

I visitor:

uno sguardo all'Area metropolitana di Firenze, Prato e Pistoia attraverso i dati di telefonia mobile

Valentina Marchi

Progetto Tirocinio IRPET
Anno Accademico 2016-2017



Master Universitario II livello
BIG DATA ANALYTICS E SOCIAL MINING

Indice

INTRODUZIONE	3
DATI	3
RISULTATI	4
ANALISI PRESENZA PER SINGOLA ANTENNA	5
FIRENZE	5
PRATO	7
PISTOIA	8
ANALISI ANTENNE AGGREGATE	9
FIRENZE	9
PRATO	10
PISTOIA	11
CO-PRESENZA DEI TURISTI NELL'AREA DI FIRENZE	12
CONCLUSIONI	16

Introduzione

«Il primo di luglio prossimo andrò all'estero, in Italia, dove resterò un po' di tempo», così annunciava Dickens nel maggio del 1844 all'amico Macvey Napier.

Era l'Europa dei Grand Tour e lo scrittore inglese si apprestava a compiere il suo viaggio alla scoperta dell'Italia. La corrispondenza epistolare avrebbe dato vita all'opera 'Lettere dall'Italia', raccolta di scritti in cui Dickens documentava il proprio viaggio.

Oggi, maggio 2018, lo scrittore al suo arrivo alla stazione di Santa Maria Novella a Firenze non dovrebbe far altro che, impugnare il proprio smartphone per comunicare a Macvey Napier il proprio arrivo nella città di Dante.

Se da un lato questo ci avrebbe privato di uno dei capolavori letterari del 1800, l'illustre scrittore, così come migliaia di altri turisti in arrivo nella culla del Rinascimento, collegandosi all'antenna GSM avrebbero lasciato l'impronta digitale del proprio passaggio.

L'utilizzo delle moderne tecnologie di comunicazione ci consente di disporre di un enorme mole di dati da raccogliere, analizzare e interpretare per comprendere meglio le dinamiche sociali, culturali ed economiche.

Se nel 1800 solo nobili e letterati potevano permettersi di affrontare viaggi per l'Europa a fine ludici e culturali, oggi in un mondo sempre più interconnesso milioni di persone ogni giorno si spostano per piacere, lavoro o necessità. Diventa pertanto di fondamentale importanza per istituzioni, aziende e policy maker approfondire la conoscenza delle dinamiche di tali flussi e degli effetti che essi hanno sul territorio urbano.

Il nostro lavoro ha come obiettivo quello di interpretare i dati GSM, raccolti nell'area metropolitana di Firenze, Prato e Pistoia, per identificare i comportamenti di coloro che, più avanti definiremo 'visitatori' con l'intento di contribuire al patrimonio informativo a disposizione delle istituzioni preposte al governo del territorio.

Dati

Negli anni recenti la diffusione dei dispositivi mobili ha portato ad un incremento del volume dei dati relativi a "dove" e "quando" le persone si spostano.

L'analisi è stata compiuta su un campione di dati GSM, ovvero dati di telefonia mobile raccolti dall'operatore telefonico durante l'uso del telefono cellulare con chiamate, SMS o traffico dati. Il fornitore dei dati è uno dei maggiori operatori nel settore delle telecomunicazioni in Italia. Per compiere le analisi di nostro interesse sono stati utilizzati i Call Data Record (CDR), ovvero record registrati durante le attività di chiamata e che contengono l'informazione temporale, un identificatore anonimo dell'utente, la cella telefonica e la durata della chiamata. Pertanto, essendo le attività dei CDR riferite a chiamate telefoniche, le analisi sono state compiute escludendo le attività di dati inerenti alla connessione Internet.

La nostra analisi vuole concentrarsi su flussi e spostamenti nell'area metropolitana di interesse da parte di soggetti che non sono ivi residenti. Per individuare nei nostri dati tali soggetti, è stato interpellato il KDDLab (ISTI – CNR, Pisa), il quale ha messo a disposizione

lo strumento del Sociometro, strumento di *data mining* che, attraverso l'analisi delle abitudini di chiamata degli utenti classifica gli stessi in categorie comportamentali.

La nostra analisi si concentra su coloro che, dal Sociometro sono classificati come "Visitor", ovvero coloro che non vivono e non lavorano/studiano nell'area di interesse, ma che occasionalmente la visitano.

Il campione risultante è composto da circa 9 milioni di record, nel range temporale che va dal 05 Ottobre al 30 novembre 2015.

User_id	Data	Ora	Antenna
---------	------	-----	---------

Per ciascuna antenna presa in esame disponiamo delle coordinate geografiche (latitudine e longitudine) che abbiamo utilizzato per le rappresentazioni su mappa.

Risultati

In questa sezione presentiamo i risultati ottenuti dai dati a nostra disposizione, cercando di rispondere alle domande: come si muovono i visitatori nell'area metropolitana di Firenze, Prato e Pistoia? Esistono differenze fra centro e periferia? Come cambia la fruizione del territorio a seconda del giorno o dell'orario?

In questo lavoro, la presenza sul territorio di un visitatore in uno specifico luogo è determinato dal fatto, che questi si colleghi ad una certa antenna telefonica.

Il limite di tali dati è quello di non fornire una precisa collocazione geografica, ma di indicarci la presenza in un'area, non consentendo alla nostra analisi di scendere a un livello di maggior dettaglio.

Nei seguenti paragrafi si propongono analisi che aiutino a comprendere la mobilità di tali utenti nell'area metropolitana di Firenze, Prato e Pistoia, quali:

1. analisi per ogni antenna della presenza media oraria nei diversi giorni della settimana, al fine di identificare differenze di comportamento dei visitatori e la presenza di eventi eccezionali;
2. Abbiamo ripetuto l'analisi aggregando le antenne sulla base di caratteristiche qualitative comuni, quali la vicinanza geografica, la vicinanza ad arterie di comunicazione, la presenza di attrattori, seguendo criteri suggeriti da esperti di dominio;
3. Ci siamo infine concentrati sulla sola area di Firenze, per rappresentare su mappa le aree 'toccate' da ciascun visitatore durante la propria permanenza nella città (grafo di co-occorrenza).

Analisi presenza per singola antenna

L'obiettivo è quello di fornire una panoramica a livello statistico delle presenze medie orarie e giornaliere degli utenti identificati dal sociometro come *visitatori*.

L'analisi è stata compiuta, attraverso lo strumento 'R', un linguaggio di programmazione e un ambiente di sviluppo specifico per l'analisi statistica dei dati. Sono state oggetto di analisi solo le antenne con il maggior numero di presenze, ovvero il maggior numero di visitatori che si sono agganciati a quell'antenna.

Di seguito riportiamo i grafici denominati '*Call intensity*', che rappresentano la media oraria delle chiamate, divise per giorno della settimana. Tale rappresentazione consente di verificare le differenze nell'utilizzo del territorio da parte dei visitatori della città.

Per esemplificare abbiamo riportato per ciascuna area in esame il grafico relativo alle antenne in cui le differenze sono maggiormente evidenti.

Firenze

Per l'area di Firenze abbiamo selezionato le antenne:

- FI001 centro storico della città
- FI023 area stadio comunale Artemio Franchi
- FI033 area di Novoli
- FI037 area Cascine

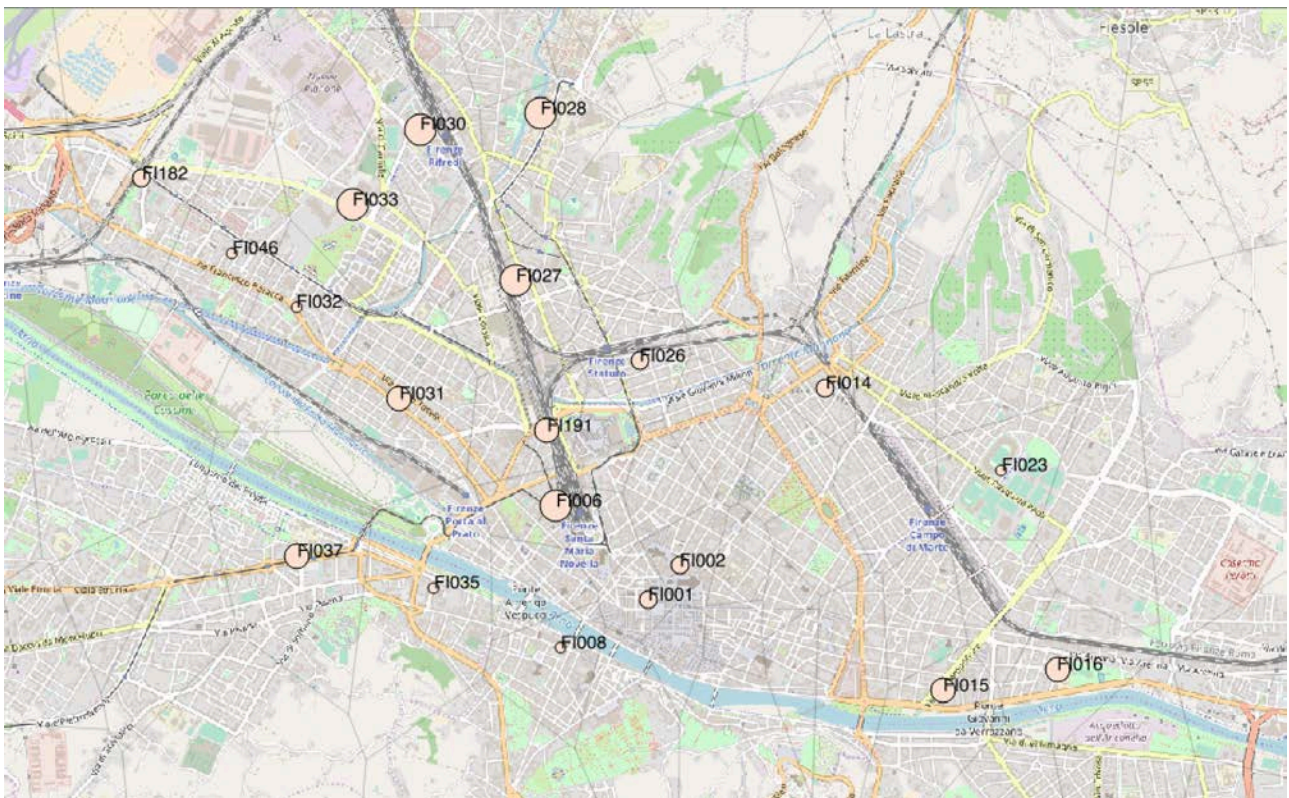


Figura 1 Top antenne di Firenze

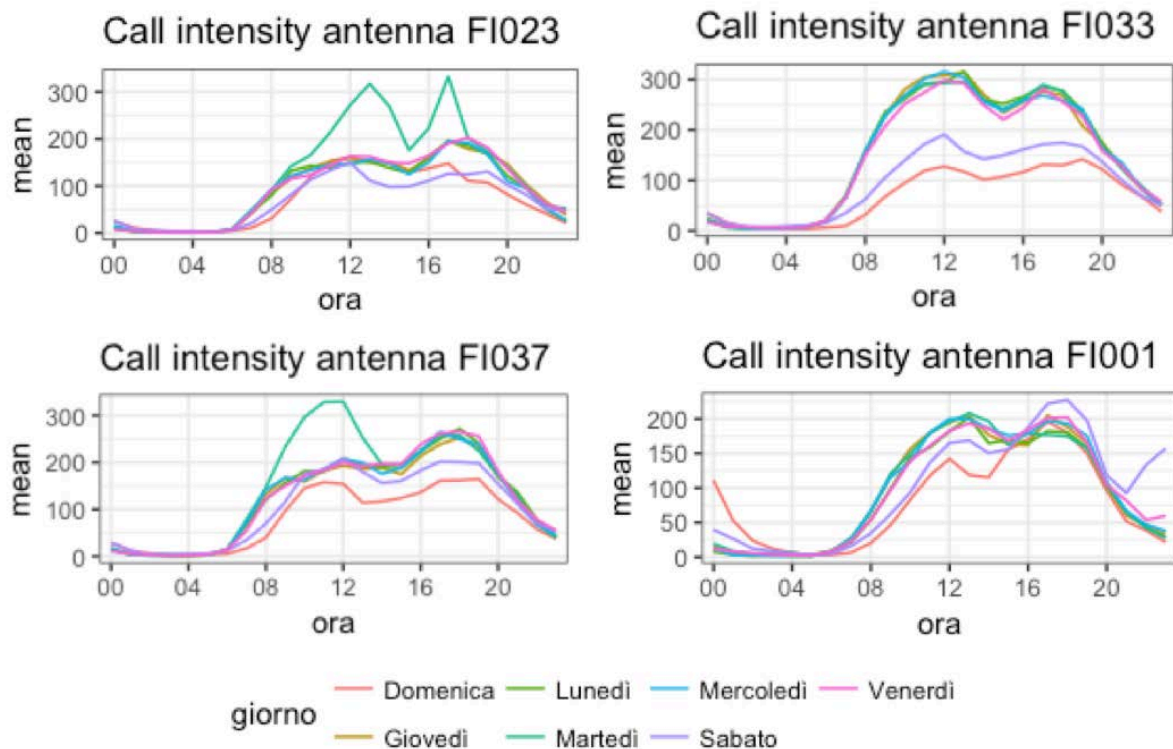


Figura 2 Call Intensity - Focus antenne Firenze

I quattro grafici mostrano differenze rispetto alle presenze di visitatori, sia per quanto riguarda i giorni della settimana che, in relazione all'orario considerato.

Nella zona del centro storico (FI001) non sembrano esserci differenze rilevanti nella presenza di visitatori fra i giorni lavorativi e week-end. È interessante notare che nel fine settimana, a differenza di altre aree, si registrano presenze di visitatori in tarda serata/notte, probabilmente trattasi di residenti al di fuori dell'area metropolitana che, raggiungono i locali del centro di Firenze per trascorrere le serate del fine settimana.

L'area di Novoli (FI033), sede sia del polo universitario che di uffici, mostra un'importante differenza fra le presenze nei giorni lavorativi e il fine settimana.

Per l'area delle Cascine (FI037) il grafico evidenzia un andamento che differisce rispetto alle altre aree nella mattinata del martedì. L'area infatti è sede del più grande mercato all'aperto di Firenze che si tiene tutti i martedì mattina.

L'antenna relativa alla zona dello stadio (FI023) riporta un andamento inatteso nel giorno di martedì. Ad una più attenta analisi tale scostamento è da imputarsi alla visita pastorale del Santo Padre tenutasi martedì 10 novembre 2015, che ha rappresentato un evento eccezionale, richiamando per l'occasione 50 mila persone durante la Santa Messa.

Prato

Per l'area di Prato abbiamo selezionato le antenne:

- PO007 centro storico della città
- PO011 SS719

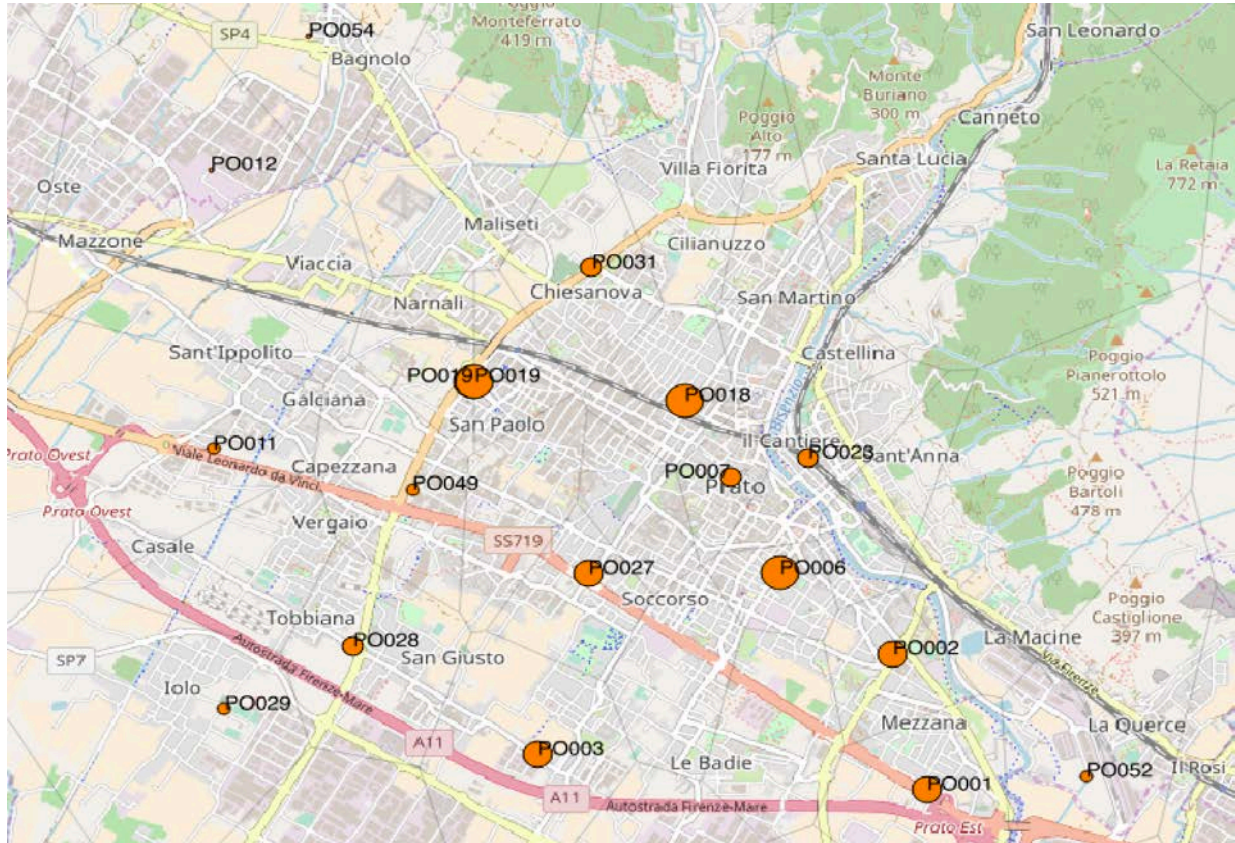


Figura 3 Top antenne di Prato

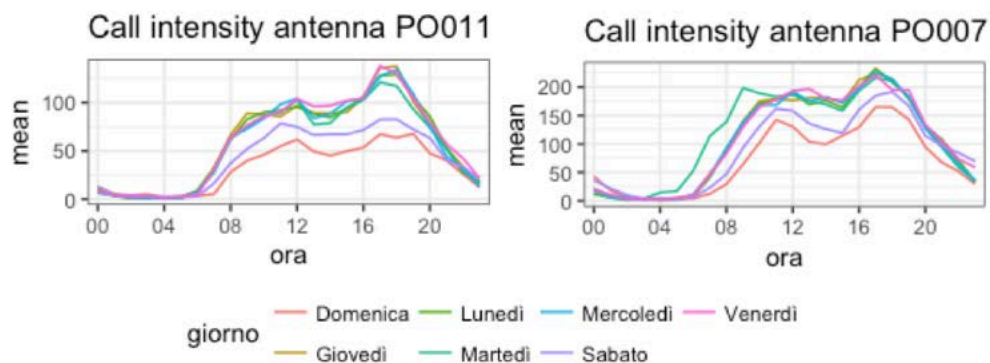


Figura 4 Call Intensity – Focus antenne Prato

La zona del centro storico di Prato (PO007) è anch'essa stata interessata dalla visita Pastorale del Santo Padre, nella primissima mattinata di martedì 10 novembre 2015. Dai dati riportati dal comune di Prato si stima fossero presenti 30 mila persone, 5 mila solo in Piazza Duomo.

Dall'area di grande comunicazione (PO011) si può osservare nei giorni lavorativi il picco di presenze dalle 16 alle 19 circa, presumibilmente l'orario di maggior congestione del traffico, specie per immettersi sull'autostrada A11.

Pistoia

Per l'area di Pistoia abbiamo selezionato le antenne:

- PT005 centro storico della città di Pistoia
- PT015 centro storico Montecatini Terme

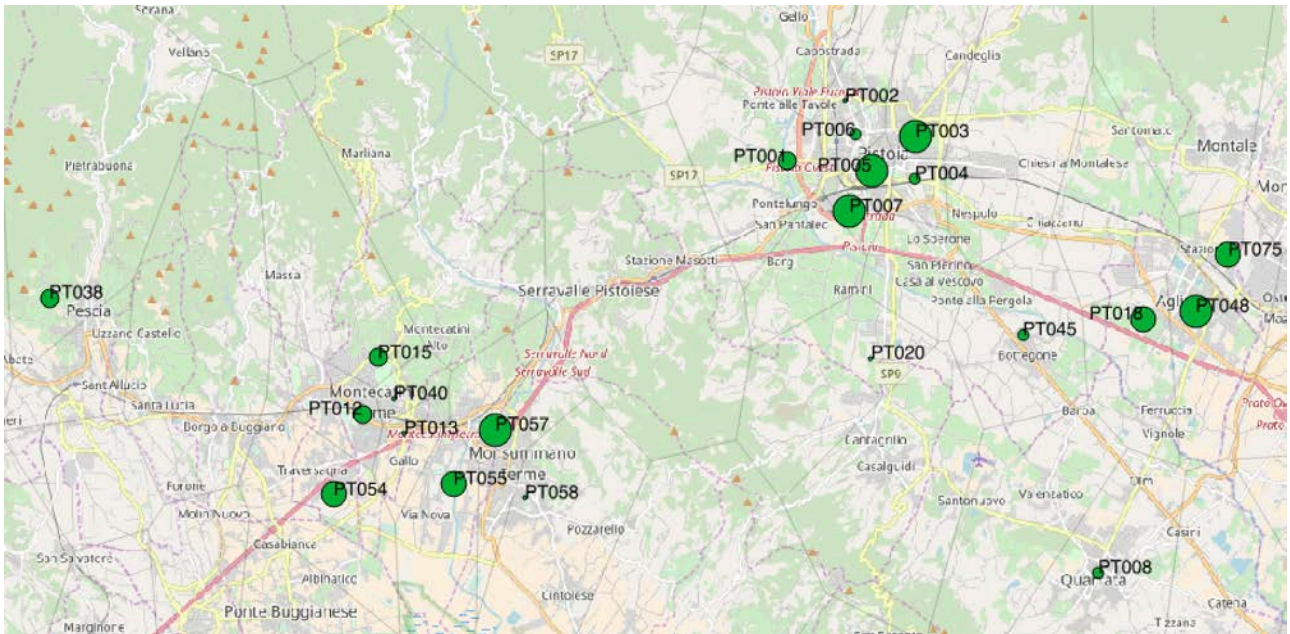


Figura 5 Top antenne Pistoia

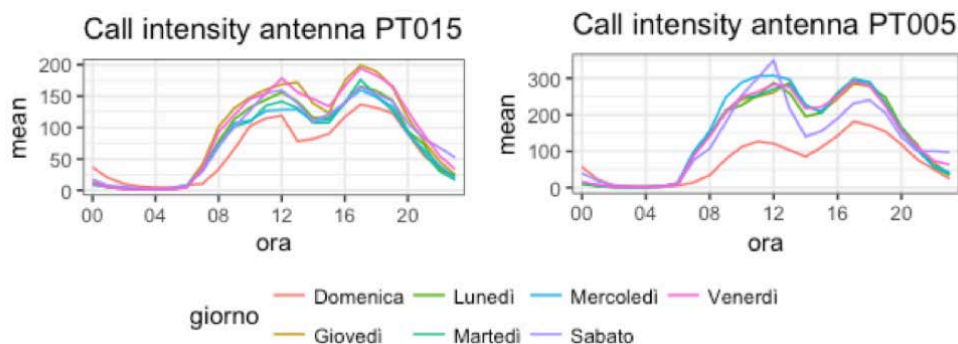


Figura 6 Call Intensity - Focus antenne Pistoia

Il centro storico di Pistoia (PT005) mostra un picco di presenze il sabato, specie nella mattina, in coincidenza del mercato settimanale.

Di particolare interesse, l'area del centro di Montecatini Terme (PT015), località a forte vocazione turistica, in cui contrariamente alle altre aree non si osservano comportamenti diversi fra i giorni lavorativi e il fine settimana, possibile evidenza che i visitatori della città si identificano con la categoria dei turisti che fanno base a Montecatini Terme per visitare la Toscana.

Analisi antenne aggregate

In considerazione del gran numero di antenne analizzate, si è scelto di aggregare le antenne sulla base di caratteristiche qualitative, quali la vicinanza geografica e l'appartenenza ad aree in cui si collocano specifiche attività socio-economiche (i dettagli dell'aggregazione sono specificati in appendice).

Abbiamo poi rappresentato tramite *heatmap* la 'Call Intensity' media oraria per giorno della settimana.

Firenze

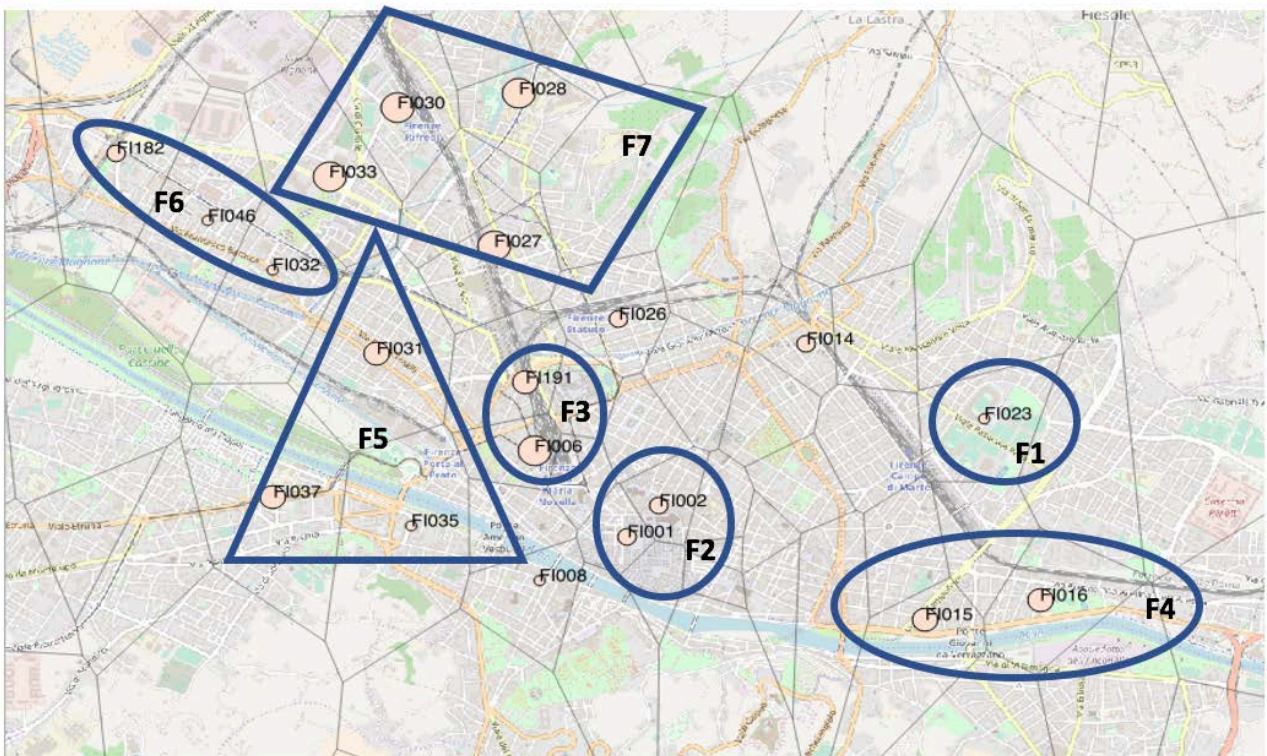


Figura 7 Antenne aggregate a Firenze

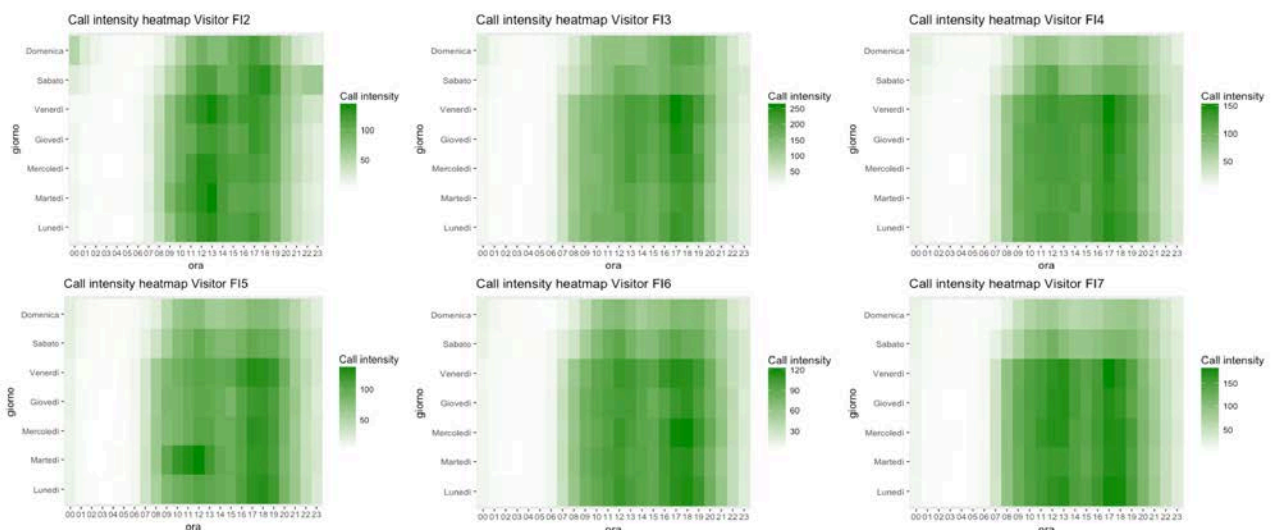


Figura 8 Heatmap Visitor Firenze

Prato

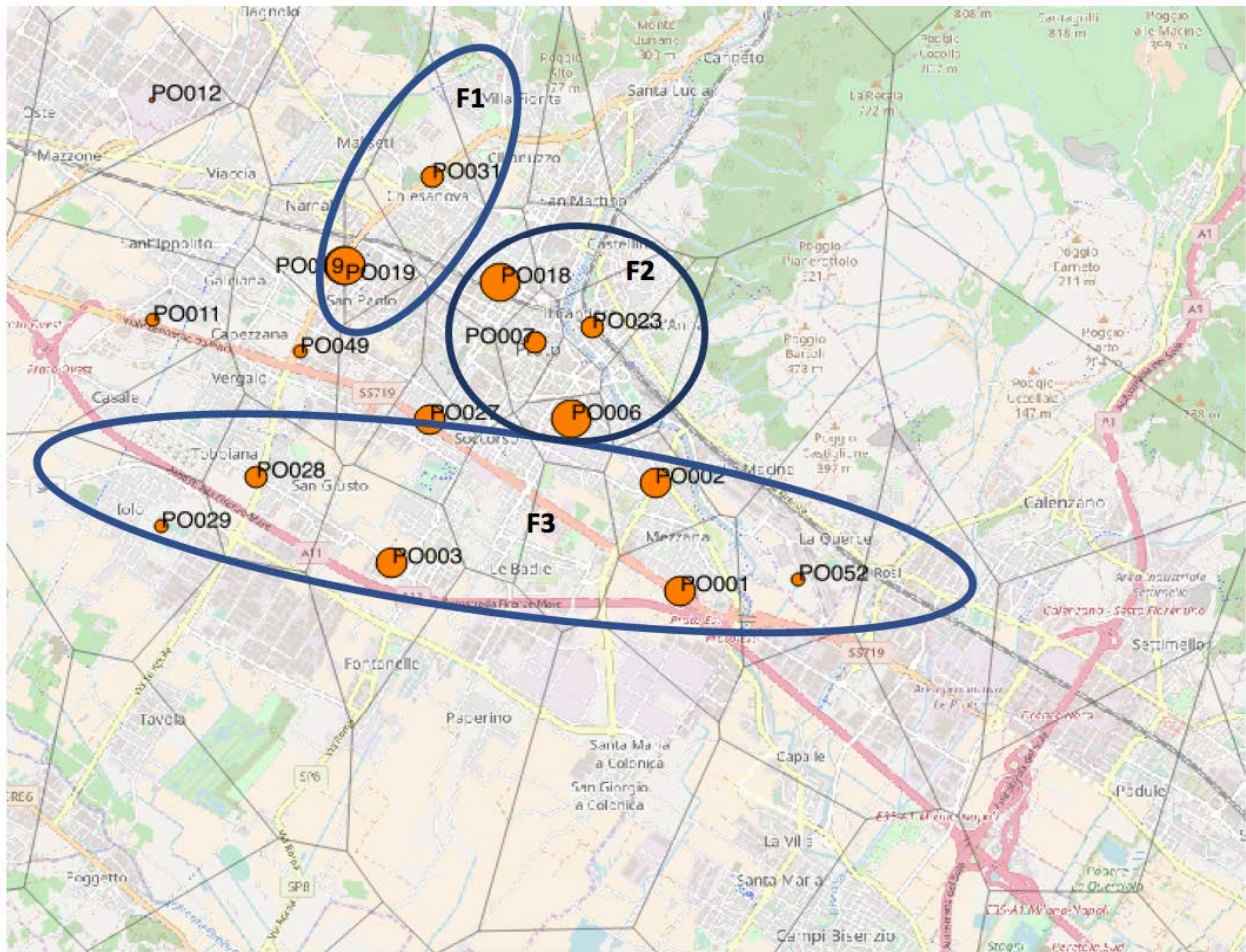


Figura 9 Antenne aggregate Prato

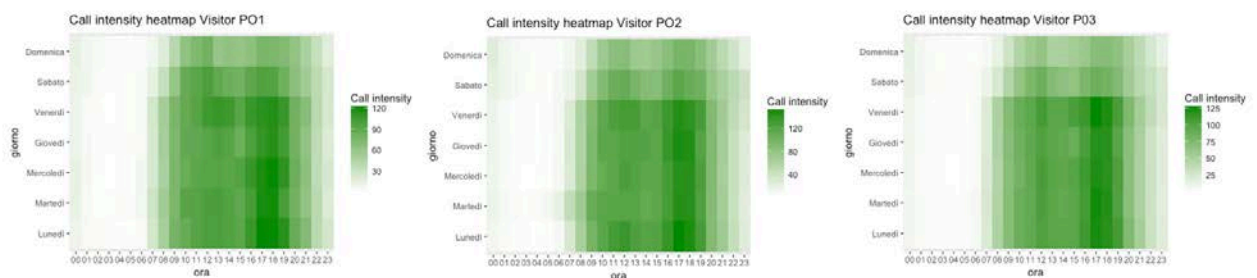


Figura 10 Heatmap Visitor Prato

Pistoia

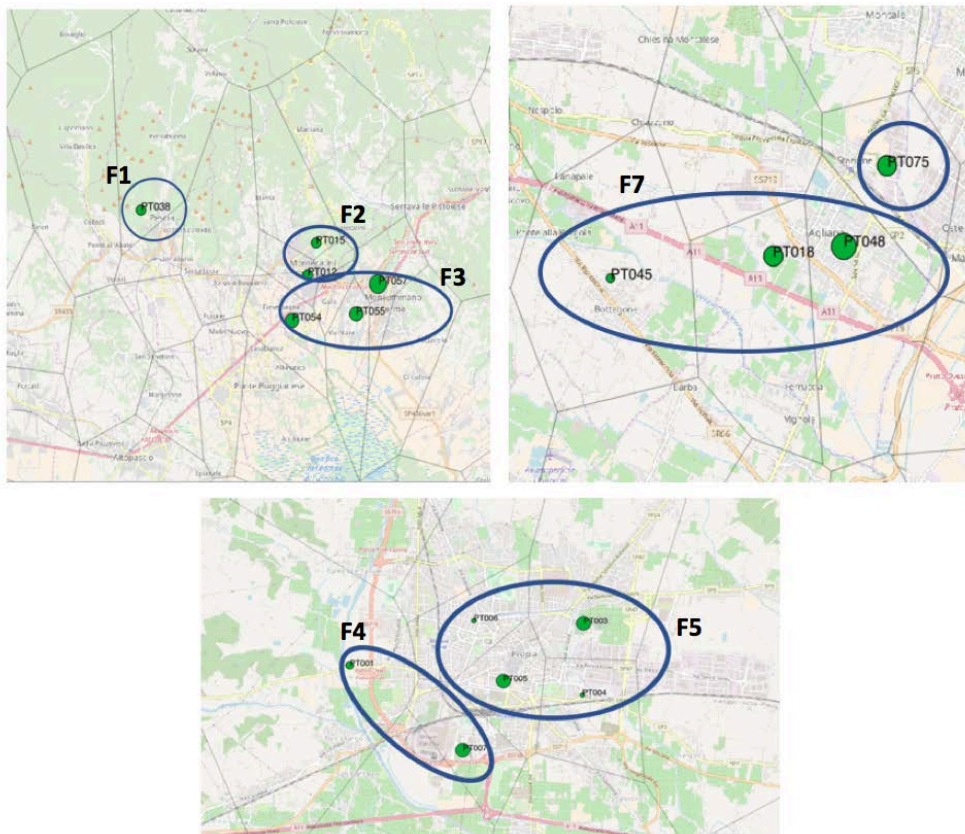


Figura 11 Antenne aggregate Pistoia

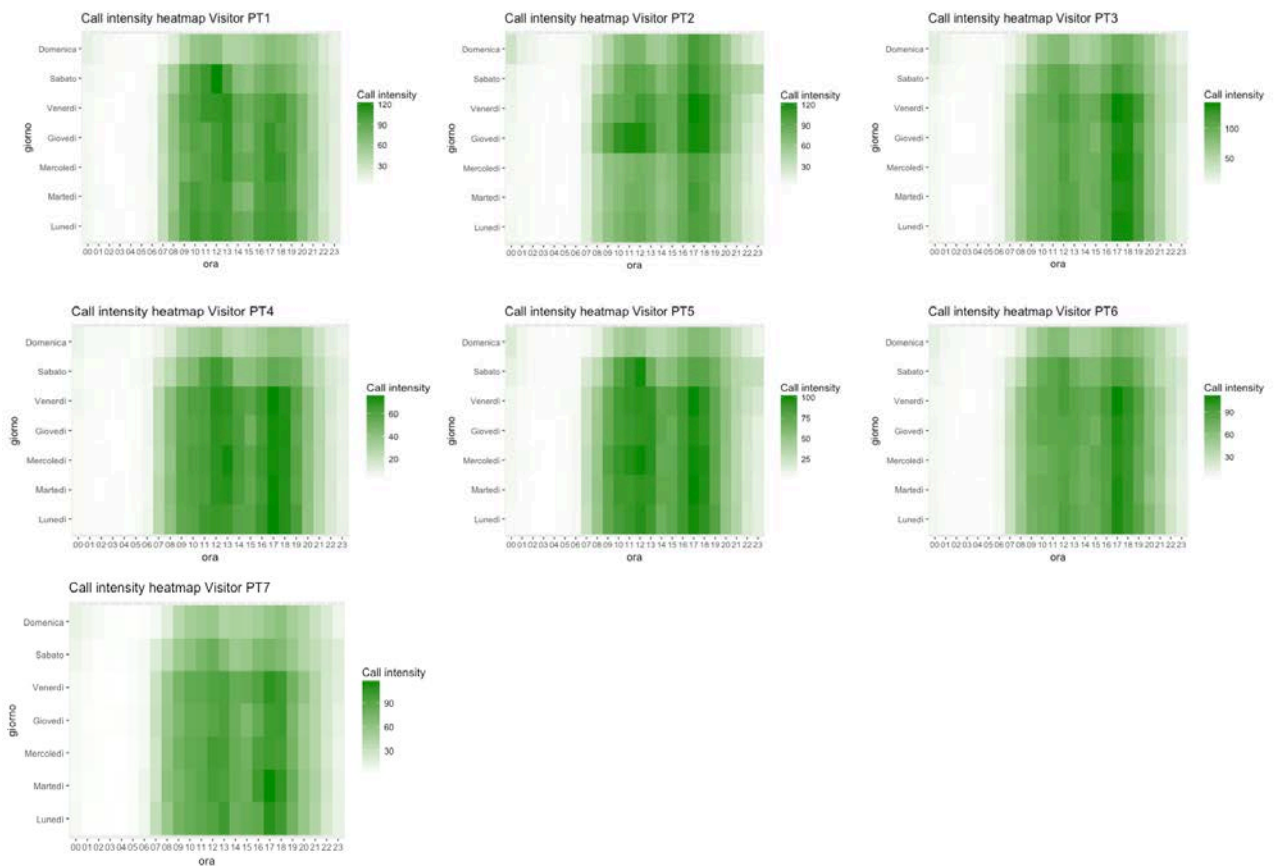


Figura 12 Heatmap visitor Prato

Co-presenza dei turisti nell'area di Firenze

Per comprendere il comportamento degli utenti classificati come *visitatori*, nell'area di Firenze, è stata compiuta un'analisi per individuare le aree della città 'toccate' dai visitatori durante la loro permanenza.

Sono stati utilizzati i dati relativi al mese di Ottobre 2015 e per ciascun utente sono state individuate le antenne a cui si sono collegati durante il loro soggiorno.

Le osservazioni sono state rappresentate tramite un grafo di co-occorrenza, i cui nodi sono le antenne unite fra di loro, sulla base del fatto che un determinato utente, nel proprio periodo di permanenza, si sia collegato ad esse. Il peso degli archi del grafo è rappresentato dal numero di utenti che si sono agganciati ad entrambe le antenne durante il proprio soggiorno a Firenze.

È stato deciso di verificare se il comportamento dei visitatori e l'utilizzo dello spazio urbano di questi, cambi all'aumentare del periodo di permanenza e se all'interno della città si possono individuare aree all'interno delle quali il comportamento del visitatore risulti omogeneo. Per indagare tali aspetti è stata effettuata una partizione del grafo massimizzando la *modularity (Louvain Method)*, che ci ha permesso di misurare come la rete possa essere scomposta in comunità, che consentono di raggruppare le antenne nella città Firenze.

Ogni community è stata rappresentata con un colore diverso, i colori attribuiti su ciascun grafo rappresentano la numerosità della comunità stessa (ad esempio il colore blu rappresenta sempre la comunità con la dimensione maggiore).

Di seguito rappresentiamo i grafi ottenuti, considerando il periodo di permanenza di almeno 1 giorno e almeno 5 giorni.

¹ Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10), P1000.

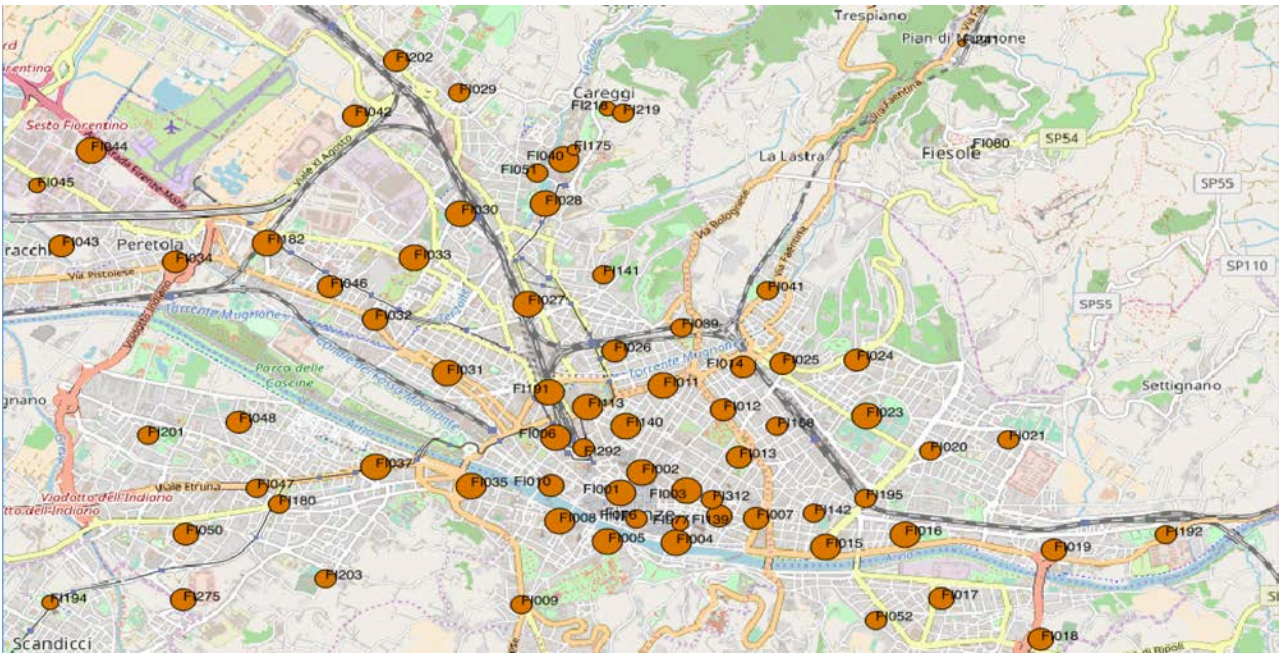


Figura 13 Count presenza antenne 1 giorni

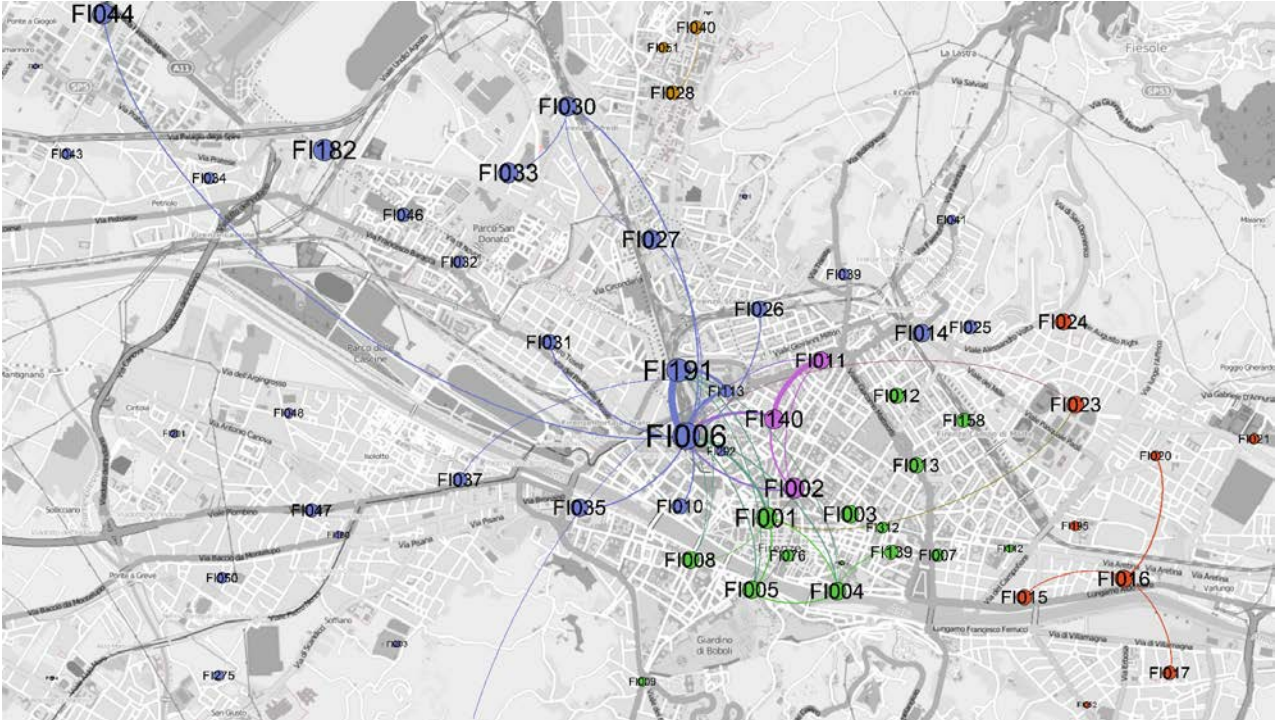


Figura 14 Grafo visitor 1 giorno

Nodes: 139
Edges: 1712
Undirected graph
Modularity: 0,253
Modularity with resolution: 0,253
Number of Communities: 15

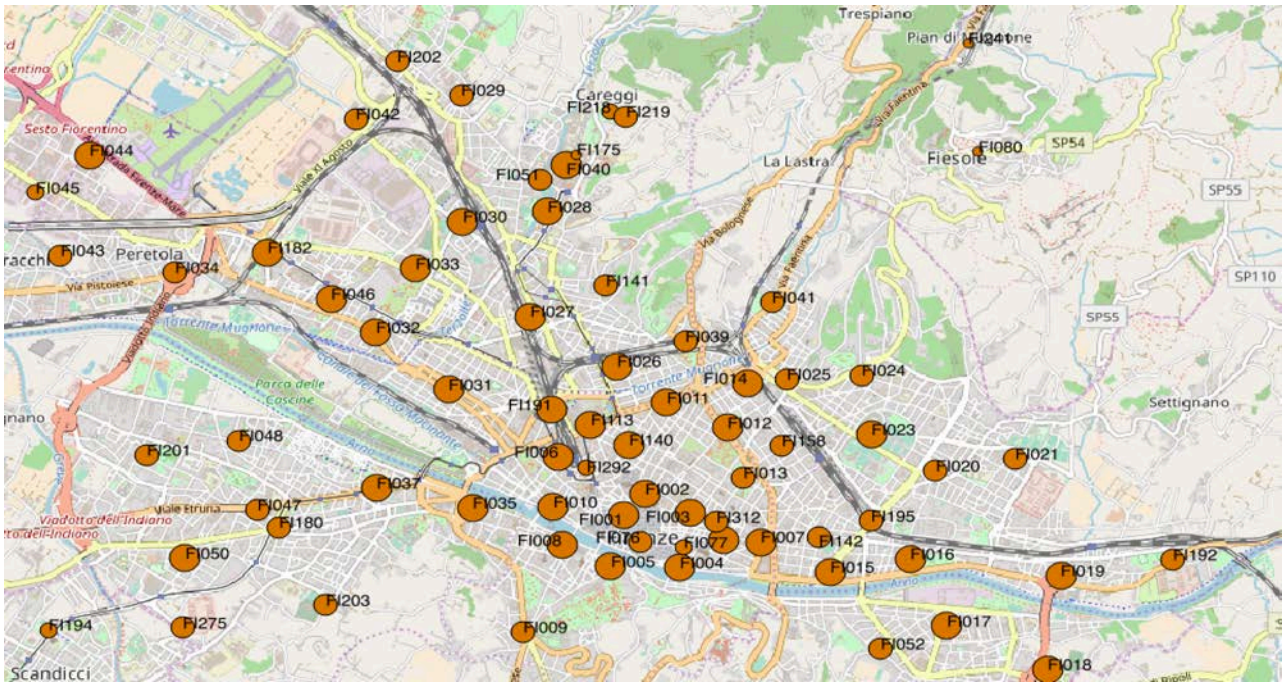


Figura 15 Count presenza antenna 5 giorni

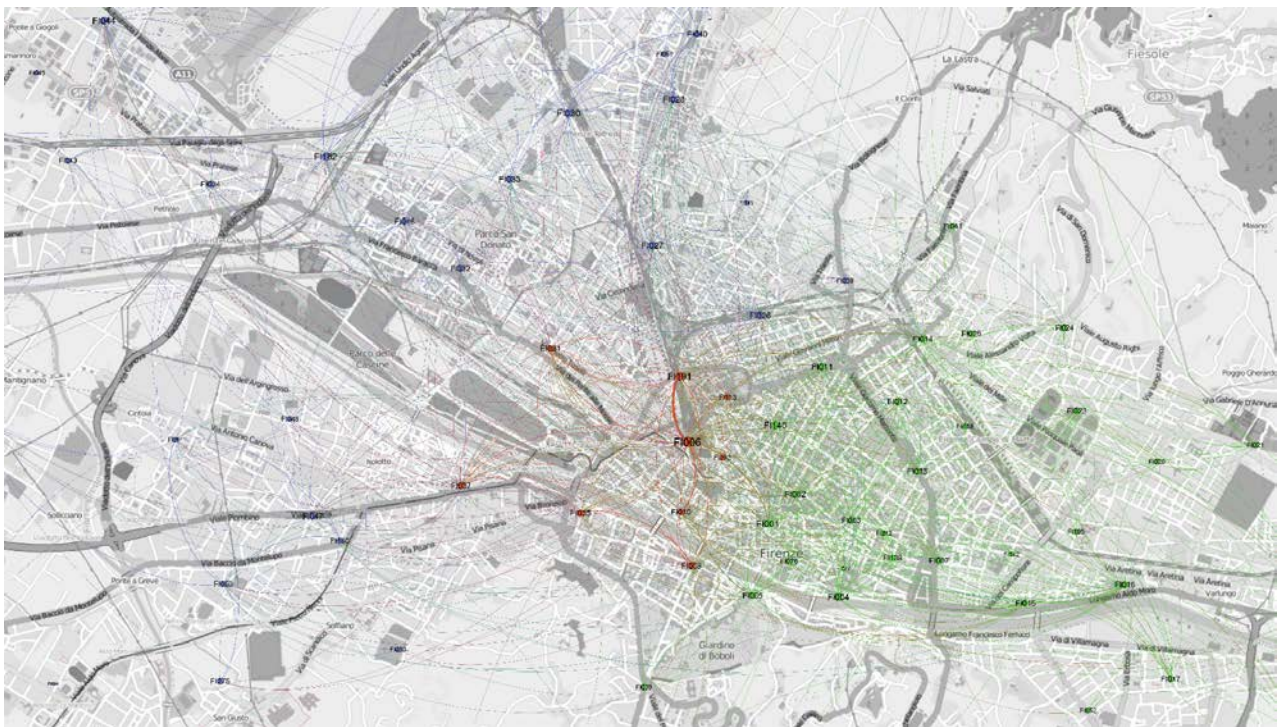


Figura 16 Grafo visitor 5 giorni

Nodes: 155
 Edges: 3814
 Undirected graph
 Modularity: 0,232
 Modularity with resolution: 0,232
 Number of Communities: 6

La scelta di ripetere l'osservazione su periodi di permanenza incrementali (da 1 a 5 giorni) è funzionale a verificare se al prolungarsi della permanenza in città, il comportamento dei visitatori cambi e vengano coinvolte aree urbane diverse.

I grafi ottenuti sembrano confermare l'ipotesi per cui, al crescere della durata del soggiorno, cresca di conseguenza l'estensione urbana coinvolta dal flusso dei visitatori.

Pare che i visitatori che soggiornano per un periodo più breve, concentrino i propri spostamenti nel centro storico, nei dintorni della stazione e delle principali attrazioni turistiche. Al crescere della permanenza aumentano gli spostamenti che coinvolgono zone più periferiche di Firenze.

Come si può osservare dai grafi, un punto nevralgico della città è la stazione di Santa Maria Novella (antenna FI006) a testimonianza del fatto che questa ricopre un ruolo di porta di accesso per molti visitatori. La stazione risulta il nodo con il grado maggiore, ma allo stesso tempo si può notare che, al crescere della durata della permanenza, l'antenna della stazione non risulta più appartenente alla community di dimensioni maggiori.

Questo a parziale conferma, che al crescere della durata di permanenza, vengono coinvolte dal flusso dei visitatori aree diverse.

L'analisi di *community detection* pare individuare aree della città in cui i visitatori potrebbero appartenere a categorie diverse. L'area del centro storico, fortemente collegata alla stazione centrale, per la quale i principali fruitori sono visitatori che si soffermano e risultano interessanti alle principali attrazioni turistiche della città.

L'area di Firenze sud/Campo di Marte in cui si collocano i principali luoghi destinati ad eventi, come ad esempio lo stadio Artemio Franchi, l'OBI Hall e il Mandela Forum, in cui i visitatori probabilmente appartengono alla categoria di coloro che si recano a Firenze in occasione di specifici eventi.

L'area di Firenze Nord in cui si collocano sia la principale struttura ospedaliera di Careggi, che i poli Universitari, il Nuovo Palazzo di Giustizia e il principale accesso autostradale – A11, che porta a ipotizzare che i visitatori appartengono a categorie che si recano a Firenze principalmente per motivi di lavoro/studio o motivi connessi alla presenza del polo ospedaliero.

Conclusioni

Il lavoro svolto si propone di dimostrare come i dati generati attraverso l'utilizzo dei dispositivi mobili, possa rappresentare un valido strumento per la comprensione del comportamento dei visitatori, come essi si spostino sul territorio nei diversi giorni della settimana e in fasce orarie diverse, come il comportamento possa variare in funzione delle aree coinvolte e delle caratteristiche socio-economiche di queste.

I dati GSM rappresentano solo un esempio, essi potrebbero essere integrati, con quelli provenienti da dispositivi dotati di GPS, che possono fornire un livello di granularità molto fine (ad esempio a livello di singola attrazione turistica). Dobbiamo comunque considerare che i dati GSM raccolgono una popolazione diversa rispetto ai GPS, permettendo di vedere l'uso della città a prescindere dai vincoli della rete stradale.

Anche i dati pubblici provenienti dall'utilizzo dei Social Network possono rappresentare sia una valida alternativa, in quanto meno costosi da reperire, sia un'integrazione in grado di fornire informazioni aggiuntive, ad esempio relative alle caratteristiche del visitatore (paese di provenienza, genere).

In aree urbane, che sempre più rappresentano poli di attrazione interessati da flussi di persone che si spostano per motivi di piacere, lavoro o necessità, diventa indispensabile approfondire la conoscenza e le dinamiche di tali fenomeni per poter elaborare le corrette policy di gestione, monitoraggio e governo del territorio.