persone da tutti gli altri stati, ad eccezione di una trentina di nazioni.

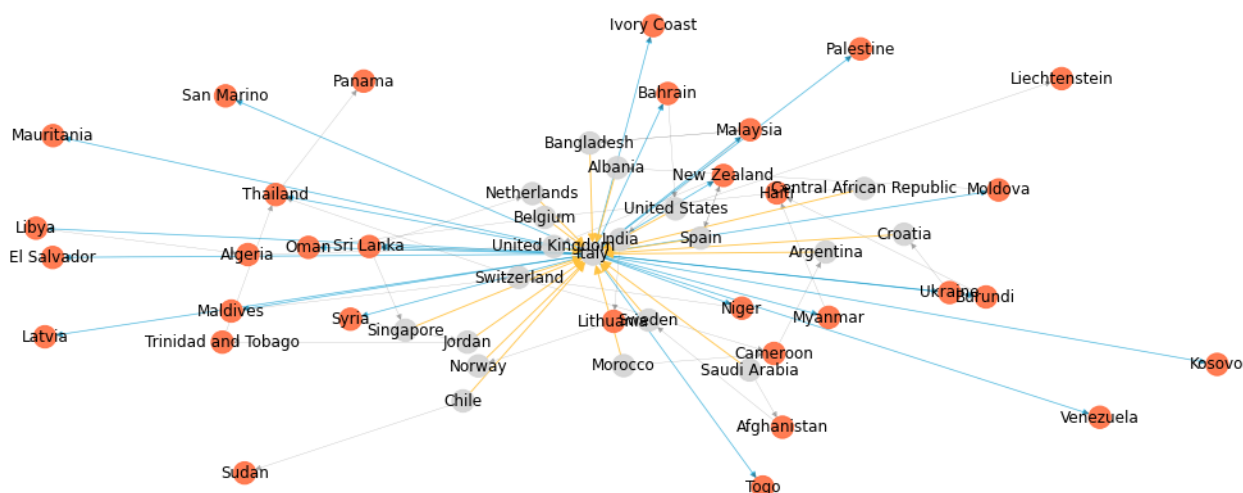


FIGURE 5.21: Nazioni non collegate direttamente con l'Italia (2) .

Tuttavia, questi dati sono poco significativi senza considerare la nazionalità e la nazione di partenza. Considerando che nell' 85,1% dei profili di chi parte la prima nazione indicata cronologicamente¹⁸ è l'Italia, crediamo si possa supporre che la maggioranza degli studenti analizzati sia di nazionalità italiana. Questo ci porterebbe a supporre che i movimenti registrati verso l'Italia si riferiscano principalmente al rientro in patria degli studenti e che siano solo 350 (il restante 14.9%) gli studenti stranieri che invece si spostano dall'estero verso l'Italia per studio e/o lavoro.

Tra le altre nazioni che compaiono come prima località di uno studente¹⁹ primeggiano Regno Unito (61), Spagna (35) e Stati Uniti (22). Se visti in proporzione al numero complessivo di studenti che hanno fatto esperienze in queste nazioni, questi numeri sono piuttosto esigui, indicando che è più probabile che un nostro connazionale si sposti verso queste nazioni che viceversa. Le nazioni da cui invece arrivano più studenti rispetto a quanti ne partono dall'Italia sono invece India (23, cioè il 73% degli utenti che hanno fatto esperienze lì), Iran (26, cioè il 73%) e Turchia (21, cioè il 51%).

La maggior parte degli studenti parte quindi dall'Italia e solo una piccola

¹⁸Questo dato è stato ottenuto non considerando la località indicata nell'attributo "Location".

¹⁹Il conteggio è riportato in Tabella A.10 in Appendice A.

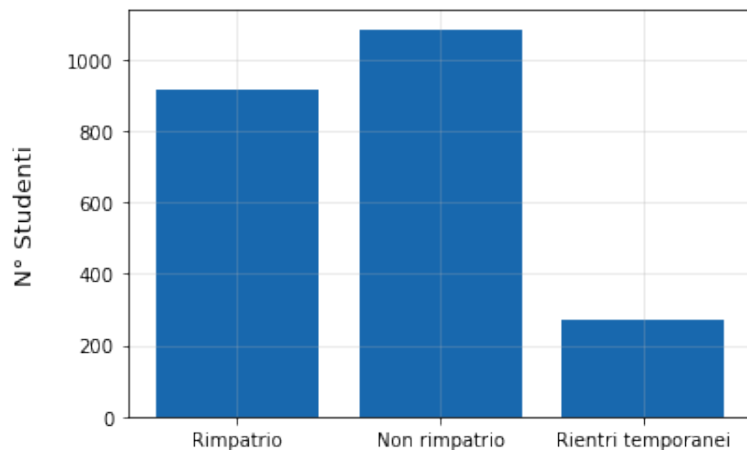


FIGURE 5.22: Numero studenti che rimpatriano.

percentuale si muove verso l'Italia. È quindi corretto supporre che la maggior parte degli archi diretti verso il nodo "Italy" indichino dei rientri in patria. Come mostrato in Figura 5.22, gli studenti che non ritornano nella prima nazione indicata nel loro profilo, sempre escludendo la località generica del profilo, sono il 45,8%, mentre il 38,8% ritorna permanentemente e il 11,6% continua a muoversi tra la nazione di partenza e altre.

5.3.1 - Differenze a livello accademico

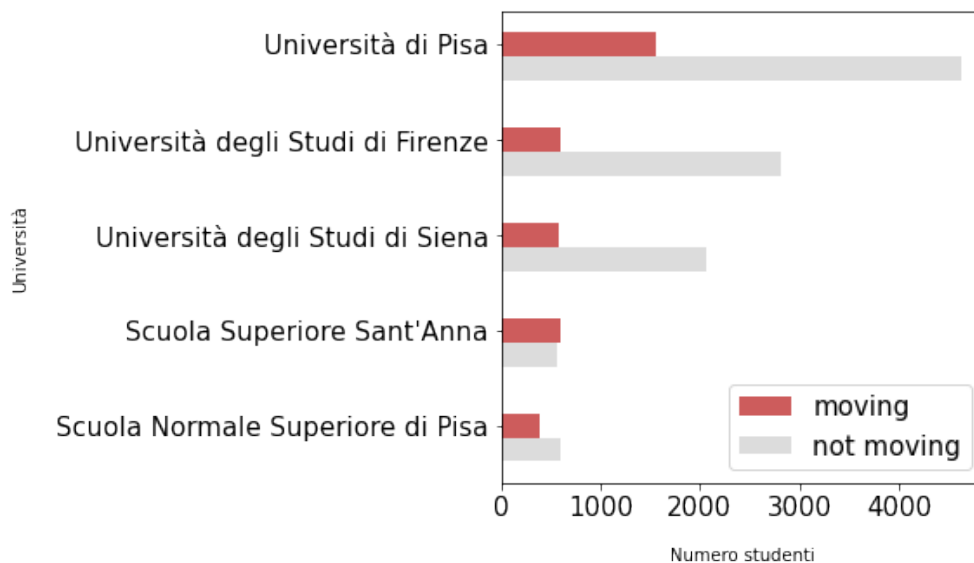


FIGURE 5.23: Numero di studenti che partono per ateneo.

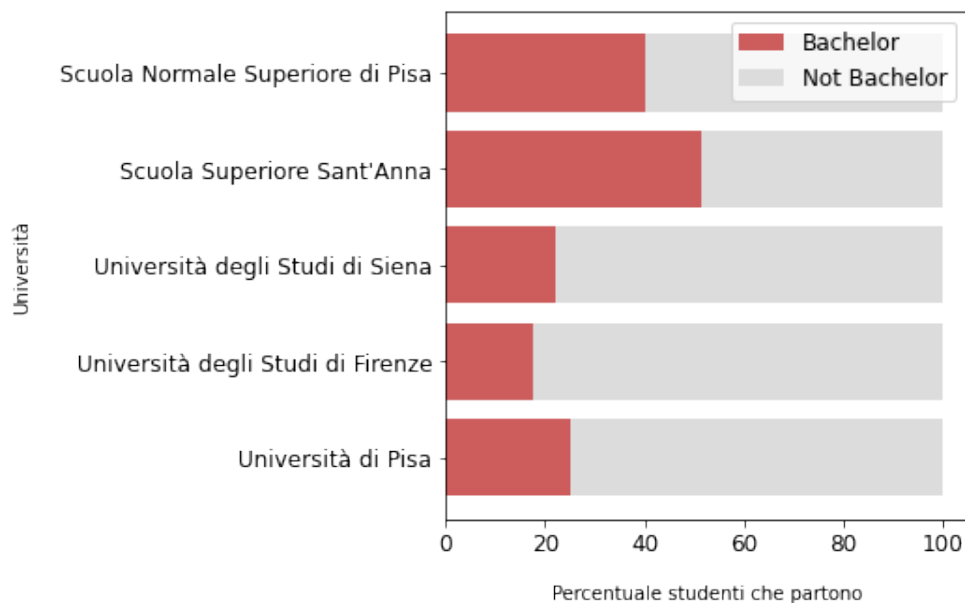


FIGURE 5.24: Percentuali di studenti che partono per ateneo.

Da un punto di vista quantitativo (Figura 5.23), l'ateneo da cui escono più studenti che in seguito si sposteranno all'estero è l'Università di Pisa (1544), seguito dall'Università degli Studi di Firenze (601), Scuola Superiore Sant'Anna (597), Università degli Studi di Siena (588) e infine dalla Scuola Normale Superiore di Pisa (397). Tuttavia, rispetto al numero totale di utenti che vi hanno studiato, gli atenei con la più alta percentuale di studenti che partono sono le due scuole universitarie superiori, 51,4% la Scuola Superiore Sant'Anna e 40,2% la Scuola Normale Superiore di Pisa, mentre gli altri tre atenei si aggirano intorno al 20% (Figura 5.24).

Tra gli studenti che si spostano all'estero, i possessori di una laurea di primo livello sono 927, i possessori di una laurea di secondo livello sono 1622 e i possessori di un dottorato sono 964. Sul totale dei possessori di una laurea di primo e di secondo livello, quelli che decidono di partire sono solo una piccola parte (24,6% per i primi e 29,2% per i secondi), mentre per i possessori di un dottorato sono poco meno della metà, il 43,8%).

Guardando gli stessi dati in relazione al numero complessivo di chi parte e chi resta invece che al totale dei possessori di un certo titolo (Figure 5.25), si osserva che:

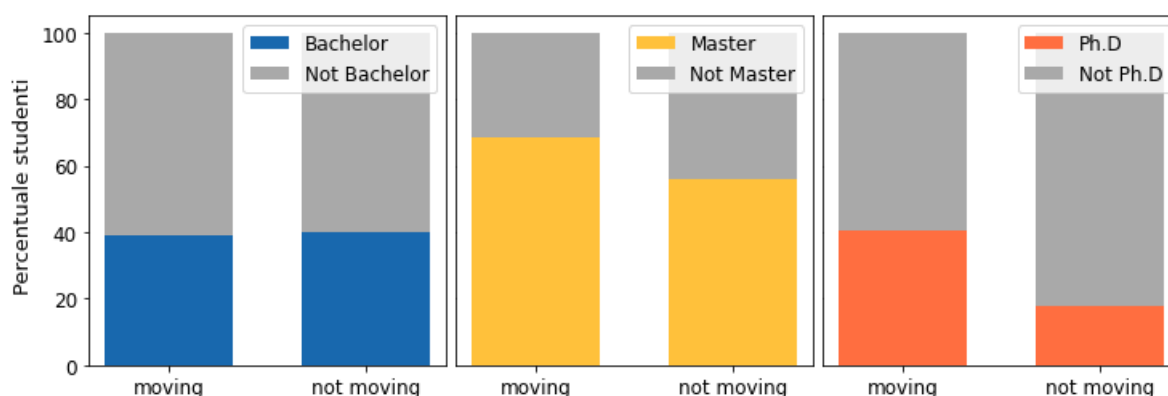


FIGURE 5.25: Percentuali possessori tra laurea tra studenti che partono e che restano.

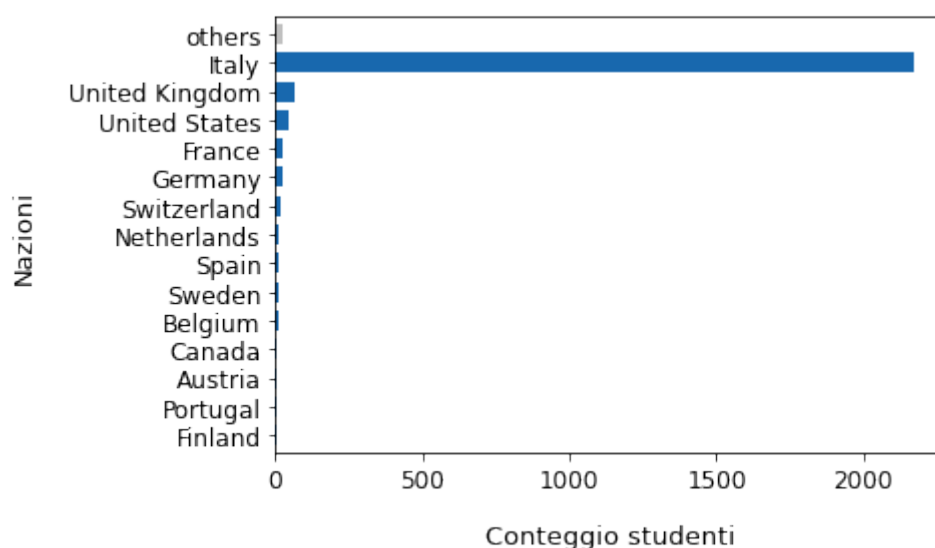


FIGURE 5.26: Luoghi in cui hanno fatto il dottorato gli studenti che partono.

- vi è una similitudine nelle percentuali dei possessori di una laurea di primo livello tra chi parte e chi resta;
- il 70,0% di chi parte ha una laurea di secondo livello, contro il 55,9% di chi resta;
- il 40,6% di chi parte ha un dottorato, contro il 17,6% di chi resta.

La Figura 5.26 mostra le località in cui hanno ottenuto il titolo di dottorato gli studenti che partono²⁰. La maggior parte sono stati ottenuti in Italia e in par-

²⁰Il conteggio è riportato in Tabella A.11 in Appendice A. Si noti che il conteggio fa riferimento al numero di titoli di dottorato ottenuti.

ticolar modo nelle università toscane:

- Scuola Superiore Sant'Anna (254);
- Università di Pisa (216);
- Scuola Normale Superiore (186);
- Università degli Studi di Siena (97);
- Università degli Studi di Firenze (68);

Quelli ottenuti all'estero sono stati presi nelle università:

- University of Cambridge (7);
- Ecole Polytechnique Fédérale de Lausanne (6);
- Harvard University (6);
- University of Oxford (5);

Per quanto riguarda gli argomenti di studio (Figura 5.27) per ogni argomento la percentuale di studenti e studentesse che si spostano all'estero si aggira intorno al $25.8 \pm 10\%$, in linea con quanto visto sulla totalità degli studenti. Si allontanano da questa stima solo tre argomenti: "Health Profession", per cui nessuno studente è andato a fare esperienza all'estero; "Veterinary Studies", di cui vanno all'estero il 40% degli studenti e "Psychology", con il 12%. Va tuttavia considerato che per tutti e tre questi campi di studio si posseggono pochi dati e che quindi non sono attendibili statisticamente.

5.3.2 - Differenze percorsi professionali

Come mostrato in Figura 5.28 per gli studenti che vanno all'estero e in Figura 5.29 per quelli che restano in patria, è possibile tracciare alcune sottili differenze anche al livello delle esperienze lavorative.

Relativamente ai tirocini, nelle prime esperienze lavorative si registra una differenza di 8 punti percentuali tra i due gruppi. Circa il 27,6% di chi parte ha

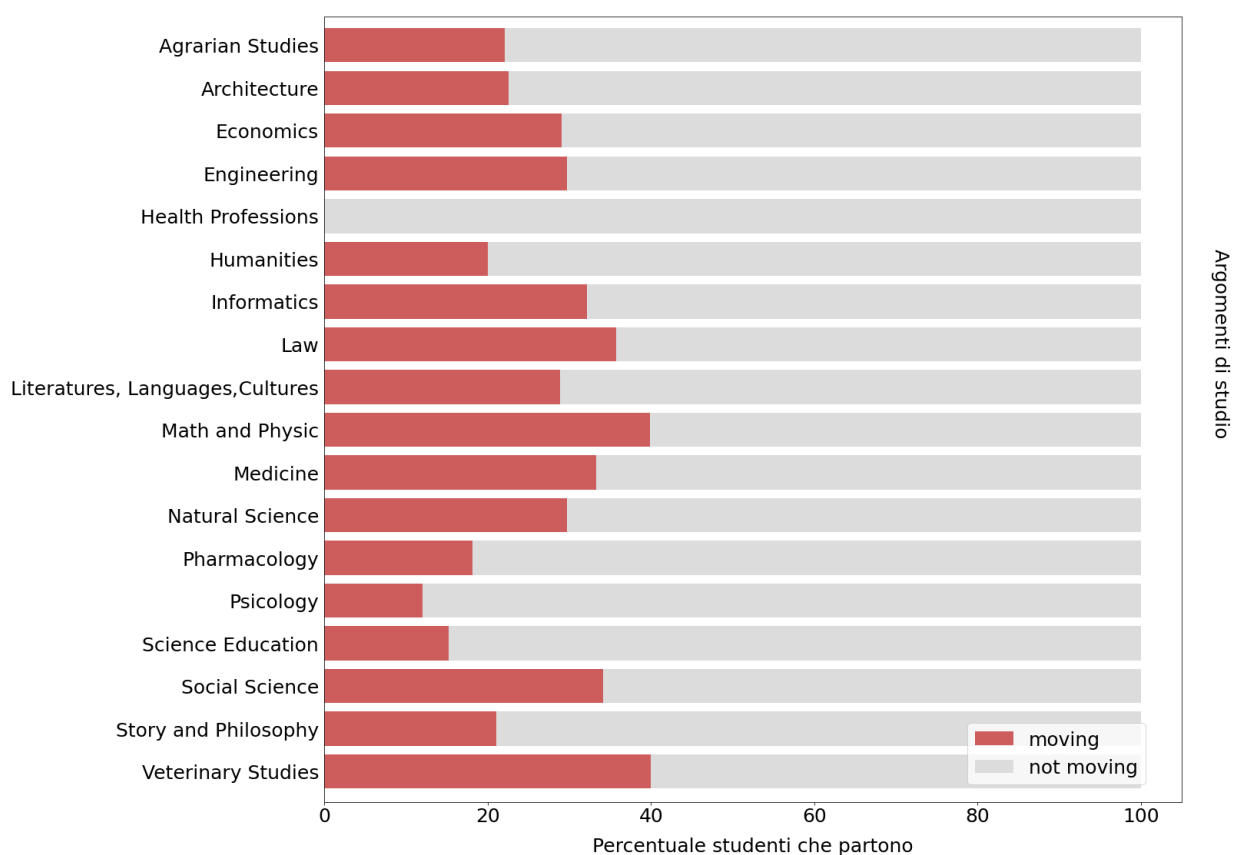


FIGURE 5.27: Percentuali di mobilità per settore accademico.

un tirocinio come prima esperienza lavorativa e il 22,3% come seconda. Invece, tra quelli che restano in Italia solo il 19,4% lo ha come prima esperienza e il 14,8% come seconda. In entrambi i gruppi il numero di tirocini decresce all'avanzare del numero di esperienze, sebbene più rapidamente tra gli studenti che partono. Alla quinta esperienza lavorativa le percentuali sono al 6,6% per la prima categoria e al 5,2% per la seconda e alla sesta scendono entrambe intorno al 3%.

Diversamente, le percentuali dei lavoratori in prima fascia sono maggiori tra chi resta e questo dato rimane costante più o meno a prescindere dal numero di esperienze lavorative, assestandosi su una differenza di circa 5 punti percentuali. Le percentuali restano più o meno stabili nelle prime quattro esperienze lavorative, aggirandosi intorno al 55,6% per gli studenti che partono e al 61,1% per l'altro gruppo, con un lieve picco intorno alla terza esperienza professionale, dato nella nostra opinione dalla diminuzione dei numeri di tirocini. Dalla quinta in poi si registra anche in questo caso una progressiva diminuzione delle percentuali.

Il 12,2% degli studenti che si spostano e il 12,4% degli studenti che restano in Italia dichiarano di aver svolto una professione di seconda fascia come prima esperienza professionale. In entrambi i gruppi le percentuali relative a questa fascia aumentano con il numero di esperienze professionali, ma in generale nel primo gruppo si registra un aumento più rapido che nel secondo. La differenza tra i due gruppi è di circa un punto percentuale fino alle settima esperienza professionale, nell'ottava la differenza sale a 10 punti percentuali (38,9% per gli studenti che partono e 28,2% per quelli che restano), ma nella nona e nella decima la tendenza si inverte, con una percentuale maggiore in chi resta. Per la lettura di questi dati è comunque importante considerare la distribuzione delle esperienze professionali riportata nel Capitolo 5.2: nel dataset, il numero di esperienze professionali successive alla sesta è molto basso.

Infine, le percentuali relative alle professioni di terza fascia sono più alte tra chi resta. Il 6,8% di chi resta dichiara di aver svolto una professione di terza fascia come prima esperienza professionale, contro il 5,7% di chi parte. Le percentuali salgono nella quinta esperienza professionale al 15,0% per chi resta e al 12,7% per chi parte.

Un'ultima osservazione riguarda le professioni svolte dagli utenti appena

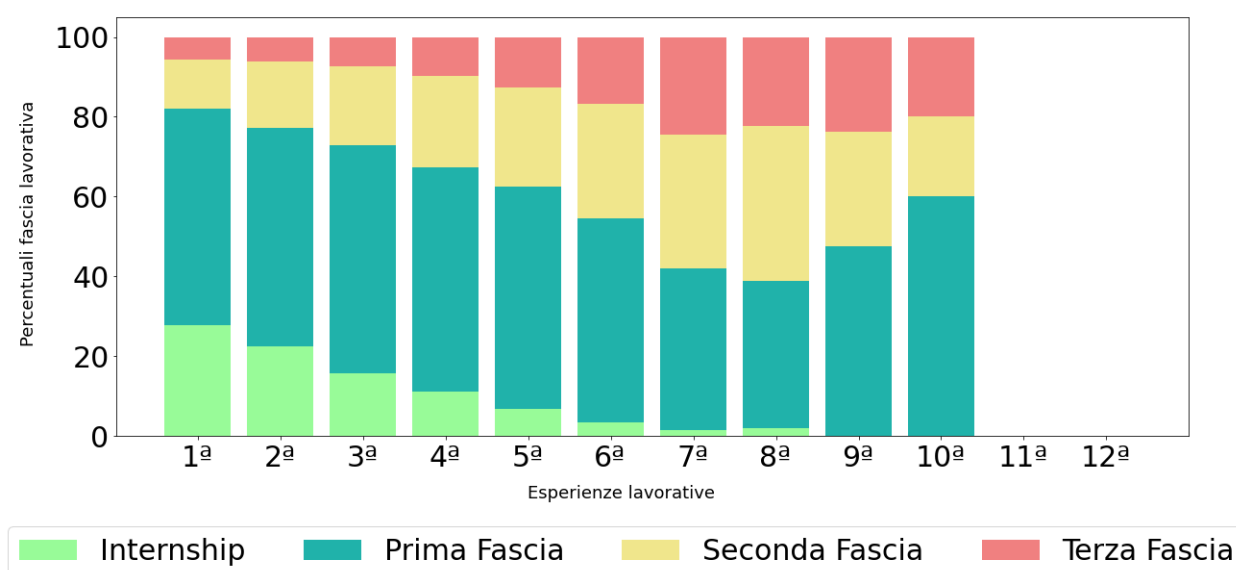


FIGURE 5.28: Percentuali livello professionale per studenti che partono.

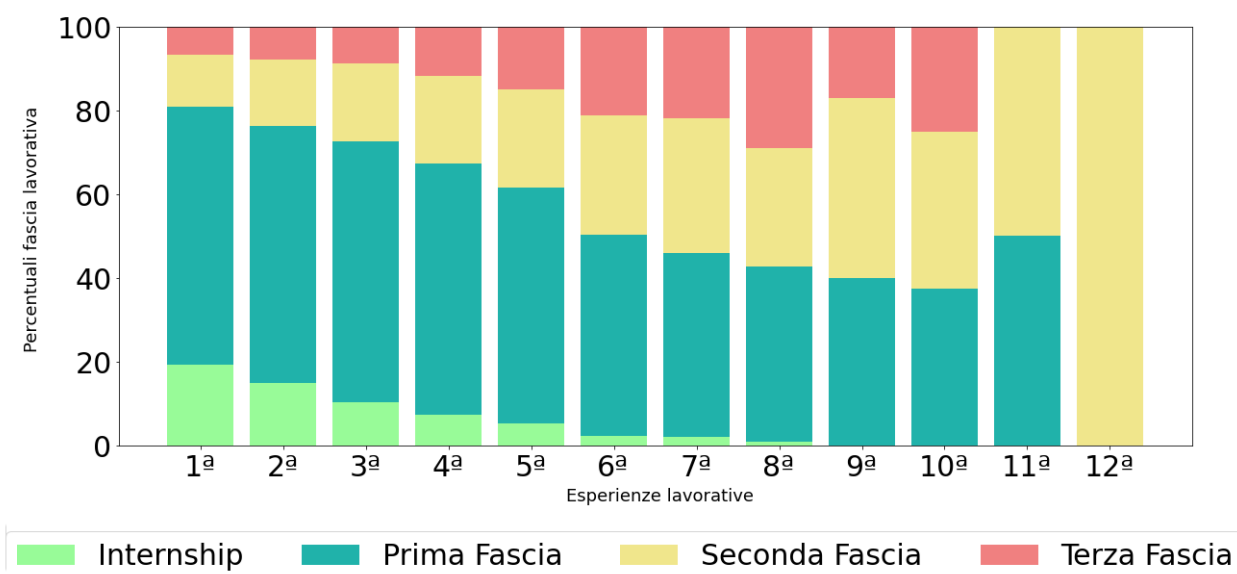


FIGURE 5.29: Percentuali livello professionale per studenti che restano.

Internship	564
Prima Fascia	932
Seconda Fascia	281
Terza Fascia	86

TABLE 5.2: Prima professione svolta fuori dall'Italia.

usciti dall'Italia (Tabella 5.2) e quelle degli utenti appena rientrati in patria. La prima professione svolta dal 39,3% degli studenti appena giunti all'estero appartiene alla prima fascia, il 23,8% fa un tirocinio, l' 11,8% un lavoro di seconda fascia e soltanto il 3,6% un lavoro di terza fascia. Il restante 21,5% si sposta solo per motivi di studio e non fa all'estero alcuna esperienza professionale.

Come mostrato in Figura 5.30, il rientro in patria coincide con un avanzamento di carriera solo nel 31,6% dei casi. La maggior parte di quelli che hanno un avanzamento di carriera hanno svolto all'estero un tirocinio e, una volta rientrati in Italia, hanno svolto una professione di prima fascia (47,9%). Il 31,0% torna in Italia per svolgere una professione di seconda fascia e il 21,1 per cento un lavoro di terza.

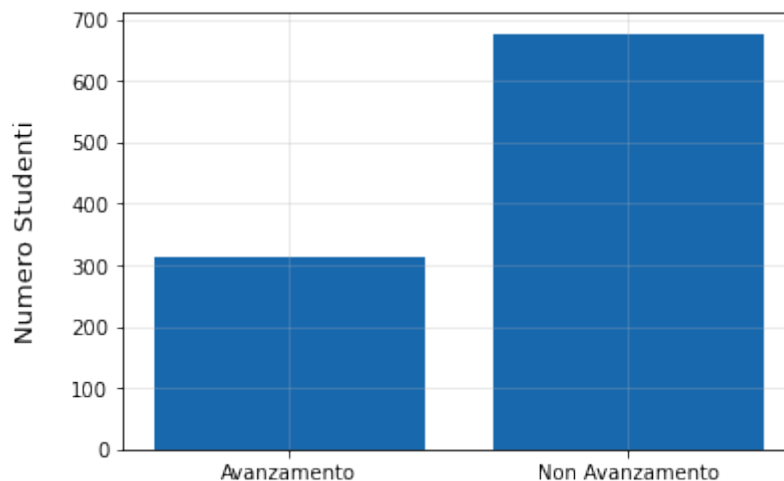


FIGURE 5.30: Quanti avanzano di carriera una volta rientrati in patria.

5.4 | *Discussione*

Data la fine risoluzione dei dati e delle analisi proposte, di seguito sono riassunte le principali informazioni che sono state ottenute.

Dall'analisi esplorativa condotta si evince che, in media, gli alunni e le alumnae degli atenei toscani tendono a conseguire tutti i loro percorsi di studi all'interno della regione. In particolare, la maggioranza di questi ha un titolo di istruzione terziaria di secondo livello (laurea magistrale o a ciclo unico) ottenuto in uno dei tre atenei di Firenze, Pisa o Siena e si è specializzata nel settore economico, spesso dopo aver frequentato una triennale in un ambito differente. Ciononostante, i settori che attirato più studenti in Toscana sono Ingegneria e Scienze Naturali, mentre la maggior parte degli studi in Scienze della Formazione sono fatti fuori dalla regione.

Meno di un quarto del campione indagato ha un titolo di dottorato. Gli atenei che producono il maggior numero di dottori sono l'Università di Pisa, la Scuola Normale Superiore e la Scuola Superiore Sant'Anna, soprattutto nel settore delle Scienze Naturali. Tuttavia, in proporzione al numero di studenti che hanno conseguito un titolo terziario di secondo livello nel settore, gli studenti di Matematica e Fisica sono più propensi a prendere un dottorato sull'argomento.

Relativamente alle esperienze lavorative, si nota tra gli alunni un lungo pe-

riodo di gavetta. In generale, le prime quattro mansioni svolte sono come tirocinante, stagista o figura junior. Sono quindi necessari, in media, almeno cinque passaggi di carriera prima di ottenere un titolo di seconda fascia.

Circa un terzo del campione si muove all'estero, probabilmente per cercare lavoro. La maggior parte di chi si muove lo fa dopo aver completato il percorso di studi in Italia e circa la metà di questi dopo aver ottenuto un titolo di dottorato in uno dei due istituti di studi superiori universitari. I settori più interessati dal fenomeno sono quelli STEM, cioè Scienze, Tecnologie, Ingegneria e Matematica, e particolarmente il settore di Matematica e Fisica.

Le mete preferite dagli alumni sono Stati Uniti e Regno Unito, ma in generale possiamo affermare che gli spostamenti fuori dall'Europa Occidentale (ad eccezione di quelli verso gli USA) sono una minima parte.

In generale, le mansioni ricoperte dagli alumni e dalle alumnae una volta fuori dall'Italia non sono molto alte: anche in questo caso tirocini, stage e figure junior. In generale però, a prescindere che rimangano fuori o ritornino in patria (circa la metà), chi decide di fare esperienze lavorative fuori dall'Italia raggiunge più velocemente una mansione di seconda fascia. Viceversa, chi resta in Italia generalmente raggiunge prima una professione di terza fascia, cioè una mansione dirigenziale. Una possibile ipotesi è che ciò sia dovuto ad maggior numero di alumni che, restando in Italia, decidono di mettersi in proprio e aprire la propria attività.

6 || Problematiche e Sviluppi futuri

Da un punto di vista strutturale, un aspetto migliorabile riguarda il tempo di esecuzione del modello di preparazione dei dati. Allo stato attuale, l'algoritmo esegue un ciclo sul dataset per ogni operazione di pulizia descritta nel Paragrafo 4.2. Ciò allunga considerevolmente i tempi di esecuzione, specialmente per dataset molto grandi. Un possibile futuro miglioramento potrebbe quindi riguardare la riduzione del numero delle iterazioni necessarie alla pulizia. Tuttavia, una soluzione più performante sarebbe l'utilizzo di tecniche di vettorializzazione o l'utilizzo di Cython¹.

Un altro problema è riscontrabile nella standardizzazione dei nomi delle università. Come si è visto nel Paragrafo 4.2.3, questo processo richiede l'utilizzo di strumenti di controllo (liste e dizionari contenenti i nominativi standardizzati delle varie istituzioni) creati sulla base dei dati raccolti. Questi strumenti devono infatti essere "esaustivi", cioè contenere i nominativi degli istituti interessati dall'analisi, ma non devono essere troppo "estesi", cioè non devono contenere nominativi di troppe università: questo perché da una parte l'utilizzo di una lista incompleta causerebbe inevitabilmente la perdita di informazioni, non riuscendo a ricollegare efficientemente le nomenclature; ma dall'altra parte, una lista troppo estesa favorirebbe l'insorgenza di errori, a causa della somiglianza dei nominativi di molte università. Per questa stessa ragione è cruciale anche il margine di similitudine impostato su Fuzzywuzzy. Un margine troppo alto potrebbe impedire il corretto riconoscimento dell'università, ma un margine troppo basso potrebbe condurre ad una associazione erronea. Ad esempio, con un margine di similitudine del 97% "University of Ghent" sarebbe associato erroneamente a "University of Trento", ma con la stessa percentuale "Scuola Normale Superiore di Pisa" non verrebbe correttamente riconosciuta come "Scuola Normale Superiore".

Gli strumenti creati e utilizzati in questo lavoro fanno quindi riferimento esclusivamente alle università italiane e alle università più grandi degli stati esteri. Sebbene il nostro algoritmo supervisionato permetta di aggiornare le liste già create, per successive analisi consigliamo la creazione di strumenti di controllo appositi.

¹<https://pandas.pydata.org/pandas-docs/stable/userguide/basics.html>

Un altro sviluppo futuro potrebbe riguardare l'analisi temporale annuale degli spostamenti degli utenti. Considerando il dataset come una rete dinamica, i cui nodi sono le diverse nazioni e gli archi i movimenti degli utenti, sarebbe possibile studiare come cambiano gli spostamenti tra le nazioni anno per anno. Utilizzando questo approccio, sarebbe eventualmente possibile utilizzare degli algoritmi di *link prediction* per fare una previsione degli spostamenti nel futuro.

Infine, per la visualizzazione dei dati potrebbe essere creata una piattaforma interattiva che permetta di navigare agevolmente tra i risultati ottenuti, evidenziando le statistiche ottenute per le varie categorie.

7 || Conclusioni

In questa tesi è stato presentato un approccio per la collezione e analisi di dati non tradizionali relativi alle carriere professionali di persone altamente formate. Il framework permette di collezionare dati, in maniera gratuita, direttamente dalla pagina web del professional network LinkedIn. I dati ottenibili hanno copertura globale, sono aggiornati e sempre disponibili, ma generalmente poco uniformi dal punto di vista della terminologia usata e della tipologia di informazioni presenti. Attraverso il processo di pulizia è stato ottenuto un dataset di utenti con informazioni sul settore di specializzazione, gli enti presso cui ha studiato e/o lavorato e le coordinate geografiche e cronologiche delle esperienze effettuate. Il dataset è stato utilizzato per proporre una analisi dettagliata dal punto di vista degli utenti e delle istituzioni. I risultati ottenuti forniscono importanti informazioni utili per il monitoraggio e l'analisi dello scambio culturale tra istituzioni e nazioni.

Tale studio è particolarmente importante al giorno d'oggi a causa dell'aumento in scala globale delle migrazioni di persone altamente qualificate [Meyer, 2001], con le conseguenze politiche, sociali ed economiche che queste comportano. Questo studio nasce dalla necessità di aggirare i limiti imposti dai dati tradizionali che non permettono di inquadrare interamente il fenomeno, a causa della mancanza di informazioni sui paesi in via di sviluppo, della mancanza di informazioni sui settori professionali più colpiti e sulla mancanza di dati costantemente aggiornati.

Un esempio di utilizzo dell'approccio proposto è stato applicato ad un caso di studio rivolto alle carriere degli alumni e delle alumnae degli atenei toscani, con particolare attenzione ai fenomeni di mobilità internazionale. Questo lavoro si distingue da [Johnson et al.,] nell'obiettivo dichiarato di offrire un esempio di analisi dei movimenti internazionali con l'utilizzo di dati accessibili gratuitamente da chiunque abbia un profilo LinkedIn, dove l'altro studio propone un esempio di analisi con l'utilizzo di dati accessibili esclusivamente con la sottoscrizione di un abbonamento a LinkedIn, cioè con il pagamento del piano Recruiter.

Ciò che si denota dai dati raccolti è che la maggior parte delle informazioni presenti sul professional network si riferiscono ad esperienze accademiche e pro-

fessionali tenute in Italia nell'arco temporale che va dall'anno 2000 all'anno 2021, cioè a dati più recenti rispetto a quelli ottenibili dalle fonti tradizionali. Il 59,0% degli alumni e delle alumne degli atenei toscani iscritti a LinkedIn ha un titolo di studio terziario di secondo livello, cioè una laurea magistrale o a ciclo unico, e che circa il 49,0% si è specializzato nel settore STEM ed in particolare in Economia (17,3%), Scienze Naturali (7,8%) e Ingegneria (6,8%). L'elevato numero di studenti STEM presente nel dataset potrebbe spiegarsi con una affluenza maggiore di studenti a questa tipologia di facoltà piuttosto che a quelle non-STEM. Tuttavia, data la natura dei dati, è possibile ipotizzare anche una maggiore iscrizione nella piattaforma di LinkedIn di studenti STEM piuttosto che di studenti non-STEM. In questa prospettiva, i numeri relativamente bassi di studenti del settore sanitario individuati nel dataset si potrebbero interpretare come una minore iscrizione di questi nel professional network. In considerazione dello studio [Baruffaldi et al., 2017], l'alta percentuale nel dataset di dottori in Scienze Naturali e di dottori in Matematica e Fisica porterebbe a pensare che questi siano quelli più propensi ad iscriversi su LinkedIn, e quindi a cercare collaborazioni internazionali ed esperienze lavorative fuori dal mondo accademico.

Il 25,2% del campione studiato è interessato da fenomeni di mobilità internazionale. Le principali mete individuate sono quelle tradizionali del fenomeno del brain drain [Carrington and Detragiache, 1998] [Docquier and Marfouk, 2004], ovvero Stati Uniti d'America e le nazioni dell'Europa Occidentale, soprattutto Regno Unito, Francia, Germania. A queste si aggiunge la Spagna che, come suggerito da [Minnecci, 2015] e [Johnson et al.,], attira specificatamente molti italiani grazie alle similitudini culturali che intercorrono tra i due paesi.

Sebbene i dati di LinkedIn non forniscano indicazioni precise sulla nazionalità degli utenti, con queste nazioni sembrerebbe evidenziarsi un fenomeno di *brain exchange*, in quanto sono presenti sia spostamenti che iniziano dall'Italia verso queste nazioni, sia spostamenti che iniziano da queste nazioni verso l'Italia, e quindi verso gli atenei toscani. Invece, gli spostamenti dall'Asia, e in particolar modo da Iran, India, Turchia e Pakistan, verso l'Italia si connotano come *brain gain*, dato lo scarso afflusso di italiani verso queste nazioni.

Inoltre, circa il 50% dei casi di mobilità internazionale si configurano come “brain circulation”, cioè spostamenti finalizzati alla specializzazione professionale con conseguente rientro in patria.

In linea con i dati ISTAT in [Minnecci, 2015], nel dataset emerge un’alta percentuale (40,6%) di possessori di un titolo di dottorato tra gli studenti che lasciano il paese, e in particolare di dottori in Matematica e Fisica. Secondo lo studio, ciò è dovuto alla mancanza di investimenti da parte dell’Italia in ricerca e sviluppo, con una conseguenziale carenza di laboratori e strumentazioni adeguate.

In conclusione, le evidenze ottenute nel nostro lavoro a partire dai dati di LinkedIn sono in linea con gli studi precedenti condotti con l’utilizzo di dati tradizionali, particolarmente in relazione all’alta percentuale di dottori che decidono di lasciare il paese e alla scelta delle mete. A nostro parere, dunque, LinkedIn potrebbe rappresentare una valida fonte di dati sulla mobilità internazionale, utilizzabile per colmare le carenze dei database tradizionali e la mancanza di informazioni sui paesi non appartenenti all’OECD.

A || Tabelle

Valore di Location	Conteggio
Italy	8858
United Kingdom	81
United States	72
Germany	55
France	53
Switzerland	50
Spain	35
Netherlands	33
Belgium	28
Denmark	9
Ireland	9
Austria	8
Luxembourg	8
China	5
Brazil	5
Singapore	4
Canada	4
Sweden	4
Malta	4
United Arab Emirates	3
Serbia	3
Finland	3
Albania	3
Mexico	3
Turkey	3
Iran	3
Colombia	3
Norway	3
Greece	2

Valore di Location	Conteggio
Peru	2
Poland	2
Tunisia	2
Kenya	2
Argentina	2
Chile	2
Qatar	2
Bangladesh	2
Hungary	2
Pakistan	2
Slovakia	2
Australia	2
Kazakhstan	2
Philippines	2
India	2
Somalia	2
Japan	2
Ethiopia	1
Bosnia and Herzegovina	1
Azerbaijan	1
South Sudan	1
Portugal	1
Bulgaria	1
Croatia	1
Cambodia	1
Romania	1
New Zealand	1
Ecuador	1
Central African Republic	1

Valore di Location	Conteggio
Georgia	1
Egypt	1
Vietnam	1
Czech Republic	1
Sierra Leone	1

TABLE A.1: Conteggio valori di Location.

Anno	N° di percorsi
2022	2
2021	308
2020	1050
2019	1324
2018	1392
2017	1380
2016	1385
2015	1388
2014	1311
2013	1193
2012	1153
2011	1072
2010	994
2009	838
2008	803
2007	707
2006	616
2005	541
2004	435
2003	408
2002	326

Anno	N° di percorsi
2001	338
2000	287
1999	254
1998	214
1997	199
1996	214
1995	172
1994	173
1993	165
1992	166
1991	139
1990	118
1989	111
1988	75
1987	84
1986	88
1985	92
1984	78
1983	62
1982	62
1981	39
1980	51
1979	67
1978	45
1977	35
1976	28
1975	34
1974	31
1973	24

Anno	N° di percorsi
1972	11
1971	15
1970	13
1969	14
1968	18
1967	9
1966	3
1965	4
1964	9
1963	3
1962	5
1961	2
1960	5
1957	1
1956	1
1955	1
1952	1
1948	1
1946	2
1943	1

TABLE A.3: Numero di percorsi per data di inizio.

Anno	N° di percorsi - Toscana
2022	1
2021	94
2020	500
2019	717
2018	836
2017	836

Anno	N° di percorsi - Toscana
2016	927
2015	919
2014	867
2013	716
2012	657
2011	596
2010	526
2009	428
2008	476
2007	397
2006	371
2005	332
2004	255
2003	244
2002	202
2001	210
2000	170
1999	163
1998	134
1997	127
1996	143
1995	119
1994	116
1993	113
1992	121
1991	100
1990	78
1989	76
1988	52

Anno	N° di percorsi - Toscana
1987	59
1986	65
1985	56
1984	52
1983	41
1982	47
1981	26
1980	35
1979	49
1978	34
1977	21
1976	20
1975	33
1974	19
1973	14
1972	9
1971	14
1970	9
1969	9
1968	13
1967	8
1966	1
1965	1
1964	8
1963	3
1962	1
1961	2
1960	4
1957	1

Anno	N° di percorsi - Toscana
1948	1
1946	1
1943	1

TABLE A.4: Numero di percorsi in Toscana per data di inizio.

Nazione	N° Utenti Lavoratori
Italy	14331
United Kingdom	502
United States	328
France	249
Germany	213
Spain	186
Switzerland	175
Belgium	168
Netherlands	105
Ireland	62
China	51
Iran	47
Austria	40
Denmark	37
Brazil	36
India	36
Russia	35
Turkey	35
Australia	35
Portugal	33
Sweden	30
Hungary	29
Pakistan	23

Nazione	N° Utenti Lavoratori
Canada	22
Malta	21
Poland	21
Japan	19
United Arab Emirates	16
Luxembourg	14
Albania	11
Czech Republic	11
Singapore	11
Bangladesh	10
Serbia	8
Chile	7
South Africa	7
Israel	7
Finland	7
Romania	7
Norway	7
Lebanon	7
Nigeria	6
Colombia	6
South Korea	6
Tunisia	6
Qatar	6
Greece	5
Ghana	5
Ethiopia	5
Costa Rica	5
Azerbaijan	5
Peru	5

Nazione	N° Utenti Lavoratori
Georgia	5
Saudi Arabia	5
Cyprus	5
Sri Lanka	5
Kazakhstan	5
South Sudan	5
Thailand	5
Zambia	4
New Zealand	4
Morocco	4
Mexico	4
Bulgaria	4
Croatia	4
Jordan	4
Kosovo	4
Central African Republic	4
Argentina	3
Estonia	3
Somalia	3
Ecuador	3
Malaysia	3
Guatemala	3
Cameroon	3
Holland	3
Philippines	2
Ukraine	2
Vietnam	2
Syria	2
Bolivia	2

Nazione	N° Utenti Lavoratori
Kenya	2
Angola	2
Samoa	2
Slovakia	2
Uzbekistan	2
Eritrea	2
Armenia	2
Belarus	2
Iceland	2
Benin	2
Haiti	2
Slovenia	2
Bosnia and Herzegovina	2
Algeria	2
Niger	2
Senegal	2
Egypt	2
Oman	2
Venezuela	2
Maldives	1
Bahrain	1
Iraq	1
Sierra Leone	1
Burundi	1
Burkina Faso	1
Vatican City	1
Indonesia	1
Myanmar	1
Libya	1

Nazione	N° Utenti Lavoratori
Afghanistan	1
Palestine	1
Montenegro	1
Togo	1
Trinidad and Tobago	1
Lithuania	1
Latvia	1
Tanzania	1
San Marino	1
Moldova	1
Mauritania	1
Liechtenstein	1
Ivory Coast	1
El Salvador	1
Panama	1
Sudan	1

TABLE A.7: Numero di utenti che hanno lavorato in una nazione.

Anno	N° di esperienze professionali
2021	1698
2020	3183
2019	3835
2018	3685
2017	3163
2016	2586
2015	2293
2014	1897
2013	1559
2012	1346

Anno	N° di esperienze professionali
2011	1148
2010	961
2009	788
2008	710
2007	564
2006	518
2005	460
2004	368
2003	304
2002	279
2001	317
2000	276
1999	172
1998	182
1997	120
1996	106
1995	102
1994	54
1993	54
1992	87
1991	63
1990	72
1988	51
1989	36
1987	36
1986	28
1985	16
1984	19
1983	16

Anno	N° di esperienze professionali
1982	27
1981	20
1980	25
1979	4
1978	11
1977	6
1976	5
1975	11
1974	4
1973	3
1972	2
1971	3
1970	6
1969	4
1968	2
1967	1
1965	1
1964	5
1960	1
1959	4
1956	1
1929	2
1900	1

TABLE A.8: Esperienze lavorative per anno.

Nazione	N°
United Kingdom	61
Spain	35
United States	22

Nazione	N°
Iran	19
France	17
India	17
Germany	15
Turkey	11
Switzerland	9
Russia	8
Belgium	8
Brazil	8
China	7
Netherlands	7
Ireland	7
Pakistan	7
Denmark	5
Poland	5
Portugal	4
Canada	4
Bangladesh	4
Austria	4
Australia	3
United Arab Emirates	3
Hungary	3
Ghana	2
Chile	2
Morocco	2
Ethiopia	2
Albania	2
Azerbaijan	2
Colombia	2

Nazione	N°
Mexico	2
Sweden	2
Cyprus	2
Saudi Arabia	2
Serbia	2
Lebanon	2
Sri Lanka	2
Israel	2
Malaysia	2
Philippines	1
Nigeria	1
Czech Republic	1
Ukraine	1
Romania	1
Senegal	1
Malta	1
Costa Rica	1
South Africa	1
Somalia	1
Ecuador	1
Georgia	1
Bahrain	1
Sierra Leone	1
Singapore	1
Finland	1
Tunisia	1
Croatia	1
Kazakhstan	1
Uzbekistan	1

Nazione	N°
Jordan	1
Myanmar	1
South Korea	1
Syria	1
Libya	1
Thailand	1
Armenia	1
Belarus	1

TABLE A.10: Numero di utenti che hanno la nazione X come prima nazione.

Nazione	N° Percorsi
Italy	908
United Kingdom	64
United States	44
France	23
Germany	22
Switzerland	18
Netherlands	13
Spain	12
Sweden	8
Belgium	8
Canada	5
Austria	5
Portugal	5
Finland	5
Australia	4
Luxembourg	3
Japan	3
Denmark	2

Nazione	N° Percorsi
Greece	2
Ireland	2
Norway	2
Hungary	2
Iran	1
Singapore	1
Argentina	1
Holland	1
Israel	1
Brazil	1

TABLE A.11: Località e numero di dottorati ottenuti dagli studenti che partono.

Nazione	N° studenti
United Kingdom	205
Spain	125
United States	105
France	65
Germany	56
Belgium	42
Netherlands	37
Ireland	25
Switzerland	24
Denmark	23
Portugal	16
Canada	12
Austria	11
Sweden	8
Finland	5
Australia	4
Luxembourg	3
Japan	3
Greece	2
Norway	2
Hungary	2
China	2
Iran	1
Singapore	1
Mexico	1
Argentina	1
Holland	1
Israel	1
Brazil	1

TABLE A.2: Numero di studenti per nazioni estere

Argomento Studi	Bachelor's Degree	Master's Degree	Ph.D	Others	TOT
Agrarian Studies	70	102	11	120	303
Architecture	24	97	20	98	239
Economics	886	1790	188	1297	4161
Engineering	469	555	153	461	1638
Health Professions	20	23	1	27	71
Humanities	496	623	68	708	1895
Informatics	284	305	210	363	1162
Law	49	365	106	425	945
Liter., Lang. & Cult.	249	209	57	525	1040
Math and Physic	174	267	168	341	950
Medicine	12	51	63	159	285
Natural Science	444	654	272	560	1930
Pharmacology	8	72	20	87	187
Psicology	47	52	3	98	200
Science Education	24	57	16	211	308
Social Science	283	479	79	560	1401
Story and Philosophy	89	158	65	272	584
Veterinary Studies	1	8	1	5	15
No Value	254	865	1460	3203	5782

TABLE A.5: Numero di percorsi di studio per argomento.

Argomento Studi	TOT
Agrarian Studies	132
Architecture	115
Economics	1849
Engineering	1019
Health Professions	35
Humanities	871
Informatics	635
Law	382
Liter., Lang. & Cult.	353
Math and Physic	486
Medicine	87
Natural Science	1120
Pharmacology	82
Psicology	71
Science Education	69
Social Science	607
Story and Philosophy	260
Veterinary Studies	9
No Value	1269

TABLE A.6: Numero di percorsi di studio in Toscana per argomento.

	Intership	1° Livello	2° Livello	3° Livello
1^a Esperienza	1875	5171	1071	546
2^a Esperienza	1282	4489	1211	559
3^a Esperienza	757	3815	1192	527
4^a Esperienza	431	2970	1099	556
5^a Esperienza	215	2120	901	538
6^a Esperienza	37	688	402	273
7^a Esperienza	8	198	151	105
8^a Esperienza	2	66	52	44
9^a Esperienza	0	24	21	11
10^a Esperienza	0	6	4	3
11^a Esperienza	0	1	1	0
12^a Esperienza	0	0	1	0

TABLE A.9: Conteggio livello professionale per esperienza lavorativa.

B || Tabella Mobilità

La seguente tabella riporta i dati ottenuti relativamente alla mobilità internazionale degli utenti e alle attività svolte.

Per ogni nazione sono riportati 7 dati:

- **TOT:** Il numero totale di utenti che, almeno una volta, hanno indicato la nazione X come località generica del profilo, località degli studi o località presso cui si è lavorato.
- **From IT:** Il numero di utenti che, secondo quanto riportato nel profilo, hanno studiato o lavorato almeno una volta in Italia prima di muoversi per la prima volta nella nazione X.
- **Educ:** Il numero di utenti che hanno indicato almeno una volta la nazione X come luogo di studi.
- **Work:** Il numero di utenti che hanno indicato almeno una volta la nazione X come luogo di lavoro.
- **To IT.:** Il numero di utenti che, dopo aver studiato o lavorato nella nazione X, hanno lavorato o studiato in Italia. Questo dato non tiene in considerazione eventuali ritorni nella nazione X ed è indipendente da eventuali spostamenti intermedi tra la nazione X e l'Italia.
- **Others:** Il numero di utenti che dalla nazione X si sposta verso una nazione diversa dall'Italia. Questo dato tiene in considerazione solo gli spostamenti diretti.
- **Stay:** Il numero di utenti che, dopo aver studiato o lavorato nella nazione X, non compiono altri spostamenti.

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
Afghanistan	1	0	0	1	1	0	0
Albania	9	6	0	8	6	1	2
Algeria	1	0	0	1	1	0	0

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
Angola	2	2	0	2	1	1	0
Argentina	6	1	1	3	6	0	0
Armenia	2	1	0	2	1	1	0
Australia	26	23	4	22	14	6	6
Austria	50	46	11	38	30	6	14
Azerbaijan	2	0	0	2	2	0	0
Bahrain	1	0	0	1	1	0	0
Bangladesh	6	0	0	5	6	0	0
Belarus	1	0	0	1	1	0	0
Belgium	180	169	42	138	105	15	60
Benin	2	2	0	2	2	0	0
Bolivia	2	2	0	2	1	0	1
Bosnia & Herzegovina	1	1	0	1	1	0	0
Brazil	27	19	1	26	16	3	8
Bulgaria	2	2	0	2	0	0	2
Burkina Faso	1	1	0	1	1	0	0
Burundi	1	1	0	1	0	1	0
Cambodia	1	0	0	0	1	0	0
Cameroon	2	1	0	2	1	0	1
Canada	34	29	12	19	18	4	12
C.A.R¹	3	3	0	3	0	1	2
Chile	6	4	0	6	3	1	2
China	42	34	2	39	31	7	4
Colombia	6	3	0	4	4	2	0
Costa Rica	4	3	0	4	4	0	0
Croatia	2	1	0	2	2	0	0
Cyprus	5	3	0	5	5	0	0
Czech Republic	11	10	0	10	9	1	1

¹Central African Republic.

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
Denmark	55	49	23	31	36	6	13
Ecuador	3	2	0	2	2	0	1
Egypt	3	2	0	2	1	1	1
El Salvador	1	1	0	1	0	0	1
Eritrea	1	1	0	1	1	0	0
Estonia	3	3	0	3	2	0	1
Ethiopia	5	3	0	4	4	0	1
Finland	13	12	5	7	3	1	9
France	251	227	65	188	140	21	90
Georgia	4	3	0	4	2	0	2
Germany	232	214	56	174	136	18	78
Ghana	2	0	0	2	2	0	0
Greece	6	6	2	5	2	1	3
Guatemala	2	2	0	2	1	0	1
Haiti	2	1	0	2	1	1	0
Hungary	20	17	2	18	14	0	6
Iceland	2	2	0	2	2	0	0
India	23	5	0	23	22	0	1
Indonesia	1	1	0	1	1	0	0
Iran	26	6	1	24	24	1	1
Iraq	1	1	0	1	1	0	0
Ireland	77	69	25	52	47	10	20
Israel	7	5	1	6	5	0	2
Ivory Coast	1	1	0	1	0	0	1
Japan	22	21	3	19	15	3	4
Jordan	4	3	0	4	3	1	0
Kazakhstan	4	3	0	4	1	2	1
Kenya	4	4	0	2	4	0	0
Kosovo	2	2	0	2	0	0	2

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
Latvia	1	1	0	1	0	0	1
Lebanon	3	1	0	3	3	0	0
Libya	1	0	0	1	1	0	0
Liechtenstein	1	1	0	1	0	0	1
Lithuania	1	1	0	1	1	0	0
Luxembourg	20	19	3	12	12	1	7
Malaysia	2	0	0	2	2	0	0
Maldives	1	1	0	1	0	1	0
Malta	18	17	0	18	10	1	7
Mauritania	1	1	0	1	0	0	1
Mexico	7	3	1	3	5	1	1
Moldova	1	1	0	1	0	1	0
Montenegro	1	1	0	1	1	0	0
Morocco	4	2	0	4	4	0	0
Myanmar	1	0	0	1	1	0	0
Netherlands	128	120	37	87	70	13	45
New Zealand	5	5	0	4	0	2	3
Niger	2	2	0	2	0	1	1
Nigeria	4	3	0	4	3	1	0
Norway	7	7	2	5	2	1	4
Oman	2	2	0	2	1	1	0
Pakistan	9	2	0	8	7	1	1
Palestine	1	1	0	1	0	0	1
Panama	1	1	0	1	0	0	1
Peru	6	4	0	4	4	1	1
Philippines	4	1	0	2	3	0	1
Poland	16	11	0	16	10	2	4
Portugal	46	42	16	30	28	3	15
Qatar	7	6	0	6	2	1	4

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
Romania	6	5	0	6	2	1	3
Russia	24	16	0	24	21	0	3
Samoa	2	2	0	2	2	0	0
San Marino	1	1	0	1	0	0	1
Saudi Arabia	4	2	0	4	4	0	0
Senegal	2	1	0	2	2	0	0
Serbia	7	4	0	6	7	0	0
Sierra Leone	1	0	0	1	1	0	0
Singapore	10	9	1	8	4	0	6
Slovakia	4	4	0	2	4	0	0
Slovenia	2	2	0	2	2	0	0
Somalia	2	0	0	1	2	0	0
South Africa	6	5	0	6	4	0	2
South Korea	3	2	0	3	2	0	1
South Sudan	3	3	0	3	1	1	1
Spain	267	227	125	151	174	23	70
Sri Lanka	3	1	0	3	3	0	0
Sudan	1	1	0	1	0	0	1
Sweden	38	35	8	28	11	6	21
Switzerland	185	173	24	148	86	20	79
Syria	2	1	0	2	2	0	0
Tanzania	1	1	0	1	1	0	0
Thailand	3	2	0	3	1	1	1
Togo	1	1	0	1	0	0	1
Trinidad & Tobago	1	1	0	1	0	1	0
Tunisia	5	3	0	4	3	1	1
Turkey	21	9	0	20	18	0	3
Ukraine	1	0	0	1	1	0	0
United Arab Emirates	13	10	0	13	7	1	5

	TOT	From IT	Educ.	Work	To IT.	Others	Stay
United Kingdom	565	502	205	382	371	49	145
United States	372	341	105	262	211	38	123
Uzbekistan	1	0	0	1	1	0	0
Vatican City	1	1	0	1	1	0	0
Venezuela	1	1	0	1	0	0	1
Vietnam	2	2	0	2	1	0	1
Zambia	3	3	0	3	3	0	0

Bibliography

- [Backstrom et al., 2012] Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. (2012). Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42.
- [Balduzzi and Rosina, 2011] Balduzzi, P. and Rosina, A. (2011). Giovani talenti che lasciano l’italia: fonti, dati e politiche di un fenomeno complesso.
- [Balmer et al., 2009] Balmer, B., Godwin, M., and Gregory, J. (2009). The royal society and the ‘brain drain’: natural scientists meet social science. *Notes and Records of the Royal Society*, 63(4):339–353.
- [Baruffaldi et al., 2017] Baruffaldi, S. H., Di Maio, G., and Landoni, P. (2017). Determinants of phd holders’ use of social networking sites: An analysis based on linkedin. *Research policy*, 46(4):740–750.
- [Becker et al., 2004] Becker, S. O., Ichino, A., and Peri, G. (2004). How large is the” brain drain” from italy? *Giornale degli economisti e annali di economia*, pages 1–32.
- [Beltrame, 2007] Beltrame, L. (2007). Realtà e retorica del brain drain in italia. *Stime statistiche, definizioni pubbliche e interventi politici*.
- [Bhagwati and Hamada, 1974] Bhagwati, J. and Hamada, K. (1974). The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. *Journal of Development Economics*, 1(1):19–42.

- [Boffo and Gagliardi, 2017] Boffo, S. and Gagliardi, F. (2017). I costi della nuova mobilità internazionale dei giovani laureati italiani: un tentativo di stima. *Migrazioni e integrazioni nell'Italia di oggi*. Rome: IRPPS, pages 87–100.
- [Bogdan et al., 2014] Bogdan, S., Mario, R., Drik, H., and Emilio, Z. (2014). Migration of professionals to the us evidence from linkedin data. In *International Conference on Social Informatics*, pages 531–543.
- [Bönisch-Brednich, 2016] Bönisch-Brednich, B. (2016). *Migrants on campus: Becoming a local foreign academic*. Routledge.
- [Boulier, 1999] Boulier, D. (1999). La migration des compétences: enjeu de justice et solidarité internationale. *People on the Move*, 27(81):69–79.
- [Brandi, 2017] Brandi, M. C. (2017). L'emigrazione dei ricercatori italiani in un mondo che cambia. *IRPPS Monografi e*, page 73.
- [Brandi and Avveduto, 2004] Brandi, M. C. and Avveduto, S. (2004). Le migrazioni qualificate in italia. *Studi Emigrazione*, (156):797–829.
- [Brockmann and Theis, 2008] Brockmann, D. and Theis, F. (2008). Money circulation, trackable items, and the emergence of universal human mobility patterns. *IEEE Pervasive Computing*, 7(4):28–35.
- [Carrington and Detragiache, 1998] Carrington, W. J. and Detragiache, E. (1998). How big is the brain drain?
- [dei Conti, 2021] dei Conti, C. (2021). Referto sul sistema universitario.
- [Docquier and Marfouk, 2004] Docquier, F. and Marfouk, A. (2004). Measuring the international mobility of skilled workers (1990-2000): release 1.0. *World Bank Policy Research Working Paper*, (3381).
- [Ferrari et al., 2011] Ferrari, L., Rosi, A., Mamei, M., and Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 9–16.

- [Gaillard and Gaillard, 1997] Gaillard, J. and Gaillard, A. M. (1997). Introduction: the international mobility of brains: exodus or circulation? *Science, Technology and Society*, 2(2):195–228.
- [Giroud and Ivarsson, 2020] Giroud, A. and Ivarsson, I. (2020). World investment report 2020: International production beyond the pandemic.
- [Granovetter, 2018] Granovetter, M. (2018). *Getting a job: A study of contacts and careers*. University of Chicago press.
- [Granovetter, 1973] Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.
- [Grubel and Scott, 1966] Grubel, H. B. and Scott, A. D. (1966). The international flow of human capital. *The American Economic Review*, 56(1/2):268–274.
- [Hui et al., 2010] Hui, P., Mortier, R., Piórkowski, M., Henderson, T., and Crowcroft, J. (2010). Planet-scale human mobility measurement. In *Proceedings of the 2nd ACM international workshop on hot topics in planet-scale measurement*, pages 1–5.
- [Johnson and Regets, 1998] Johnson, J. M. and Regets, M. C. (1998). International mobility of scientists and engineers to the united states—brain drain or brain circulation?. *SRS Issue brief*.
- [Johnson et al.,] Johnson, S. C., Grow, A., Perrotta, D., Theile, T., de Valk, H., and Zagheni, E. Openness to migrate internationally for a job: Evidence from linkedin data.
- [LinkedIn, 2021a] LinkedIn (2021a). About linkedin. <https://about.linkedin.com/>. Accessed: 2021-10-25.
- [LinkedIn, 2021b] LinkedIn (2021b). LinkedIn’s self service advertising tool. <https://business.linkedin.com/marketing-solutions/ads>. Accessed: 2021-10-25.

- [Logan,] Logan, B. I. The reverse transfer of technology from sub-saharan africa: Some lessons from a preliminary assessment of the zimbabwean case.
- [Lutz, 1961] Lutz, V. C. (1961). Alcuni aspetti strutturali del problema del mezzogiorno: la complementarità dell'emigrazione e dell'industrializzazione.(some structural aspects of the southern problem: the complementarity of “emigration” and industrialisation). *Moneta e credito*, 15(56).
- [Mas-Bleda et al., 2014] Mas-Bleda, A., Thelwall, M., Kousha, K., and Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, 101(1):337–356.
- [Meyer, 2001] Meyer, J.-B. (2001). Network approach versus brain drain: lessons from the diaspora. *International migration*, 39(5):91–110.
- [Milio et al., 2012] Milio, S., Lattanzi, R., Casadio, F., Crosta, N., Raviglione, M., Ricci, P., and Scano, F. (2012). Brain drain, brain exchange and brain circulation. il caso italiano nel contesto globale. *Aspen Institute*.
- [Ministero dell'Istruzione, dell'Università, della Ricerca, 2021] Ministero dell'Istruzione, dell'Università, della Ricerca (2021). Dati atenei toscani. <http://ustat.miur.it/dati/didattica/toscana/atenei-statali>. Accessed: 2021-11-9.
- [Minnecci, 2015] Minnecci, F. (2015). If there were a ‘highly skilled red octopus’? the case of italian highly skilled mobility at times of crisis. *Economics and Sociology (Ternopil)*, 8(3):170–182.
- [Parlamento europeo e Consiglio europeo, 2016] Parlamento europeo e Consiglio europeo (2016). Regolamento (ue) 2016/679 del parlamento europeo e del consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/ce. <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=uriserv:>

[OJ.L.2016.119.01.0001.01.ITA&toc=OJ:L:2016:119:TOC](#). Accessed: 2021-10-27.

- [Pugliese, 2017] Pugliese, E. (2017). Giovani e anziani nella nuova emigrazione italiana. *IRPPS Monografi e*, page 45.
- [Pultar and Raubal, 2009] Pultar, E. and Raubal, M. (2009). A case for space: physical and virtual location requirements in the couchsurfing social network. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 88–91.
- [Rodriguez, 1975] Rodriguez, C. A. (1975). Brain drain and economic growth: a dynamic model. *Journal of Development Economics*, 2(3):223–247.
- [Rudolph and Hillmann, 1998] Rudolph, H. and Hillmann, F. (1998). The invisible hand needs visible heads: Managers, experts and professionals from western countries in poland. In *The New Migration in Europe*, pages 60–84. Springer.
- [We Are Social, 2021] We Are Social (2021). Digital 2021. <https://wearesocial.com/uk/blog/2021/01/digital-2021-the-latest-insights-into-the-state-of-digital/>. Accessed: 2021-10-25.
- [Zagheni et al., 2014] Zagheni, E., Garimella, V. R. K., Weber, I., and State, B. (2014). Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 439–444.
- [Zagheni and Weber, 2012] Zagheni, E. and Weber, I. (2012). You are where you e-mail: using e-mail data to estimate international migration rates. In *Proceedings of the 4th annual ACM web science conference*, pages 348–351.