

POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

Pattern di co-localizzazione spaziale delle dotazioni urbane:
analisi empirica dei quartieri di Torino



Relatori

Prof. Giulio Zotteri

Prof. Luigi Buzzacchi

Prof. Antonio De Marco

Candidato

Marco Del Nibletto

Anno Accademico 2020/2021

Sommario

Introduzione.....	8
1. Rassegna della letteratura.....	9
2. Dati.....	16
3. Descrizione del campione.....	19
4. Costruzione dei Quartieri	32
4.1 Dataset principale	41
4.2 Dataset istituzionale	53
5. Costruzione dello spazio delle amenities	64
6. Caratterizzazione dei quartieri	73
7. Conclusioni.....	82
Bibliografia.....	83
Appendice A	87
Appendice B	96

Indice figure e tabelle

Figura 3.1. Area contenente le amenities del dataset.	19
Figura 3.2. Rappresentazione delle amenities delle categorie "amenity" e "shop".	20
Tabella 3.1. Riclassificazione tipologie amenities nel dataset	20
Tabella 3.2. Tipologie amenities scartate.....	23
Tabella 3.3. Amenities presenti nel dataset.....	26
Tabella 3.4. Tipologie inserite nel dataset di amenities istituzionali ...	31
Figura 4.1. Andamento Indice di Accessibilità in funzione della distanza con parametro di decadimento 12.....	34
Figura 4.2. Andamento Indice di Accessibilità in funzione della distanza con parametro di decadimento 20.....	34
Figura 4.3. Rappresentazione delle amenities in base al valore dell'indice di accessibilità.	36
Figura 4.4. Rappresentazione del 10% dei punti con indice di accessibilità più basso.....	38
Figura 4.5. Rappresentazione dei centri individuati con parametro di decadimento 16.....	39
Figura 4.6. Visualizzazione dei quartieri con parametro di decadimento 16.	40
Figura 4.1.1. Visualizzazione dei quartieri con parametro di decadimento 12.	42
Figura 4.1.2. Visualizzazione dei quartieri con parametro di decadimento 20.	43
Figura 4.1.3. Forme dei quartieri con parametro di decadimento 12..	44

Figura 4.1.4. Forme dei quartieri con parametro di decadimento 20..	45
Figura 4.1.5. Numero di quartieri in funzione del parametro di decadimento.	46
Figura 4.1.6. Tipologie di amenities per quartiere in funzione del parametro di decadimento.....	47
Figura 4.1.7. Dati sulle tipologie di amenities per quartiere.	48
Figura 4.1.8. Distanza media minima tra il centro e quello più vicino ad esso.....	48
Figura 4.1.9. Media distanza media tra i centri in funzione del parametro di decadimento.	49
Figura 4.1.10. Area media dei quartieri in funzione del numero di quartieri	50
Figura 4.1.11. Perimetro medio dei quartieri in funzione del numero di quartieri.	51
Figura 4.1.12. Densità media dei quartieri in funzione del parametro di decadimento.	52
Figura 4.2.1. Quartieri da amenities istituzionali con parametro di decadimento 7.....	53
Figura 4.2.2. Quartieri da amenities istituzionali con parametro di decadimento 12.....	54
Figura 4.2.4. Forme dei quartieri da amenities istituzionali con parametro di decadimento 12.	56
Figura 4.2.5. Circoscrizioni della città di Torino.....	56
Figura 4.2.6. Numero di quartieri individuati in funzione del parametro di decadimento.	57

Figura 4.2.7. Numero medio di tipologie di amenities in funzione del parametro di decadimento.....	58
Figura 4.2.8. Distanza dal media tra i centri più vicini al variare del parametro di decadimento.....	59
Figura 4.2.9. Media delle distanze medie tra i centri al variare del parametro di decadimento.....	60
Figura 4.2.10. Area media dei quartieri in funzione del numero di quartieri individuati.	61
Figura 4.2.11. Perimetro medio dei quartieri in funzione del numero di quartieri individuati.	62
Figura 4.2.12. Densità media dei quartieri in funzione del numero di quartieri individuati.	63
Figura 5.1. Spazio delle amenities.....	65
Figura 5.2. Amenity space con correlazioni superiori a 0.5.	67
Figura 5.3. Amenity space filtrato dai quartieri con meno amenities..	68
Figura 5.4. Rappresentazione dei quartieri centrali.....	69
Figura 5.5. Rappresentazione dei quartieri esterni.....	69
Figura 5.6. Rappresentazione dell'amenity space dei quartieri centrali.	70
Figura 5.7. Rappresentazione dell'amenity space dei quartieri esterni.	70
Figura 5.8. Rappresentazione dei quartieri da dataset con amenities più diffuse.....	71
Figura 5.9. Rappresentazione dell'amenity space delle amenities più presenti.....	72
Tabella 6.1.....	73

Figura 6.3. Quartieri caratterizzati per le tipologie clothes, shoes e bag.	79
Figura 6.3. Quartieri caratterizzati per la tipologia restaurant.	80

Introduzione

Nell'evoluzione delle società, le città rappresentano un polo di attrazione della popolazione, offrendo a molti la possibilità di migliorare la propria qualità di vita. Questi luoghi evolvendosi sviluppano numerose attività e servizi differenti per rispondere alle esigenze dei propri abitanti e crescendo si generano internamente sottogruppi per soddisfare le diverse aree cittadine. I cluster generati devono provvedere ai bisogni locali con attività comunemente utili ma possono caratterizzarsi per alcune particolari tipologie di amenities, che potrebbero trarre maggiori vantaggio dal posizionarsi in quelle zone.

In questo lavoro si propone una definizione dei possibili cluster della città di Torino, indagando sulla distribuzione delle attività e dei servizi contenuti al loro interno. Si propone inoltre una possibile caratterizzazione delle diverse aree, in base alle amenities contenute al loro interno.

Per raggiungere questi obiettivi si è replicata la metodologia applicata da Hidalgo nell'articolo scientifico "*The amenity mix of urban neighborhoods*" (2018). Questo studio raccoglie i dati di amenities di 47 città statunitensi e propone un modello che raggruppa i dati in quartieri ed identifica quali tipologie di amenities hanno maggiori probabilità di localizzarsi in prossimità l'una dell'altra.

La tesi è articolata in sette capitoli. Nel primo capitolo viene fornita una rassegna della letteratura riguardante le scelte di localizzazione di attività e servizi all'interno delle città. Il secondo capitolo si sofferma sul confronto tra due servizi alternativi che forniscono i dati utili ad effettuare questo lavoro. Il terzo capitolo illustra i dati utilizzati in questo studio e le operazioni svolte per ottenere il campione con cui vengono svolte le analisi. Nei capitoli quattro e cinque sono descritte alcune fasi operative e metodologiche. Il quarto capitolo illustra l'algoritmo con cui sono stati clusterizzate le amenities ed esamina i possibili risultati del modello ottenuti con diversi parametri. Il quinto capitolo utilizza gli output della fase precedente per comprendere come le amenities siano legate tra loro da un punto di vista spaziale. Nel sesto capitolo si indaga

sulla possibile caratterizzazione delle diverse aree cittadine sulla base delle attività presenti. Le conclusioni ed i possibili sviluppi futuri sono descritti nel capitolo finale.

Questo lavoro fornisce una nuova visione degli spazi cittadini torinesi, attraverso un'inedita definizione dei quartieri. L'approccio utilizzato permette di evidenziare il legame tra le amenities ma anche tra determinate aree e particolari attività commerciali, come verrà mostrato nel capitolo riguardante le conclusioni del lavoro.

1. Rassegna della letteratura

"*A new theory of urban design*" (Alexander, 1987) costituisce una pietra miliare nell'evoluzione della progettazione degli spazi cittadini. Questo lavoro suggerisce un approccio "generativo" della materia, dovendo trattare processi complessi in cui non è possibile definire adeguatamente a priori le specifiche dell'oggetto conclusivo. Questa visione mette da parte la progettazione come sequenza schematica, favorendo un metodo che prevede l'evoluzione degli obiettivi durante le fasi, in cui il gruppo di lavoro deve essere in grado di interpretare il contesto in cui opera, facendo emergere la migliore soluzione in linea con le finalità progettuali.

Questo avanzamento si sposa con la progettazione parametrica, che negli ultimi anni viene sempre più utilizzata dall'urbanistica per comprendere l'evoluzione del territorio e realizzare soluzioni su vasta scala (Schumacher, 2009).

Gli sviluppi di questa disciplina sono favoriti dalla maggior potenza di calcolo offerta dalla tecnologia, affiancata dalle competenze sviluppate nel campo dei big data. Applicazioni di queste modalità di lavoro hanno permesso la costruzione di modelli per migliorare l'efficienza energetica (Nagy et al., 2018), la forma (Alonso et al., 2019; Noyman et al., 2019) e la pedonalità dei quartieri (Rakha & Reinhart, 2012; Sonta & Jain, 2019, pp. 454–461).

L'articolo "*The amenity mix of urban neighborhoods*" (Hidalgo et al., 2018) è uno studio che offre un esempio di questo genere di sviluppi. L'autore applica le nuove conoscenze per comprendere come queste possano essere estese a caratteristiche non quantitative dei quartieri, come individuare il mix di amenities presenti nelle aree cittadine più vivibili. Per sviluppare le analisi, vengono raccolti i dati di oltre 1 milione di amenities relative a 47 città statunitensi e studiato il loro pattern di co-localizzazione. Analizzando la correlazione delle diverse tipologie di amenities all'interno dei quartieri è possibile sviluppare un modello che fornisce indicazioni su quali amenities avranno una maggiore o minore probabilità di sviluppo, in relazione a quelle già presenti nell'area di riferimento.

L'articolo di Hidalgo si basa sull'applicazione di un concetto empirico, il principio della *relatedness*: la probabilità che un Paese, o una regione geografica, intraprenda con successo un'attività economica dipende dalle attività ad essa correlate presenti in quell'area (Hidalgo et al., 2018). Allo stesso modo, anche i risultati negativi sono soggetti a questo criterio.

Gli studi per affermare questo principio hanno indagato sulla presenza di cluster di attività economiche a livello internazionale e regionale, per poi allargare gli studi su scale di livello inferiore, fino ad aree metropolitane. In questo lavoro, come in "*The amenity mix of urban neighborhoods*", si cerca di estendere l'applicazione a livello cittadino, affrontando alcune difficoltà che questo ampliamento comporta.

La ragione per cui il principio della *relatedness* presenta un elevato grado di riproducibilità, sia su scala geografica che per temi trattati, risiede negli svariati modi con cui le economie e le organizzazioni sono in grado di apprendere (Hidalgo et al., 2018).

Questo principio non è legato a banali rapporti di input-output, fondamentali in contesti che presentano costi di trasporto relativamente alti, come poteva accadere alcuni decenni fa, ma nelle dinamiche attuali, fornitore e produttore non devono essere necessariamente vicini. Ad esempio, si è passati da diversi quotidiani cittadini consegnati manualmente ad avere poche testate principali

per ogni Paese, raggiungibili da tutti attraverso i principali motori di ricerca (Hidalgo et al., 2018).

Negli ultimi decenni il rapporto tra costi di spostamento della conoscenza e costi di trasporto dei suoi frutti, ossia i prodotti, è aumentato di pari passo ai progressi tecnologici. Questo cambiamento è dettato dai miglioramenti raggiunti nel trasporto delle merci in senso ampio, ma un analogo sviluppo non si è verificato per la diffusione della conoscenza. Nonostante l'accesso ad essa sia stato facilitato e si siano ridotti i tempi di trasporto delle informazioni, il sapere generato risulta spesso legato alle persone e ai rapporti umani. Inoltre i benefici derivanti dalle esternalità di conoscenza decadono rapidamente all'aumentare della distanza dal luogo dove sono generate e ciò diventa particolarmente rilevante in ambiti e settori tecnologici caratterizzati da un'elevata presenza di informazioni non codificate (Hidalgo et al., 2018).

Il diffondersi del principio della "*relatedness*" potrebbe suggerire miglioramenti nel campo delle politiche industriali. Un suo approccio pragmatico, in questo senso, migliorerebbe l'utilizzo delle risorse economiche, evitando investimenti in attività non in grado di sostenersi senza aiuti o che non ne necessitano sensibilmente. Per raggiungere questi obiettivi, le politiche industriali non dovrebbero limitarsi a identificare le attività economiche più promettenti. Il principio suggerisce di favorire i meccanismi che permettono il diffondersi delle conoscenze tra industrie e regioni, spingendo verso la diversificazione e non la specializzazione in un'unica direzione (Hidalgo et al., 2018).

Il valore del principio è stato evidenziato da una moltitudine di lavori basati sull'analizzare gli effetti sulle strutture economiche di attività legate tra loro (Boschma, 2005; Guevara et al., 2016; Hidalgo et al., 2018, 2007; Jara-Figueroa et al., 2018; Neffke et al., 2011; Neffke e Henning, 2013).

Uno studio del 2015 (Boschma et al.) attraverso l'analisi di 366 diverse città statunitensi tra il 1981 e il 2010 ha evidenziato come innovazioni e brevetti siano un motore per il cambiamento tecnologico a livello cittadino. In dati questo si traduce in un aumento della probabilità del 30% di sviluppare la nuova tecnologia se il legame con quelle presenti aumenta del 10%.

Analogamente, si riduce la probabilità di uscire delle tecnologie esistenti dell'8%.

Inoltre, esternalità tra industrie e settori diversi ma complementari tra loro permettono una crescita del tasso di occupazione, mentre altri tipi di attività non permettono di avere risultati analoghi (Frenken et al., 2007).

Applicando il principio su una dimensione cittadina il problema a cui si va incontro è l'individuazione di aree dai confini ben definiti. Cluster urbani sono necessari per capire come le amenities si rapportano tra loro, come sono raggruppate e che benefici offrono ai cittadini.

Negli ultimi anni, gli edifici situati in zone che rendono disponibili una maggiore varietà di servizi (ad esempio trasporti, negozi, aree pedonali e parchi) vengono ritenute sempre più di maggior valore da parte degli abitanti. Ciò si riflette sul prezzo degli immobili ed è evidenziato anche dallo svolgimento di un maggior numero di iniziative locali in queste aree (Hidalgo et al, 2018).

I poli di amenities esercitano una forte attrazione sulle scelte di localizzazione, sia di attività economiche che amministrative. Nelle città di Washington D.C. e di Atlanta, nonostante l'area pedonale costituisca solo l'1% del suolo, tra il 2009 e il 2013 ha attirato il 50% dei nuovi uffici, negozi, hotel e appartamenti (Hidalgo et al., 2018). Inoltre, la possibilità di avere nelle vicinanze della propria abitazione svariate amenities, spinge una vasta fascia della popolazione ad avvicinarsi ai centri cittadini (Florida, 2019). Molti cittadini risultano disposti a pagare un prezzo aggiuntivo per godere dei vantaggi presenti in quelle aree (Leinberger & Lynch, 2014).

La letteratura a disposizione che illustra le scelte di localizzazione risulta spesso troppo astratta e non permette di addentrarsi sufficientemente nelle dinamiche di quartiere per comprendere come i vari servizi generino benefici (Christaller e Baskin, 1966; Eaton e Lipsey, 1982; Fujita e Krugman, 2004; Hotelling, 1929; Krugman, 1993; Losch, 1954; Marshall, 1890; Mulligan, 1984). Questo rende difficoltoso comprendere come l'inserimento di una particolare attività commerciale o uno spazio pubblico influiscano sul tessuto cittadino. Gli studi disponibili eseguono una classificazione grossolana delle

amenities in macrocategorie, dividendone gli effetti tra esogeni ed endogeni (Hidalgo et al, 2018).

Gli effetti esogeni si rilevano nel momento in cui più amenities si localizzano attorno ad una risorsa esterna di comune interesse per le clientele (Berry, 1967). Esempi sono offerti da città come Londra che presentano una fitta rete di trasporto metropolitano, dove le trafficate stazioni e fermate della metro attraggono un gran numero di individui. L'elevata circolazione di pubblico in queste aree spinge servizi e negozi a posizionarsi nelle vicinanze, attirando ulteriore traffico cittadino e locale, conseguente all'effetto di spillover dei numerosi spostamenti (Hidalgo et al, 2018).

Scelte di localizzazione dettate da effetti esogeni possono verificarsi sia nel caso di negozi rivali, che nel caso di negozi che offrono beni complementari. La quotidianità ci offre molti esempi in questo senso, come incroci di strade trafficate o centri cittadini che attirano abitanti e turisti, anche in presenza di effetti endogeni (Hidalgo et al., 2018).

La presenza di un fattore esterno che genera un vantaggio non garantisce successo nel lungo periodo. La possibilità che un punto di attrazione esterno generi un cluster di amenities attorno a sé dipende da vari fattori. Primi tra tutti sono la raggiungibilità da parte dei consumatori e la posizione di cluster rivali che potrebbero deviare la clientela (Hidalgo et al., 2018).

Un modello per questo genere di competizione è stato sviluppato da Christaller (1966), facendo riferimento alla "*central place theory*". Secondo questa teoria, gli elementi si distribuiscono naturalmente in maniera centralistica. Questa distribuzione è ben osservabile nella materia organica e inorganica, ma anche in elementi astratti come le organizzazioni. Nelle città è possibile evidenziare come gli organi che la compongono siano distribuiti, secondo il modello, attorno a un "*central place*". La posizione del centro e l'area che si sviluppa attorno influiscono sulla competizione con i cluster confinanti (Christaller e Baskin, 1966).

Un posto definito centrale non è il punto che minimizza le distanze dagli abitanti dal centro ma deve svolgere la funzione di centro. In questo senso, assume la funzione di centro, se gli abitanti si recano in quel luogo per

svolgere le loro mansioni, "*central professions*", generando i "*central products*" e i "*central services*". Questo genere di beni viene prodotto e offerto nelle zone centrali per poi essere consumato nei vari cluster cittadini. Analogamente, sono presenti i "*dispersed goods*" e i "*dispersed services*", che non vengono prodotti e consumati nelle aree centrali. Esistono anche beni e servizi che sono indifferenti all'area di produzione e consumo, gli "*indifferent goods*" e gli "*indifferent services*" (Christaller e Baskin, 1966).

Un'altra classificazione derivante dalla teoria di Christaller divide beni e servizi in base alla loro frequenza d'acquisto. Ad esempio, beni acquistati poco frequentemente, come elementi d'arredo o elettrodomestici, saranno presenti in un minor numero di cluster ma che godono di un maggiore bacino d'affluenza (Christaller e Baskin, 1966).

Ci sono anche effetti endogeni che spingono le amenities a co-localizzarsi e possono essere legati alla complementarità dei beni ma anche alla competizione tra essi.

La localizzazione dovuta a complementarità fa riferimento a beni e servizi non direttamente in competizione tra loro e che vengono generalmente acquistati nella stessa occasione d'acquisto, come un gelataio presente davanti ad una pizzeria. Questa scelta di posizionamento può essere motivata da un vantaggio che viene offerto al consumatore, riducendo i suoi spostamenti e il tempo impiegato per ottenere quei prodotti (Eppli e Benjamin, 1994).

Cluster di negozi complementari possono nascere sia per attirare il maggior traffico prodotto dai ridotti costi di trasporto della clientela (sul lato razionale) ma anche per sfruttare le esternalità che si generano a seguito dell'effetto di spillover (Hernandez e Bennison, 2000). In questa maniera si sfrutta il maggior pubblico dei negozi più popolari, che in parte sarà deviato su quelli vicini. Il classico esempio è fornito dai centri commerciali e dai negozi ancora. A questo genere di esercizi commerciali vengono offerti benefici, come posizione all'interno del complesso, e vantaggiosi contratti di locazione, per attirare un pubblico il più vasto possibile. L'esternalità prodotta viene venduta ai negozi minori, che non sarebbero in grado di generare un analogo afflusso di clientela verso di loro se fossero isolati (Brueckner, 1993).

Attività in competizione possono generare aree dedicate a una specifica categoria merceologica o di servizi, posizionandosi nello stesso cluster e rendendo non vantaggioso restare fuori da esso. Possibili esempi sono forniti dalle attività di ristorazione disposte lungo la stessa via: i consumatori visitano sì un ristorante per occasione di consumo, ma, in una circostanza analoga successiva, è maggiormente probabile che decidano di visitare un'altra attività dello stesso cluster (Hidalgo et al, 2018).

Un modello che esprime queste logiche di posizionamento è quello di Hotelling (1929), che spiega come la scelta migliore in determinate situazioni sia quella che presenta il minor rischio. In queste circostanze, i competitor scelgono di posizionarsi in modo tale che per gli altri non sia conveniente attuare azioni atte a rubare traffico dai vicini. Queste dinamiche spingono a localizzarsi nelle vicinanze gli uni degli altri e si traduce come possibilità di vantaggio per i consumatori, a cui si offre la possibilità di confrontare i prodotti, riducendo i loro costi di ricerca.

Si è evidenziato come mettere a disposizione della clientela la possibilità di esaminare beni simili spinga i competitor a formare cluster (Eaton e Lipsey, 1975; Scitovsky, 2013). La varietà di negozi a disposizione è un importante driver di scelta per i consumatori e studi hanno mostrato che centri commerciali che presentano un'ampia gamma al loro interno hanno maggiori guadagni (Hise et al., 1983).

La competizione all'interno del cluster ha conseguenti effetti sui prezzi che vengono offerti al pubblico. Per aumentare la propria quota di visitatori, i commercianti offrono prezzi vantaggiosi tentando di aggiudicarsi una fetta più ampia. Come conseguenza di questa competizione sui prezzi, il numero complessivo di visitatori del cluster aumenta, generando un effetto a catena che spinge sempre più i negozi a ridurre i prezzi, in quanto viene facilitata la possibilità dei consumatori di confrontare i rivali (Dudey, 1990). La scelta dei commercianti se entrare o meno nel cluster e competere in questa maniera, è relativa ai possibili vantaggi generati dall'aumento delle vendite, confrontato con la riduzione dei guadagni delle singole transazioni (DiPasquale e Wheaton, 1995).

2. Dati

Il principio della *relatedness* fornisce una possibile previsione delle attività economiche in grado di svilupparsi in una determinata zona, in base alla correlazione con le attività presenti. Le applicazioni riguardano generalmente aree internazionali, Paesi e regioni che presentano confini definiti.

Per replicare i concetti di questo principio su una scala geografica minore è presente la problematica dell'individuazione di zone definite. Le aree urbane offrono divisioni in circoscrizioni e quartieri, che forniscono un'idea di come una città sia strutturata, da un punto di vista geografico e amministrativo, ma non permettono di comprendere le dinamiche e gli spostamenti che intercorrono tra gli abitanti e la città.

Dato l'obiettivo della tesi di applicare il principio della *relatedness* per comprendere come la presenza di attività economiche e sociali influiscono su un vicinato, viene applicato un algoritmo che identifica dei nuovi quartieri in base alla distribuzione delle amenities sul suolo cittadino.

Nell'articolo "*The amenity mix of urban neighborhoods*" vengono raccolte le informazioni di 1,26 milioni di amenities di 47 città statunitensi. Ogni amenity è caratterizzata dalla propria posizione attraverso latitudine e longitudine, e dalla tipologia, in base all'attività di cui si occupano. La raccolta dati per la costruzione dei quartieri viene eseguita tramite Google Places API, un servizio a pagamento di Google che permette di raccogliere queste informazioni tramite richieste HTTP.

Esistono vari servizi che permettono di ottenere le coordinate geografiche di punti sulla superficie terrestre, tra questi vi è OpenStreetMap. OpenStreetMap (OSM) è un progetto mondiale open source che rende disponibili una vastità di dati geografici gratuitamente. Le informazioni presenti all'interno di OSM riguardano numerose categorie: dagli elementi geografici naturali, come corsi d'acqua, alle infrastrutture artificiali. I dati possono essere inseriti direttamente da utenti, che decidono di diventare membri dell'organizzazione OSM pagando una quota, ma anche da fonti governative e imprese commerciali.

Per recuperare i dati di OSM si può ricorrere a vari mezzi. Overpass Turbo è uno strumento di data mining per OSM che permette di generare automaticamente query che restituiscono dati in base agli input di ricerca. Come in Google Places API, anche Overpass Turbo permette di selezionare l'area su cui effettuare la ricerca. La tipologia dei nodi di interesse può essere selezionata tra le numerose categorie. In Overpass Turbo non sono presenti limitazioni riguardo all'esportazione dei dati disponibili, al contrario di Google Places API.

Google Places API e OpenStreetMap presentano entrambi delle peculiarità, con conseguenti punti di forza e problematiche.

Trattandosi di un servizio di un'impresa commerciale da una parte, e di una fondazione non a scopo di lucro dall'altra, bisogna precisare che si tratta di due progetti con finalità molto diverse. Le informazioni raccolte su Google Places API sono rivolte principalmente a soddisfare le esigenze dei consumatori finali, raccogliendo dati su negozi e punti di interesse rivolti a questi ultimi. Il differente scopo della fondazione OSM: *"incoraggiare la crescita, lo sviluppo e la distribuzione di dati geospaziali liberi e fornirli a chiunque per l'uso e la condivisione"* ha conseguenze anche sulla struttura del database, che include molte più categorie di nodi, rischiando di apparire confusionario a chi si avvicina al servizio per la prima volta.

La copertura del territorio di OSM risulta meno omogenea del servizio di Google. Le finalità e i mezzi a disposizione di Google Places API permettono di offrire una banca dati molto ampia. OpenStreetMap non offre una numerosità paragonabile ed inoltre i dati non risultano distribuiti in maniera omogenea sul territorio. Questa problematica può essere dovuta sia dalla modalità di inserimento dei dati che alla disponibilità dei dati stessi. Ad esempio, in alcune zone potrebbero essere presenti maggiori informazioni, in base alla distribuzione dei membri di OSM sul territorio. Un possibile esempio di distorsione dovuta alla disponibilità dei dati può verificarsi per amenities la cui posizione è regolamentata, come ad esempio le farmacie, o che sono di particolare rilevanza per un territorio, come i *torèt*, tipiche fontanelle pubbliche della città di Torino. Confrontando la copertura sul territorio di Torino, OSM e Google Places API offrono un numero paragonabile di farmacie

nel database. Eseguendo un confronto con una categoria affine ma non regolamentata per quanto riguarda la posizione, come gli ambulatori medici, il database di Google Places API risulta molto più completo rispetto a OSM.

I due servizi presentano alcune problematiche comuni, come non fornire informazioni sulle date di apertura e chiusura delle attività nel tempo. Questo comporta la presenza di amenities che sono state rimosse o che sono state spostate in altre località. Questa non risulta essere una limitazione per questo lavoro ma non permette di svolgere studi di approfondimento analoghi ad altre applicazioni del principio della *relatedness*. Altra questione presente in OSM e Google Places API è la presenza di alcune amenities non esattamente categorizzate, come "bar" che rientrano nella voce "café". Si tratta di casi isolati che non influiscono sui risultati.

Google Places API è un servizio a pagamento che richiede di sostenere dei costi in base al numero di richieste effettuate. Vi è la possibilità di acquistare pacchetti che riducono i costi unitari, accessibili a fronte di una spesa molto consistente. Per coprire un territorio ampio è necessario effettuare un gran numero di richieste HTTP e conseguente impiego di risorse per ridurre i tempi per compilarle. Questa sfida è superabile attraverso script che programmano multiple richieste API ma si rischia di andare incontro alle licenze di Google nel caso si faccia un utilizzo improprio dei dati raccolti.

OSM, d'altra parte, elimina questi problemi, offrendo i dati gratuitamente e con una costruzione veloce delle query che permette di raggiungere immediatamente le informazioni di interesse, anche su vasta scala, con tempi di elaborazione minimi.

In base alla disponibilità economica e a eventuali problematiche sulla licenza, il servizio utilizzato per il reperimento dei dati per il lavoro in questione è OSM.

3. Descrizione del campione

Il dataset contenente le informazioni sui nodi che formeranno i quartieri viene costruito recuperando i dati da OSM. Dall'ampia classificazione offerta dal servizio, la scelta ricade sui nodi che rispondono alla classificazione "amenity" e "shop". Vengono scelte queste due categorie in quanto al loro interno sono contenuti i nodi che rispondono meglio agli obiettivi del lavoro e inoltre presentano una similarità con la classificazione dell'articolo di Hidalgo (2018).

Lo strumento di data mining utilizzato è Overpass turbo, a cui bisogna fornire in input il centro e il raggio del cerchio dell'area di interesse. La scelta del posizionamento del centro ricade sulla coordinata (45.06; 7.66) in corrispondenza della fermata metro Vinzaglio. Questo punto, con la sua centralità geografica e estendendo il raggio fino a 10.000 metri, permette di includere l'intero territorio comunale della città di Torino, come visibile in figura 3.1. Dato che le scelte di posizionamento delle attività commerciali e gli spostamenti dei cittadini non si limitano a confini definiti, l'estensione del raggio permette di includere anche possibili poli di interesse presenti nelle zone limitrofe, come l'area attorno al centro commerciale *Le Gru*.

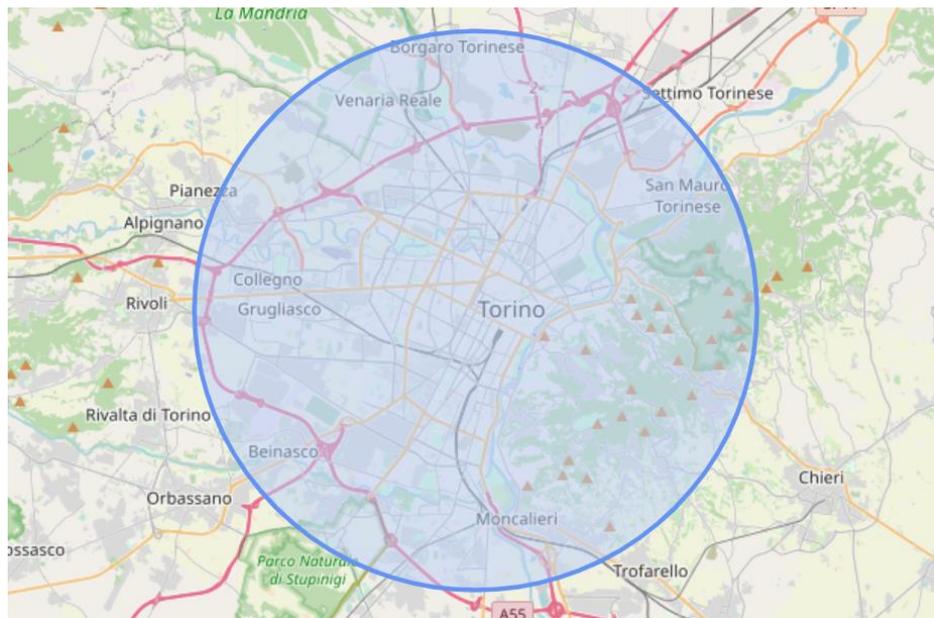


Figura 3.1. Area contenente le amenities del dataset.

Dall'area analizzata si ottengono 6.544 nodi dal tag "amenity", distribuiti in 92 tipologie e 2.086 nodi dal tag "shop", distribuiti all'interno di 133 tipologie. Si ottiene un dataset contenente 8.630 nodi distribuiti in 225 categorie, visibili nella figura 3.2.

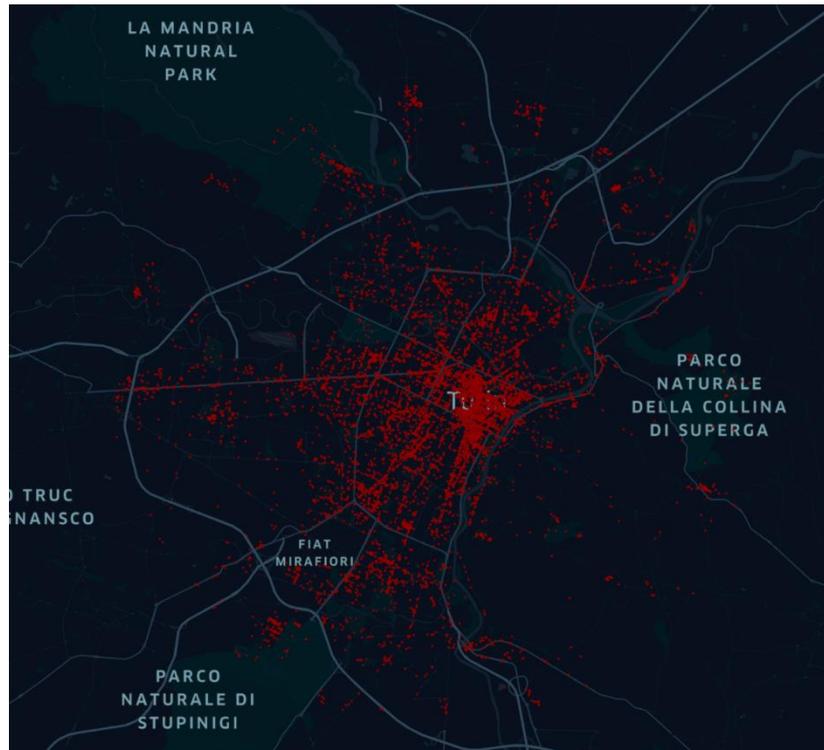


Figura 3.2. Rappresentazione delle amenities delle categorie "amenity" e "shop".

Dato l'elevato numero di tipologie presenti nel dataset, viene svolta una riclassificazione accorpando quelle che risultano simili e/o svolgono una stessa funzione, analogamente a quanto fatto nell'articolo originale. Nella tabella 3.1. vengono elencate le tipologie che sono state accorpate e la nuova tipologia ad esse assegnata.

Tabella 3.1. Riclassificazione tipologie amenities nel dataset

Tipologie originali	Presenza	Nuove tipologie	Presenza
Alcohol	39		
Wine	14	Alcohol	54
Brewing_supplies	1		
Bar	267		
Cafe	458	Bar	725
Bookmaker	4		
Gambling	1	Bookmaker	5
Clothes	202		
Boutique	1	Clothes	203
Social_facility	8		
Nursing_home	1	Social_facility	9
Photo	7		
Camera	1	Photo	8
Office	5		
Coworking_space	6	Office	11
Beauty	27		
Solarium	1		
Spa	1	Beauty	32
Massagge	3		
Art	5		
Art_centre	15	Art	21
Art_gallery	1		
Interior_decoration	3		
Window_blind	2		
Carpet	1	Interior_decoration	19
Curtain	4		
Antiques	9		
Furniture	22		
Kitchen	6		
Bed	5	Furniture	38
Bathroom_furnishing	4		
Doors	1		

Fuel	313	Fuel	336
Charging_station	23		
Convenience	127	Convenience	131
General	4		
Confectionery	52	Confectionery	57
Chocolate	5		
Community_centre	8	Community_centre	12
Conference_centre	4		
Electronics	17		
Hifi	5		
Computer; mobile-phone	1	Electronics	52
Computer	16		
Mobile_phone	13		
Pet	22	Pet	23
Pet_grooming	1		
Laundry	15	Laundry	31
Dry_cleaning	16		
Parking	152		
Parking_space	14	Parking	169
Motorcycle_parking	3		
Post_office	64	Post_office	65
Post_office;ATM	1		
Cafe	456		
Cafe; Restaurant	1	Cafe	458
Café; Fast_food	1		
Vehicle_inspection	1		
Tyres	9		
Car_repair	61	Car_service	83
Car_parts	12		
Motorcycle	12	Motorcycle	13
Motorcycle_repair	1		

Newsagent	43	Newsagent	45
Newspaper	2		

Un'altra operazione svolta, per evitare ambiguità nelle fasi di analisi successive, è la rimozione di alcune categorie di nodi. Analogamente a quanto fatto nel lavoro di Hidalgo (2018), vengono eliminate le amenities poco presenti o che non risultano essere adeguate ai fini del lavoro. Vengono eliminati 2.885 nodi appartenenti a 69 tipologie, indicate nella tabella 3.2, indicando la rispettiva presenza percentuale rispetto al dataset completo.

Tabella 3.2. Tipologie amenities scartate

Amenity	Punti	Percentuale
Drinking_water	778	9,02%
Bench	698	8,09%
Bicycle_parking	422	4,89%
Waste_basket	273	3,16%
Chair	93	1,08%
Parking_entrance	90	1,04%
Vending_machine	87	1,01%
Toilets	70	0,81%
Post_box	68	0,79%
Recycling	67	0,78%
Telephone	54	0,63%
Fountain	29	0,34%
Clock	15	0,17%
Bbq	14	0,16%

Waste_disposal	13	0,15%
Table	10	0,12%
Vacant	7	0,08%
Photo_booth	5	0,06%
Shelter	4	0,05%
Public_bath	3	0,03%
Agency	3	0,03%
Watering_place	2	0,02%
Water_point	2	0,02%
Payment_terminal	2	0,02%
Frame	2	0,02%
Rescue_station	1	0,01%
Fire_station	1	0,01%
Food_court	1	0,01%
Music_school	1	0,01%
Bicycle_repair_station	1	0,01%
Dojo	1	0,01%
Social_centre	1	0,01%
Animal_shelter	1	0,01%
Compressed_air	1	0,01%
Piano	1	0,01%
Research_institute	1	0,01%
Planetarium	1	0,01%
Lavoir	1	0,01%
Numismatics	1	0,01%
Collector	1	0,01%
Frozen_food	1	0,01%

Model	1	0,01%
Hobby	1	0,01%
Craft	1	0,01%
Religion	1	0,01%
Locksmith	1	0,01%
Money_lender	1	0,01%
Food	1	0,01%
Leather	1	0,01%
Scuba_diving	1	0,01%
Fabric	1	0,01%
Party	1	0,01%
Fair_trade	1	0,01%
e-cigarette	1	0,01%
Health_food	1	0,01%
Games	1	0,01%
Cheese	1	0,01%
Wholesale	1	0,01%
Fashion_accessories	1	0,01%
Printer_ink	1	0,01%
Cannabis	1	0,01%
Plaques	1	0,01%
Nutrition_supplements	1	0,01%
Dairy	1	0,01%
Sewing	1	0,01%
Rental	1	0,01%
Hairdresser_supply	1	0,01%
Appliance	1	0,01%

Assicurance	1	0,01%
-------------	---	-------

Le scelte che hanno portato all'eliminazione di queste amenities sono diverse. Si è scelto di eliminare i tipi presenti con un solo nodo, per evitare classificazioni che costituiscono eccezioni nello sviluppo organico della città. Sono state rimosse anche amenities molto presenti all'interno del dataset originale, come "drinking_water" e "bench", che risultano essere rispettivamente la prima e la terza tipologia più presenti. In questi casi si è ritenuto che la loro influenza nelle operazioni successive fosse troppo elevata, non rispondendo adeguatamente alla definizione di amenity di questo lavoro.

Dalle operazioni svolte in precedenza si arriva ad avere un dataset composto da 5.775 amenities in 113 tipologie. Nella tabella 3.3 sono inserite le tipologie con la rispettiva frequenza e percentuale rispetto alle amenities che hanno superato le operazioni precedenti.

Tabella 3.3. Amenities presenti nel dataset.

Amenity	Punti	Percentuale
Restaurant	725	12.55%
Bar	725	12.55%
Pharmacy	415	7.19%
Fuel	336	5.82%
Clothes	203	3.52%
fast_food	202	3.50%
Bank	181	3.13%
bicycle_rental	178	3.08%
Parking	169	2.93%
supermarket	163	2.82%

convenience	131	2.27%
hairdresser	116	2.01%
Bakery	115	1.99%
Pub	93	1.61%
ice_cream	86	1.49%
car_service	83	1.44%
Tobacco	79	1.37%
post_office	65	1.13%
Atm	64	1.11%
Bicycle	64	1.11%
Butcher	63	1.09%
Books	59	1.02%
confectionery	57	0.99%
Alcohol	54	0.94%
Electronics	52	0.90%
School	47	0.81%
Newsagent	45	0.78%
Furniture	38	0.66%
greengrocer	36	0.62%
Taxi	36	0.62%
Beauty	32	0.55%
Car	32	0.55%
Laundry	31	0.54%
University	31	0.54%
Optician	30	0.52%
kindergarten	29	0.50%
Kiosk	29	0.50%

Florist	28	0.48%
Library	28	0.48%
car_wash	26	0.45%
Hardware	26	0.45%
Jewelry	26	0.45%
Pastry	26	0.45%
place_of_worship	24	0.42%
Police	24	0.42%
Pet	23	0.40%
Shoes	23	0.40%
Sports	23	0.40%
Clinic	22	0.38%
Stationery	22	0.38%
Art	21	0.36%
Doctors	21	0.36%
doityourself	21	0.36%
Veterinary	21	0.36%
travel_agency	20	0.35%
Deli	19	0.33%
interior_decoration	19	0.33%
Theatre	18	0.31%
Cinema	16	0.28%
Dentist	16	0.28%
Cosmetics	15	0.26%
Hospital	14	0.24%
driving_school	13	0.23%
Motorcycle	13	0.23%

second_hand	13	0.23%
community_centre	12	0.21%
Nightclub	12	0.21%
marketplace	11	0.19%
Office	11	0.19%
Copyshop	10	0.17%
Gift	10	0.17%
Paint	10	0.17%
Tattoo	10	0.17%
car_rental	9	0.16%
Chemist	9	0.16%
Coffee	9	0.16%
department_store	9	0.16%
funeral_directors	9	0.16%
Houseware	9	0.16%
social_facility	9	0.16%
Herbalist	8	0.14%
Photo	8	0.14%
car_sharing	7	0.12%
Music	7	0.12%
Seafood	7	0.12%
Toys	7	0.12%
Bag	6	0.10%
Childcare	6	0.10%
Outdoor	6	0.10%
Tailor	6	0.10%
Tea	6	0.10%

variety_store	6	0.10%
bookmaker	5	0.09%
garden_centre	5	0.09%
musical_instrument	5	0.09%
Perfumery	5	0.09%
bureau_de_change	4	0.07%
Erotic	4	0.07%
language_school	4	0.07%
Mall	4	0.07%
baby_goods	3	0.05%
estate_agent	3	0.05%
Farm	3	0.05%
medical_supply	3	0.05%
Pasta	3	0.05%
radiotechnics	3	0.05%
Studio	3	0.05%
Townhall	3	0.05%
Trade	3	0.05%
College	2	0.03%
Electrical	2	0.03%
Hearing_aids	2	0.03%
Ticket	2	0.03%

Dal dataset viene formato un sotto dataset con amenities "istituzionali". Con questo termine si intende definire quei servizi gestiti o regolati pubblicamente che si riferiscono alla collettività. All'interno sono contenuti 957 nodi appartenenti a 18 tipologie. Nella tabella 3.4 sono contenute le amenities che

rispondono alla definizione di istituzionale, con la rispettiva frequenza e presenza percentuale rispetto alla numerosità di questo dataset.

Tabella 3.4. Tipologie inserite nel dataset di amenities istituzionali

Amenity	Punti	Percentuale
Pharmacy	415	43.36%
Parking	169	17.66%
post_office	65	6.79%
School	47	4.91%
Taxi	36	3.76%
University	31	3.24%
kindergarten	29	3.03%
Library	28	2.93%
Police	24	2.51%
place_of_worship	24	2.51%
Clinic	22	2.30%
Art	21	2.19%
Hospital	14	1.46%
community_centre	12	1.25%
social_facility	9	0.94%
Childcare	6	0.63%
Townhall	3	0.31%
College	2	0.21%

4. Costruzione dei Quartieri

Ottenuti i dataset, si procede con la fase successiva di costruzione dei quartieri. Questa parte viene eseguita tramite un algoritmo sviluppato con *Python*. Si è deciso di utilizzare questo linguaggio per sfruttarne la semplicità e le elevate performance fornite.

I dataset contengono al loro interno il codice identificativo, la latitudine, la longitudine e la tipologia di ogni amenity. Queste informazioni permettono di svolgere la costruzione dei quartieri.

I dataset vengono forniti ad un algoritmo di *clustering*, che aggrega le amenities in base al rispettivo indice di accessibilità, tenendo conto delle distanze con le amenities vicine. L'algoritmo individua i punti con maggiore indice di accessibilità, chiamati "picchi", attorno ai quali vengono distribuite le amenities formando i quartieri. La procedura è simile alle metodologie che clusterizzano le amenities basandosi sui pattern con cui vengono visitate (Cranshaw et al., 2012).

L'indice di accessibilità fornisce per ogni amenity una misura di quanto essa sia inserita nel tessuto cittadino. Questo valore permette di tradurre il concetto astratto di accessibilità, definendo il numero di località raggiungibili partendo da un punto sulla mappa (Handy e Niemeier, 1997).

La letteratura mette a disposizione un'ampia scelta di metodologie per il calcolo dell'indice di accessibilità, rispondendo ad esigenze diverse. In questo lavoro, per cercare di replicare quanto svolto nell'articolo di Hidalgo (2018), si utilizza la formula (4.1) che restituisce una misura *gravity-based* dell'indice di accessibilità (Handy e Niemeier, 1997).

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

(4.1)

Nella formula (4.1): A_i rappresenta l'indice di accessibilità dell'amenity i -esima, N è il numero di amenities presenti nel dataset, γ è il parametro di decadimento e d_{ij} rappresenta la distanza tra l'amenity i -esima e j -esima.

Data la struttura della formula, il contributo che viene fornito dalle N amenities all'indice di accessibilità dell'amenity i -esima, dipende dal parametro di decadimento e dalla distanza con le altre amenities. Ogni amenity può offrire un contributo all'indice di accessibilità delle altre che varia tra 0 e 1. L'indice di accessibilità dell'amenity i -esima è infine ottenuto sommando i contributi delle N amenities presenti nel dataset.

Due amenities situate nella stessa coordinata presenteranno distanza nulla e offriranno vicendevolmente un contributo pari a 1 all'indice di accessibilità dell'altra. All'aumentare della distanza, l'apporto offerto all'indice di accessibilità si riduce. Ad esempio, due amenities situate a una distanza $d_{ij} = \ln(2)/\gamma$ contribuiscono per un valore pari a $1/2$ ai loro indici di accessibilità.

La scelta del parametro di decadimento influenza i risultati che si ottengono. Il valore dell'indice di accessibilità è inversamente proporzionale al parametro di decadimento, in quanto aumentando questo valore si sconta maggiormente l'apporto delle amenities.

Nelle figure 4.1 e 4.2 viene mostrato come il variare del parametro di decadimento influenzi il valore dell'indice di accessibilità al variare delle distanze.

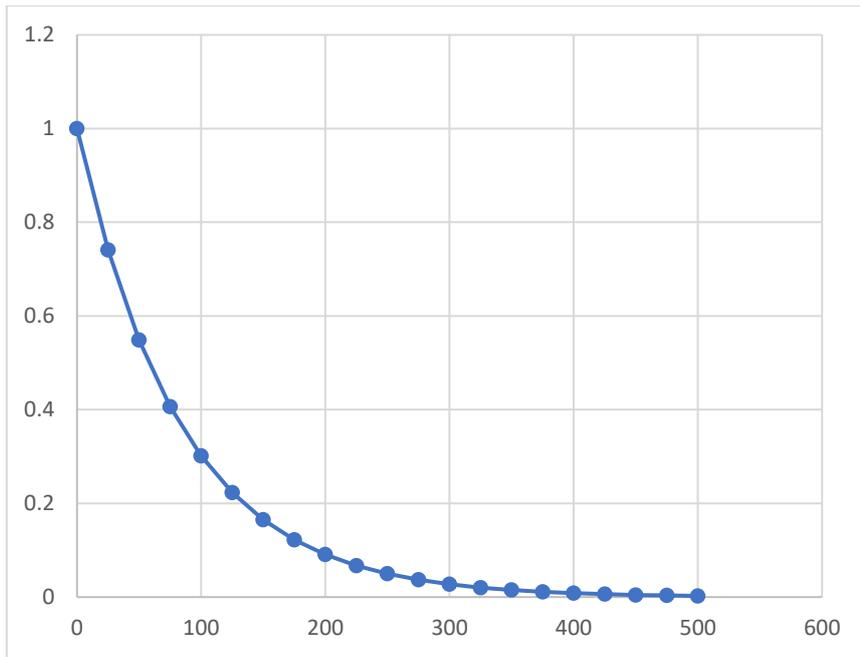


Figura 4.1. Andamento dell'indice di accessibilità in funzione della distanza con parametro di decadimento 12.

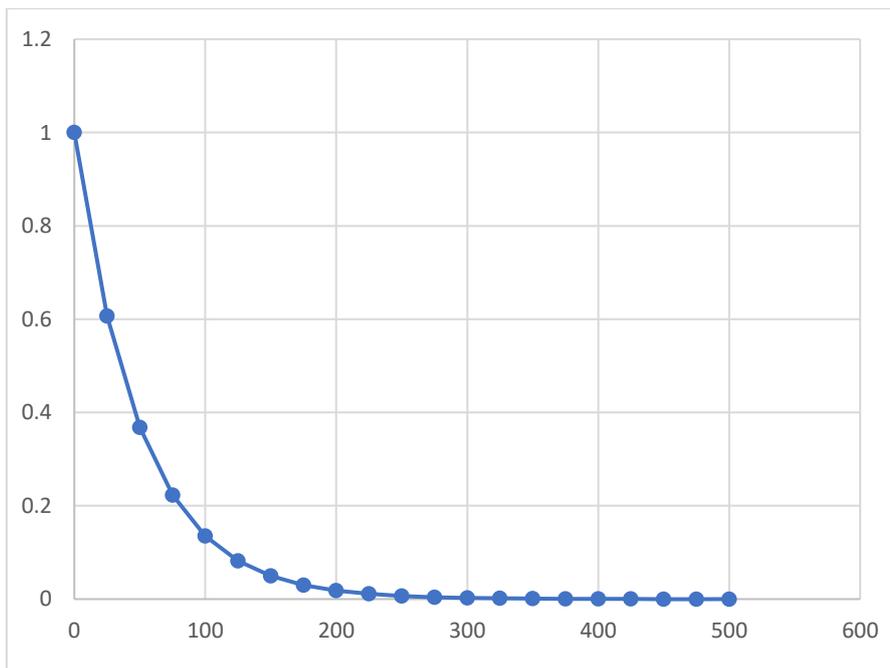


Figura 4.2. Andamento dell'indice di accessibilità in funzione della distanza con parametro di decadimento 20.

Dai grafici si osserva come i valori di accessibilità tendano rapidamente a contributi nulli, a partire da una distanza di circa 400 metri, essendo il risultato di un'esponenziale negativa. Da notare come i valori si riducano più velocemente nel grafico con parametro di decadimento più elevato. Questo comporta che parametri di decadimento alti sentano maggiormente l'influenza di amenities vicine tra loro, rendendo non significativi i contributi delle amenities mediamente distanti.

Nell'articolo originale (Hidalgo et al., 2018) viene utilizzato un parametro di decadimento pari a 16. Questo valore dimezza il contributo all'indice di accessibilità ad ogni aumento della distanza tra le amenities di 62.5 metri, e rende significativi i contributi delle amenities all'interno di un raggio di 500 metri. Nel lavoro di Hidalgo (2018), per facilitare lo svolgimento dell'algoritmo, per ogni amenity presente nel dataset viene sommato il contributo delle 2.000 località più vicine ad essa. Questa semplificazione è accettabile in quanto il valore dell'indice di accessibilità converge prima di sommare il 2.000-esimo contributo. Nell'algoritmo sviluppato, non presentando difficoltà in questa fase, per ogni amenity viene sommato il contributo delle restanti. Nella figura 4.3 vengono rappresentati i punti con diverse gradazioni in base al valore dell'indice di accessibilità. I punti che presentano valori più alti sono visivamente più gialli, per poi scurirsi abbassando il relativo valore.

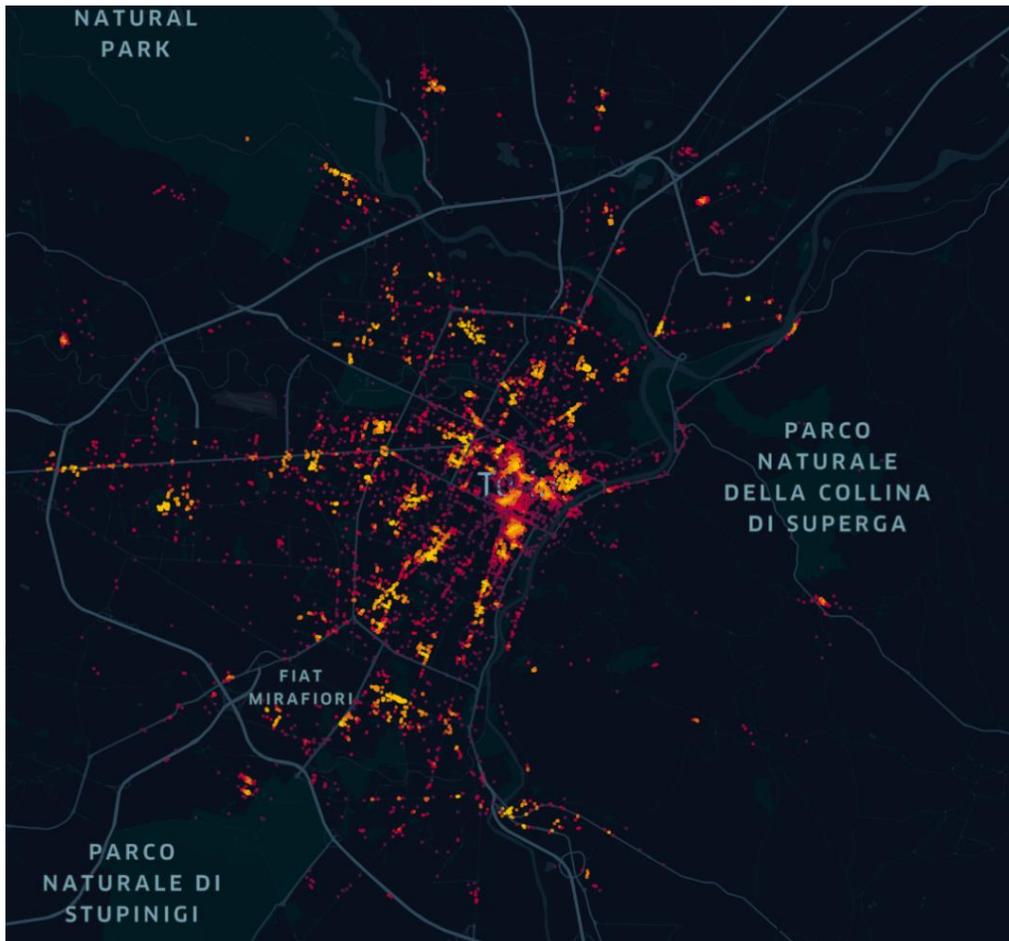


Figura 4.3. Rappresentazione delle amenities in base al valore dell'indice di accessibilità.

Per ottenere il calcolo dell'indice di accessibilità di ogni amenity è necessario conoscere le distanze con gli altri punti. Partendo da latitudine e longitudine si può ricorrere a diverse formule o espedienti per ottenere questa misura. Il metodo più preciso consiste nel misurare la distanza geodetica, che misura la distanza tra due punti su uno sferoide. Sono presenti librerie di Python che, con le dovute approssimazioni, permettono di calcolare le distanze geodetiche fornendo in input le coordinate terrestri. Questa modalità risulta molto dispendiosa dal punto di vista computazionale, richiedendo lunghi tempi per ricavare la matrice quadrata contenente le distanze tra le amenities.

Questo aspetto è stato risolto introducendo una semplificazione: le coordinate geografiche contenute nel dataset vengono convertite in coordinate planari, ossia coordinate cartesiane su un piano, in modo da identificare ogni amenity con la propria ascissa e ordinata. A questo punto è possibile calcolare la distanza euclidea tra le amenities, riducendo notevolmente i tempi di esecuzione dell'algoritmo. La semplificazione porta delle imprecisioni ma il grado di approssimazione risulta accettabile, essendo presente una differenza tra le due modalità di calcolo di circa 2 metri per ogni chilometro.

La fase successiva dell'algoritmo ha come obiettivo individuare i "picchi" attorno ai quali si andranno a formare i quartieri. Dato che l'interesse in questa fase è focalizzato sulle aree maggiormente dense, si prendono in considerazione le amenities con un indice di accessibilità più elevato. Come nell'articolo di Hidalgo (2018) vengono escluse dall'individuazione dei centri il 10% delle amenities che presentano l'indice più basso, mostrate nella figura 4.4.



Figura 4.4. Rappresentazione del 10% dei punti con indice di accessibilità più basso.

Per evitare che si generino quartieri schiacciati tra loro, viene applicata un'euristica che permette di individuare centri dei quartieri che siano adeguatamente distanziati. Un'amenity, per risultare il centro del proprio quartiere deve possedere un indice di accessibilità sufficientemente elevato. La soglia da superare non è definita in maniera assoluta, ma è legata ai valori delle amenities vicine. Viene fatto questo ragionamento in quanto le zone più dense di amenities presentano i nodi con gli indici di accessibilità più elevati e si otterrebbero centri a distanze minime.

L'indice di accessibilità del nodo che costituisce il centro del proprio quartiere è il più elevato delle n amenities più vicine ad esso. Il valore di n è definito attraverso l'euristica nella formula 4.2 utilizzata anche nell'articolo di Hidalgo (2018). Attraverso quest'euristica si riesce ad evitare picchi secondari.

$$n_i = 3A_i + 50 \quad (4.2)$$

Nella formula 4.2 n_i rappresenta il numero di amenities da superare, A_i il valore di accessibilità dell'amenity i -esima.

Una volta individuati i centri si conosce il numero di quartieri in cui è diviso il dataset e si prosegue con la distribuzione delle amenities. Nella figura 4.5 vengono rappresentati i picchi individuati con parametro di decadimento pari a 16.

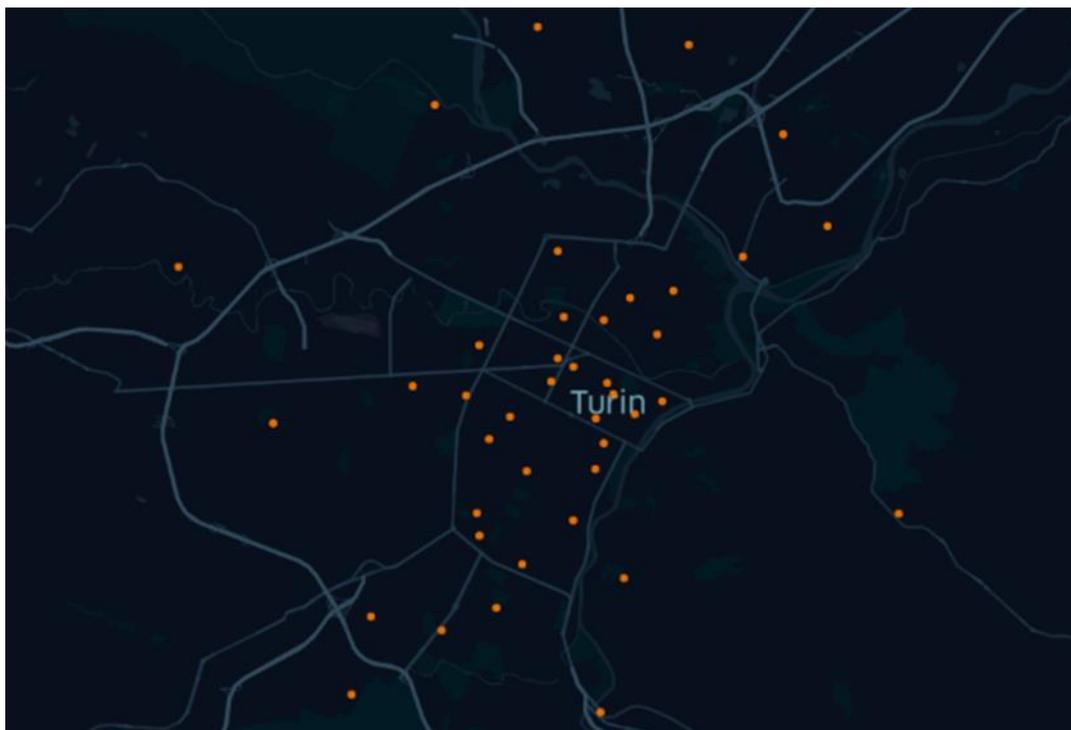


Figura 4.5. Rappresentazione dei centri individuati con parametro di decadimento pari a 16.

La prima operazione in questa fase è l'assegnazione delle amenities più vicine ai centri. Ad ogni quartiere vengono attribuite le amenities che distano meno di 500 metri dal centro. Nel caso in cui siano presenti 2 centri distanziati meno di 1 kilometro, le amenities contese vengono attribuite al centro più vicino. Questa prima distribuzione facilita la costruzione delle zone interne del quartiere e permette di porre maggiore attenzione sulla definizione dei limiti, che richiede un elevato sforzo computazionale.

Le restanti amenities vengono assegnate ai quartieri attraverso un processo iterativo. Come prima cosa viene calcolata la distanza tra queste amenities libere e quelle che sono già entrate a far parte di un quartiere. Attraverso queste misure viene individuata l'amenity libera più prossima ad una appartenente ad un quartiere. L'amenity non assegnata entrerà a far parte dello stesso quartiere. Questo processo viene ripetuto tante volte quante sono le amenities libere, in modo da assegnarle tutte.

Completata la formazione dei quartieri, ognuno di essi viene esportato su un file CSV, in modo da essere rappresentati ed analizzati nelle fasi successive.

I quartieri individuati vengono rappresentati graficamente con Kepler, uno strumento di analisi di dati geospaziali. Le amenities vengono raffigurate all'interno dello spazio cittadino con colorazioni diverse, in base al quartiere di appartenenza. Nella figura 4.6 è presente la raffigurazione dei quartieri che si ottengono utilizzando un parametro di accessibilità pari a 16.

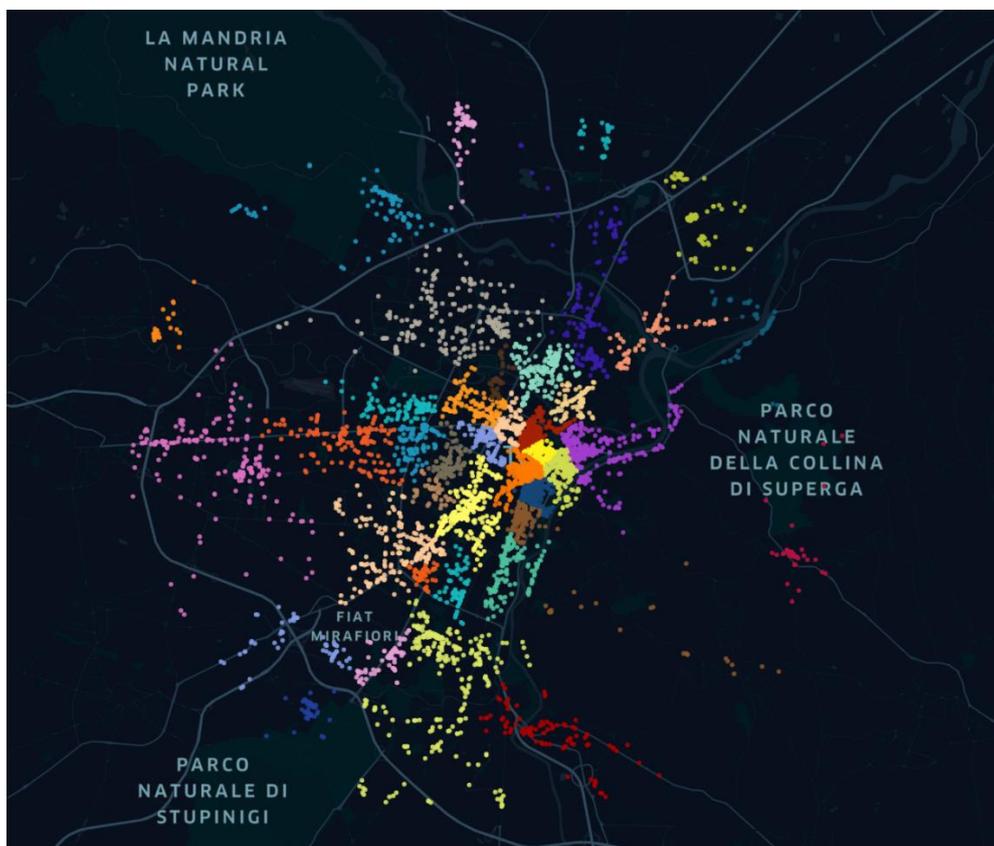


Figura 4.6. Visualizzazione dei quartieri con parametro di decadimento 16.

Analizzando l'immagine, si osserva come la maggior parte dei quartieri sia concentrata nell'area centrale, che corrisponde al centro cittadino. Si tratta della zona di maggiore interesse nei confronti del pubblico e conseguentemente risulta molto attrattiva per un gran numero di amenities. La scelta di molti nodi di localizzarsi in quest'area si traduce, dal punto di vista dell'algoritmo, in punti che presentano un elevato indice di accessibilità.

Nonostante nell'area centrale siano presenti i nodi con gli indici di accessibilità più elevati, non tutti possono diventare centri di un quartiere. L'euristica mostra qui il suo effetto, individuando centri adeguatamente distanziati. La presenza di diversi picchi nell'area centrale comporta quartieri in lotta tra di

loro per raccogliere le amenities che li circondano. Questo si manifesta visivamente in quartieri non molto estesi ma che presentano un gran numero di amenities al loro interno.

Allontanandosi in direzione delle zone periferiche, si nota come i quartieri si distribuiscano lungo le principali vie cittadine. Dalla figura 4.6 si nota come questi risultino molto più ampi e allungati lungo le vie principali, risultando meno uniformi. Questa è una conseguenza del modo con cui vengono assegnate le amenities che distano più di 500 metri dai centri dei quartieri. Queste amenities più distanti dai picchi vengono assegnate iterativamente ai cluster, in base alla vicinanza con quelle già assegnate, che possono aumentare non incontrando rivali sul proprio cammino.

4.1 Dataset principale

Per cercare la migliore distribuzione delle amenities, e conseguente forma dei quartieri, sono state effettuati diversi tentativi, modificando ogni volta il valore dell'indice di accessibilità. Questo parametro viene fatto variare tra 12 e 20, ottenendo diversi risultati. Non si è andati oltre questi valori in quanto si è ritenuto essere dei limiti al di là dei quali i valori degli indici di accessibilità sarebbero risultati poco rappresentativi dei fenomeni analizzati. Con un parametro di decadimento troppo basso si rischia che le amenities più distanti abbiano un impatto troppo elevato, rispetto alla loro reale accessibilità. Analogamente, un aumento del parametro di decadimento avrebbe comportato di ridurre il valore di amenities realmente prossime tra loro.

Sono state eseguite anche delle prove combinate variando il raggio di assegnazione delle amenities attorno ai picchi. Modifiche di questo parametro non hanno evidenziato sostanziali modifiche nella realizzazione dei cluster ma solo un modo per velocizzare i tempi di esecuzione dell'algoritmo.

Di seguito, nelle figure 4.1.1 e 4.1.2, viene visualizzata la mappa di Torino con i cluster che si ottengono rispettivamente con parametro di decadimento 12 e 20.



Figura 4.1.1. Visualizzazione dei quartieri con parametro di decadimento 12.

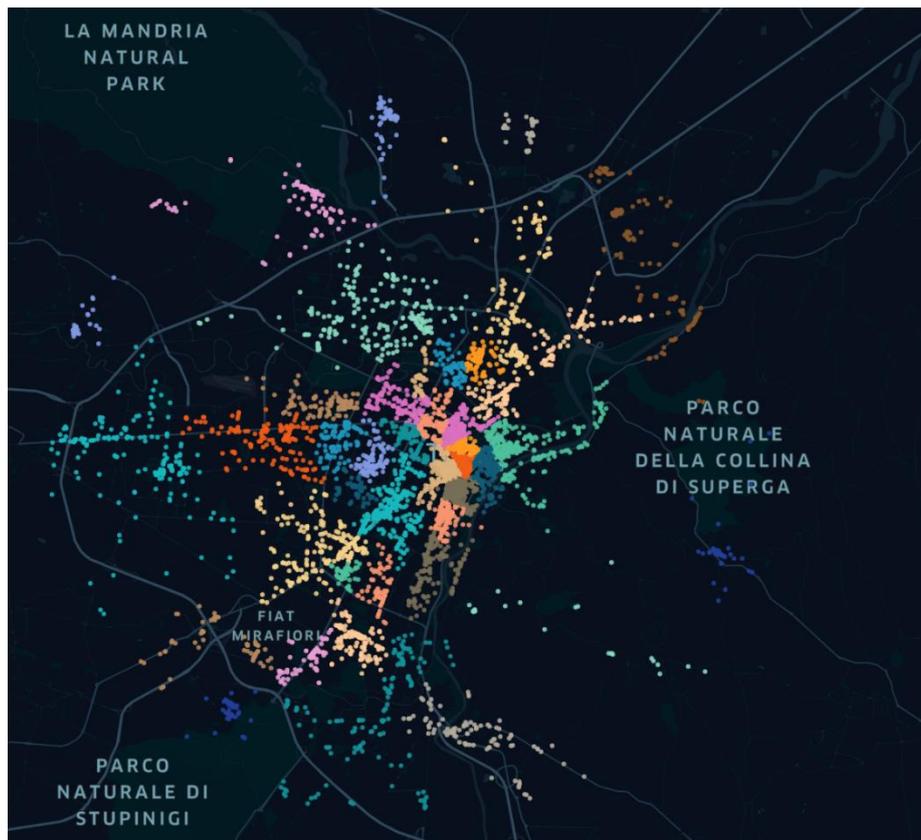


Figura 4.1.2. Visualizzazione dei quartieri con parametro di decadimento 20.

Le rappresentazioni dei quartieri riflettono le aspettative per quanto riguarda le forme che assumono. I maggiori cambiamenti sono presenti nell'area maggiormente densa di amenities, quella del centro cittadino. Applicando un valore 12 al parametro di decadimento i quartieri risultano complessivamente più ampi, come conseguenza del minor numero di picchi individuati. Questo può essere dettato dalla forma dell'euristica, che è funzione degli indici di accessibilità. Con un valore 12 gli indici di accessibilità risultano nel complesso più elevati e ciò comporta che, per diventare centro del proprio quartiere, un'amenity dovrà superare l'indice di un maggior numero di amenities.

Nella figura 4.1.2 con il parametro di decadimento 20, il numero di quartieri individuati aumenta. I nuovi cluster si formano per la maggior parte nell'area centrale, spezzettando ulteriormente l'area più densa di nodi.

Nelle figure 4.1.3 e 4.1.4 vengono rappresentate le forme dei poligoni dei quartieri che si ottengono con i parametri di decadimento 12 e 20.

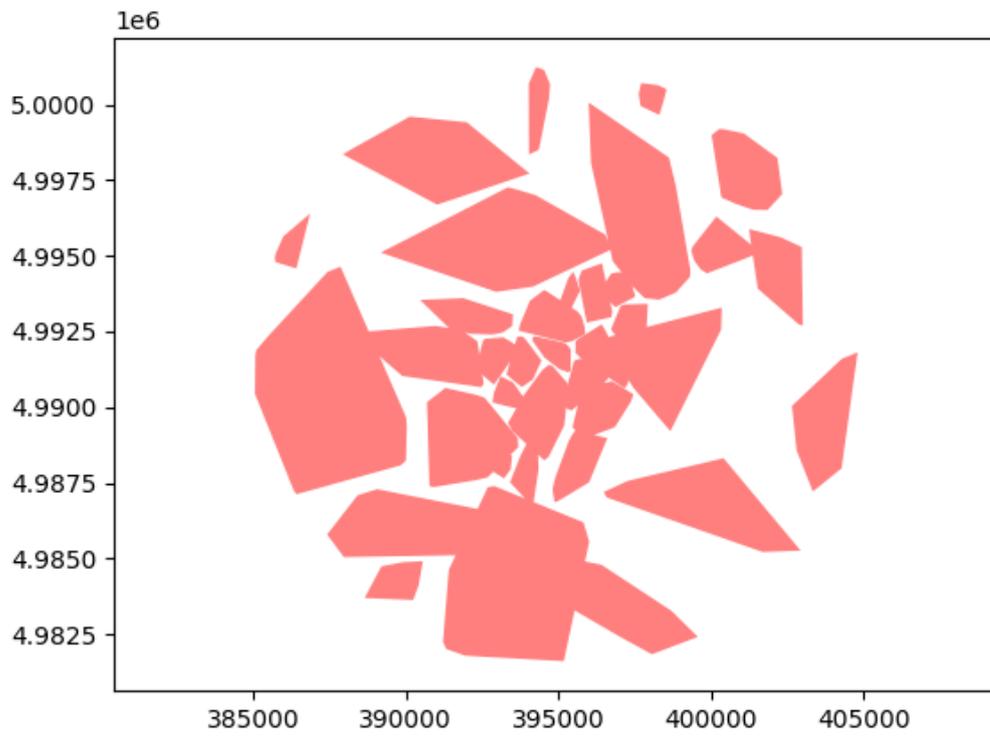


Figura 4.1.3. Forme dei quartieri con parametro di decadimento 12.

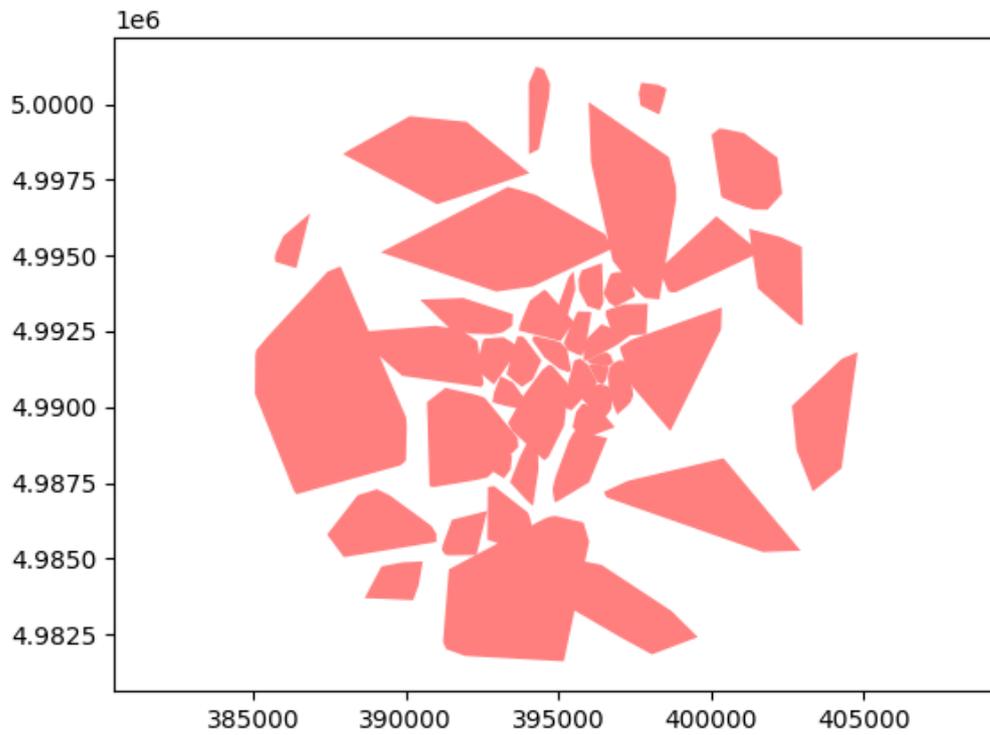


Figura 4.1.4. Forme dei quartieri con parametro di decadimento 20.

Analizzando le due immagini si osserva come nelle zone centrali più ricche di amenities si concentrino il maggior numero di quartieri. Questi risultano poco estesi al contrario di quelli che si creano nelle zone periferiche che occupano la maggior parte dello spazio. Questa differenza visiva risiede nell'assegnazione delle amenities ai quartieri, che permette ai quartieri periferici di inglobare le amenities esterne non contese.

La rappresentazione contiene al proprio interno numerosi spazi bianchi. Queste aree costituiscono zone prive di amenities, come l'area ferroviaria o quella fluviale, ma anche strade di confine tra i quartieri. Le vaste aree vuote tra i quartieri esterni sono dovute alla bassa presenza di amenities in queste aree. Va sottolineato che i quartieri esterni, nonostante siano rappresentati come dei grandi blocchi, presentano una bassa densità di amenities al proprio interno.

Confrontando le figure 4.1.3 e 4.1.4 si nota bene come le maggiori differenze nella formazione dei quartieri risiedono in quelli interni, mentre quelli perimetrali mantengono le stesse forme. Questa dinamica è dettata dal variare dell'indice di accessibilità che influenza soprattutto i valori di amenities vicine tra loro.

Per mostrare i cambiamenti che intercorrono utilizzando diversi valori, nelle figure seguenti vengono mostrati gli andamenti di alcuni elementi al variare del parametro di decadimento e altri fattori.

Per comprendere come il parametro di decadimento influisca sull'individuazione dei centri nella, figura 4.1.5 viene mostrato l'andamento del numero di quartieri che si ottengono al variare del parametro.

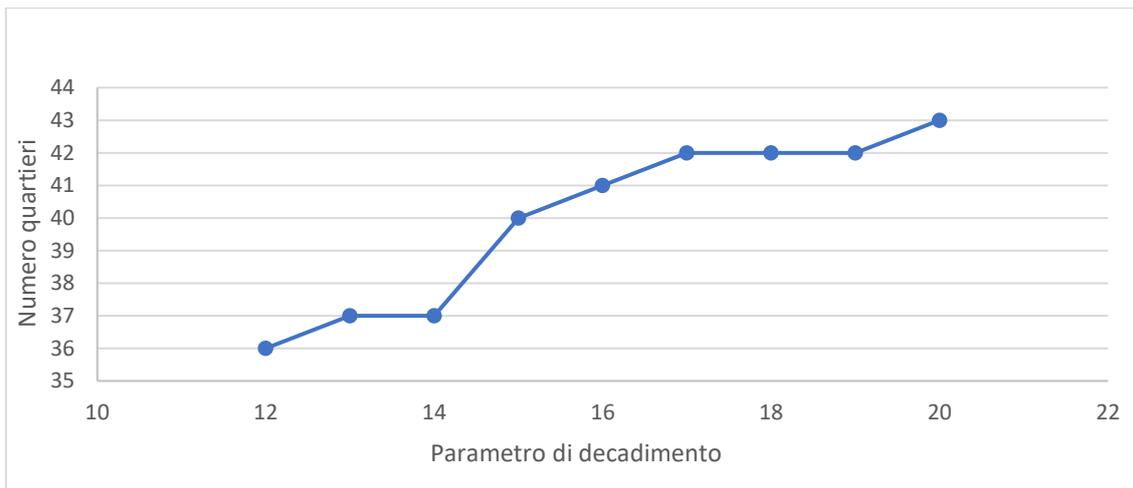


Figura 4.1.5. Numero di quartieri in funzione del parametro di decadimento.

Come anticipato precedentemente, il numero di quartieri è direttamente proporzionale al parametro di decadimento. Abbassando il parametro si riduce il valore dell'indice di accessibilità con conseguente effetto sul valore dell'euristica. Questo comporta che ogni nodo riduce il numero di valori da superare, aumentando le proprie probabilità di diventare centro. Indicativamente l'aumento unitario del parametro di decadimento ha un effetto analogo sul numero di quartieri.

La numerosità dei quartieri parte da un minimo di 36 quartieri fino ad arrivare a 43, in corrispondenza del parametro più alto analizzato. Con parametro di decadimento pari a 17 e 18 l'algorithmo individua gli stessi 42 centri. Il numero di picchi con parametro 19 rimane lo stesso ma cambiano le loro coordinate dei centri attorno ai quali di formano i cluster.

Nella figura 4.1.6 viene mostrato l'andamento del numero di tipologie di amenities contenute in ciascuno quartiere al variare del parametro di decadimento. Viene preso in considerazione questo aspetto per verificare la varietà di tipi che ogni quartiere presenta al proprio interno.

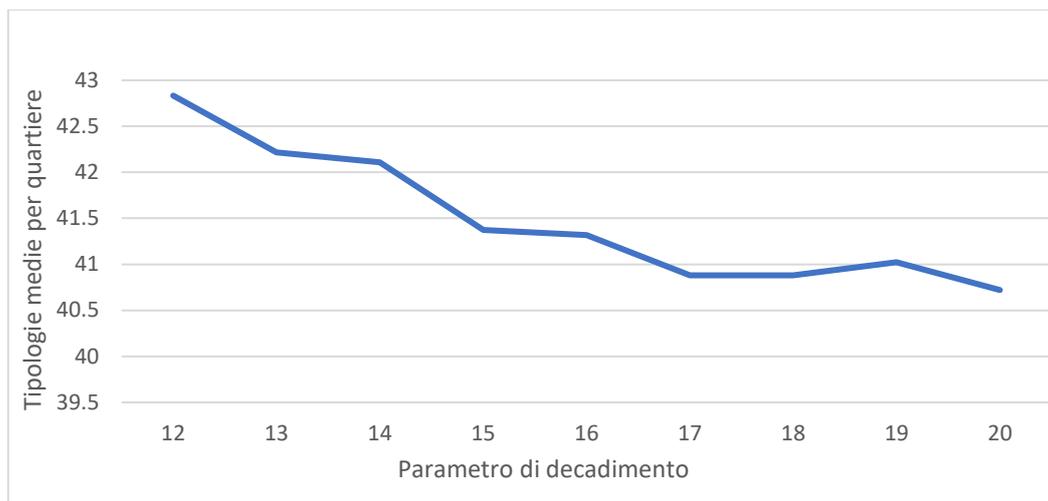


Figura 4.1.6. Tipologie di amenities per quartiere in funzione del parametro di decadimento

Il numero di tipologie di amenities contenute in ogni quartiere risulta essere decrescente con l'aumentare del parametro. In questo andamento va considerato il numero di quartieri, che aumenta con il parametro di decadimento, dovendo dividere il numero totale di amenities tra più quartieri. Complessivamente, la varietà nella distribuzione delle 113 tipologie non cambia significativamente al variare del parametro di decadimento, rimanendo sempre compresa tra 40 e 43.

Nella tabella 4.1.7 sono riportati dati supplementari per le tipologie di amenities contenute in ciascun quartiere.

Parametro di decadimento	Numero medi di tipologie di amenities per quartiere	Mediana	Deviazione standard	Varianza
12	42.83333333	40	15.008331	225.25
13	42.21621622	39	14.739422	217.2505
14	42.10810811	39	14.596897	213.0694
15	41.375	38	13.480889	181.7344
16	41.31707317	37	12.533738	157.0946
17	40.88095238	36.5	12.369706	153.0096
18	40.88095238	36.5	12.369706	153.0096

19	41.02380952	37.5	12.12335	146.9756
20	40.72093023	38	11.680486	136.4337

Figura 4.1.7. Dati sulle tipologie di amenities per quartiere.

Come è possibile osservare, il valore della mediana è sempre minore della media, indicando che la distribuzione è asimmetrica e presenta una coda a destra.

Nella figura 4.1.8 viene mostrata la media della distanza minima tra il centro di ogni quartiere e quello più vicino ad esso. Viene rappresentata questa misura per verificare se i picchi risultavano dispersi all'interno della mappa.

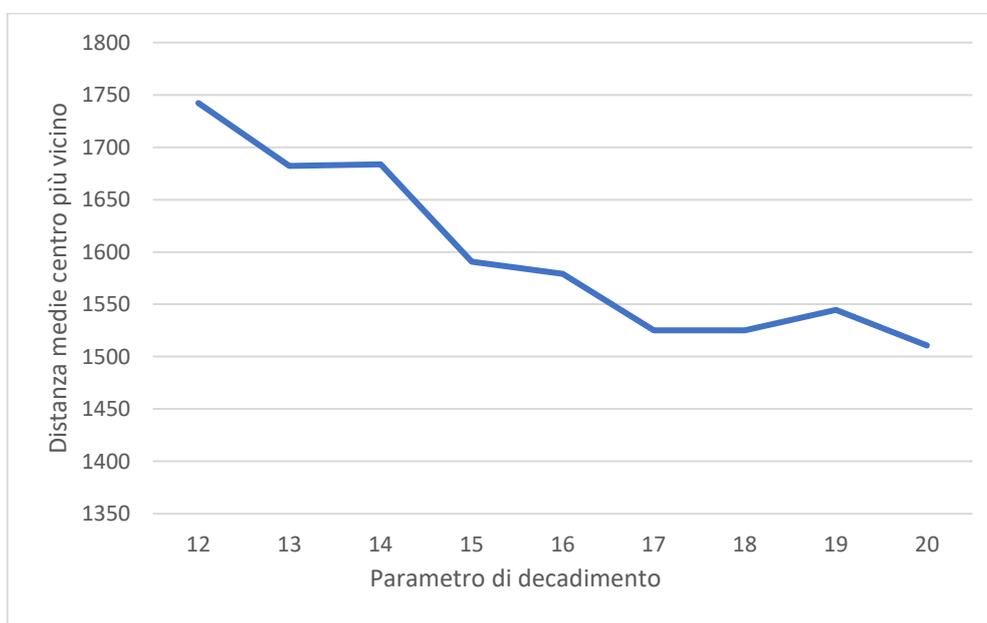


Figura 4.1.8. Distanza media minima tra il centro e quello più vicino ad esso.

Per ottenere questi valori, si è individuato per ogni centro quello più vicino ad esso, misurata la distanza tra i due punti e calcolata la media. La distanza media massima, pari a 1.742 metri, si ottiene in corrispondenza del parametro di decadimento 12. Il valore minimo invece si raggiunge alzando il parametro a 20, con 1.510 metri. Tra i due valori il range è di 232 metri, pari al 13% della distanza massima.

All'aumentare del parametro, i centri sembrano ridurre la distanza da quello più vicino. Dai dati, aumentando di un'unità il parametro di decadimento si

riduce la distanza di circa 30 metri. Nell'analizzare queste variazioni bisogna tenere conto dell'aumento del numero di quartieri che corrisponde all'aumento del parametro di decadimento.

Nella figura 4.1.9 viene mostrato l'andamento della media delle distanze medie tra i centri. Al variare del parametro di decadimento, per ogni centro viene calcolata la distanza media dagli altri centri, per poi eseguire la media di queste medie. In questo modo il valore che si ottiene è pesato per il numero di centri per quel parametro di decadimento. Questo valore è preso in considerazione per analizzare come i centri occupino lo spazio cittadino.

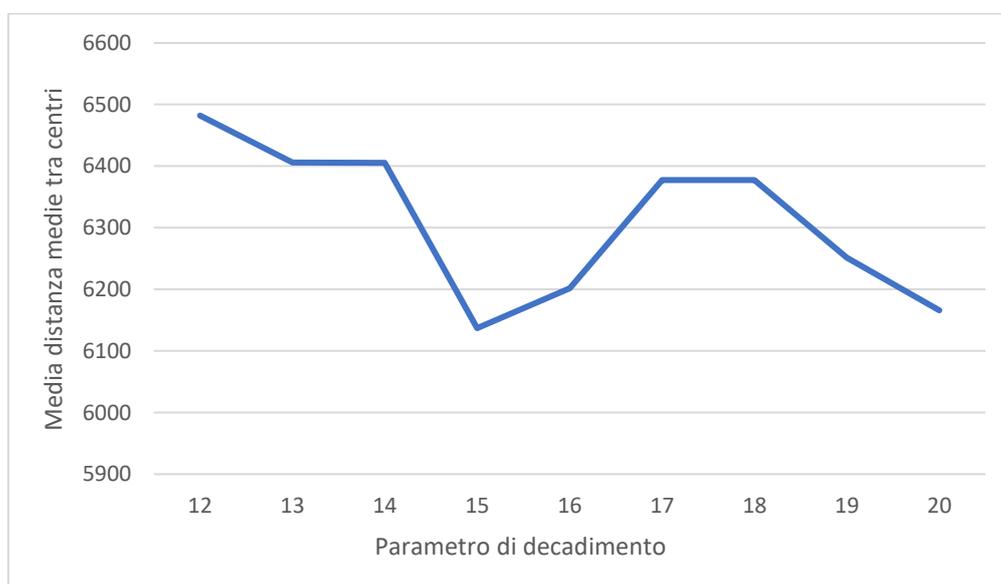


Figura 4.1.9. Media distanza media tra i centri in funzione del parametro di decadimento.

La massima distanza media tra le amenities è pari a 6.400 metri. Questo valore si ottiene in corrispondenza del parametro di decadimento 12 che produce il minimo numeri di quartieri. L'andamento decresce velocemente, raggiungendo il minimo di 6.130 metri con il parametro di decadimento 15, a cui però non corrisponde il numero minimo di quartieri. Le distanze medie salgono nonostante con più quartieri i centri dovrebbero essere più raccolti.

Indicativamente la distanza media si riduce di 25 metri all'aumentare del parametro di decadimento. L'andamento irregolare del grafico potrebbe indicare che tra 12 e 15 i centri dei quartieri tendono a essere raccolti verso il centro, per poi andare ad allargarsi verso le zone periferiche. Con i

parametri di decadimento 19 e 20 e il conseguente aumento dei quartieri individuati, i centri risultano più fitti sull'area analizzata.

Nella figura 4.1.10 viene mostrato l'andamento dell'area media dei quartieri rispetto al parametro di decadimento. Questo valore fornisce un'idea di quanto i quartieri cambino dimensione variando il parametro.

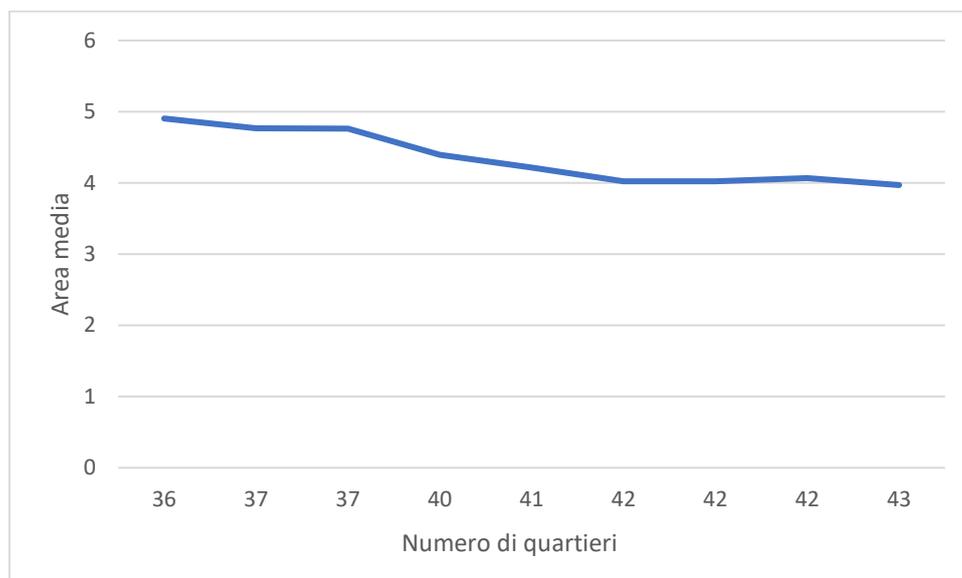


Figura 4.1.10. Area media dei quartieri in funzione del numero di quartieri

Per calcolare l'area dei singoli quartieri è stato applicato un algoritmo che costruisce il poligono contenente tutti i punti del quartiere. Il bordo esterno viene definito dalla forma convessa del poligono, di cui si calcola l'area.

L'andamento dell'area media dei quartieri in funzione del parametro di decadimento risulta essere decrescente, passando da 5 km² a 4 km² tra gli estremi presi in considerazione.

In linea teorica, ci si aspetterebbe un andamento monotono decrescente, in quanto aumentando il numero di quartieri l'area complessiva viene divisa in più zone. Tuttavia, questo non succede in quanto la somma dei singoli quartieri non è costante ma dipende da come vengono assegnate le amenities ai quartieri e la forma esterna che si ottiene. Ad esempio, tra i parametri 18 e 19, nonostante il numero di quartieri sia lo stesso, l'area media aumenta leggermente.

Nella figura 4.1.11 viene mostrato l'andamento del perimetro medio dei quartieri al variare del parametro di decadimento.



Figura 4.1.11. Perimetro medio dei quartieri in funzione del numero di quartieri.

Come per il calcolo delle aree, viene calcolato il perimetro del poligono esterno generato dai quartieri.

In questo caso il perimetro medio risulta inversamente proporzionale al numero di quartieri, passando da circa 8 km per 36 quartieri a poco più di 7 km con 43 quartieri, riducendo la propria lunghezza di circa 100 metri al variare del parametro di decadimento.

Nella figura 4.1.12 viene mostrata la densità media di amenities per km² dei quartieri al variare del parametro di decadimento. Viene mostrato questo grafico per comprendere la relazione tra numero di quartieri e distribuzione delle amenities.

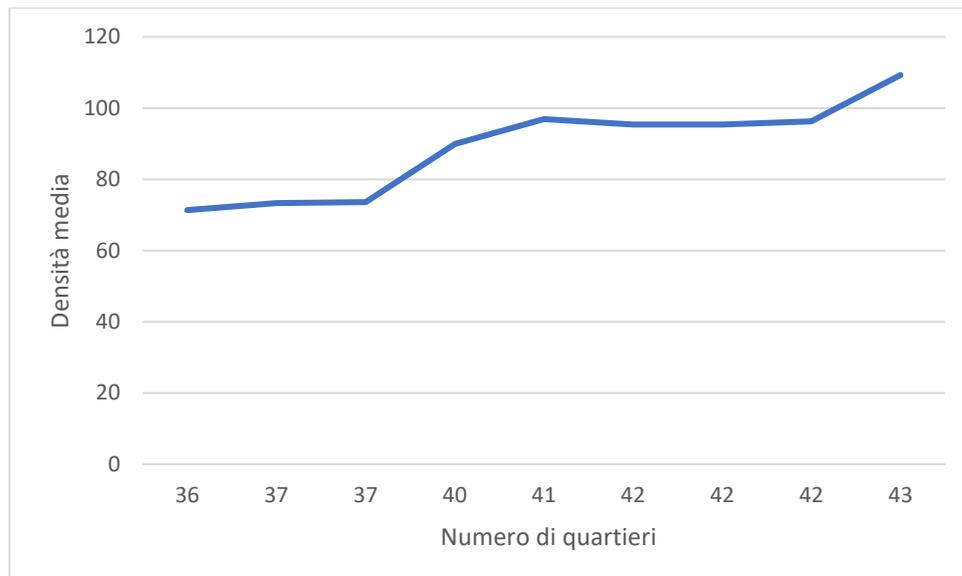


Figura 4.1.12. Densità media dei quartieri in funzione del parametro di decadimento.

La relazione tra densità e numero di quartieri presenta un andamento crescente, presentando i valori più bassi in corrispondenza del minor numero di quartieri. Il massimo di 110 amenities per kilometro è presente per con i 43 quartieri.

La relazione risulta in linea con il grafico che presenta l'area media dei quartieri. Il numero di amenities nel dataset rimane costante ma l'area media dei cluster si riduce, generando quartieri più fitti di amenities.

Nell'articolo di Hidalgo (2018) viene utilizzato un parametro di decadimento pari a 16, che permette di ottenere dei quartieri ragionevoli. Con questo valore il contributo che due amenities si offrono vicendevolmente si dimezza ogni 62.5 metri e risulta essere trascurabile oltre i 500 metri. Questo risulta essere in linea con ricerche che mostrano come le passeggiate pedonali raramente superano i 10 minuti di cammino (Handy e Niemeier, 1997; Sevtsuk, 2014).

Considerando i grafici visti in precedenza e il significato che viene attribuito ai quartieri nel paper originale, vengono scelti i quartieri che si ottengono con il parametro di decadimento pari a 16 per le fasi successive.

4.2 Dataset istituzionale

Per il dataset contenente le amenities istituzionali viene fatta un'analisi analoga alla precedente ma allargando a dei parametri di decadimento tra 7 e 20. Viene fatta questa scelta in quanto queste tipologie di amenities coinvolgono spostamenti maggiori.

Nelle figure 4.2.1 e 4.2.2 vengono mostrati i quartieri ottenuti con i parametri di decadimento 7 e 12 con il database di amenities istituzionali.

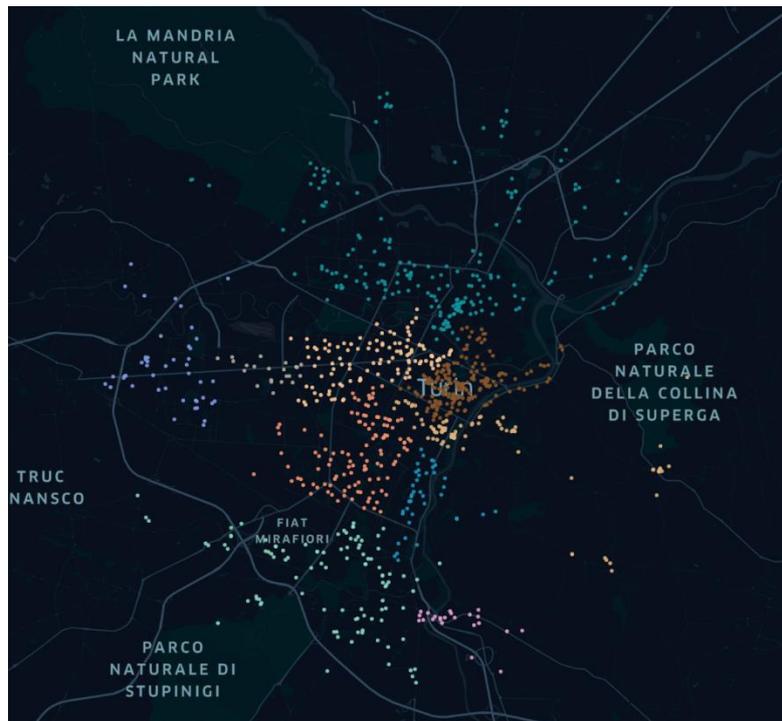


Figura 4.2.1. Quartieri da amenities istituzionali con parametro di decadimento 7.



Figura 4.2.2. Quartieri da amenities istituzionali con parametro di decadimento 12.

Sono rappresentati i quartieri con parametri di decadimento 7 e 12 in quanto ad essi corrispondono il minimo e il massimo numero di quartieri che si ottengono.

Visivamente si nota subito come il minor numero di amenities presenti su una superficie analoga a quella precedente si traduce in un minor numero di quartieri individuati. Le zone del centro cittadino risultano comunque spezzettate in quartieri minori, mentre le altre aree sono separate abbastanza omogeneamente tra gli altri cluster.

Le figure 4.2.3 e 4.2.4 mostrano le forme assunte dai quartieri con i parametri di decadimento 7 e 12.

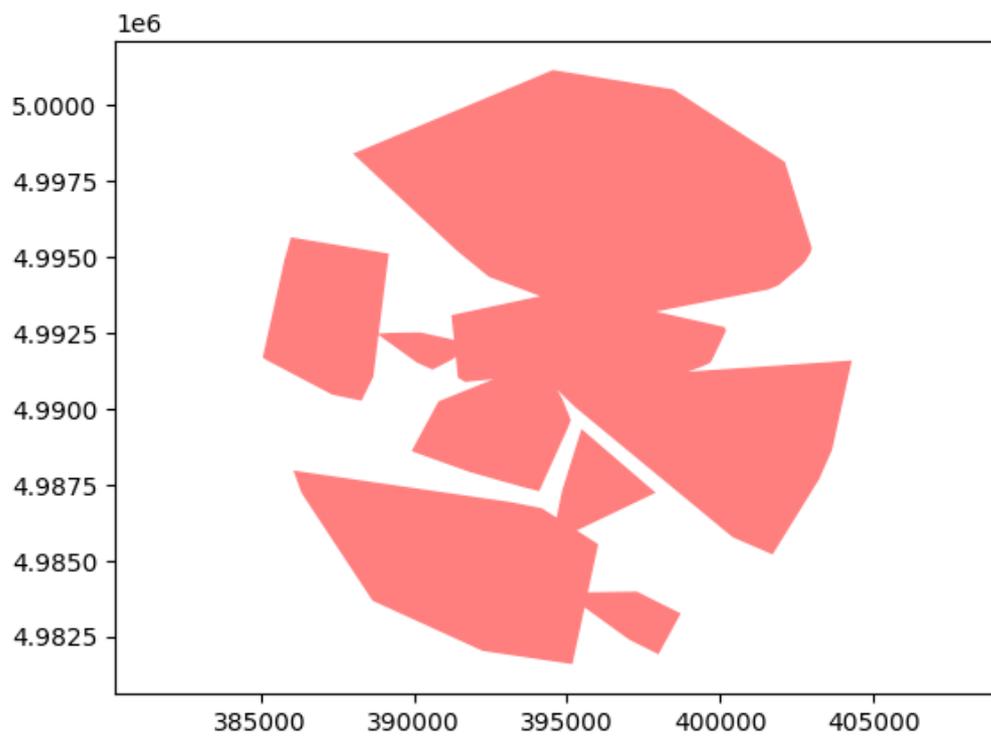


Figura 4.2.3. Forme dei quartieri da amenities istituzionali con parametro di decadimento 7.

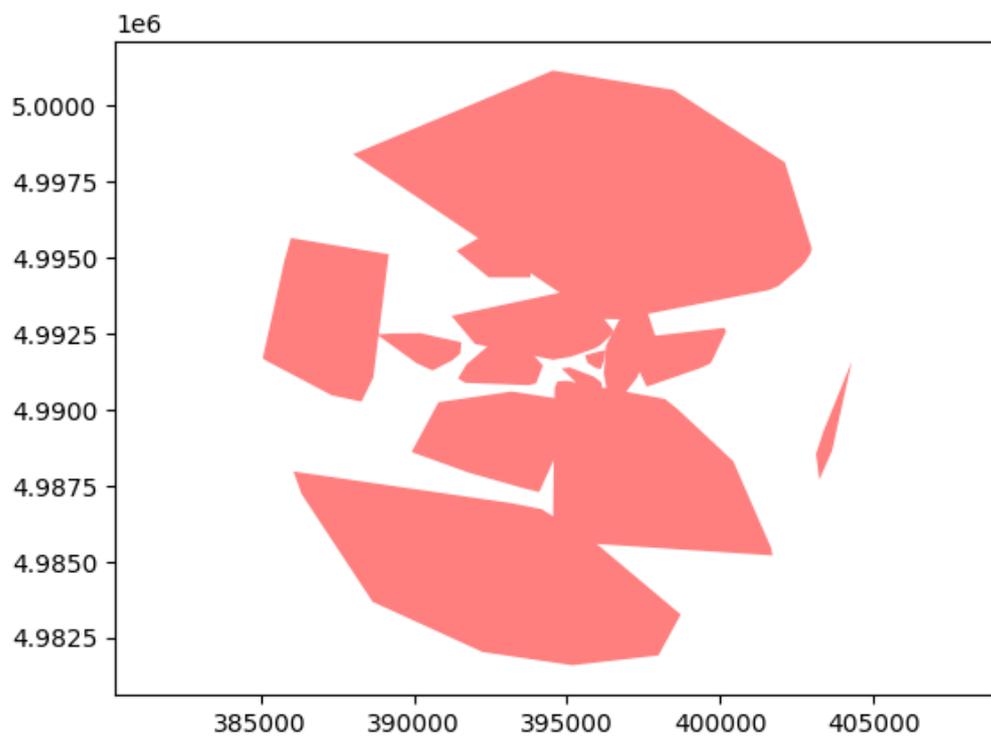


Figura 4.2.4. Forme dei quartieri da amenities istituzionali con parametro di decadimento 12.

Dalle due immagini si osserva bene come le zone centrali siano spezzettate mentre le aree esterne siano più ampie.

Inoltre, è possibile notare come sia presente una somiglianza tra gli spazi delineati dalle forme dei quartieri e le circoscrizioni torinesi, mostrate nella figura 4.2.5.

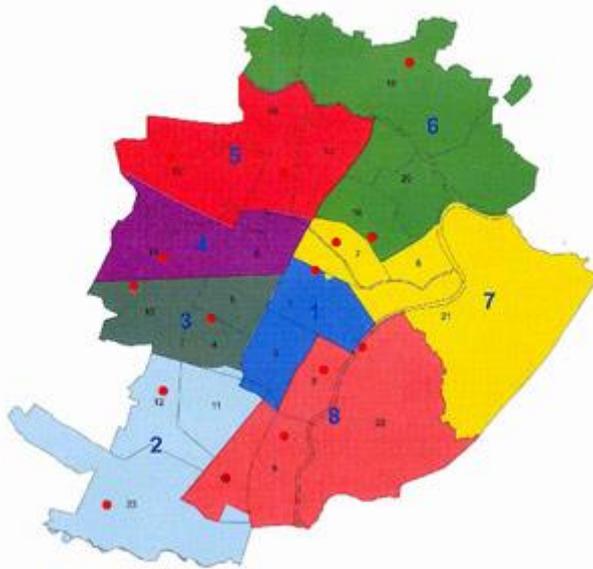


Figura 4.2.5. Circoscrizioni della città di Torino.

Confrontando la figura 4.2.4 e la figura 4.2.5, si possono notare delle analogie nella divisione degli spazi. La prima circoscrizione è assimilabile ai quartieri frammentati del centro. Le vaste aree individuate sopra e sotto contengono rispettivamente le circoscrizioni 6 e 2. Nello spazio del lato sinistro si incastrano le circoscrizioni 3, 4 e 5 e analogamente in alto a destra si inserisce la settima circoscrizione. La circoscrizione 8 è assimilabile al trapezio rettangolo in basso a destra.

Queste somiglianze non dovrebbero sorprendere, dato il modo con cui sono stati ottenuti questi quartieri. Trattandosi di amenities istituzionali, la loro distribuzione sul territorio dovrebbe essere omogenea, in modo da garantire la raggiungibilità agli abitanti.

Significativo è anche il parametro di decadimento con cui sono stati ottenuti, 12. Con questo valore si rendono significativi i contributi delle amenities ad una distanza superiore rispetto all'utilizzo del parametro di decadimento 16. Infatti, per raggiungere le tipologie inserite nel database istituzionale, si potrebbe essere disposti a compiere uno spostamento superiore ai 500 metri, limite di significatività con un parametro di decadimento 16.

Nelle figure successive viene mostrato l'andamento di alcuni valori al variare del parametro di decadimento.

Nella figura 4.2.6 viene mostrato come varia il numero di quartieri individuati dall' algoritmo rispetto al parametro di decadimento applicato, per capire come questo influenzi il numero di centri individuati.

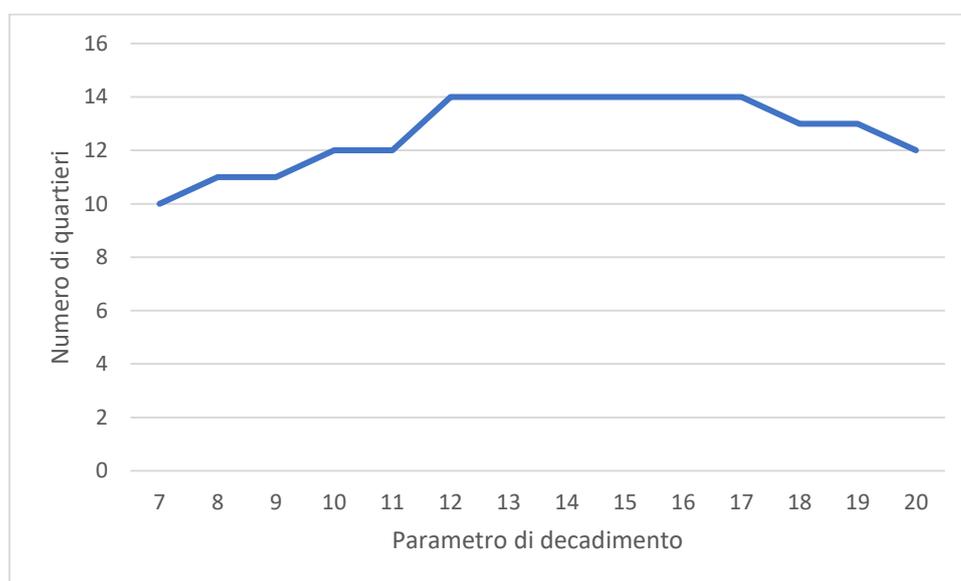


Figura 4.2.6. Numero di quartieri individuati in funzione del parametro di decadimento.

Al contrario delle aspettative, il numero di quartieri ottenuti non è direttamente proporzionale al parametro di decadimento. In corrispondenza del parametro 12 viene individuato il minimo numero di quartieri, 10, per poi aumentare fino a 14 con i parametri tra 12 e 17. All'interno di questo intervallo di parametri si assegnano le amenities ai quartieri alla stessa maniera, pur individuando centri diversi. Superato il parametro 17, il numero di centri individuati si riduce, al contrario di quanto accadeva in precedenza con un dataset maggiormente fitto di amenities.

Nell figura 4.2.7 viene mostrato il numero medio di tipologie presenti nei quartieri. Questa misura permette di comprendere se tutti i quartieri sono ugualmente serviti dalle amenities istituzionali.

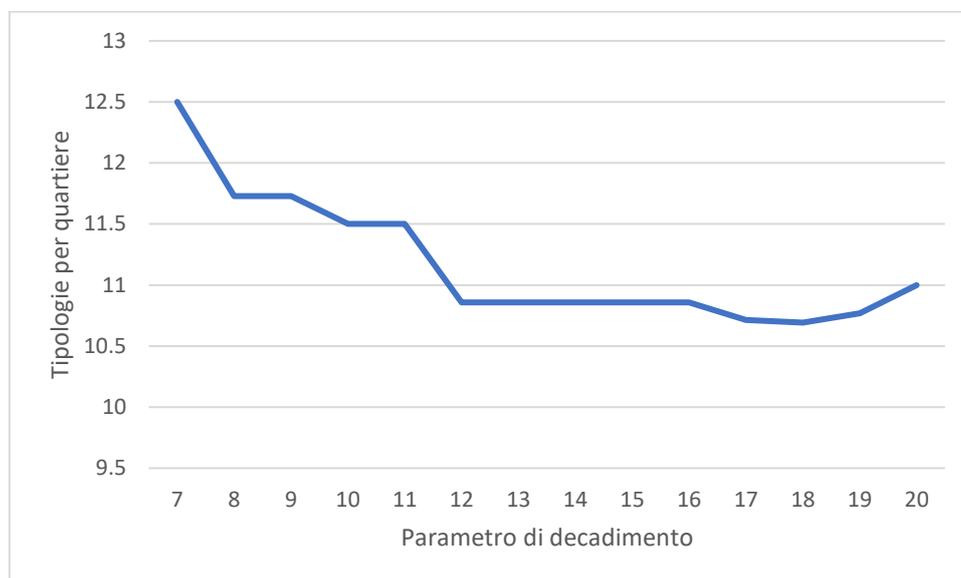


Figura 4.2.7. Numero medio di tipologie di amenities in funzione del parametro di decadimento

Analizzando l'immagine, il parametro che offre una maggiore varietà all'interno dei quartieri è quello che individua il minor numero di quartieri, dovendo dividere le amenities in meno gruppi. Aumentando il valore, i quartieri risultano meno variegati, dovendo dividere lo stesso numero di amenities tra più cluster. Con i parametri 17 e 18, nonostante il numero di quartieri diminuisca, la varietà continua a ridursi, per poi aumentare con i parametri più elevati.

Considerando che 5 delle 18 tipologie di amenities istituzionali non possono essere distribuite adeguatamente in quanto troppo poco numerose, con tutti i parametri si riesce ad avere una discreta distribuzione della varietà nei quartieri.

Nella figura 4.2.8 viene mostrata la relazione tra le distanze medie dei centri più vicini al variare del numero di quartieri individuati. Questo valore permette di comprendere quanto i centri siano dispersi sulla superficie analizzata.

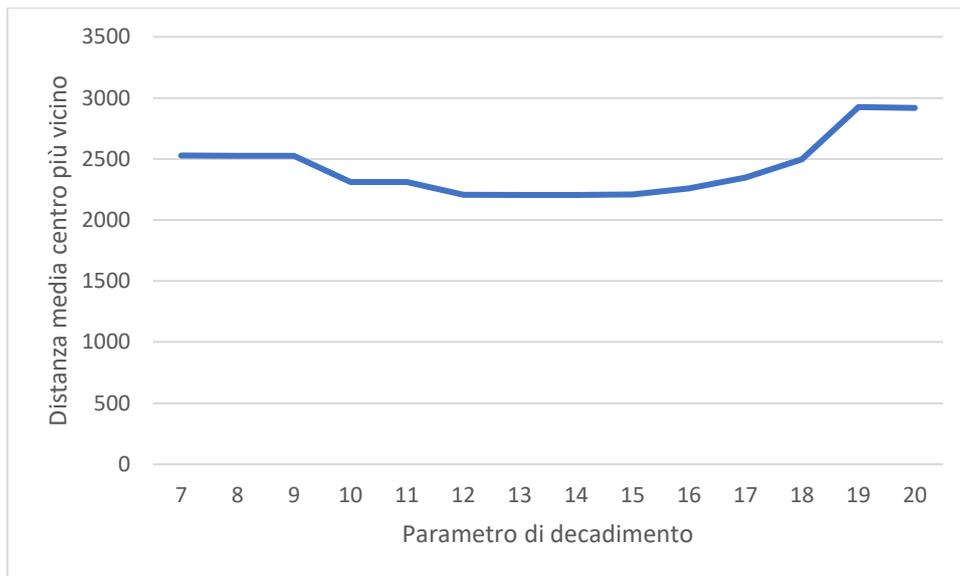


Figura 4.2.8. Distanza media tra i centri più vicini al variare del parametro di decadimento.

Come in precedenza, per ottenere questi valori si è individuato per ogni centro quello più vicino ad esso, misurata la distanza tra i due punti e calcolata la media.

In questo caso i centri individuati risultano tutti più o meno ugualmente distribuiti tra i parametri di decadimento 7 e 18, restando in un range di 300 metri tra 2.200 e 2.500 metri. Con i parametri 19 e 20, nonostante il numero di centri si riduca, si raggiunge la massima distanza media dal centro più vicino con quasi 3.000 metri.

Nella figura 4.2.9 viene mostrata la relazione tra la distanza media pesata tra i centri e il variare del parametro di decadimento. La misurazione offre una misura della posizione dei centri sulla mappa.

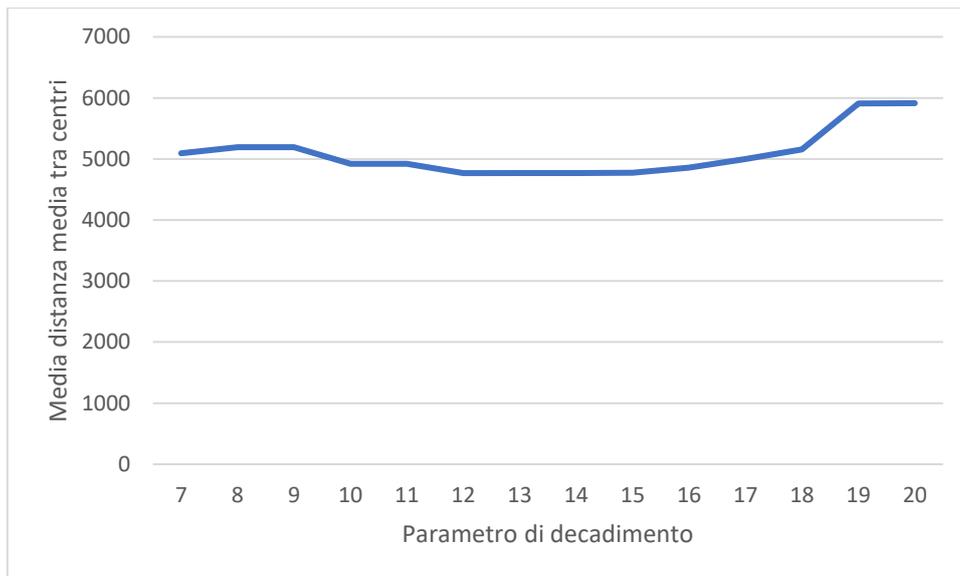


Figura 4.2.9. Media delle distanze medie tra i centri al variare del parametro di decadimento.

Anche in questo caso, al variare del parametro di decadimento, per ogni centro viene calcolata la distanza media dagli altri centri, per poi eseguire la media di queste medie. In questo modo il valore che si ottiene è pesato per il numero di centri individuati per il rispettivo parametro di decadimento.

Come si osserva, in maniera analoga alla figura precedente, i centri risultano ad una distanza paragonabile nell'intervallo tra 7 e 18, attorno ai 5.000 metri. I centri individuati dai parametri di decadimento 19 e 20 risultano molto più dispersi, raggiungendo quasi una distanza media di 6.000 metri.

Nella figura 4.2.10 viene mostrata la relazione tra area media dei quartieri e numero di quartieri individuati. Questa misura permette di comprendere come venga distribuito lo spazio analizzato.

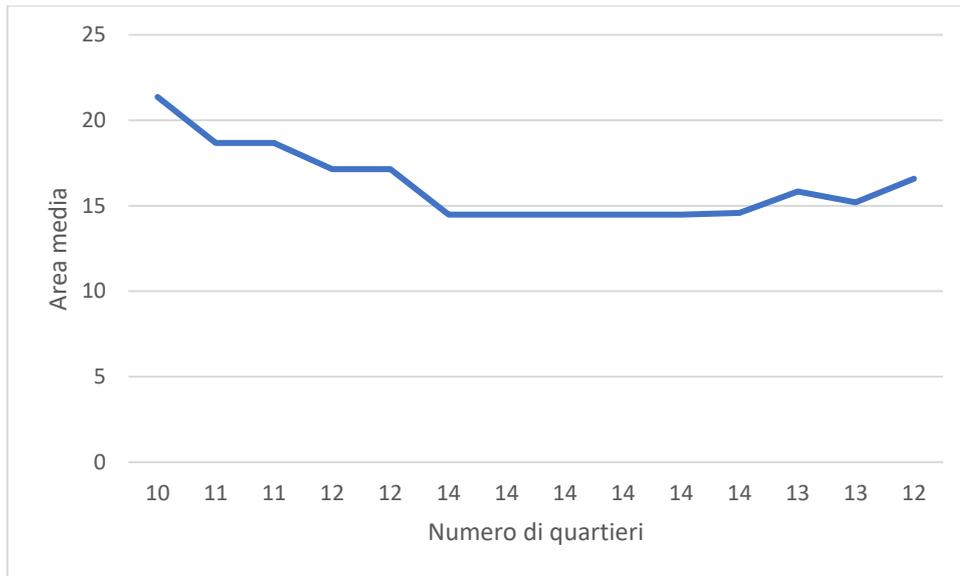


Figura 4.2.10. Area media dei quartieri in funzione del numero di quartieri individuati.

Come osservato nel dataset principale, l'area complessivamente coperta non è costante ma dipende da come vengono distribuite le amenities ai quartieri. In linea di massima, l'area media dei quartieri si riduce all'aumentare del numero di cluster individuati. Un'eccezione è rappresentata dal parametro di decadimento 19, che nonostante individui lo stesso numero di quartieri del parametro 18, ottiene un'area media inferiore di 600 metri quadrati.

Nella figura 4.2.11 viene mostrata la relazione tra perimetro medio dei quartieri e numero di quartieri individuati.

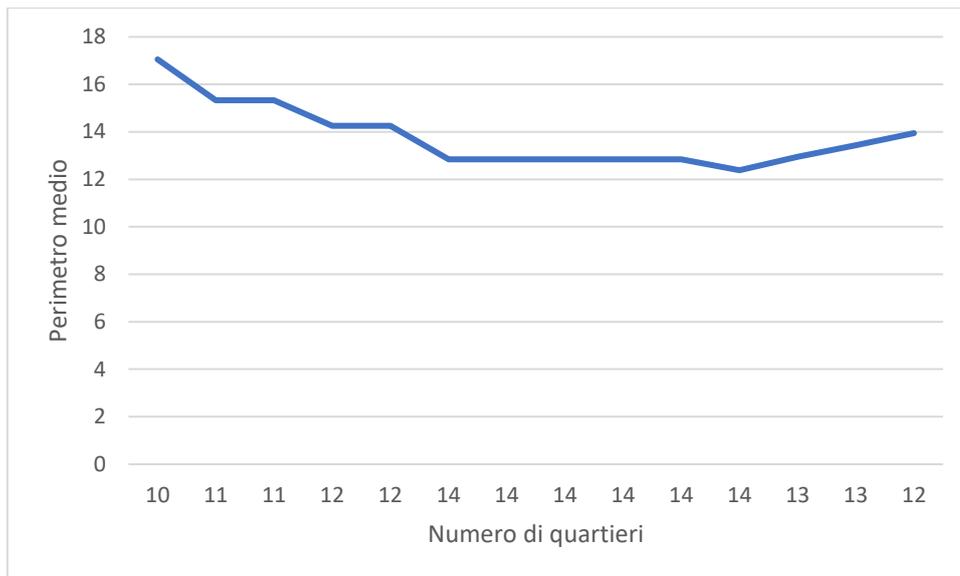


Figura 4.2.11. Perimetro medio dei quartieri in funzione del numero di quartieri individuati.

In questo caso, il perimetro medio dei quartieri risulta inversamente proporzionale al numero di quartieri individuati. In corrispondenza del parametro di decadimento 7 che individua 10 cluster si ottiene il maggior perimetro medio con circa 17 mila metri. Il valore si riduce fino al parametro 18 con i suoi 14 quartieri e 12 mila e 400 metri, per poi aumentare nei valori finali.

Nella figura 4.2.12 viene mostrata la relazione tra densità media dei quartieri e numero di centri individuati. Questa misura permette di comprendere se i cluster che si ottengono risultano diversamente ricchi di amenities all'interno.

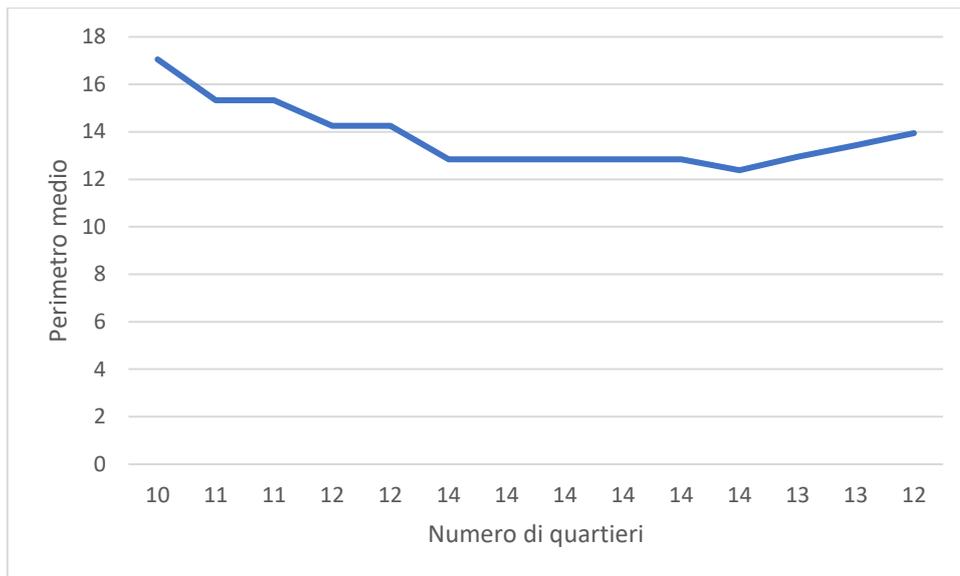


Figura 4.2.12. Densità media dei quartieri in funzione del numero di quartieri individuati.

Come è possibile osservare, la densità media risulta variare molto tra i parametri analizzati. Da una distribuzione delle amenities in 10 quartieri, che presentano la maggiore area media, deriva la minima densità (meno di 8 amenities per km quadrato). La misura sale fino a 18,3 amenities per km quadrato, con 13 amenities, nonostante a questo valore non corrisponda l'area media minore. Inoltre, nonostante anche con il parametro di decadimento 19 vengano individuati 13 quartieri, la densità media risulta inferiore di 2 amenities per km quadrato.

5. Costruzione dello spazio delle amenities

Ottenuti i quartieri nella fase di clusterizzazione, si procede con la realizzazione dello “spazio delle amenities”. Si tratta di una rappresentazione in forma di grafo che indica quali tipologie di amenities hanno maggiore probabilità di trovarsi nello stesso quartiere. L’algoritmo che ha permesso di realizzare i grafi è stato realizzato con l’ambiente di lavoro *R* ed è inserito nell’appendice B.

Per capire la relazione tra le varie tipologie viene calcolato un indice di correlazione tra le amenities presenti nei quartieri individuati. Per fare questo viene calcolato il coefficiente di correlazione per ranghi di Spearman tra le 113 tipologie amenities presenti nel dataset.

Per ogni coppia di tipo di amenities, viene calcolato il rispettivo coefficiente di correlazione di Spearman come indicato nella formula 5.1.

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

W_{ij} rappresenta il grado di correlazione tra la tipologia i e la tipologia j , N_{ic} è il numero di volte che la tipologia i appare nel cluster c , N_{jc} è il numero di volte che la tipologia j appare nel cluster c . Il calcolo del coefficiente di correlazione viene esteso a tutti i quartieri individuati.

Viene calcolato un coefficiente di correlazione per ranghi per evitare problemi che potrebbero sorgere con altre modalità di calcolo, come con il coefficiente di correlazione di Pearson. Eventuali problematiche possono sorgere dall’asimmetria della distribuzione delle tipologie di amenities, ad esempio amenities come ristoranti sono estremamente diffuse sul territorio, mentre altre come cinema e profumerie sono meno presenti.

Un primo aspetto che si nota sono due coppie di nodi separate dall'albero principale dovuto al fatto che non stati rappresentati i legami che le tenevano legate in quanto non rispettavano i requisiti previsti.

Lo spazio delle amenities che si ottiene si sviluppa su una serie di nodi che rappresentano il tronco (che parte dalla tipologia *restaurant* passa per *bar*, *fast_food*, *sports*, *coffe* e termina in *perfumery*) da cui si allargano gran parte dei rami.

Attorno alle due tipologie più presenti, bar e ristoranti, è possibile osservare il cluster delle amenities legate al cibo, di colore verde. Questo gruppo raccoglie *bar*, *restaurant*, *fast_food*, *deli*, *confectionery*, *ice_cream* e *pub* a cui potrebbe essere aggiunta anche la tipologia *bakery*.

Il nodo bar entra a far parte anche di un sotto cluster che coinvolge anche le tipologie *pharmacy*, *supermarket*, *post_office* e *fuel* che riguardano acquisti frequenti e ripetuti. Questo genere di amenity è distribuito omogeneamente sul territorio per ridurre i costi di ricerca dei consumatori, come verrà osservato nel capitolo successivo attraverso il coefficiente di variazione.

Le amenities rivolte all'abbigliamento risultano legate tra loro, essendo acquistati spesso nella stessa occasione e questo si manifesta nelle scelte di localizzazione di queste attività. In basso a destra è osservabile il ramo *bag-shoes-mall* mentre in alto a sinistra *sports-clothes-department_store*. A quest'ultimo ramo sono anche legati due amenities particolari, *jewelry* e *furniture*, che sono accumulate da un processo d'acquisto complesso e accompagnato da figure specializzate.

Da notare come alcune coppie di amenity, nonostante facciano parte dello stesso cluster, tendano a respingersi tra loro. Questo si manifesta nelle triadi non transitive come quella tra *tobacco*, *newsagent* e *kindergarten*. In questo caso tabaccherie e scuole dell'infanzia risultano legate a edicole ma non correlate tra loro a sufficienza per entrare nell'albero ricoprente massimo.

Per facilitare la visione e la comprensione di possibili legami tra le amenities, ed evitare confusione visiva, nella figura 5.2 viene rappresentato l'*amenity space* alzando il filtro del grado di correlazione a 0.5.

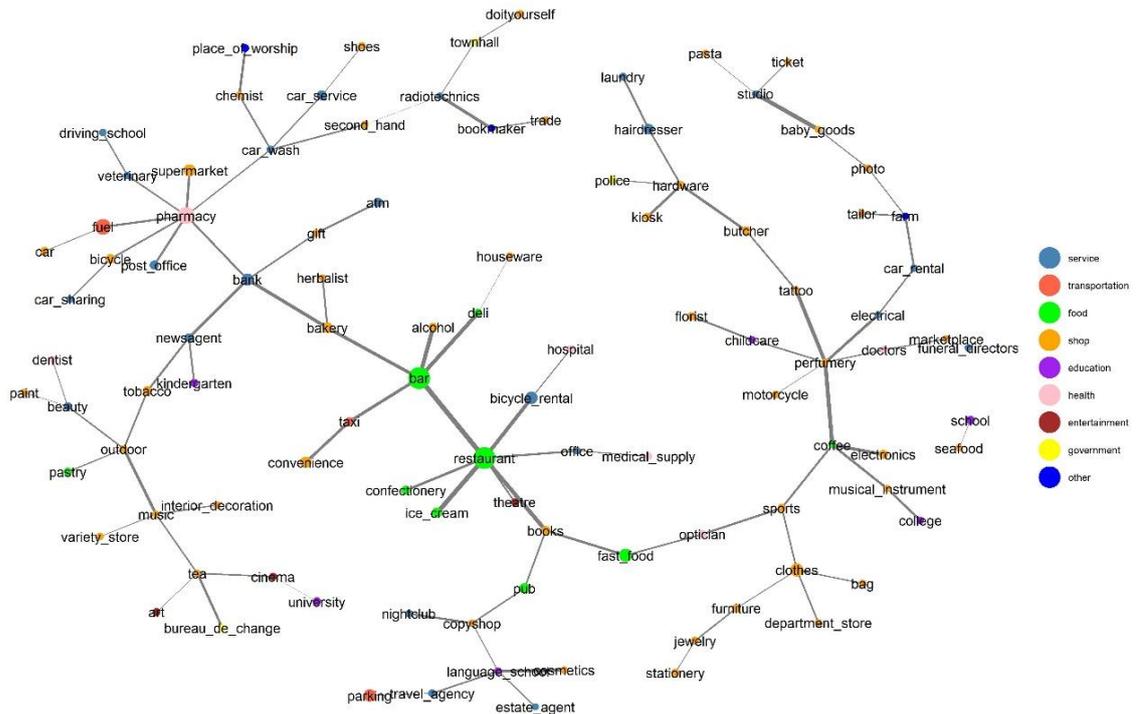


Figura 5.3. Amenity space filtrato dai quartieri con meno amenities.

In questo caso vengono eliminati dall'albero 11 dei 113 nodi. L'operazione permette di definire meglio il ramo dell'abbigliamento aggiungendo l'amenity al cluster in questione.

Anche il ramo legato a beni acquistati frequentemente è più chiaro in alto a sinistra, incorporando la sezione che si allarga dal nodo *car_wash*.

Per comprendere possibili differenze tra i quartieri centrali e periferici è svolta un'ulteriore analisi. Le due classificazioni sono visibili nelle immagini 5.3 e 5.4.

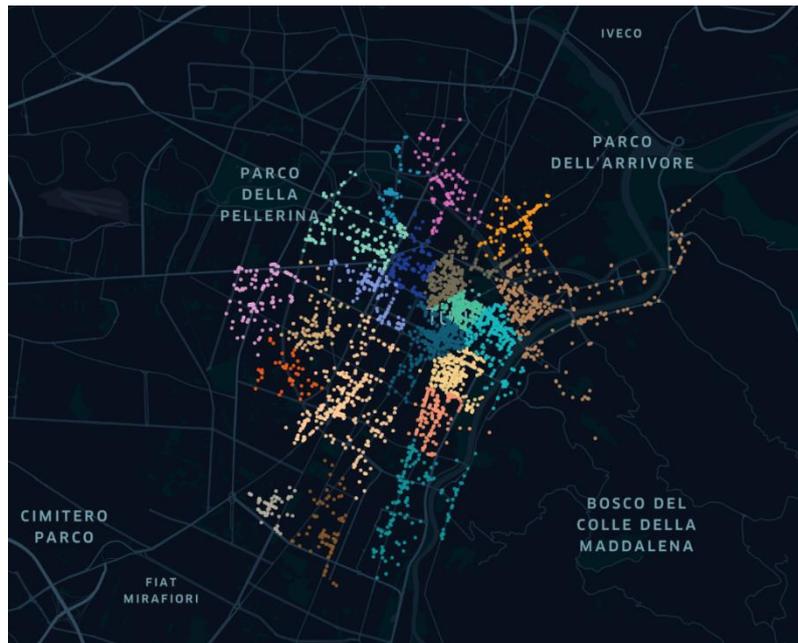


Figura 5.4. Rappresentazione dei quartieri centrali.

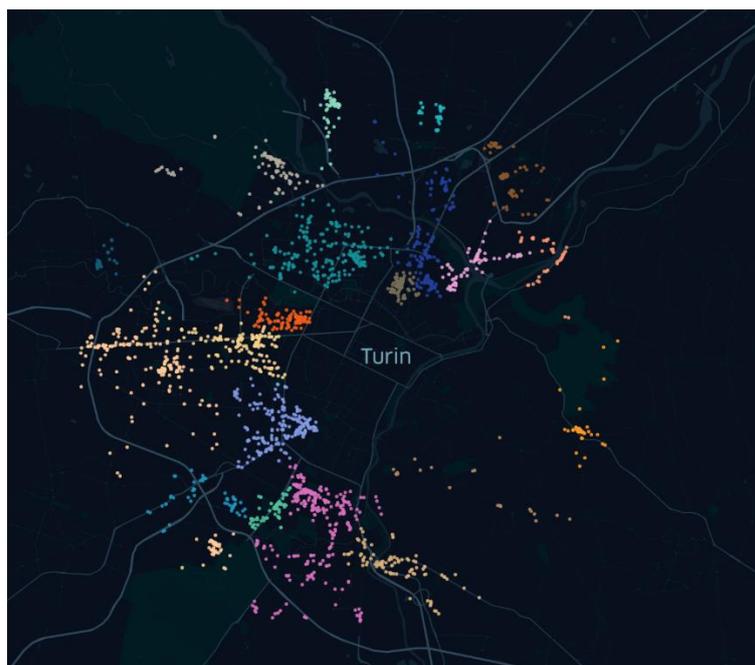


Figura 5.5. Rappresentazione dei quartieri esterni.

Nelle figure 5.6 e 5.7 sono rappresentati gli *amenity space* che si ottengono rispettivamente dai due raggruppamenti.

Per cercare la relazione tra le categorie più presenti viene svolta un'ulteriore analisi eliminando dal dataset principale le amenities con meno di 30 nodi. Si ottiene un dataset con 4836 amenities distribuiti su 35 tipologie. Applicando un parametro di decadimento pari a 16 si ottengono 44 quartieri, mostrati nella figura 5.8.

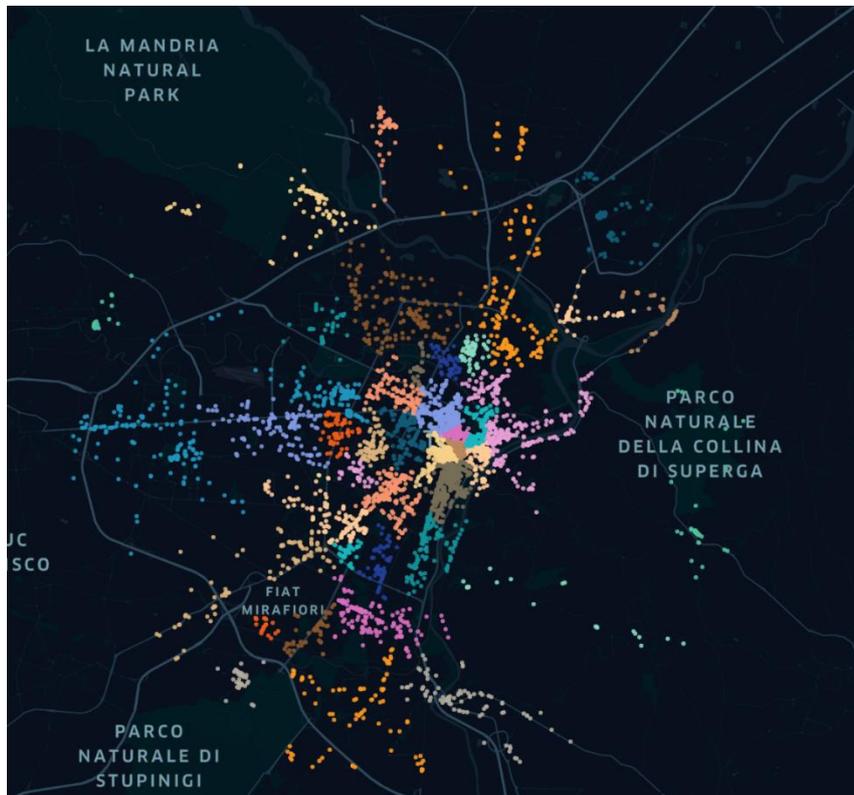


Figura 5.8. Rappresentazione dei quartieri da dataset con amenities più diffuse.

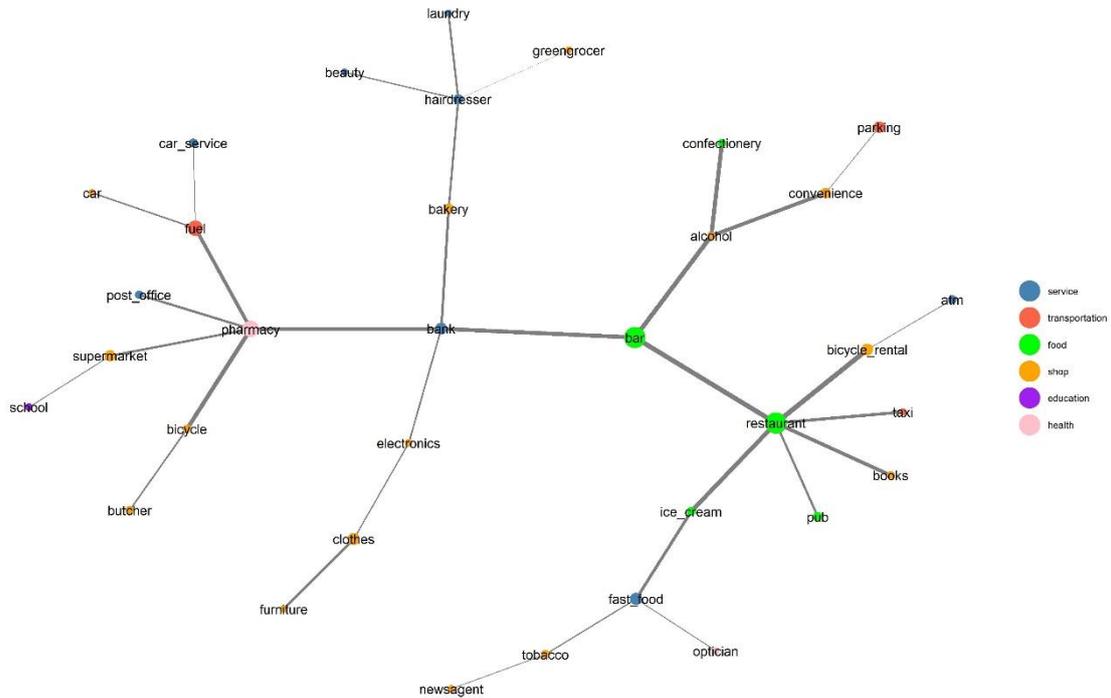


Figura 5.9. Rappresentazione dell'amenity space delle amenities più presenti.

Esaminando l'*amenity space* delle amenity più presenti nel dataset si possono osservare 4 cluster principali: a sinistra beni e servizi che richiedono un basso coinvolgimento del consumatore, a destra il cluster legato al cibo, in alto i servizi e in basso acquisti che richiedono un maggiore grado di interazione con i consumatori.

6. Caratterizzazione dei quartieri

Dalla fase di clusterizzazione delle amenities sono stati individuati nuovi quartieri per la città di Torino, che hanno permesso di delineare lo spazio delle amenities per comprendere come queste sono legate tra loro nello spazio cittadino.

Avendo a disposizione questa nuova definizione dei quartieri, viene effettuato un tipo di analisi non presente sul lavoro originale di Hidalgo (2018). Dato che i cluster ottenuti non sono delle esatte repliche in miniatura della città, si cerca di comprendere se determinate aree sono caratterizzate dalla presenza di particolari tipologie di amenities.

Per comprendere come le varie tipologie sono disposte, più o meno ordinatamente nei quartieri, viene calcolato per ognuna di esse il coefficiente di variazione. Questo valore permette di comprendere se la presenza della tipologia nei quartieri si avvicina o si discosta dalla media attesa. Nella tabella 6.1 vengono riportate le tipologie con il rispettivo numero medio per quartiere, deviazione standard e coefficiente di variazione. Questi valori sono derivati dai quartieri ottenuti con parametro di decadimento 16.

Tabella 6.1

Tipologia	Media	Deviazione standard	Coefficiente di variazione
studio	0.073	0.341	4.667
electrical	0.049	0.215	4.416
college	0.049	0.215	4.416
hearing_aids	0.049	0.215	4.416
ticket	0.049	0.215	4.416

erotic	0.098	0.370	3.791
farm	0.073	0.260	3.559
pasta	0.073	0.260	3.559
medical_supply	0.073	0.260	3.559
trade	0.073	0.260	3.559
townhall	0.073	0.260	3.559
radiotechnics	0.073	0.260	3.559
estate_agent	0.073	0.260	3.559
baby_goods	0.073	0.260	3.559
perfumery	0.122	0.395	3.237
Bag	0.146	0.472	3.223
university	0.756	2.346	3.102
language_school	0.098	0.297	3.041
mall	0.098	0.297	3.041
bureau_de_change	0.098	0.297	3.041
community_centre	0.293	0.862	2.946
tailor	0.146	0.417	2.848
Tea	0.146	0.417	2.848
musical_instrument	0.122	0.327	2.683
bookmaker	0.122	0.327	2.683
garden_centre	0.122	0.327	2.683
marketplace	0.268	0.699	2.606
photo	0.195	0.505	2.586
toys	0.171	0.436	2.556

music	0.171	0.436	2.556
car_sharing	0.171	0.436	2.556
copyshop	0.244	0.616	2.526
paint	0.244	0.616	2.526
shoes	0.561	1.362	2.429
outdoor	0.146	0.353	2.415
variety_store	0.146	0.353	2.415
childcare	0.146	0.353	2.415
coffee	0.220	0.519	2.362
chemist	0.220	0.519	2.362
funeral_directors	0.220	0.519	2.362
car_rental	0.220	0.519	2.362
gift	0.244	0.575	2.358
clothes	4.951	11.546	2.332
herbalist	0.195	0.454	2.325
cosmetics	0.366	0.848	2.319
seafood	0.171	0.376	2.204
office	0.268	0.585	2.182
tattoo	0.244	0.531	2.177
motorcycle	0.317	0.679	2.140
department_store	0.220	0.469	2.137
social_facility	0.220	0.469	2.137
Art	0.512	1.085	2.118
hospital	0.341	0.719	2.107

travel_agency	0.488	1.015	2.081
nightclub	0.293	0.594	2.031
cinema	0.390	0.793	2.031
jewelry	0.634	1.283	2.024
second_hand	0.317	0.602	1.900
doityourself	0.512	0.966	1.886
doctors	0.512	0.966	1.886
houseware	0.220	0.414	1.886
interior_decoration	0.463	0.829	1.789
driving_school	0.317	0.560	1.768
deli	0.463	0.799	1.725
furniture	0.927	1.536	1.658
school	1.146	1.881	1.641
car_wash	0.634	1.030	1.625
clinic	0.537	0.858	1.599
theatre	0.439	0.700	1.595
beauty	0.780	1.220	1.563
Car	0.780	1.220	1.563
alcohol	1.317	2.053	1.559
stationery	0.537	0.829	1.545
sports	0.561	0.857	1.527
dentist	0.390	0.579	1.484
confectionery	1.390	2.047	1.472
books	1.439	2.107	1.465

veterinary	0.512	0.737	1.438
electronics	1.268	1.808	1.426
pastry	0.634	0.904	1.426
kiosk	0.707	0.994	1.405
kindergarten	0.707	0.969	1.370
police	0.585	0.796	1.359
optician	0.732	0.963	1.317
place_of_worship	0.585	0.764	1.306
hardware	0.634	0.819	1.292
car_service	2.024	2.513	1.241
laundry	0.756	0.931	1.232
taxi	0.878	1.064	1.211
bicycle_rental	4.341	5.257	1.211
parking	4.122	4.890	1.186
florist	0.683	0.810	1.186
library	0.683	0.810	1.186
Pet	0.561	0.664	1.184
fast_food	4.927	5.824	1.182
bicycle	1.561	1.835	1.176
greengrocer	0.878	1.017	1.158
pub	2.268	2.576	1.136
ice_cream	2.098	2.206	1.052
newsagent	1.098	1.143	1.042
fuel	8.195	8.503	1.038

restaurant	17.683	18.270	1.033
bakery	2.805	2.689	0.959
tobacco	1.927	1.840	0.955
convenience	3.195	3.038	0.951
Atm	1.561	1.449	0.928
butcher	1.537	1.416	0.921
post_office	1.585	1.431	0.903
pharmacy	10.122	8.107	0.801
hairdresser	2.829	2.240	0.792
bank	4.415	3.457	0.783
Bar	17.683	13.709	0.775
supermarket	3.976	2.542	0.639

Osservando i risultati, le amenities che presentano un coefficiente di variazione maggiore dovrebbero essere concentrate sul territorio e propense a caratterizzare alcuni quartieri.

Il fondo della tabella indica quali amenities presentano nei quartieri un numero di nodi vicino alla media. Analizzando questo genere di amenities si nota che si tratta di negozi e servizi che coprono il territorio per facilitare la loro raggiungibilità e ridurre i costi di ricerca come supermercati, farmacie, distributori di carburante, tabaccherie.

Per comprendere se alcuni quartieri sono caratterizzati dalla presenza di alcune tipologie di amenities viene applicato il test del chi-quadro, per esaminare la differenza tra i valori osservati e attesi. Per fare questo viene applicata la formula 6.1.

$$\chi^2 = \frac{((a*d-b*c)^2)*n}{(a+b)*(a+c)*(b+d)*(c+d)} \quad (6.1)$$

Una prima analisi viene svolta confrontando il numero di nodi appartenenti ad un quartiere con quelli presenti nel resto del dataset. Applicando la formula 6.1 a rappresenta il numero di nodi nel quartiere della tipologia analizzata, b il numero di nodi nel quartiere escludendo quelli della particolare tipologia analizzata, c rappresenta il numero di nodi nel dataset della tipologia analizzata escludendo quelli presenti nel quartiere, d il numero di nodi nel dataset escludendo quelli presenti nel quartiere analizzato, n rappresenta il numero di amenities nel dataset.

Per evidenziare quartieri contrassegnati da particolari tipi di amenities, questi devono superare due condizioni: presentare un chi-quadro significativo al 95% e presenza della particolare tipologia superiore al valore atteso. Per ottenere il valore atteso viene calcolata la percentuale della tipologia nel dataset e moltiplicata per il numero di amenities nel quartiere.

Nella figura 6.2 sono riportati i quartieri che si caratterizzano per le tipologie *clothes* (in verde), *shoes* (blu) e *bag* (arancione).

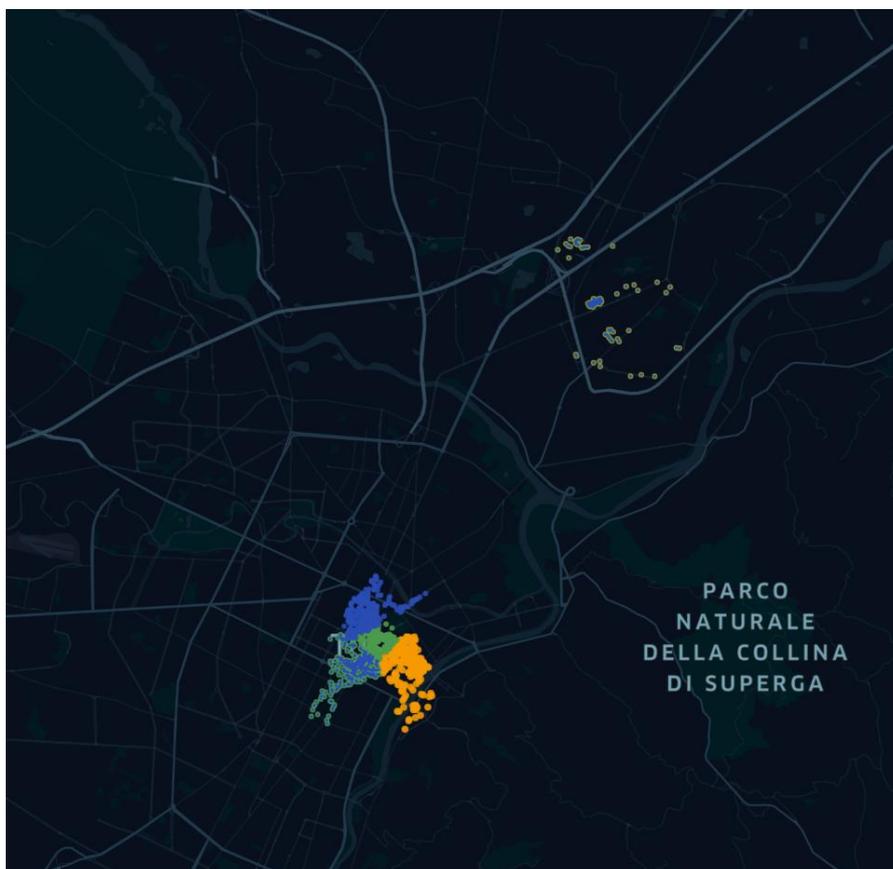


Figura 6.3. Quartieri caratterizzati per le tipologie clothes, shoes e bag.

Come è osservabile nell'immagine i quartieri del centro torinese formano un cluster che riguarda queste tipologie di negozi. Il quartiere in alto a destra risulta caratterizzato per tutte e tre questi beni in quanto è presente un centro commerciale.

Per comprendere se sono presenti possibili cluster che si caratterizzano non semplicemente per il numero, ma per la densità delle amenities sulla superficie del quartiere, viene svolta una ulteriore analisi.

Sempre utilizzando la formula del chi-quadro indicata nella formula 6.1 in questo caso viene applicata con le informazioni sull'estensione dei quartieri. In questo caso a rappresenta la densità della tipologia analizzata all'interno del quartiere, b la densità delle altre amenities all'interno del quartiere, c la densità della tipologia analizzata nel resto dei quartieri, d la densità delle restanti amenities nel resto dei quartieri, n la somma delle densità.

In questo caso viene riportata la figura 6.3 contenente i quartieri caratterizzati per la presenza delle amenities restaurant.

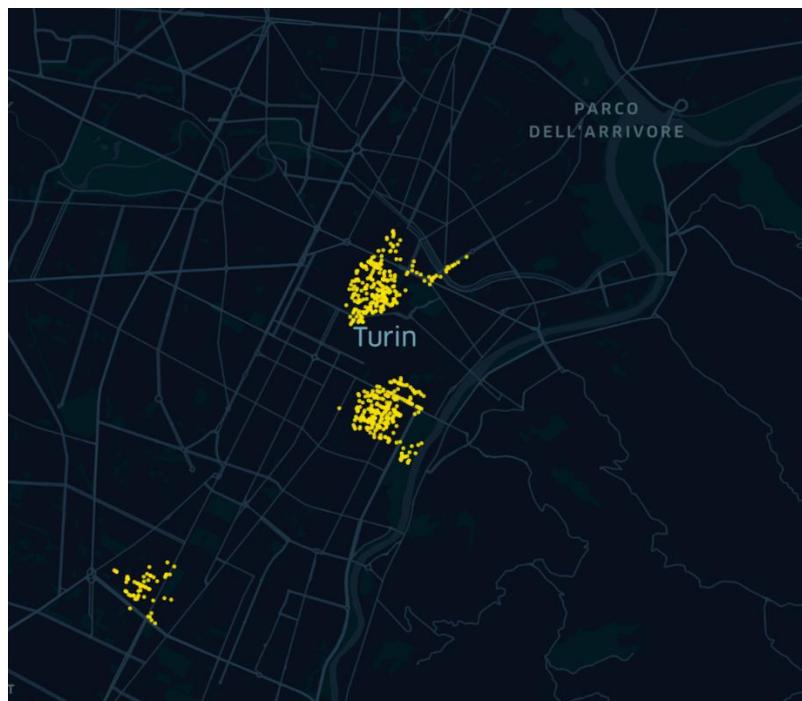


Figura 6.3. Quartieri caratterizzati per la tipologia restaurant.

Le aree che stanno alla densità, presentano un elevato numero di ristoranti risultano essere la zona del quadrilatero romano e della parte nord di San

Salvario. Questo non dovrebbe sorprendere essendo zone con molti locali che attirano cittadini e turisti di Torino, sfruttando la vicinanza con i maggiori punti di interesse della città.

7. Conclusioni

Il lavoro svolto si è posto come obiettivo indagare sulla divisione organica degli spazi cittadini di Torino e comprendere i legami tra le attività presenti. Per fare questo sono state raccolte le informazioni di più di 5.000 amenities situate nel comune di Torino e nelle aree limitrofe. Attraverso un algoritmo di clustering con questi dati sono stati individuati dei nuovi quartieri con l'intenzione di manifestare le aree di spostamento degli abitanti.

I nuovi cluster individuati hanno evidenziato una distribuzione centrica degli spazi cittadini. Le aree del centro storico risultano maggiormente dense di amenities e meno estese, mentre allontanandosi verso le aree periferiche si riscontra un trend opposto.

Al fine di comprendere come le scelte di localizzazione delle tipologie di amenities siano influite dalla presenza di altre è stato realizzato lo spazio delle amenities. Si tratta di un albero ricoprente massimo che riflette le correlazioni tra le tipologie di amenities all'interno dei quartieri ottenuti. Questo oggetto mostra come alcune attività che offrono una stessa categoria merceologica o beni consumati o acquistati nella stessa occasione tendano a disporsi nelle vicinanze e a formare cluster, in linea con il modello di Hotelling e gli studi di Hernandez e Bennison.

Per comprendere come questi cluster siano presenti sul territorio si è cercato di evidenziare aree che contengono i vari tipi di amenities in maniera superiore al resto della mappa. Questo è stato possibile attraverso il test del chi-quadro che ha mostrato come l'area del centro cittadino rappresenti il cluster legato all'abbigliamento, come lecito aspettarsi dagli studi sui fattori esogeni e all'effetto spillover.

Possibili evoluzioni e miglioramenti dello studio riguardano l'estensione della base di dati da cui partire, per indagare maggiormente le relazioni tra

amenities. Questa estensione è percorribile sia cambiando lo strumento per la raccolta dati, ma anche estendo l'analisi a più città. Un avanzamento in questa direzione è percorribile anche attraverso studi sui parametri forniti all'algoritmo, per comprendere la bontà dei cluster individuati.

Bibliografia

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-retail-distribution/dp/B0000CNPHS>

Boschma, R. A. (2005). Does geographical proximity favour innovation? *Economie et institutions*, 6–7, 111–128 <https://doi.org/10.4000/ei.926>

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>

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

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

Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012). *The livelihoods project: Utilizing social media to understand the dynamics of a city*. Sixth International AAAI Conference on Weblogs and Social Media

DiPasquale, D., & Wheaton, W. C. (1995). *Urban economics and real estate markets* (1 edition). Pearson.

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

Eaton, B. C., & Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies*, 42(1), 27–49. <https://doi.org/10.2307/2296817>. JSTOR

Eaton, B. C., & Lipsey, R. G. (1982). An economic theory of central places. *The Economic Journal*, 92(365), 56–72. <https://doi.org/10.2307/2232256>. JSTOR

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>

Florida, R. (2019). *The rise of the creative class*. Basic books.

Frenken, K., Oort, F. V., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697. <https://doi.org/10.1080/00343400601120296>

Fujita, M., & Krugman, P. (2004). The new economic geography: Past, present and the future. In R. J. G. M. Florax, & D. A. Plane (Eds.), *Fifty years of regional science* (pp. 139–164). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-07223-3_6

Guevara, M. R., Hartmann, D., Aristarán, 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>

Hernández, 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

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482–487. <https://doi.org/10.1126/science.1144581>

Hise, R., Kelley, P., Gable, M., & McDonald, J. B. (1983). Factors affecting the performance of individual chain store units: An empirical analysis. *Journal of Retailing*, 59(2), 22–39

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153), 41–57. <https://doi.org/10.2307/2224214>. JSTOR

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences*, 115(50), 12646–12653. <https://doi.org/10.1073/pnas.1800475115>

Krugman, P. (1993). On the number and location of cities. *European Economic Review*, 37 (2), 293–298. [https://doi.org/10.1016/0014-2921\(93\)90017-5](https://doi.org/10.1016/0014-2921(93)90017-5).

Leinberger, C. B., & Lynch, P. (2014). Foot traffic ahead: Ranking walkable urbanism in America's largest metros. <https://trid.trb.org/view/1314064>

Losch, A. (1954). *The economics of location: A pioneer book in the relations between economic goods and geography*. Yale University Press

Marshall, A. (1890). *Principles of economics*. Macmillan and Company.

Mulligan, G. F. (1984). Agglomeration and central place theory: A review of the literature. *International Regional Science Review*, 9(1), 1–42. <https://doi.org/10.1177/016001768400900101>

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.

Appendice A

```
import csv
import pandas as pd
import geopy.distance
import pickle
import math
import decimal
import utm
from scipy.spatial import distance

def extractDigits(x):
    return [[el] for el in x]

#parametri
database_da_analizzare='database_pronto.csv'
delimitatore=', '
coordinate_da_inserire=_in='coordinate.pickle'
nuovariga=''
separatore=', '
raggio_minimo=500

dati_spearman=[]

#importare dati puliti precedentemente
with open(database_da_analizzare, newline='') as csvfile:
    lettore = csv.reader(csvfile, delimiter=delimitatore)

#stampo header, per vedere cosa corrisponde alle colonne e
controllare gli indici
    header=next(lettore)
    print (header)
#creazione della matrice con le coordinate di ogni punto
    coordinate=[[riga[1],riga[2]] for riga in lettore]
    print('elementi in matrice coordinate')
    print (len(coordinate))

#trasformare coordinate in float
pointList = [[float(coordinata[0]), float(coordinata[1])] for
coordinata in coordinate]

#trasformo le coordinate in cartesiane
mod= list(utm.from_latlon(coordinata[0],coordinata[1]) for
coordinata in pointList)
```

```

#genera per ogni coordinata una lista con quattro elementi ma
interessano solo i primi 2
cartesiane=list((elemento[0],elemento[1]) for elemento in mod)

#controllo che siano presenti lo stesso numero di elementi
print('elementi in matrice coordinate cartesiane')
print(len(cartesiane))

#creo matrice delle distanze con le coordinate cartesiane
distanze = distance.cdist(cartesiane, cartesiane, 'euclidean')

#controllo len della lista distanze e anche della sottolista
print('lunghezza delle matrice delle distanze e delle sottoliste')
print(len(distanze))
print(len(distanze[0]))

##seconda fase

#matrice dei contributi di accessibilità #capire se usare km o
metri per le distanze
contributi_accessibilita=list(list(math.exp(distanze[a][b]*(-
16)*(0.001)) for a in range(0,len(distanze)))for b in range
(0,len(distanze)))

#controllo len della lista distanze e anche della sottolista
print('lunghezza delle matrice dei contributi di accessibilità e
delle sottoliste')
print(len(contributi_accessibilita))
print(len(contributi_accessibilita[0]))

#lista indici di accessibilità
indici_accessibilita=list(sum(contributi_accessibilita[a]) for a
in range(0,len(distanze)))

#controllo len degli indici di accessibilità
print ('len lista indici di accessibilità')
print (len(indici_accessibilita))

#devo trovare il 90% delle amenities di interesse
#creo un listone con tutte i singoli indici di accessibilità
#rendo ognuno una matrice
indici_accessibilita_matrciati=extractDigits(indici_accessibilit
à)
print ('len lista in cui ogni elemento è reso sotto lista')
print (len(indici_accessibilita_matrciati))

#genero lista per associare a ogni indice di access. la posizione
posizioni=[i for i in range(0,len(distanze))]

#a ognuno associo il numero
lista_accessibilita_posizione=[[indici_accessibilita[indice],ind
ice] for indice in posizioni]
#controllo il len
print('lunghezza della lista con indice di access e posizione')

```

```

print(len(lista_accessibilita_posizione))

#li ordino in ordino decrescente per prendere i maggiori
lista_accessibilita_posizione.sort(reverse=True)

#prendo il 90% di questi
quota=int(0.9*(len(distanze)))
quota_100=int(1*(len(distanze)))

#lista con le posizioni di interesse
posizioni_notevoli=[lista_accessibilita_posizione[numero][1] for
numero in range(0,quota)]
posizioni_notevoli_100=[lista_accessibilita_posizione[numero][1]
for numero in range(0,quota_100)]
#controllo il numero di posizioni
print ('numero delle posizioni notevoli')
print(len(posizioni_notevoli))
print(len(posizioni_notevoli_100))

#esporto la lista delle posizioni notevoli
posizioni_notevoli_out=open('posizioni_notevoli_90.pickle','wb')
pickle.dump(posizioni_notevoli,posizioni_notevoli_out)
posizioni_notevoli_out.close()

#esporto la lista delle posizioni notevoli
posizioni_notevoli_out=open('posizioni_notevoli.pickle','wb')
pickle.dump(posizioni_notevoli_100,posizioni_notevoli_out)
posizioni_notevoli_out.close()

#ricreo la lista delle coordinate di interesse
cartesiane_notevoli=[cartesiane[posizione] for posizione in
posizioni_notevoli]
#controllo la lunghezza
print('lunghezza lista cartesiane notevoli')
print(len(cartesiane_notevoli))

#ricreo la lista delle coordinate di interesse per i quartieri
cartesiane_notevoli_100=[cartesiane[posizione] for posizione in
posizioni_notevoli_100]
#controllo la lunghezza
print('lunghezza lista cartesiane notevoli 100')
print(len(cartesiane_notevoli_100))

#ricreo la lista delle coordinate di interesse
coordinate_notevoli=[coordinate[posizione] for posizione in
posizioni_notevoli]
#controllo la lunghezza
print('lunghezza lista coordinate notevoli')
print(len(coordinate_notevoli))

#ricreo la lista delle coordinate di interesse 100
coordinate_notevoli_100=[coordinate[posizione] for posizione in
posizioni_notevoli_100]
indici_accessibilita_100=[(indici_accessibilita[posizione]*2)
for posizione in posizioni_notevoli_100]

```

```

#controllo la lunghezza
print('lunghezza lista coordinate notevoli 100')
print(len(coordinate_notevoli_100))

#ricalcolo matrice delle distanze
distanze_notevoli=distance.cdist(cartesiane_notevoli,
cartesiane_notevoli, 'euclidean')
#controllo della lunghezza
print('lunghezza delle matrice delle distanze notevoli e delle
sottoliste')
print(len(distanze_notevoli))
print(len(distanze_notevoli[0]))

#creo un listone con tutte le singole distanze notevoli
lista_distanze_notevoli=list(distanze_notevoli[b][c] for b in
range (0,len(distanze_notevoli)) for c in range
(0,len(distanze_notevoli)))
print('len della lista con tutte le singole distanze_notevoli')
print (len(lista_distanze_notevoli))

#rendo ogni singola distanza notevole una sottolista in modo da
poter aggiungere il relativo indice di accessibilità
lista_elementi_matriciati=extractDigits(lista_distanze_notevoli)
print ('len lista in cui ogni elemento è reso sotto lista')
print (len(lista_elementi_matriciati))

#lista con gli indici di acces notevoli
accessabilita_notevoli=[indici_accessibilità[posizione] for
posizione in posizioni_notevoli]
#controllo
print('lunghezza della lista degli indici di accessibilità
notevoli')
print(len(accessabilita_notevoli))

#ricreo le sottoliste
composite_list
=[lista_elementi_matriciati[x:x+len(distanze_notevoli)] for x in
range(0, len(lista_elementi_matriciati),len(distanze_notevoli))]
#controllo
print('lunghezza della lista ricomposta con le sottoliste')
print(len(composite_list))

#ad ogni elemento aggiungo il relativo indice accesaabilità
for e in range (0,len(distanze_notevoli)):
    for g in range (0,len(distanze_notevoli)):
        composite_list[g][e].append(accessabilita_notevoli[e])
#controllo con l'aggiunta degli indici di access
print('lunghezza della lista ricomposta con gli indici di access')
print (len(composite_list))

#ordino ogni sottolista in base al primo elemento, cioè le
distanze
for c in range (0,len(distanze_notevoli)):

```

```

    composite_list[c].sort()
#controllo
print('len composite list ordinata con indici di access')
print (len(composite_list))

#calcolo il valore dell'euristica
euristica=list((accessabilita_notevoli[a]*3+50) for a in range
(0,len(accessabilita_notevoli)))
euristica=[round(x) for x in euristica]
#controllo dell'euristica
print('len della lista dei valori dell euristica')
print (len(euristica))

#per trovare i massimi di ogni riga
massimi=[]
for a in range (0,len(euristica)):
    #nella prova viene troppo alta l'euristica e dà problemi,
    eseguo la ricerca orizzontale fino al valore indicato
    dall'euristica
    massimi.append(max(composite_list[a][b][1] for b in range
(0,euristica[a])))
    #massimi.append(max(composite_list[a][b][1] for b in range
(0,9)))

#controllo lunghezza lista centro
print('lunghezza della lista massimi')
print(len(massimi))

#lista che mi indicherà le posizioni dei centri
centro=[]
for a in range(0,len(euristica)):

    if composite_list[a][0][1]>=massimi[a]:
        centro.append('centrooooo')
    else:
        centro.append('no')

#controllo lunghezza lista centro
print('lunghezza della lista centro')
print(len(centro))

#ottengo le posizioni dei centri sulla lista delle coordinate
notevoli
indices = [i for i, s in enumerate(centro) if 'centrooooo' in s]
print('numero di centri')
print(len(indices))

#esporto la lista delle posizioni notevoli
indices_out=open('indices.pickle','wb')
pickle.dump(indices,indices_out)
indices_out.close()

#creo lista con le coordinate dei centri
coordinate_centri=list(coordinate_notevoli[i] for i in indices)
print('numero di coordinate centri')
print(len(coordinate_centri))

```

```

#esporto le coordinate dei centri
with open("coordinate_centri.csv","w") as f:
    wr = csv.writer(f,delimiter="\n")
    wr.writerow(coordinate_centri)

cartesiane_centri=list(cartesiane_notevoli[i] for i in indices)
print('numero di cartesiane centri')
print(len(cartesiane_centri))

#ricalcolo matrice delle distanze
distanze_notevoli=distance.cdist(cartesiane_notevoli_100,
cartesiane_notevoli_100, 'euclidean')
#controllo della lunghezza
print('lunghezza delle matrice delle distanze notevoli e delle
sottoliste con il 100% delle amenities')
print(len(distanze_notevoli))
print(len(distanze_notevoli[0]))

#creo lista con tante sottoliste quanti i quartieri
torino=[[[] for x in range (0,len(indices))]

#creo una lista dove ho gli indici di tutti i punti da assegnare
che mano a mano si svuota
punti_da_assegnare=[i for i in range
(0,len(coordinate_notevoli_100))]
#la lista dei punti assegnati è vuota inizialmente
punti_assegnati=[]

#assegno le amenities meno distanti di 500 metri, dovrei anche
verificare se ci sono 2 indici distanti meno di 1km
amenities_comuni=[]

for indice in indices:
    for punto_da_assegnare in punti_da_assegnare:
        if
distanze_notevoli[indice][punto_da_assegnare]<raggio_minimo:

torino[indices.index(indice)].append(punto_da_assegnare)
    if punto_da_assegnare in punti_assegnati:
        amenities_comuni.append(punto_da_assegnare)
    else:
        punti_assegnati.append(punto_da_assegnare)

```

```

#devo gestire il caso in cui uno stesso punto sia stato assegnato
a più quartieri
#prima di tutto voglio sapere quali punti si ripetono
amenities_comuni = list(dict.fromkeys(amenities_comuni))

#poi li metto in una lista
for elemento in amenities_comuni:
    for quartiere in torino:
        if elemento in quartiere:
            quartiere.remove(elemento)

#poi verifico per ognuno di essi rispetto a quale quartiere
abbiano la minima distanza
for elemento in amenities_comuni:
    distanze_quartieri=[distanze_notevoli[elemento][indice] for
indice in indices]
    minima_dist=min(distanze_quartieri)

torino[distanze_quartieri.index(minima_dist)].append(elemento)

#rimuovo questi punti da quelli da assegnare
for elemento in punti_assegnati:
    punti_da_assegnare.remove(elemento)

#devo riformare la matrice delle distanze perchè con l'array non
si puo usare list.index()
distanze=list(list(distanze_notevoli[a][b] for a in range
(0,len(coordinate_notevoli_100))) for b in range
(0,len(coordinate_notevoli_100)))

while punti_da_assegnare!=[]:
    #per le amenities da assegnare voglio calcolare le loro
distanze rispetto a quelle assegnate
    minimi=[]
    #per ogni amenity da assegnare identifico la distanza minore
rispetto alle amenity assegnate
    for num in punti_da_assegnare:
        minimi.append(min(list((distanze[num][numero]) for
numero in punti_assegnati)))
    minimo=min(minimi)
    #indice del punto a cui assegnare
    indice_da_assegnare=minimi.index(minimo)

    numero_da_assegnare=punti_da_assegnare[indice_da_assegnare]

numero_a_cui_assegnare=distanze[numero_da_assegnare].index(minim
o)

for quartiere in torino:
    if numero_a_cui_assegnare in quartiere:
        quartiere.append(numero_da_assegnare)
        punti_assegnati.append(numero_da_assegnare)
        punti_da_assegnare.remove(numero_da_assegnare)

```

```

print('aggiornamento')
print(len(punti_da_assegnare))
print(len(punti_assegnati))
print(len(punti_da_assegnare)+len(punti_assegnati))

#esportazione
coordinate_notevoli = [[float(coordinata[0]),
float(coordinata[1])] for coordinata in coordinate_notevoli]

#importare dati puliti precedentemente
with open(database_da_analizzare, newline='') as csvfile:
    lettore = csv.reader(csvfile, delimiter=delimitatore)

#stampo header, per vedere cosa corrisponde alle colonne e
controllare gli indici
    header=next(lettore)
    print (header)
#creazione della matrice con le coordinate di ogni punto
    amenity=[riga[3] for riga in lettore]

amenity_notevoli=[amenity[posizione] for posizione in
posizioni_notevoli_100]

for quartiere in torino:
    quartiere1=[[coordinate_notevoli_100[i],amenity_notevoli[i]]
for i in quartiere]
    with
open((str(torino.index(quartiere))+"_quartiere.csv","w") as f:
    wr = csv.writer(f,delimiter="\n")
    wr.writerow(quartiere1)

dato=[[torino.index(quartiere), amenity_notevoli[i]] for
quartiere in torino for i in quartiere]

#listone per creare albero
with open("dati_spearman_prof.csv","w") as f:
    wr = csv.writer(f,delimiter="\n")
    wr.writerow(dato)

#pulizia dei dati

import csv
import string
import pickle
from pandas import read_csv

indices_in=open('indices.pickle','rb')
indices=pickle.load(indices_in)

```

```

for i in range (0,len(indices)):
    input_file = open((str(i))+'_quartiere.csv', 'r')
    output_file = open((str(i))+'_quartiere_pulito.csv', 'w')
    data = csv.reader(input_file)
    writer = csv.writer(output_file)
    specials = '['
    specials1=']'
    specials2=""
    specials3=" "

    for line in data:
        line = [value.replace(specials,'') for value in line]
        line = [value.replace(specials1,'') for value in line]
        line = [value.replace(specials2,'') for value in line]
        line = [value.replace(specials3,'') for value in line]
        writer.writerow(line)

    input_file.close()
    output_file.close()

df = read_csv((str(i))+"_quartiere_pulito.csv")
df.loc[-1] = df.columns # adding a row
df.index = df.index + 1 # shifting index
df = df.sort_index() # sorting by index
df.columns = ['lat', 'lon','type']
df.to_csv((str(i))+"_quartiere_pulito.csv")

```

Appendice B

```
# import libraries
library("igraph")
library("NetworkToolbox")
library("intergraph")
library("network")
library("ggplot2")
library("ggnet")
library("sna")

# importo i dati su cui eseguire la correlazione
dati <- read.table("dati_spearman_nuovo.csv", header = TRUE, sep
= ",", colClasses = c("type" = "character", "neighborhood" =
"character"))

# ottiene la lista delle variabili
colonne <- names(dati)

# ottiene la dimensione del vettore
dimensione <- length(colonne)

# disabilita la notazione scientifica
options(scipen=999)

# inizializza le matrici
coefficienti <- matrix(NA, nrow = dimensione, ncol = dimensione)
valorip <- matrix(NA, nrow = dimensione, ncol = dimensione)

# calcola il coefficiente della correlazione ed il valore p
for (i in 1:dimensione) {
  for (j in 1:dimensione) {
    a <- cor.test(dati[[colonne[i]]], dati[[colonne[j]]], method
= "spearman")
    coefficienti[i, j] <- a$estimate
    valorip[i, j] <- a$p.value
  }
}

#realizzo matrice del max spanning tree
albero <- MaST(coefficienti, normal = TRUE)

# imposto le soglie
scoeff <- 0.3
svalop <- 0.01

# calcolo il valore assoluto
mcoeff <- abs(coefficienti)

# calcolo le matrici booleane dei vincoli
c1 <- (mcoeff > scoeff) + 0
```

```

c2 <- (valorip < svalop) + 0

# vincolo finale
c <- c1 * c2

#valori di correlazione dell'albero
archi<-albero*c

#matrice booleana sugli archi
archil<-(abs(archi)>0)+0

#riassegno il nome ai nodi
colnames(archil)<-colonne
rownames(archil)<-colonne

#realizzo il grafo del mast
rete <- graph_from_adjacency_matrix(archil, mode = "undirected",
diag = FALSE)

#importo il file dove sono assegnate le categorie di amenities
tipi <- read.table("tipologie.csv", header = TRUE, sep = ",")
#trasformo la tabella in un vettore
tipi<- unlist(tipi)
#assegno la caratteristica ai nodi
V(rete)$type<- tipi

#imposto le dimensioni dei nodi
dim <- read.table("dimensioni.csv", header = TRUE, sep = ",")
dime<-unlist(dim)
#normalizzo
max_dim<-max(dime)
min_dim<-min(dime)
dim_grafo<-(((dime-min_dim)/(max_dim-min_dim))*2)+0.2)
dim_grafo<-signif(dim_grafo,2)

#lista con i valori di correlazione dell'albero
spessori<-archil*albero
#seleziono triangolo inferiore
spessori<- spessori[lower.tri(spessori,diag=F)]
#lo trasformo in vettore
spessori_vettore<-unlist(spessori)

#recupero i valori positivi e diversi da 1
spessori_edge<-spessori_vettore[spessori_vettore>0
spessori_vettore!=1]

#normalizzo i valori per renderli più distinguibili
massimo<-max(spessori_edge)
minimo<-min(spessori_edge)
spessori_grafo<-(((spessori_edge-minimo)/(massimo-
minimo))*2)+0.2)

#eliminazione nodi non legati

```

```

somme_colonne<-colSums(archil)
binario<- ifelse(somme_colonne>1,1,0)

V(rete)$t1<- binario
V(rete)$t1<- ifelse(V(rete)$t1,1,NA)

colori =c("service" = "steelblue", "transportation" = "tomato",
"food"="green", "shop"="orange",
"education"="purple", "health"="pink", "entertainment"="brown", "go
vernment"="yellow", "other"="blue")

# build chart
p <- ggnet2(rete,
            label= TRUE,
            size= rep(dim_grafo,1),
            color = tipi,
            palette= colori,
            na.rm = "t1",
            edge.size = rep(spessori_grafo,1))+ guides(size =
FALSE)

# view chart
p

# set image size
wimage <- 15
himage <- 10

# save chart
ggsave("grafo_pieno.jpg", height = himage, width = wimage)

```