# Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale



## Modelli di co-localizzazione e spazio delle amenities: un confronto tra città italiane

Relatori:

Prof. Luigi Buzzacchi Prof. Antonio De Marco Prof. Roberta Taramino Candidati: Virginia Barba Maria Elena Vetrugno

#### Sommario

Introduzione	
1. Rassegna Letteratura	2
2. Dati	9
3. Descrizione del Campione	15
4. Costruzione dei quartieri	25
5. Analisi dei Quartieri	33
5.1 Caratterizzazione dei quartieri	33
5.2 Poligoni dei Quartieri	37
6. Costruzione dello Spazio delle Amenities	42
6.1 Spazio delle Amenities	42
6.2 Analisi del Revealed Comparative Advantage (RCA)	45
Conclusioni	47
Bibliografia e Sitografia	50

#### Introduzione

Le città, da sempre, risultano essere il fulcro della vita e degli affari degli uomini. Le trasformazioni ambientali, le vicende politiche, economiche e sociali si intrecciano con le abitudini dei cittadini e plasmano l'aspetto di questi luoghi.

Le città si evolvono, quindi, di pari passo con l'evolversi degli uomini al fine di rispondere ai mutamenti delle loro necessità.

Lo scopo di questo lavoro è quello di identificare dei possibili cluster all'interno di alcune città italiane, analizzando le attività che li costituiscono, al fine di identificare similitudini e differenze.

Al fine di perseguire questo obiettivo si è replicata l'analisi descritta da Hidalgo nel paper scientifico "The amenity mix of urban neighborhoods" (2018). In questo lavoro l'autore utilizza un dataset realizzato su 47 città americane, sulla base del quale realizza un modello che identifica quali amenities hanno maggiore probabilità di localizzarsi in un determinato quartiere.

La tesi è costituita da sei capitoli. Nel primo capitolo è presente una rassegna della letteratura relativa ai modelli di localizzazione. Il secondo capitolo descrive gli step effettuati per la raccolta dei dati e la conseguente creazione del dataset. Il terzo descrive, nel dettaglio, i dati utilizzati come base di partenza per l'analisi e nel quarto capitolo si descrive il funzionamento dell'algoritmo. Il quinto e il sesto capitolo, infine, descrivono i risultati prodotti dall'algoritmo e propongono una analisi dei quartieri.

#### 1. Rassegna Letteratura

Per anni gli studiosi hanno svolto ricerche sulla localizzazione delle imprese e sulle motivazioni che ne influenzano l'aggregazione ponendosi la seguente domanda: "In che modo le aziende scelgono dove posizionarsi?". In questo lavoro di tesi si è quindi deciso di approfondire gli studi relativi a questo argomento e di definire dei possibili cluster di amenities, all'interno delle città analizzate.

Uno tra i pionieri nello sviluppo della progettazione degli spazi urbani è Christopher Alexander che, con il suo "A new theory of urban design" (1987), ne offre una visione innovativa. Alexander, infatti, non considera più la progettazione come un susseguirsi di step predefiniti da completare sequenzialmente, ma come una serie di sviluppi da delineare in corso d'opera. La progettazione, quindi, richiede una continua analisi e una costante attenzione al contesto, al fine di ottenere la soluzione migliore in relazione alle necessità del progetto.

Grazie all'avvento delle nuove tecnologie e al dilagante impiego dei big data, è stato possibile introdurre questa linea di pensiero all'interno della progettazione parametrica, la quale ha iniziato a diffondersi nell'ambito dell'urbanistica (Schumacher, 2009). In questo modo, è stato quindi possibile approfondire temi come la pedonabilità (Rakha & Reinhart, 2012; Sonta & Jain, 2019), l'efficienza energetica (Nagy et al., 2018) e la forma dei quartieri (Alonso et al., 2019; Noyman et al., 2019).

Un esempio di studio incentrato su queste tematiche è offerto dall'articolo "The amenity mix of urban neighborhoods" (Hidalgo et al., 2018). Hidalgo fa tesoro di tutta la letteratura teorica preesistente, al fine di realizzare uno studio avente come obiettivo la determinazione di leggi empiriche che regolano la collocazione dei servizi.

Vengono quindi raccolte 1,26 milioni di amenities relative a 47 città americane e se ne osservano le dinamiche di co-localizzazione. La loro correlazione permette di elaborare dei modelli che predicono quali amenities potrebbero svilupparsi maggiormente e quali invece rischierebbero di dover uscire dal mercato, prendendo in considerazione le attività già presenti in uno specifico territorio.

Su questo si basa il principio della *relatedness*, ovvero il successo o l'insuccesso di un'attività dipende da quelle ad essa correlate e già presenti nella stessa area (Hidalgo et al., 2018). I relativi studi sono partiti da un'analisi più ampia, realizzata a livello internazionale. Restringendo man mano il raggio si è poi arrivati ad avere come riferimento i diversi centri cittadini. In questa tesi, si vuole replicare tale analisi utilizzando un dataset costituito da diverse città italiane al fine di avere uno scenario ampio ed eterogeneo.

La relatedness è un parametro altamente versatile, sia a livello di scala geografica che in termini di tematiche di riferimento, grazie "alla varietà dei meccanismi attraverso i quali le economie e le organizzazioni apprendono" (Hidalgo et al., 2018).

Tale principio non si basa su scontate relazioni input-output (come, ad esempio, la necessità di procurarsi del ferro per poter produrre acciaio) in quanto questi legami sono prettamente tipici di qualche decennio fa in cui i costi di trasporto erano relativamente elevati; ai giorni nostri, invece, non è più indispensabile che fornitore e produttore siano geograficamente vicini tra loro. Un esempio è offerto dalla distribuzione dei giornali: precedentemente ogni città aveva un suo giornale, consegnato da un fattorino che seguiva un tragitto assegnatogli; ora ogni Paese ha pochi giornali di riferimento, raggiungibili con un paio di click tramite pochi motori di ricerca che riescono a servire il mondo intero (Hidalgo et al., 2018).

Partendo da queste considerazioni, si evince come il costo relativo allo spostamento della conoscenza sia aumentato rispetto al costo dello spostamento dei suoi prodotti, grazie al continuo avanzamento degli strumenti informatici e tecnologici. Oramai si ha una conoscenza concentrata in pochi luoghi che però risulta assiduamente connessa in reti di persone, mentre i prodotti percorrono lunghe distanze, viaggiando in tutto il mondo, in maniera sempre più rapida. Ci troviamo, quindi, in una situazione in cui il trasporto delle merci ha avuto modo di migliorarsi e svilupparsi in larga scala, mentre il sapere risulta strettamente ancorato alle persone e al capitale umano, nonostante si siano fatti progressi in termini di rapidità e facilità nell'accesso delle informazioni (Hidalgo et al., 2018).

Un'ulteriore applicazione del principio della *relatedness* si ha nell'ambito delle politiche industriali. Questo parametro può essere utilizzato al fine di apportare miglioramenti in

tale settore, come ad esempio evitare di costruire le cosiddette "Cattedrali nel deserto", ovvero fare dei grandi investimenti in imprese industriali che si rivelano poi inutilizzabili perchè collocate in zone inadeguate, prediligendo solo l'apparenza e non la sostanza. O ancora fare delle valutazioni prima che le imprese del settore privato possano ottenere investimenti pubblici rivolti a settori già largamente sviluppati. Per poter applicare il principio in questi termini, non basterà più individuare le industrie promettenti, ma sarà necessaria l'identificazione di meccanismi che rendano più agevole i flussi di conoscenza tra industrie e regioni, mettendo da parte l'eccessiva specializzazione, ma costruendo percorsi unici che conducono alla diversificazione (Hidalgo et al., 2018).

Numerosi studiosi hanno analizzato i benefici del principio di relatività in diversi contesti economici (Hidalgo et al., 2018; Neffke et al., 2011; Zhu et al., 2017; Kogler et al., 2013; Boschma et al., 2015; Guevara et al., 2016).

Gli studi di Neffke et al. nel 2011 e quelli successivi di Zhu et al. nel 2017 hanno dimostrato, rispettivamente per la Svezia e per la Cina, che la probabilità che un'azienda entri in un territorio cresce all'aumentare della presenza di altre aziende correlate ad essa.

Kogler et al. (2013) e Boschma et al. (2015) hanno svolto un'analisi che ha spiegato le dinamiche legate ai brevetti negli Stati Uniti: le città che avevano maggiori probabilità di presentare nuovi brevetti in un determinato settore tecnologico erano quelle che avevano già esperienza nei settori correlati.

Guevara et al. (2016) ha condotto uno studio affine, prendendo in considerazione le aree di ricerca di un'università o di un Paese, affermando che ci fossero più probabilità che venisse pubblicato un particolare studio in quelle zone in cui si era già specializzati in ambiti attinenti.

Volendo applicare il principio della relatedness a livello urbano, un problema da risolvere è la definizione dei confini da considerare. Per questo motivo si ricorre alla configurazione di cluster, essenziali per comprendere le modalità di aggregazione delle amenities.

La localizzazione delle amenities all'interno delle città, e più nello specifico, all'interno dei quartieri, è oggetto di un crescente interesse in diversi settori. Infatti, collocando in maniera strategica le amenities all'interno dei quartieri si ha un impatto:

- a livello sociale: si incrementa l'interazione tra cittadini migliorando i rapporti all'interno della comunità;
- a livello economico: si favorisce lo sviluppo del commercio al dettaglio locale;
- a livello ambientale: si invogliano i cittadini a ridurre l'utilizzo di mezzi di trasporto, riducendo conseguentemente l'inquinamento urbano.

Negli ultimi anni i cluster di servizi hanno iniziato ad espandersi e ad attirare un numero crescente di persone. Ad esempio, i distretti pedonali presenti a Washington D.C. rappresentano meno dell'1% della superficie totale, ma tra il 2009 e il 2013 hanno attratto circa il 50% dei nuovi uffici, negozi e hotel dell'area metropolitana (Hidalgo et al., 2018). L'interesse per queste zone "a misura d'uomo" non fa distinzione d'età. I sondaggi e i dati relativi agli affitti mostrano infatti che sia i millennial che gli anziani prediligono vivere in luoghi con facile accesso alla vendita al dettaglio e ai servizi.

Attualmente gli studi disponibili risultano prettamente teorici ed è difficile immaginare i concreti effetti che può avere l'aggiunta di un nuovo servizio all'ambiente urbano. La letteratura parte quindi dalla distinzione approssimativa delle amenities in due macrocategorie, suddividendole in:

- Agglomerazioni causate da clustering esogeno;
- Agglomerazioni causate da clustering endogeno.

Si ha un clustering esogeno quando più aziende si localizzano circondando una risorsa esterna comunemente attraente da parte della clientela (Berry, 1967). Ne sono esempio le zone in cui sono presenti fermate della metro e stazioni, nelle quali è presente un alto traffico di persone. Ed è proprio queste zone che sono predilette da titolari di bar e ristoranti che sfruttano l'alto traffico cittadino, il quale tende ad intensificarsi.

Questa forma di raggruppamento può emergere sia tra imprese concorrenti sia tra imprese complementari. Altri esempi di casi esogeni che attraggono servizi sono incroci stradali

altamente collegati, strutture pubbliche visitate di frequente e attrazioni turistiche (Hidalgo et al., 2018).

D'altro canto, non occorre tener conto esclusivamente dei fattori esterni, dal momento che non è detto che portino a dei benefici che perdurano nel lungo periodo, ma considerare anche altri elementi. Il sostentamento economico di un cluster nei pressi di un fattore esogeno dipenderà in parte dal livello di accessibilità e in parte dalla localizzazione dei cluster concorrenti (Hidalgo et al., 2018).

Tramite lo studio dell'indice di accessibilità, è possibile valutare le opportunità che un territorio può offrire in funzione della distribuzione delle attività commerciali e i consumatori. Solitamente, in letteratura si è affermato che le aree con un'alta densità di potenziali clienti influenzassero la localizzazione dei negozi; tramite invece modelli empirici (Seim, 2006; Waldfogel, 2008) si è dimostrato che gli esercizi tendono a posizionarsi in aree con alte densità di persone e quindi prediligono zone che siano facilmente raggiungibili dai clienti. Al fine di valutare l'accessibilità di un luogo si è quindi deciso di considerare un raggio di copertura di circa 600 metri corrispondente ad uno spostamento di circa 10-15 minuti a piedi, che si è osservato coprire la maggior parte dei viaggi a piedi dei vari consumatori (Buzzacchi et al., 2021).

Un modello che invece va ad approfondire il posizionamento di cluster concorrenti, i quali riducono l'area di mercato che ogni aggregato può rivendicare, lo si può trovare in "Central Place Theory" (Christaller & Baskin, 1966), secondo il quale, considerando il lungo periodo, il mercato disponibile sarà suddiviso in bacini uguali, a seguito di una distribuzione centralistica degli elementi. Il punto in cui si trova il centro e la relativa area circostante incidono sulla concorrenza dei cluster adiacenti (Christaller & Baskin, 1966). È importante sottolineare che quando si parla di centro non si intende il punto che minimizza le distanze degli abitanti da quel riferimento, ma invece proprio di un luogo che ha quella funzione di centralità. Si identifica quindi un centro a cui la popolazione fa riferimento per svolgere le proprie principali attività, "central professions", producendo i relativi "central products" e i "central services". Parallelamente, vi saranno dei "dispersed goods" e dei "dispersed services", che però non faranno capo alle zone centrali. Il modello prevede inoltre gli

"indifferent goods e gli "indifferent services", che si caratterizzano come i beni e servizi indifferenti all'area di produzione e consumo (Christaller & Baskin, 1966). La teoria di Christaller dà inoltre grande importanza alla frequenza di acquisti dei vari beni e servizi, portandolo ad affermare che i beni acquistati con minore frequenza (ad es. mobili) fossero presenti in un numero inferiore di cluster ma con bacini di utenza più ampi.

Passando invece all'agglomerazione dovuta a clustering endogeno, si presenta con una duplice natura: "complementare" o "competitivo".

Il raggruppamento complementare di servizi si riferisce alla co-ubicazione di attività che non competono direttamente tra loro ma che spesso vengono visitate dai clienti durante la stessa uscita o durante lo svolgimento dello stesso evento: teatri e gelaterie, ad esempio, sono considerate attività complementari (Eppli & Benjamin, 1994).

Questo tipo di raggruppamento è generalmente spiegato da un risparmio in termini di costi e tempo (Hernandez & Bennison, 2000; Nelson, 1958). I clienti, quindi, hanno un incentivo a visitare i cluster che offrono una più ampia scelta di prodotti complementari e questo a sua volta motiva la localizzazione complementare.

Questo porta all'instaurarsi di un processo di sfruttamento della popolarità: attività meno popolari attirano un maggior numero di clienti grazie alla popolarità dei negozi situati nei dintorni. L'esempio principe è rappresentato dai centri commerciali (Brueckner 1993).

Il clustering competitivo, invece, nasce dal bisogno del cliente di confrontare la qualità, i prezzi delle merci e dei servizi prima di effettuare un acquisto (Scitovsky 2013). Il cliente è solitamente più propenso ad effettuare questo particolare modus operandi per l'acquisto di prodotti "shopping goods" rispetto al caso in cui debba acquistare "convenience goods" (Hidalgo, Castañer, e Sevtsuk 2020).

Per shopping goods si intendono quei prodotti per i quali il consumatore è disposto ad investire del tempo nella ricerca di informazioni e per i quali percepisce un livello di rischio nell'acquisto in relazione alle prestazioni attese del prodotto; mentre con convenience goods si definiscono tutti quei beni di largo consumo che il cliente acquista in maniera ricorrente.

Sono esempio di clustering competitivo le stradine turistiche dei centri città, affollate di ristoranti.

Un aspetto cruciale da tener in conto quando si parla di cluster competitivi è la scelta dei prezzi offerti al consumatore. A causa dell'elevata concorrenza, i commercianti sono portati ad offrire prezzi sempre più vantaggiosi, al fine di aumentare il loro bacino d'utenza. Questo comporta l'instaurarsi di un circolo vizioso: il numero di clienti aumenta e il rivenditore abbassa sempre più i prezzi (Dudey, 1990). In questo modo, ogni singolo negoziante si troverà a dover valutare se entrare o meno nel cluster e prender parte alla competizione, considerando i benefici che potrebbe trarre dall'aumento delle vendite in relazione alla diminuzione dei prezzi di ogni transazione.

La formalizzazione di quanto spiegato finora è contenuta nel "classic model of ice cream vendors on a beach" proposto da Hotelling nel 1929.

Hotelling ha infatti dimostrato che è possibile raggiungere l'equilibrio ottimale posizionando due venditori dello stesso bene a distanze pari rispettivamente a 1/3 e ¾ ma è altresì preferibile adottare dei raggruppamenti al centro della spiaggia, i quali richiedono ai clienti di camminare per distanze maggiori ed eliminano il rischio esercitato dai concorrenti.

#### 2. Dati

Il presente capitolo è volto a illustrare la metodologia impiegata per l'elaborazione del dataset utilizzato.

Grazie ai progressi nel campo dell'urbanistica computazionale negli ultimi anni sono stati introdotti nuovi approcci alla progettazione urbana, con particolare attenzione ad aspetti quali pedonabilità, efficienza energetica e struttura urbanistica dei quartieri.

Di particolare interesse risulta la possibilità di analizzare i modelli di co-localizzazione delle amenties all'interno delle città col fine ultimo di identificare i quartieri sovra e sottospecializzati e il mix ottimale di servizi che dovrebbero avere al loro interno.

Questa analisi risulta possibile grazie all'utilizzo del principio della relatedness, il quale permette di determinare le attività economiche e i servizi maggiormente propense a svilupparsi in un determinato luogo.

Partendo da questo principio ci si pone come obiettivo l'analisi di vari aspetti caratterizzanti diverse città italiane: la struttura geografica, la distribuzione delle attività economiche e il loro reciproco influenzarsi.

Per ogni città in esame grazie all'utilizzo di un algoritmo si identificano, a partire dalla distribuzione delle amenities, i quartieri.

Al fine di dar vita all'analisi è stato necessario, come primo passo, realizzare un database. I dati di interesse riguardano la latitudine e longitudine delle amenities all'interno del suolo italiano e, per ciascuna coordinata, la tipologia di amenities.

Nel paper "The amenity mix of urban neighborhoods" come oggetto di analisi sono state selezionate 47 città statunitensi e circa 1,2 milioni di amenities. La realizzazione del dataset di partenza è stata possibile grazie all'impiego di Google Places API.

In questa tesi, invece, la raccolta dei dati è stata possibile grazie all'impiego di OpenStreetMap. Nata nel 2004 da un'idea di Steve Coast, OSM è una fondazione senza scopo di lucro avente come fine la creazione di un database di dati geografici accessibili a chiunque gratuitamente.

Come si può leggere sul sito web ufficiale:

«OpenStreetMap è una mappa liberamente modificabile dell'intero pianeta. OpenStreetMap permette a chiunque sulla Terra di visualizzare, modificare ed utilizzare dati geografici con un approccio collaborativo».

Essendo un progetto open source, concepito secondo una logica collettiva, le mappe offerte da OpenStreetMap sono editabili da chiunque decida di parteciparvi, a seguito di una semplice registrazione al sito web. Tali mappe sono rilasciate con licenza Creative Commons la quale permette di liberalizzare la loro modifica e riproduzione, a patto che venga citata la fonte.

Dalla sua nascita ad oggi OSM ha visto crescere rapidamente il suo impiego nei settori più disparati. Durante il terremoto di Haiti del 2010, la tempestiva creazione di mappe aggiornate a seguito della catastrofe si è rivelata fondamentale al fine di rispondere più rapidamente alla situazione di emergenza.

A partire dal 2019, invece, l'Istat ha iniziato ad utilizzare questo strumento per le proprie analisi al fine di determinare i tratti stradali più pericolosi sulla base di alcuni parametri (come le condizioni del traffico, la conformazione della strada, etc.).

Per poter estrarre i dati dalle mappe presenti in OpenStreetMap si è deciso di impiegare Overpass Turbo (<a href="https://overpass-turbo.eu/">https://overpass-turbo.eu/</a>), un tool web based che consente di visualizzare su mappa o esportare i risultati ottenuti tramite Overpass API query.

Ad esempio, al fine di estrarre le informazioni relative alla città di Bologna si è utilizzata la seguente query:

area[name='Bologna'][admin\_level=8]->.searchArea;
node[amenity](area.searchArea);
out;

Trattandosi, come già detto, di un progetto no profit la qualità e disponibilità dei dati è fortemente influenzata dal lavoro svolto dai contribuenti. Questo aspetto risulta essere, al contempo, un punto di forza ma anche un punto debole di OSM.

Tramite l'utilizzo di Overpass Turbo si sono analizzate, per ogni città in esame, le informazioni con tag "amenity". Ossia, le strutture utilizzate da visitatori e residenti, come ad esempio: servizi igienici, telefoni, banche, farmacie, bar, parcheggi e scuole.

Per ogni singola amenity OSM fornisce la possibilità di inserire svariate informazioni, tra le quali:

Latitude: latitudine dell'amenity;

Longitude: longitudine dell'amenity;

Id: identificativo dell'amenity;

- Amenity: tipologia del servizio (fuel, post office, etc.);

- Brand;

Source: identifica chi ha effettuato la registrazione;

Opening\_hours: orari di apertura del servizio.

Come già detto però, trattandosi di dati inseriti manualmente dagli utenti, la maggior parte delle amenities analizzate non disponeva di molte di queste informazioni.

Nella nostra analisi si è quindi deciso di utilizzare esclusivamente le informazioni strettamente necessarie: localizzazione geografica dell'amenity, tipologia e id univoco.

Si è quindi effettuata una prima investigazione sul dataset, confrontando il numero di amenities estrapolate e il numero di abitanti nella città. La quantità di amenities individuate non risulta direttamente proporzionale alla densità di popolazione di ciascuna città. Ad esempio, Messina con 225.711 abitanti presenta solo 720 amenities, addirittura meno di quelle presenti a Lecce che ha 93.679 abitanti (Tabella 1). Questo potrebbe essere causato dall'assenza di mappers nell'area, dalla modalità di inserimento dei dati o dalla scarsa disponibilità degli stessi.

Tabella 1 Città analizzate

Comune	Abitanti	Amenities
Napoli	941 685	2121
Torino	849 795	5629
Palermo	641 183	1433
Genova	559 684	9379
Firenze	361 057	4223
Messina	225711	720
Padova	208641	3306
Parma	200253	1201
Trieste	200121	1802
Brescia	195157	1638
Modena	188090	927
Perugia	164210	1265
Livorno	156067	2066
Rimini	148833	719
Ferrara	132336	877
Trento	120709	4896
Bergamo	119597	1202
Siracusa	118107	1423
Vicenza	109500	1209
Novara	102144	661
Arezzo	97924	655
Lecce	93679	962

Si è inoltre considerata l'estensione delle città sul territorio, per indagare se fosse presente una correlazione tra dimensione della città, in termini geografici, e numero di amenities presenti. Anche in questo caso non si è rilevata alcuna dipendenza tra queste due variabili. Ad esempio, come si può notare dai dati riportati in tabella 2, la città di Bergamo si estende

per 40,16 km² e presenta 1202 amenities quasi pari al numero di amenities registrate per la città di Vicenza che ha una superficie doppia.

Tabella 2 Città analizzate

Comune	Abitanti	Amenities
Perugia	449,51	1265
Ferrara	405,16	877
Arezzo	384,70	655
Parma	260,60	1201
Genova	240,29	9379
Lecce	238,00	962
Messina	213,75	720
Siracusa	207,78	1423
Modena	183,19	927
Palermo	160,59	1433
Trento	157,90	4896
Rimini	135,71	719
Torino	130,01	5629
Napoli	117,27	2121
Livorno	104,80	2066
Novara	103,00	661
Firenze	102,40	4223
Padova	93,03	3306
Brescia	90,34	1638
Trieste	85,11	1802
Vicenza	80,00	1209
Bergamo	40,16	1202

Guardando i dati relativi all'utilizzo di OSM il territorio con il maggior numero di collaboratori risulta essere Roma, seguita da Milano e Venezia. Gli utenti attivi in OSM presenti nel Sud Italia sono sensibilmente meno rispetto a quelli presenti nel Centro e nel Nord Italia, e questo spesso si riflette sui dati riportati nelle mappe.

Un altro punto debole di OpenStreetMap risulta essere la classificazione dei servizi. Infatti, a causa della grande varietà di amenities che possono essere utilizzate per classificare un luogo, oltre che dell'assenza di parametri certi che permettano di distinguere inequivocabilmente una tipologia di amenity dall'altra e della soggettività alla base del processo di creazione delle mappe, si può incappare in dati poco coerenti o non uniformi. Un esempio riscontrato frequentemente è stata la categorizzazione non precisa per la tipologia di amenities "bar" che spesso rientrano nelle amenities di tipo "cafè" o viceversa.

Si è comunque deciso di procedere con l'analisi utilizzando OSM grazie alla vasta disponibilità di dati offerti gratuitamente.

Per poter sfruttare al meglio le informazioni ottenute, si è deciso di rimaneggiare il dataset. Nello specifico, si sono eliminate le amenities di scarso interesse e se ne sono accorpate altre ritenute simili dal punto di vista merceologico.

#### 3. Descrizione del Campione

Nel seguente capitolo analizzeremo la metodologia seguita per definire il campione utilizzato nell'analisi successiva.

Il dataset è stato costruito ricavando da OSM le informazioni relative a 22 città italiane. Le città sono state selezionate in base alla densità di popolazione, ponendo inoltre attenzione alla relativa localizzazione (località marittima, località montuosa, etc.) e all'area geografica (Nord, Centro, Sud) in modo da avere un campione il più vario possibile. Nello specifico, si sono scelti comuni con popolazione superiore ai 92.000 e inferiore ai 950.000 abitanti, in modo da poter analizzare e confrontare i risultati dell'algoritmo su contesti differenti.

Tra l'altro le città con popolazione inferiore a 92.000 abitanti, oltre ad essere intrinsecamente meno interessanti per il tipo di ricerca che si intende svolgere, sono anche meno rappresentative in quanto il contributo offerto da OSM risultava anche men che proporzionale.

Tra le numerose classificazioni offerte da OSM si è deciso di analizzare esclusivamente i nodi con tag "amenity" in quanto maggiormente interessanti per l'analisi.

Al fine di estrapolare, per ogni città in analisi, i dati su cui lavorare si è impiegato, come strumento di data mining, Overpass turbo. Nella query elaborata si è imposto come parametri della ricerca:

- area[name='xy']: con il quale si è indicato, di volta in volta, per quale area geografica limitare la ricerca dei dati. In tal modo, tutti i punti registrati dagli utenti come appartenenti a quella data searchArea verranno estratti. Come si può osservare dalla Figura 1 questo procedimento restituisce anche punti esterni ai limiti comunali di Torino. Tale scelta è stata fatta al fine di non escludere eventuali punti di interesse posti oltre i limiti amministrativi;
- node[amenity]: con il quale si è limitata la ricerca ai soli dati con tag "amenity".

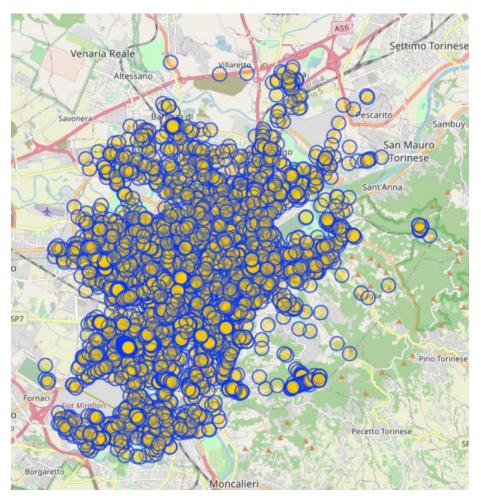


Figura 1 Risultati estrazione con Overpass Turbo per la città di Torino

Dopo aver ottenuto i dati, si è passati alla loro elaborazione. Come primo step si sono eliminate tutte le informazioni superflue ai fini dell'analisi, come ad esempio indirizzo, numero di telefono, orario di servizio, etc. Le informazioni d'interesse, che sono state mantenute all'interno del database sono:

- latitudine;
- longitudine;
- id;
- tipologia amenity.

Si è quindi passati alla pulizia vera e propria dei dati. Nello specifico, si sono analizzate le amenities presenti nelle città, e si è scelto quali amenties utilizzare e quali scartare.

Le amenities descritte in tabella 3 sono state eliminate poiché ritenute poco pertinenti all'analisi. Amenities come "bench" o "waste\_basket", nonostante risultino essere tra i nodi registrati con più alta frequenza, sono state rimosse dall'analisi poiché la tipologia di servizio offerto non risulta idoneo agli scopi di questo lavoro. Inoltre, l'elevata presenza di questi servizi avrebbe potuto influenza eccessivamente l'algoritmo, distorcendone i risultati.

Tabella 3 Amenities eliminate dall'analisi

Presenza
6792
3110
2682
1999
275
257
119
93
25

Le amenities presenti nella tabella 4 sono state anch'esse eliminate, poiché la loro presenza, all'interno del dataset, risulta essere estremamente esigua e non offre alcun contributo significativo.

Tabella 4 Amenities eliminate

Amenity	Presenza	Percentuale
gambling	18	0,037%
courthouse	17	0,035%
bureau_de_change	17	0,035%
social_centre	14	0,029%
bicycle_repair_station	14	0,029%
spa	14	0,029%

food_court	14	0,029%
watering_place	14	0,029%
compressed_air	14	0,029%
music_school	13	0,027%
street_lamp	13	0,027%
events_venue	12	0,025%
coworking_space	11	0,023%
bus_stop	10	0,021%
watering_place	14	0,029%
compressed_air	14	0,029%
music_school	13	0,027%
street_lamp	13	0,027%
events_venue	12	0,025%
coworking_space	11	0,023%
bus_stop	10	0,021%
watering_place	14	0,029%
office	10	0,021%
table	10	0,021%
money_transfer	9	0,019%
osmica	9	0,019%
nursing_home	8	0,017%
animal_shelter	8	0,017%
research_institute	8	0,017%
public_bookcase	7	0,014%
vehicle_inspection	7	0,014%
water_point	7	0,014%
lavoir	6	0,012%
biergarten	6	0,012%
boat_rental	6	0,012%
sanitary_dump_station	6	0,012%
stripclub	6	0,012%

wifi	6	0,012%
public_bath	5	0,010%
herbalist	5	0,010%
dojo	5	0,010%
telecommunication	5	0,010%
planetarium	5	0,010%
conference_centre	5	0,010%
toy_library	4	0,008%
crematorium	4	0,008%
flag_pole	4	0,008%
ticket_validator	4	0,008%
baby_hatch	3	0,006%
letter_box	3	0,006%
walking_bus	3	0,006%
mortuary	3	0,006%
tipography	3	0,006%
grave_yard	3	0,006%
telecommunication	5	0,010%
planetarium	5	0,010%
conference_centre	5	0,010%
toy_library	4	0,008%
crematorium	4	0,008%
flag_pole	4	0,008%
ticket_validator	4	0,008%
baby_hatch	3	0,006%
letter_box	3	0,006%
walking_bus	3	0,006%
mortuary	3	0,006%
tipography	3	0,006%
grave_yard	3	0,006%
animal_boarding	3	0,006%

waste_transfer_station	3	0,006%
casino	3	0,006%
smoking_area	2	0,004%
swingerclub	2	0,004%
ticket_validator	4	0,008%
baby_hatch	3	0,006%
letter_box	3	0,006%
walking_bus	3	0,006%
mortuary	3	0,006%
tipography	3	0,006%
grave_yard	3	0,006%
animal_boarding	3	0,006%
waste_transfer_station	3	0,006%
casino	3	0,006%
smoking_area	2	0,004%
swingerclub	2	0,004%
weighbridge	2	0,004%
professional_school	2	0,004%
prep_school	2	0,004%
animal_training	2	0,004%
luggage_locker	2	0,004%
wayside_shrine	2	0,004%
factory	2	0,004%
public_lift	2	0,004%
bell	2	0,004%
payment_terminal	2	0,004%
drinking_water	1	0,002%
weighbridge	2	0,004%
professional_school	2	0,004%
weighbridge	2	0,004%
professional_school	2	0,004%

prep_school	2	0,004%
animal_training	2	0,004%
luggage_locker	2	0,004%
wayside_shrine	2	0,004%
factory	2	0,004%
public_lift	2	0,004%
bell	2	0,004%
payment_terminal	2	0,004%
drinking_water	1	0,002%
prison	1	0,002%
wine	1	0,002%
caposaldo geodetico	1	0,002%
pet_sitting	1	0,002%
juice_bar	1	0,002%
binoculars	1	0,002%
financial_advice	1	0,002%
motorcycle_rental	1	0,002%
gym	1	0,002%
locker	1	0,002%
social_club	1	0,002%
lottery	1	0,002%
parking_tickets	1	0,002%
crypt	1	0,002%
pet_wash	1	0,002%
locker_room	1	0,002%
parking_exit	1	0,002%
kennel	1	0,002%
restaurant; cafeteria	1	0,002%
Technical-Construction	1	0,002%
wellness_center	1	0,002%
canteen	1	0,002%

educational_farm	1	0,002%
cabinet	1	0,002%
yes	1	0,002%
chocolate	1	0,002%
observatory	1	0,002%
neviera	1	0,002%
construction	1	0,002%
vegetables	1	0,002%
social_center	1	0,002%
embassy	1	0,002%
civic_center	1	0,002%
music_venue	1	0,002%
archive	1	0,002%
dive_centre	1	0,002%
post_depot	1	0,002%
administration	1	0,002%
rescue_station	1	0,002%
post_depot	1	0,002%
administration	1	0,002%
rescue_station	1	0,002%
tyres	1	0,002%
piano	1	0,002%
Solarium	1	0,002%
left_luggage	1	0,002%
leasing	1	0,002%

Al fine di rendere più comprensibile l'analisi si è inoltre deciso di riclassificare le amenities (tabella 5). Amenities come "cafe" e "bar" sono state aggregate in quanto relative alla stessa tipologia di servizio; amenties come "university" e "kindergarten" invece, sono state accorpate in quanto, pur essendo dei servizi differenti, appartengono alla stessa macro-

categoria. In questo caso, non avendo bisogno, ai fini dell'analisi, di un livello di dettaglio così specifico, si è preferito creare un'unica categoria.

Tabella 5 Nuove amenities

Amenities Originali	Presenza	Nuova Amenity	Presenza	
cafe	3369	bar 6203		
ice_cream	514			
bar	2320			
charging_station	372	fuel_charging	1909	
fuel	1537			
atm	489	bank & atm	1942	
payment_terminal	2			
bank	1451			
bicycle_parking	1312	parking	2705	
motorcycle_parking	176			
parking	1217			
school	485	education	942	
kindergarten	287			
language_school	27			
college	22			
childcare	20			
university	101			
dentist	126	doctors 490		
clinic	115			
doctors	249			
stripclub	6	nightclub	109	
nightclub	103			
car_rental	96	car_rental	182	
car_sharing	86			
gym	23	spa & gym	38	
spa	14			

dancing_school	1		

Le amenities così ottenute sono state accuratamente analizzate ponendo particolare attenzione alla loro presenza sul territorio.

Nello specifico, si è esaminata la coerenza dei dati ottenuti con la reale presenza rilevata nelle varie città. Grazie a questo controllo si è ritenuto opportuno affinare ulteriormente i dati procedendo con l'eliminazione delle amenities non presenti in tutte le città analizzate o l'accorpamento in gruppi omogenei in modo da poter meglio evidenziare le differenze presenti sul territorio. Di seguito le amenities utilizzate (tabella 6):

Nuova Amenity		
restaurant		
bar		
bank_ATM		
fuel_charging		
pharmacy		
education		
library		
pub		
post_office		
cultural_entertainment		

### 4. Costruzione dei quartieri

Dopo aver ottenuto il dataset, si è proceduto con la fase di analisi.

L'obiettivo dell'analisi, i cui risultati sono descritti in questo capitolo, è quello di generare dei possibili cluster all'interno di 22 città italiane, al fine di determinare caratteristiche comuni sulla base delle amenities contenute in ogni quartiere.

Per perseguire tale scopo è stato applicato al dataset l'algoritmo sviluppato nella tesi "Pattern di co-localizzazione spaziale delle dotazioni urbane: analisi empirica dei quartieri di Torino" (Del Nibletto, 2021).

Come esplicitato nel paper "The amenity mix of urban neighborhoods" (Hidalgo et. al, 2018), l'algoritmo aggrega i dati sulla base dell'indice di accessibilità. Più specificatamente, determina i punti con indice maggiore ("peaks") e assegna ad essi le amenities presenti nelle immediate vicinanze (ossia, i servizi posti a distanza non superiore di 500m). Questo procedimento è ripetuto in maniera iterativa fino a che tutti i servizi non risultano assegnati ad un cluster.

L'indice di accessibilità è determinato tramite la distribuzione spaziale delle destinazioni potenziali, la facilità di raggiungimento di ogni destinazione, la grandezza, la qualità e il carattere delle attività svolte in quelle destinazioni (Handy e Niemeier, 1997). Ossia, l'indice esplicita, per ogni amenities, la misura di quanto questa risulta inserita nel tessuto urbano.

Le varie formule relative all'indice di accessibilità, presenti in letteratura possono essere suddivise in tre categorie (Handy e Niemeier, 1997):

- isochrone-based measures: tengono conto esclusivamente del numero di amenities
   che possono essere raggiunte entro una certa distanza;
- gravity-based measures: le amenities sono valutate grazie all'impiego di una funzione di impedenza;
- utility-based measures: grazie alle quali è possibile differenziare gli utenti coinvolti.

Fra le varie formule presenti in letteratura, l'algoritmo che abbiamo utilizzato sfrutta la stessa definizione impiegata nel paper di Hidalgo, si tratta, in questo caso di una misura di tipo gravity-based. Tali misure tengono conto della quantità di attività presenti nelle

diverse destinazioni in base al costo, al tempo o alla distanza necessaria per raggiungerle, in altre parole, l'indice di accessibilità esprime in questo modo "il potenziale delle opportunità di interazione" (Hansen, 1959):

$$A_i = \sum_{i=1}^N e^{-\gamma \, d_{ij}}$$

Dove:

- $A_i$  rappresenta l'indice di accessibilità relativo all'amenities i-esima;
- $d_{ij}$  rappresenta la distanza che intercorre fra l'amenity i e l'amenity j;
- γ rappresenta il parametro di decadimento;
- *N* rappresenta il numero di amenties analizzate.

Come esplicitato dalla formula, l'indice di accessibilità dell'amenity i risulta calcolato sulla base della somma dei vari contributi offerti dalle N amenities contenute nel dataset.

Il contributo, che ogni amenity fornisce, è influenzato dal parametro di decadimento e dalla distanza che intercorre fra l'amenity e tutte le altre. Ad esempio, due amenities localizzate nella stessa posizione apporteranno un contributo pari a 1 all'indice di accessibilità dell'altra. Tale contributo risulta inversamente proporzionale alla distanza: all'aumentare della distanza che intercorre fra le amenities l'indice di accessibilità diminuisce.

Ruolo importante nell'elaborazione dei quartieri è rappresentato dall'euristica utilizzata all'interno dell'algoritmo. Grazie al suo impiego si impedisce la creazione di quartieri con centri molto ravvicinati, soprattutto in zone ad alta densità di amenities (i.e. i centri delle città). Di seguito l'euristica utilizzata:

$$n_i = 3Ai + 50$$

Dove:

 $n_i$  rappresenta il numero di amenities vicine all'i-esima

In questo modo si individuano i centri e, di conseguenza, il numero di quartieri che verranno generati. A partire da questa informazione si aggiungono, man mano, tutti i restanti dati ai vari cluster, partendo dalle amenities geograficamente più vicine a quelle individuate come centri dei quartieri.

Al fine di valutare il comportamento dell'algoritmo e la sua stabilità si è scelto di analizzare e confrontare i risultati prodotti al variare del parametro di decadimento.

Nella tabella 6, è descritto l'andamento dell'indice di accessibilità al variare del parametro di decadimento e al variare della distanza fra servizi.

Come si evince dal grafico in figura 2, l'indice di accessibilità è sempre compreso tra i valori 0 ed 1 e al crescere del parametro gamma la curva che descrive il suo andamento risulta più ripida.

Tabella 6 Andamento indice di Accessibilità

Indice di Accessibilità							
distanza [km]	γ= 5	γ = 12	γ = 16	γ = 20	γ = 30		
0	1,00	1,00	1,00	1,00	1,00		
0,03	0,86	0,70	0,62	0,55	0,41		
0,05	0,78	0,55	0,45	0,37	0,22		
0,08	0,67	0,38	0,28	0,20	0,09		
0,1	0,61	0,30	0,20	0,14	0,05		
0,12	0,55	0,24	0,15	0,09	0,03		
0,15	0,47	0,17	0,09	0,05	0,01		
0,2	0,37	0,09	0,04	0,02	0,00		
0,25	0,29	0,05	0,02	0,01	0,00		
0,3	0,22	0,03	0,01	0,00	0,00		
0,35	0,17	0,01	0,00	0,00	0,00		
0,4	0,14	0,01	0,00	0,00	0,00		
0,5	0,08	0,00	0,00	0,00	0,00		
0,7	0,03	0,00	0,00	0,00	0,00		
0,9	0,01	0,00	0,00	0,00	0,00		

Al fine di determinare il parametro ottimale per la nostra analisi si sono confrontati i risultati ottenuti per gamma compreso tra 5 e 30. Un valore di γ eccessivamente ridotto, tende ad aggregare i dati in maniera eccessiva, mentre al crescere del parametro gamma il numero di quartieri aggiuntivi non risulta particolarmente significativo e i nuovi quartieri non identificano zone di particolare interesse.

È importante sottolineare che la scelta di un valore troppo basso rischierebbe di alterare i risultati: l'impatto delle amenities molto distanti sarebbe sovrastimata. Viceversa, un valore elevato del parametro di decadimento comporterebbe una sostanziale riduzione, perfino inferiore a quella reale, dell'effettiva influenza esercitata dall'amenity.

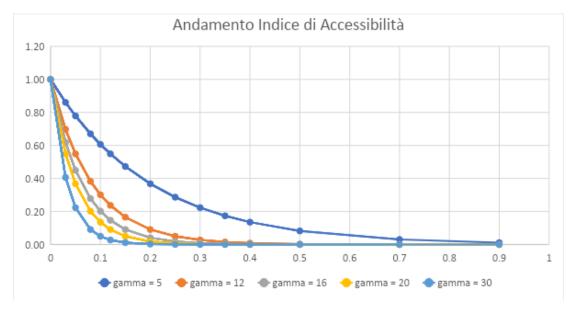


Figura 2 Andamento indice accessibilità

Si sono quindi analizzati i risultati ottenuti per la città di Torino al variare del parametro di decadimento, al fine di scegliere il parametro più adeguato per l'analisi. Per  $\gamma$ =5 l'algoritmo ha prodotto 7 quartieri (figura 3); la distribuzione delle amenities risulta essere poco omogenea. In particolare, il quartiere 0 è costituito da 1660 amenities, e il quartiere 5 da 36:

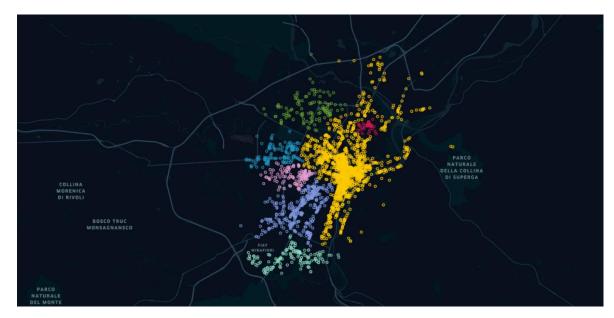


Figura 3 Clustering per γ=5

Per gamma pari a 12 l'algoritmo identifica il doppio dei quartieri, con all'interno un numero di amenities oscillante fra 60 e 424 (Figura 4):

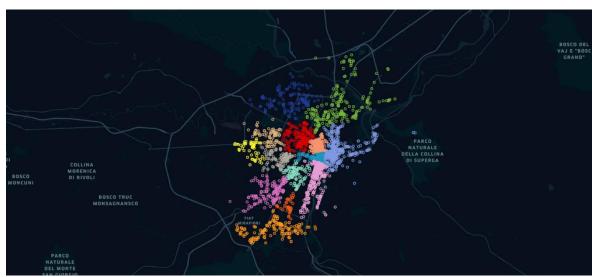


Figura 4 Clustering per γ=12

Per gamma pari a 16 si ottengono 19 quartieri (Figura 5):

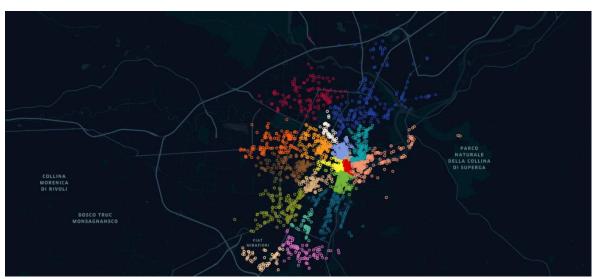


Figura 5 Clustering per γ=16

Per un valore di gamma pari a 30 l'algoritmo individua 25 quartieri, alcuni dei quali costituiti da un numero esiguo di dati; ad esempio, il quartiere 14 risulta formato da sole 6 amenities (Figura 6):

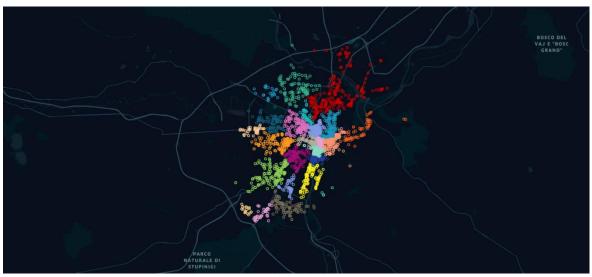


Figura 6 Clustering per γ=30

Di seguito sono riportate le forme dei quartieri ottenute con parametro di decadimento rispettivamente pari a 5 (Figura 7), 16 (Figura 8) e 30 (Figura 9), al fine di poter confrontare i casi limiti e il caso con il γ già precedentemente selezionato da Hidalgo:

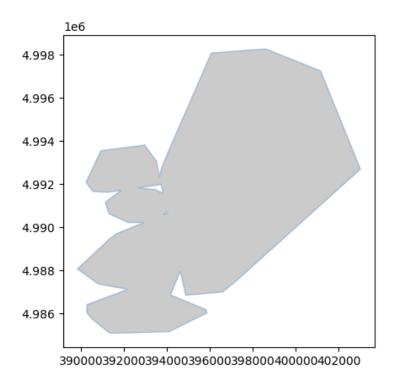


Figura 7 Poligoni quartieri con γ=5

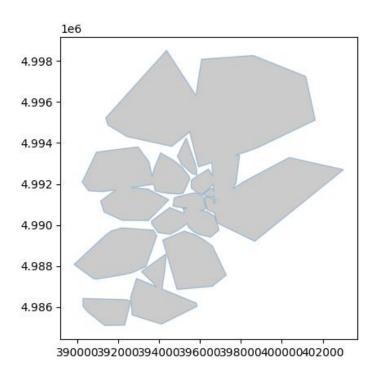


Figura 8 Poligoni quartieri con  $\gamma$ =16

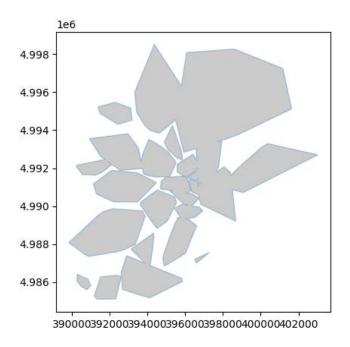


Figura 9 Poligoni quartieri con γ=30

A seguito delle considerazioni fatte finora, si è quindi deciso di utilizzare come valore del parametro di decadimento γ=16, stesso valore impiegato da Hidalgo nel paper "The amenity mix of urban neighborhoods". Sulla base di questa scelta, il contributo offerto da ciascuna amenity viene dimezzato ogni 62,5 m e risulta trascurabile per distanze superiori ai 500m. Tale semplificazione risulta coerente con quanto dimostrato da Handy & Niemeier prima e da Sevtsuk poi: gli spostamenti pedonali raramente superano i 10 minuti di tempo.

Al fine di individuare le amenities da inserire in ogni quartiere, l'algoritmo esclude dall'analisi i dati caratterizzati da un minor indice di accessibilità. Nello specifico, seguendo quanto fatto nell'articolo di Hidalgo, l'algoritmo scarta il 10% delle amenities e lavora con il restante dataset, caratterizzato da un più significativo indice di accessibilità.

#### 5. Analisi dei Quartieri

#### 5.1 Caratterizzazione dei quartieri

Questo capitolo è volto ad illustrare alcune considerazioni che si sono potute dedurre dall'analisi effettuata a seguito della costruzione dei quartieri, illustrata nel precedente capitolo.

Osservando nel complesso i risultati ottenuti dall'algoritmo, risulta evidente un pattern comune all'interno del quartiere 0 di ogni città analizzata.

Nella maggior parte delle località considerate la tendenza che ne risulta individua come prime tre tipologie di amenities per frequenza: restaurant, bar e bank ATM.

Nello specifico, i ristoranti e i bar risultano essere rispettivamente sempre al primo e al secondo posto in ordine di frequenza; al terzo posto si posiziona, in 11 su 22 città, l'amenity "bank", nelle restanti invece non si identifica un'unica alternativa (tabella 7):

Tabella 7 Amenities al terzo posto per frequenza

Amenity	Num. Città	
fuel_charging	6	
library	2	
pharmacy	2	
education	1	

È interessante notare come, solitamente, il comportamento registrato nel quartiere 0 risulti simile anche negli altri e la stessa predominanza presente nel quartiere 0 per le amenities "restaurant" e "bar" si registra in tutti i quartieri tranne in 2 quartieri a Messina e in 1 a Perugia (Allegato 1).

La distribuzione delle amenities identificata nel quartiere 0 si ripresenta anche negli altri quartieri, non solo quando nei primi tre posti sono presenti le classiche amenities, ma anche quando si identificano categorie particolari. Un esempio significativo risulta essere Firenze, dove al terzo posto si hanno le librerie, non solo nel quartiere 0 ma anche in 9 su

13 dei restanti quartieri, identificando una specializzazione non solo a livello di quartiere ma anche a livello di città che in questo modo si differenzia dalla media (Figura 10).

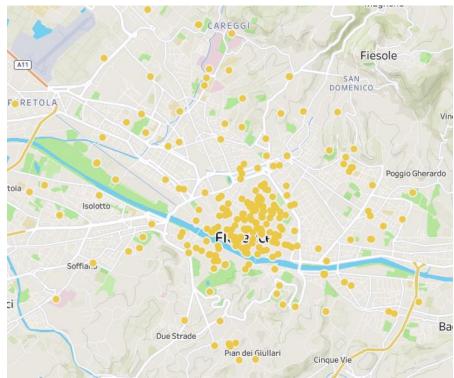


Figura 10 Rappresentazione dell'amenity "library" nella città di Firenze

Fa eccezione la città di Messina: per questa città sono stati esaminati 409 records e l'algoritmo ha prodotto 4 quartieri.

Nonostante i ristoranti e i bar siano, come percentuale sul totale, le amenities maggiormente presenti, a differenza dell'andamento generale, i quartieri 1 e 2 hanno una grande presenza di uffici postali, scuole, farmacie.

Seppur la percentuale di ristoranti e bar sia la più alta fra le varie tipologie di amenities presenti è importante notare che questo valore risulta inferiore al valore medio fra tutte le città analizzate.

Da una prima analisi della distribuzione delle amenities nei quartieri si evince come generalmente si identifica una corrispondenza tra quanto accade nel quartiere 0 e quanto accade nella città nel suo complesso. Come, infatti, si evince dalla Tabella 8 i quartieri 0 considerati comprendono quasi il 50% delle amenities analizzate.

Tabella 8 Distribuzione media delle amenities nei quartieri

Quartiere	Media frequenza amenities	
0	49%	
1	20%	
2	14%	
3	9%	
4	7%	
5	9%	
6	7%	
7	4%	
8	5%	
9	5%	
10	5%	
11	6%	
12	3%	
13	4%	
14	2%	
15	4%	
16	3%	
17	4%	
18	3%	
19	3%	
20	3%	
21	1%	
22	4%	
23	5%	

Nelle città caratterizzate da un elevato numero di quartieri la distribuzione delle amenities risulta essere più omogenea, viceversa, nelle città con pochi quartieri è il quartiere 0 ad avere, solitamente, la più alta percentuale di dati. (Figura 11).

Comune	Num. Amenities	Num. Quartieri	Abitanti	Superficie kmq città
Genova	9379	24	559.684	240,29
Torino	5629	19	849.795	130,01
Firenze	4223	13	361.057	102,40
Napoli	2121	8	941.685	117,27
Palermo	1433	7	641.183	160,59
Trieste	1802	7	200.121	85,11
Perugia	1265	5	164.210	449,51
Siracusa	1423	5	118.107	207,78
Trento	4896	5	120.709	157,90
Padova	3306	5	208.641	93,03
Brescia	1638	5	195.157	90,34
Messina	720	4	225.711	213,75
Livorno	2066	4	156.067	104,80
Vicenza	1209	4	109.500	80,00
Bergamo	1202	4	119.597	40,16
Arezzo	655	3	97.924	384,70
Modena	927	3	188.090	183,19
Rimini	719	3	148.833	135,71
Ferrara	877	2	132.336	405,16
Parma	1201	2	200.253	260,60
Novara	661	2	102.144	103,00
Lecce	962	1	93.679	238,00

Figura 11 Istogramma

Per meglio comprendere cosa influenzi il processo di definizione dei quartieri abbiamo a questo punto cercato di identificare gli aspetti più significativi e come si può evincere dal grafico risulta evidente una correlazione stretta tra numero di quartieri creati e numero di amenities. È interessante l'eccezione di Trento (Figura 12), dove nonostante l'elevato numero di amenities (4896) il numero di quartieri risulta esiguo (5). Questo risulta particolarmente evidente se si confronta con la città di Firenze la quale, pur avendo un numero di amenities leggermente inferiore presenta più del doppio dei quartieri. Una possibile interpretazione potrebbe essere legata alla diversa conformazione del territorio; infatti, seppur Trento presenti una superficie importante, gran parte delle amenities si sviluppano nel centro città piuttosto che nelle aree periferiche caratterizzate dal territorio pre-montano e montano.

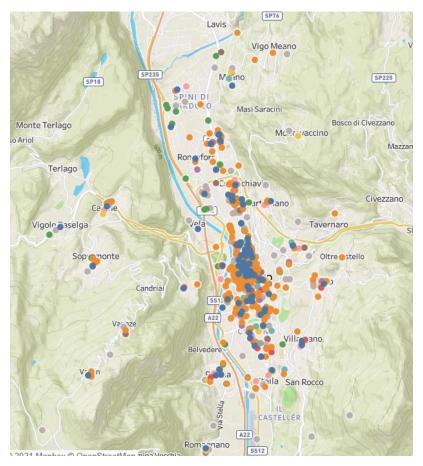


Figura 12 Amenities nella città di Trento

Come dimostrato dai dati in tabella, la correlazione tra numero di amenities e numero di quartieri generati risulta slegata dal numero di abitanti e dalla superficie geografica.

### 5.2 Poligoni dei Quartieri

Grazie all'impiego di un tool, sviluppato nel lavoro di tesi "Pattern di co-localizzazione spaziale delle dotazioni urbane: analisi empirica dei quartieri di Torino" (Del Nibletto, 2021), è stato possibile ottenere la rappresentazione dei poligoni dei quartieri.

In questo modo si è potuto analizzare nel dettaglio la morfologia dei cluster ottenuti e indagare sulla presenza di aspetti comuni intrinseci, ad esempio legati alla posizione geografica della città studiata.

Partendo dai risultati ottenuti dall'algoritmo di clustering, il tool legge le coordinate all'interno di ogni quartiere, crea una lista composta da longitudine e latitudine di ogni punto e crea quindi i singoli poligoni.

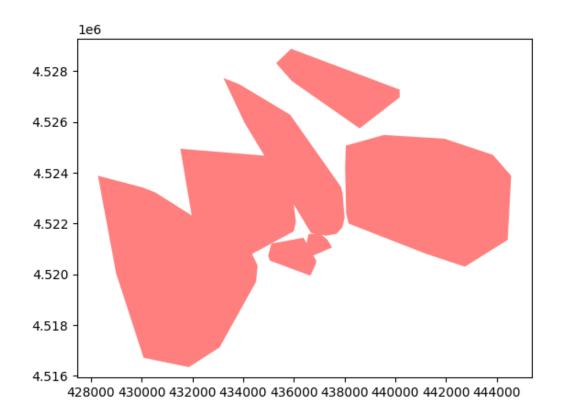


Figura 13 Poligoni dei quartieri per la città di Napoli

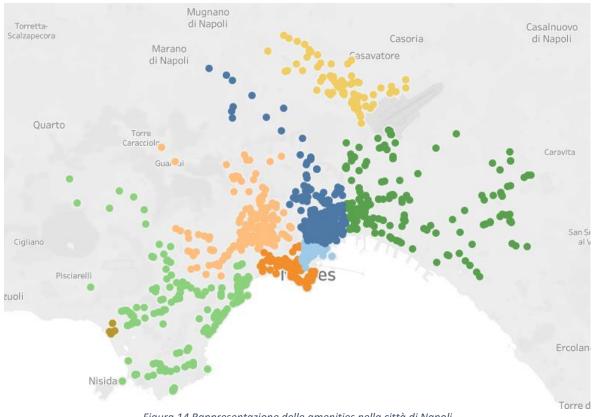


Figura 14 Rappresentazione delle amenities nella città di Napoli

In figura 13 sono riportati i poligoni dei quartieri per la città di Napoli (figura 14). Trattandosi di una località marittima la struttura dei quartieri tende ad assumere una conformazione particolare. Nello specifico, è possibile notare, per la maggior parte dei quartieri una forma allungata all'interno del quale sono raccolte amenities presenti nell'entroterra e sul mare.

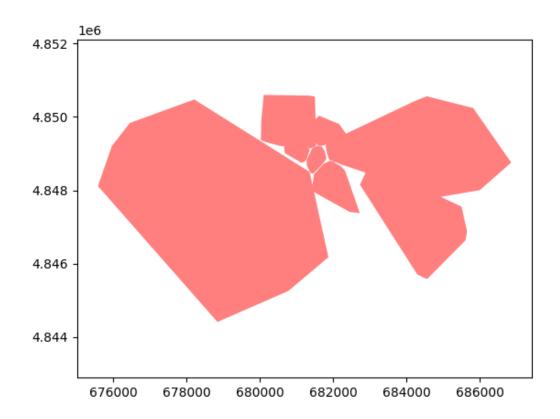


Figura 15 Poligoni dei quartieri per la città di Firenze

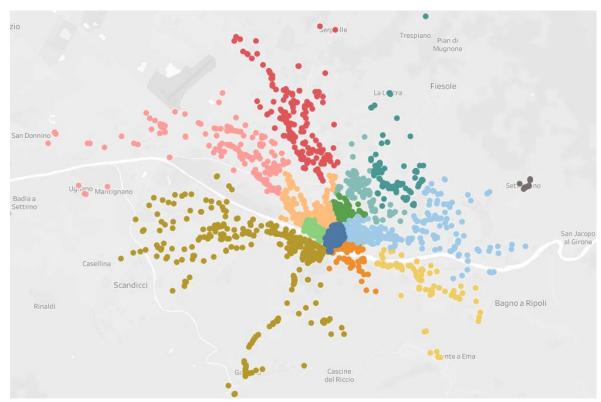


Figura 16 Rappresentazione delle amenities nella città di Firenze

Analizzando una città posizionata nell'entroterra la forma dei quartieri risulta essere estremamente differente. Ad esempio, come mostrato nelle Figure 15 e 16 per la città di Firenze l'algoritmo individua numerosi quartieri di dimensioni modeste localizzati attorno al centro città e quartieri più grandi posti esternamente.

I quartieri più centrali risultano poco estesi e le amenities al loro interno hanno una densità elevata. Al contrario, i quartieri localizzati nelle zone periferiche risultano occupare superfici più ampie e la localizzazione delle amenities risulta meno densa. Tale distribuzione fa sì che i quartieri periferici assorbano al loro interno le amenities più esterne e, quindi, non contese.

Si può notare inoltre una maggiore concentrazione di quartieri nelle zone centrali, nelle quali si ha una numerosità più rilevante di amenities. Si evince inoltre un'estensione ridotta dei quartieri centrali, a differenza di quelli periferici che accorpano anche servizi più esterni.

Si osserva in aggiunta la presenza di zone in cui non sono presenti amenities, individuate da porzioni di superficie bianche, prevalentemente tra i confini dei diversi quartieri. Si vuole comunque evidenziare che i quartieri periferici, pur essendo rappresentati come un grande agglomerato, sono caratterizzati da una bassa densità di amenities al loro interno, ma assorbono quelle attività che risulterebbero isolate.

Dal confronto delle figure delle forme dei quartieri, si possono rilevare differenze di conformazione prevalentemente nei quartieri centrali, mentre i quartieri periferici restano per lo più invariati. Questo effetto è legato alla variazione dell'indice di accessibilità che ha un importante impatto sulle amenities adiacenti.

## 6. Costruzione dello Spazio delle Amenities

#### 6.1 Spazio delle Amenities

Dopo aver individuato i quartieri attraverso il metodo di clustering, si è costruita la rete delle amenities presenti negli stessi cluster, utilizzando un algoritmo sviluppato in R. L'intero modello predittivo è basato sull'"Amenity Space" individuato.

Per la realizzazione di questa rete di servizi si è utilizzata la correlazione di Spearman, al fine di misurare la dislocazione delle attività in tutti i cluster. L'obiettivo è quello di creare un grafo pesato, calcolando il coefficiente di Spearman, con la formula 5.1, per ogni quartiere individuato:

$$W_{ij} = spearman(N_{ic}, N_{ic})$$
 (5.1)

Dove:

 $W_{ij}$  indica la correlazione tra l'amenity i e l'amenity j;

 $N_{ic}$  indica in numero di volte che l'amenity i compare nel cluster c;

 $N_{jc}$  indica in numero di volte che l'amenity j compare nel cluster c.

Si è scelto di utilizzare una correlazione basata sui ranghi (Spearman), onde evitare problematiche legate alla distribuzione dei dati come potrebbe avvenire per altre tipologie, come ad esempio la correlazione di Pearson.

In quest'ultimo caso, al fine di avere dei risultati corretti, è richiesto che i dati siano distribuiti normalmente; i nostri dati non sono distribuiti in maniera perfettamente normale, a causa della presenza di servizi molto comuni (come ristoranti, bar), e di altri più sporadici (teatri, librerie), si è optato per il coefficiente di Spearman, garantendo la robustezza del risultato.

In una prima fase di analisi si è svolta la costruzione dello spazio delle amenities su ogni singola città del dataset, ma a causa del ridotto numero di dati per alcune città, il livello di significatività in uso (66%) risultava ben lontano dai livelli standard statistici.

Per ottenere dei risultati consistenti, si è quindi deciso di aggregare i dati delle diverse città analizzate. A causa di problemi computazionali, strettamente legati ai mezzi a disposizione, si è dovuta escludere l'analisi sulla raccolta sommaria dei dati, e si è optato per

un'aggregazione basata sulla suddivisione geografica della penisola italiana: Nord, Centro e Sud.

Le figure 17, 18 e 19 mostrano le reti con le correlazioni dei servizi che si localizzano negli stessi cluster. Per permettere una più facile lettura del grafo si è rappresentato il Maximum Spanning Tree con una correlazione di Spearman a coppie pari o maggiore di 0.3 e un livello di significatività del 98%.

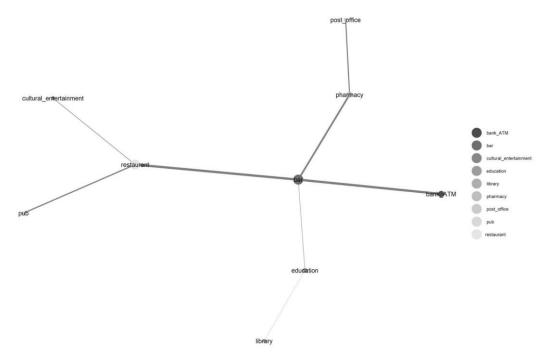


Figura 17 Amenity Space per città selezionate del Nord Italia

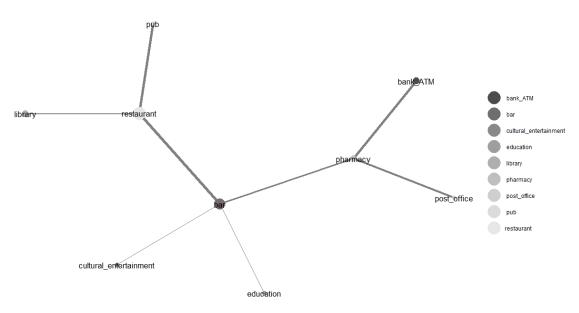


Figura 18 Amenity Space per città selezionate del Centro Italia

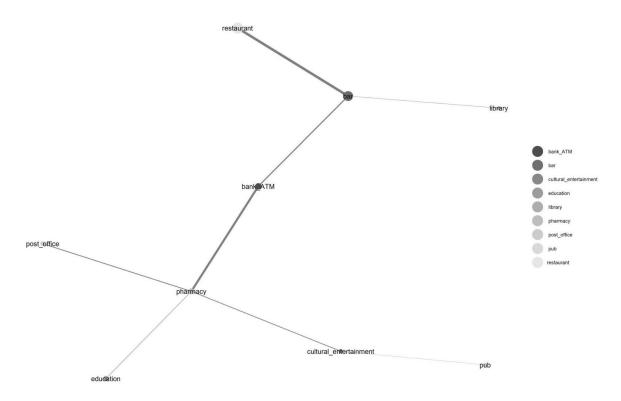


Figura 19 Amenity Space per città selezionate del Sud Italia

I nodi della rete rappresentano i diversi tipi di servizi, mentre gli archi collegano le amenities che potrebbero esser presenti nello stesso quartiere. Lo spessore degli archi è proporzionale al valore della correlazione ottenuto; la dimensione del nodo rappresenta il numero di volte che una determinata attività si presenta nella raccolta dati.

Si può notare come sia ricorrente in tutti e tre i casi la forte correlazione tra ristorante e pub, entrambi appartenenti alla stessa tipologia di amenities, identificabile come luoghi di ristorazione. Un altro risultato, per certi versi inaspettato, è la correlazione significativa tra farmacie e uffici postali, probabilmente dovuto al fatto che entrambi i servizi sono distribuiti in modo omogeneo nelle diverse città, in modo tale da ricoprire nel miglior modo possibile l'intera area urbana e minimizzare i costi di ricerca degli acquirenti.

Al centro della rete, come si può notare, si trova il cluster bar che collega i servizi di somministrazione alimentare come pub e ristoranti. A questi nodi sono, inoltre, collegate amenities come uffici postali, bancomat e farmacie che identificano acquisti ripetuti frequentemente, che come precedentemente accennato, sono distribuiti strategicamente sull'intero territorio.

Un'altra interessante osservazione riguarda le coppie di amenities che si respingono, pur facendo parte dello stesso cluster, in questo caso si parla di "triadi non transitive". Ad esempio, nel caso nello scenario del Centro Italia, troviamo i ristoranti collegati a librerie e pub, ma questi ultimi due non sono collegati tra loro, per giunta in nessuno spazio rappresentato.

Grazie a questa rete è possibile stimare quali servizi saranno presenti in una determinata area, prendendo in considerazione le amenities già rilevate. Inoltre, può essere un valido supporto per rilevare i servizi che sono sotto o sopra rappresentati in un determinato quartiere considerando il relativo livello di specializzazione. Ad esempio, si può notare, come in tutte le rappresentazioni, la rete suggerisce che i quartieri specializzati in uffici postali e farmacie, tendono a specializzarsi anche in bar, ma difficilmente in librerie.

Queste osservazioni portano quindi ad una comprensione più approfondita della complementarità tra imprese e servizi nei contesti urbani. Purtroppo, però non si ha evidenza della direzione della complementarità, ovvero se funziona in una sola o entrambe le direzioni: possiamo osservare ristoranti vicino a bar, ma dai dati non possiamo dedurre se è perché i bar preferiscano posizionarsi vicino ai ristoranti o se entrambi i tipi di servizi si suggestionino a vicenda.

#### 6.2 Analisi del Revealed Comparative Advantage (RCA)

Nel paper "The Product Space Conditions the Development of Nations" Hidalgo utilizza la misura della proximity per analizzare l'abilità di un Paese nel produrre un bene in relazione ai beni che già produce.

Per fare un esempio, un Paese produttore ed esportatore di mele, certamente avrà i requisiti necessari per produrre anche pere, in quanto il clima, il terreno e le tecnologie già in uso per coltivare mele, ben si prestano anche alla coltivazione delle pere. Se invece, si considera un prodotto totalmente differente, questo vantaggio viene meno, e tutte le competenze sviluppate per il mercato delle mele risulterebbero inutili.

Al fine di calcolare la proximity è fondamentale conoscere l'RCA, ossia il vantaggio comparato "rivelato".

Secondo l'economista Balassa (Stafforte et al. 2012), l'RCA di una nazione può essere dedotto osservando l'andamento degli scambi commerciali internazionali ex-post, i quali rivelano i vantaggi comparati dei vari Stati.

L'RCAc,i è in grado di determinare, per un Paese c che esporta una quantità di bene pari a i, se tale esportazione risulta superiore alla quota media del commercio globale. Nell'analisi di Hidalgo et al. (2007) i vantaggi comparati sono utilizzati al fine di determinare quali esportazioni risultano essere rilevanti per un Paese (RCA>1), e quali invece risultano essere marginali (RCA<1).

Un'analisi analoga può essere effettuata anche nel nostro caso di studio (Allegato 2). Ossia, grazie all'utilizzo dell'RCA è possibile individuare, per ogni quartiere di ogni città, se questo risulta sovra specializzato o sotto specializzato. Nello specifico, si è utilizzato l'indice di vantaggio comparato rivelato in forma normalizzata, l'indicatore quindi è compreso tra -1 e 1 e rappresenta:

- un quartiere specializzato per RCA > 0
- un quartiere sotto-specializzato per RCA < 0

I dati ottenuti dal calcolo dell'RCA permetto di carpire qualche interessante spunto di analisi. Ad esempio, per quanto riguarda la città di Napoli le amenities sono state suddivise in otto quartieri; i quartieri 2 e 3 risultano sovra specializzati in cinema, teatri e musei (indicati in tabella con "cultural\_entertainment") e in pub. Il quartiere 3 risulta essere quello con "più specializzazioni", oltre alle amenities già citate sono presenti in elevata percentuale anche le scuole ("education" - RCA = 0,41) e si ha una discreta presenza di stazioni di ricarica per auto, benzinai, uffici postali e farmacie (0,28). Risulta invece scarsa la percentuale di ristoranti. Quest'ultimi sono presenti soprattutto nei primi 4 quartieri (quartiere 0, 1 e 2). Le librerie risultano essere la tipologia di amenities meno presente in tutta la città, ad eccezione del quartiere 2 che ha un RCA pari a 0,62.

Nella città di Firenze invece, il quartiere più specializzato in "education" risulta essere il 12 con un RCA pari a 0,745. è interessante notare come la presenza di istituti scolastici in

questo quartiere, abbia presumibilmente influenzato la presenza di librerie che, con un RCA pari a 0,51 rendono il quartiere 12 il più specializzato in questa tipologia di servizi.

#### Conclusioni

Lo studio e l'approfondimento della progettazione urbana richiedono la possibilità di definire insiemi di individui, attività ed infrastrutture a scala sub-urbana che mostrino qualche tipo di correlazione e di funzionalità autonoma. I quartieri definiti per via amministrativa possono fungere allo scopo, tuttavia spesso la loro definizione/morfologia non rispecchia la rete di relazioni correnti e risulta quindi una modalità di partizione dello spazio imprecisa e/o superata. Individuare e ri-definire "quartieri" composti da servizi eterogenei e coerenti con le effettive aree di estensione delle attività sociale che si svolgono su scala sub-urbana richiede un impegno significativo, sia dal punto di vista concettuale, sia da quello empirico.

Facendo riferimento allo studio svolto da Hidalgo et al. (2018) che sfrutta il principio di *relatedness* in questa tesi abbiamo implementato un modello per delineare i quartieri in senso socio-economico, al fine di esprimere una valutazione e proporre dei suggerimenti di ottimizzazione sulla combinazione dei servizi.

Sono state quindi raccolte più di 23.000 amenties da 22 città italiane, con diverse caratteristiche a livello di localizzazione geografica e numero di abitanti, in modo tale da poter svolgere un'analisi che fornisca dei risultati consistenti e robusti. Tramite l'utilizzo di questi dati e di un algoritmo di clustering, si sono costruiti i quartieri così da poter individuare le principali aree economiche delle città e poter studiare le interazioni tra le amenities che ne fanno parte.

I risultati indicano che i cluster sono invariabilmente caratterizzati da un nucleo centrale in cui c'è una maggiore densità di amenities, circondato da un'area periferica in cui i servizi si presentano più dispersi e coprono delle aree più estese.

Già precedentemente si erano svolti degli studi relativi alla mappatura di reti di prodotti correlati (Hausmann et al., 2014; Hidalgo et al., 2007), industrie (Jara-Figueroa et al., 2018; Neffke et al., 2011; Neffke & Henning, 2013), occupazioni (Alabdulkareem et al., 2018;

Muneepeerakul et al., 2013), brevetti (R. Boschma et al., 2015; Kogler et al., 2013), e aree di ricerca (Guevara et al., 2016). Come si può ben immaginare, la struttura delle reti dei servizi dipende da diverse variabili come la numerosità delle amenities analizzate o ancora l'ambito in cui si decide di svilupparlo. Ad esempio, la struttura ottenuta in questo lavoro è diversa da quella derivata nell'area di ricerca che copre campi come la biologia, la chimica e la fisica (Guevara et al., 2016), che si presenta ad anello, mentre è più simile allo spazio dei prodotti, in cui si ha un nucleo ben definito, in questo caso composto prevalentemente dai servizi di ristorazione, circondato da una periferia che include servizi come librerie, centri di intrattenimento e uffici postali. Un aspetto divergente riguarda invece i dati di input che portano a questa struttura, in quanto per gli aggregati di prodotti rispecchia la presenza di input comuni, mentre per i co-agglomerato di servizi tende a riflettere la domanda condivisa, i prezzi degli immobili locali e i vincoli di zonizzazione.

I risultati ottenuti sono soggetti a limitazioni, prime fra tutte l'origine dei dati. L'analisi, infatti, è basata su un dataset costruito con l'utilizzo di un servizio di mappatura online, realizzato con il contributo dei singoli utenti e quindi non ufficialmente verificato; inoltre, l'analisi è stata eseguita solo su 22 città italiane.

Inoltre, le distanze sono calcolate in "linea d'aria" e non sfruttando informazioni sul traffico o condizioni stradali, a causa di mancanza delle risorse e conseguente dei dati relativi a una stima del tempo di viaggio tra le di coppie di punti. Questo aspetto può essere però considerato marginale dal momento che ci si è concentrati su distanze relativamente brevi (<500 m), che non richiedono spostamenti in auto e hanno una bassa probabilità di essere interrotte da caratteristiche geografiche.

Nonostante i limiti descritti, i risultati e i modelli presentati indicano nuovi interessanti risvolti di ricerca urbana e possibili azioni da intraprendere per il relativo miglioramento. Si può pensare di utilizzare i risultati proposti per assistere nella scelta di incentivi commerciali e di tipi di permessi da somministrare per ottenere un bilanciamento dei servizi presenti in una determinata area ed eventualmente come modificarli al fine di migliorare lo sviluppo dei quartieri, oltre che la pedonabilità di una zona.

Un altro importante sviluppo è lo studio dell'interazione tra i servizi di un determinato territorio e gli investimenti immobiliari privati o pubblici che hanno luogo. Questo potrebbe

portare a una migliore comprensione delle esternalità e degli effetti multipli che l'evoluzione dei cluster di servizi crea per i quartieri e viceversa.

Possibili miglioramenti potrebbero riguardare lo strumento di raccolta dati adottato, in modo tale da superare i limiti riscontrati ed avere una più alta attendibilità dei dati.

Un altro sviluppo del lavoro potrebbe interessare l'evoluzione del modello statistico adottato, che attualmente si basa su regressioni lineari, ma che potrebbero essere sostituite da modelli non lineari.

Si potrebbe inoltre pensare di aggiungere informazioni riguardo la densità della popolazione, possibili variazioni stagionali del traffico rilevate dai telefoni cellulari o ancora un'osservazione temporale delle attività nate e fallite nel corso del tempo, monitorando il dinamismo legato ai servizi registrati.

# Bibliografia e Sitografia

Alexander, C., Neis, H., Anninou, A. and King, I. 1987. A new theory of urban design, New York, NY: Oxford University Press.

Alonso, L., Zhang, Y. R., Grignard, A., Noyman, A., Sakai, Y., Elkatsha, M., Doorley, R., & Larson, K. (2019). Data-driven, evidence-based simulation of urban dynamics. use case volpe. Unifying Themes in Complex Systems, IX, 4

Berry, B. (1967). Geography of market centers and retail distribution. Prentice Hall. https://www.amazon.com/Geography-market-centers-retaildistribution/d/B0000CNPHS

Boschma, R., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. Industrial and Corporate Change, 24(1), 223–250. https://doi.org/10.1093/icc/dtu012

Brueckner, J. K. (1993). Inter-store externalities and space allocation in shopping centers. The Journal of Real Estate Finance and Economics, 7(1), 5–16. https://doi.org/10.1007/BF01096932

Buzzacchi, Luigi; Leveque, Philippe; Taramino, Roberta; Zotteri, Giulio (2021). Using betweenness metrics to investigate the geographical distribution of retailers.

César A. Hidalgo, Habitat International, https://doi.org/10.1016/j.habitatint.2020.102205 0197-3975/© 2020 Elsevier Ltd. All rights reserved. The amenity mix of urban neighborhoods

César A. Hidalgo and Elisa E. Castañer (2016). Do we need another coffee house? The amenity space and the evolution of neighborhoods

Christaller, W., & Baskin, C. W. (1966). Central places in southern Germany. Prentice-Hall.

Del Nibletto M. (2021). Pattern di co-localizzazione spaziale delle dotazioni urbane: analisi empirica dei quartieri di Torino.

Dudey, M. (1990). Competition by choice: The effect of consumer search on firm location decisions. The American Economic Review, 80(5), 1092–1104

Eppli, M., & Benjamin, J. (1994). The evolution of shopping center research: A review and analysis. Journal of Real Estate Research, 9(1), 5–32. https://doi.org/10.5555/rees.9.1.l0432475l1363870

Guevara, M. R., Hartmann, D., Aristar´an, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics, 109(3), 1695–1709. https://doi.org/10.1007/s11192-016-2125-9.

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. Environment and Planning A: Economy and Space, 29(7), 1175–1194. https://doi.org/10.1068/a291175

Hernandez, T., & Bennison, D. (2000). The art and science of retail location decisions. International Journal of Retail & Distribution Management. https://doi.org/10.1108/09590550010337391

Hidalgo C.A. et al. (2018) The Principle of Relatedness. In: Morales A., Gershenson C., Braha D., Minai A., Bar-Yam Y. (eds) Unifying Themes in Complex Systems IX. ICCS 2018. Springer Proceedings in Complexity. Springer, Cham. https://doi.org/10.1007/978-3-319-96661-8\_46

Hotelling, H. (1929). Stability in competition. The Economic Journal, 39(153), 41–57. https://doi.org/10.2307/2224214. JSTOR Nagy, D., Villaggi, L., & Benjamin, D. (2018). Generative urban design: Integrating financial and energy goals for automated neighborhood layout. Proceedings of the Symposium on Simulation for Architecture and Urban Design, 25.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. Strategic Management Journal, 34(3), 297–316.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Economic Geography, 87, 237–265

Noyman, A., Doorley, R., Xiong, Z., Alonso, L., Grignard, A., & Larson, K. (2019). Reversed urbanism: Inferring urban performance through behavioral patterns in temporal telecom data. Environment and Planning B: Urban Analytics and City Science, 46(8), 1480–1498

Rakha, T., & Reinhart, C. (2012). Generative urban modeling: A design work flow for walkability-optimized cities. Proceedings of SimBuild, 5(1), 255–262.

Schumacher, P. (2009). Parametricism: A new global style for architecture and urban design. Architectural Design, 79(4), 14–23.

Scitovsky, T. (2013). Welfare & competition. Routledge

Sevtsuk, A. (2014). Location and agglomeration: The distribution of retail and food businesses in dense urban environments. Journal of Planning Education and Research, 4(4), 374–393. https://doi.org/10.1177/0739456X14550401.

Sonta, A. J., & Jain, R. K. (2019). Optimizing neighborhood-scale walkability. Computing in civil engineering 2019: Data, sensing, and analytics. VA: American Society of Civil Engineers Reston

https://overpass-turbo.eu/

https://www.openstreetmap.org/

https://www.forumpa.it/citta-territori/open-street-map-e-lo-stivale-quanto-sono-collaborativi-i-territori-in-italia/