

Proposta di misurazione della mobilità a partire da dati di traffico telefonico

Paolo Dilda (*), Paola Pucci, Fabio Manfredini, Paolo Tagliolato

(* Dipartimento di Architettura e Pianificazione – Politecnico di Milano, via Bonardi 3, 20133 – Milano
tel. +39 0223995449, fax +39 0223995435, e-mail paolo.dilda@polimi.it

La proposta di misurazione della mobilità urbana deriva dall'elaborazione di dati di traffico telefonico provenienti dalla rete mobile di Telecom Italia, nell'ambito di una ricerca svolta dal Dipartimento di Architettura e Pianificazione del Politecnico di Milano e finalizzata a sperimentare possibili applicazioni dei dati di traffico telefonico negli studi urbani e territoriali.

Sono stati utilizzati i seguenti dati:

Erlang: descrive la densità del traffico telefonico ogni 15 minuti per aree di dimensioni pari a 250 m x 250 m;

Handover: misura il trasferimento del segnale telefonico da un'antenna all'altra quando il cellulare attivo si muove. Si può considerare quindi una informazione potenzialmente in grado di descrivere la mobilità degli utenti telefonici.

L'indicatore proposto è costruito rapportando gli Handover con gli Erlang per ciascuna cella di una porzione del territorio della regione urbana milanese con la finalità di verificare le potenzialità di un suo utilizzo per descrivere e interpretare il grado di mobilità territoriale durante diversi intervalli orari della giornata, nei giorni feriali e in quelli festivi.

In the paper we propose to measure urban mobility by using mobile phone traffic data, provided by Telecom Italia in the framework of a research carried out by the Department of Architecture and Planning, Politecnico di Milano, aimed at testing potential applications of mobile phone data to urban studies. Were used the following data:

Erlang, a measure which describes the mobile phone activity as a function of position and time, recorded at a spatial resolution of about 200 meters, every 15 minutes.

Handover, a process which happens when the phone is moving away from the area covered by one cell and entering the area covered by another cell. It may therefore represents a potential information able to describe the mobility of telephone users.

We therefore present a spatial indicator which compares the Handover with the Erlang measures for a portion of the Milano urban region in order to evaluate its ability to describe and interpret mobility patterns and intensity during different time slots of the day.

Negli ultimi anni sono stati sviluppati progetti di ricerca finalizzati a comprendere se e come i dati di traffico telefonico possano essere utilizzati come strumenti di analisi e di rappresentazione delle dinamiche urbane e di studio del movimento degli individui.

In questi studi gli utenti che generano traffico telefonico possono essere considerati come una rete di sensori in grado di fornire informazioni utili sulle forme e sui modi d'uso del territorio, difficilmente ottenibili da altre fonti di dati tradizionalmente utilizzate negli studi urbani, come ad esempio i censimenti, le indagini telefoniche o il dispiegamento di reti di sensori, che presentano limitazioni dovute ai costi elevati e allo scarso livello di aggiornamento.

In questo contesto, obiettivo della ricerca è sperimentare possibili applicazioni dei dati di traffico telefonico disponibili nella rete radiomobile Telecom, nell'ambito di studi urbani e territoriali.

La ricerca utilizza, come dati di traffico telefonico della rete mobile di Telecom Italia:

Erlang che descrive la densità del traffico telefonico ogni 15 minuti per aree di dimensioni pari a 250 m x 250 m. L'informazione relativa al traffico telefonico in Erlang registrato da ogni antenna distribuita sul territorio, è stata attribuita a ciascuna cella¹, attraverso un processo di interpolazione del dato effettuato da Telecom.

In merito al dato espresso in Erlang, che restituisce l'intensità di traffico telefonico nell'unità di tempo, è utile ricordare che questa tipologia di informazione non consente una diretta correlazione tra intensità del traffico telefonico e numero di utenti che lo generano. Bisogna inoltre considerare che le modalità d'uso del telefono cellulare variano, anche consistentemente, con l'età degli utenti e con le loro caratteristiche socio-professionali.

Il traffico che viene registrato rappresenta inoltre un campione - poiché gli utenti di Telecom sono circa il 40% del mercato, su scala nazionale, e di essi, solo coloro che generano traffico sono registrati dal dato Erlang - peraltro fortemente orientato su alcune età e condizioni socio-professionali.

Handover che misura il trasferimento del segnale telefonico da un'antenna all'altra quando il cellulare attivo si muove. Il dato presenta problemi di trattamento in quanto non è semplice associare il numero di handover fra cella e cella agli spostamenti degli utenti fra zone diverse, sia perché nel passaggio fra celle ci possono essere handover multipli, sia perché è possibile che avvengano handover fra celle anche in assenza di spostamento dell'utente (es. utente fermo nell'area di sovrapposizione fra due celle), sia perché la linea di confine fra due celle non è ben delineata ed in generale il passaggio può avvenire in ciascun punto dell'area di confine e parziale sovrapposizione delle due celle. Queste condizioni, unitamente all'assenza di applicazioni rilevate in letteratura, anche a livello sperimentale o di prototipo, per questa tipologia di dato nell'ambito degli studi sul traffico, ci ha portato a proporre approfondimenti possibili in riferimento alla utilizzabilità dei dati, piuttosto che a condurre sperimentazioni applicative con i dati stessi.

La necessità di trattare una grande mole di dati e di mettere in comunicazione diverse fonti informative, ha portato alla costruzione di un apposito sistema informativo. Questo, come sinteticamente schematizzato in figura 1, è articolato in quattro moduli intercomunicanti tra loro grazie a un modulo centrale ("piattaforma informatica" nello schema) usato per le elaborazioni.

Il modulo "dati telecom" rappresenta i dati telefonici in ingresso pervenuti da Telecom. Essi sono distinguibili in due categorie: dati organizzati in insiemi di tuple (record con attributi); serie temporali di mappe *raster* georeferenziate.

Il secondo modulo di *input* ("dati territoriali") rappresenta una varietà di basi di dati socio-demografici e di layer informativi territoriali.

Le uscite delle elaborazioni sono rappresentate nel modulo "output". Le restituzioni sono eterogenee: statistiche, grafici di vario tipo, mappe tematiche georeferenziate.

Al cuore del sistema, come anticipato, si trova un modulo capace di immagazzinare ed elaborare i dati dalle diverse fonti. Tale modulo è a sua volta costituito da alcuni sottosistemi impiegati per diversi compiti:

- un database relazionale, implementato su piattaforma Microsoft SQL 2005 Express e utilizzato per la memorizzazione e l'organizzazione degli insiemi di tuple provenienti da Telecom e per la loro elaborazione;
- un sistema di Data Mining, implementato in Matlab e corredato, all'interfaccia con i dati telecom, da una serie di algoritmi sviluppati in linguaggio AWK per pretrattare le serie di mappe georeferenziate in modo da ottimizzarne l'importazione;
- alcuni applicativi GIS, commerciali (suite ESRI) e open source (GRASS), usati sia per l'elaborazione delle mappe tematiche in output al sistema, sia per estrarre dati territoriali da fornire al database relazionale e al sistema di Data Mining.

¹ Ogni cella ha una dimensione di 0.002777778 gradi di longitudine e di 0.002083333 gradi di latitudine.

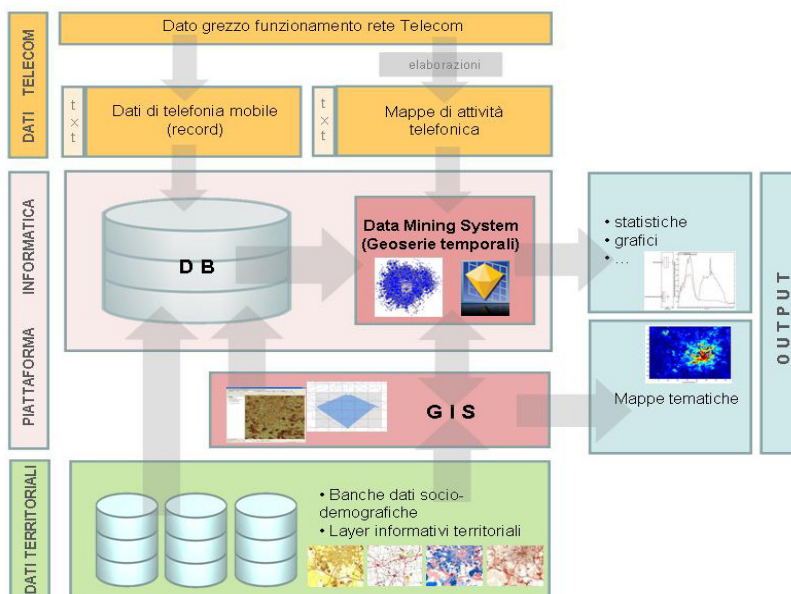


Figura 1 – Schema della struttura del sistema informativo.

La strumentazione GIS è stata utilizzata principalmente per le seguenti operazioni:

- georeferenziazione e riproiezione dei dati territoriali provenienti da diverse fonti (dati ISTAT, basi dati Regione Lombardia, ecc.);
- elaborazione di layer informativi territoriali sia di carattere qualitativo che di carattere quantitativo;
- rasterizzazione delle informazioni geografiche, finalizzata sia alla rappresentazione dei risultati (output) sia all'analisi dei dati;
- Operazioni di interpolazione (attraverso metodologia IDW) di dati territoriali puntuali al fine di rappresentare fenomeni quantitativi attraverso mappe tematiche per aree del territorio preso in esame.

In relazione alle caratteristiche intrinseche dei dati disponibili da fonte Telecom (Erlang e Handover) - in riferimento sia alla loro parametrizzazione e alla capacità di esprimere una misura, sia ai condizionamenti legati alla loro utilizzabilità rispetto alla descrizione di fenomeni d'uso del territorio - abbiamo ritenuto di interesse sperimentare un indicatore espresso come rapporto tra Handover ed Erlang.

Il parametro Handover misura il trasferimento del segnale telefonico da un'antenna all'altra quando il cellulare attivo si muove mentre il parametro Erlang misura la densità di traffico telefonico (chiamate) offerto dagli utenti o smaltito dalla rete per ogni cella. L'indicatore proposto è un tentativo di misurare gli spostamenti degli utenti in rapporto all'aspetto quantitativo rappresentato dal carico telefonico.

Il rapporto Handover/Erlang può esprimere quindi il grado di mobilità di una cella, che abbiamo definito "motilità" (Kaufmann, 2002) mutuando questo termine da indagini condotte sulle pratiche di mobilità².

² L'autore definisce infatti "motilità" la capacità a essere mobili, una disposizione alla mobilità, un potenziale di mobilità come intensità e stile della nostra mobilità che è condizionato dalle capacità di ognuno di avere accesso e di appropriarsi delle possibilità offerte da un sistema socio-economico e territoriale dato.

L'indicatore Handover/Erlang è stato costruito sui valori aggregati e medi riferiti ad un periodo temporale compreso tra il 7 ed il 20 settembre 2009; in riferimento a questo periodo sono stati presi i valori medi dei giorni feriali e domenica. Una prima sperimentazione di lettura dei risultati è stata condotta su un ambito territoriale di riferimento avente come baricentro l'area metropolitana milanese comprendente un totale di oltre 2000 antenne.

In questo caso, i dati di Handover e Erlang utilizzati per costruire il rapporto Handover/Erlang, sono riferiti entrambi al dato delle antenne e solo successivamente, il rapporto Handover/Erlang è stato territorializzato attraverso una procedura con strumentazione GIS. A partire dalle tabelle dei dati relative a ciascuna antenna, è stata effettuata una operazione di *join* con il dato territorializzato della localizzazione delle antenne (in formato *shape*) ed una interpolazione con metodo IDW (*Inverse Distance Weighting*), che produce una superficie di valori dell'indicatore in formato *grid* (con celle di dimensioni di circa 490 x 490 m.) che copre in modo continuo tutto il territorio in questione.

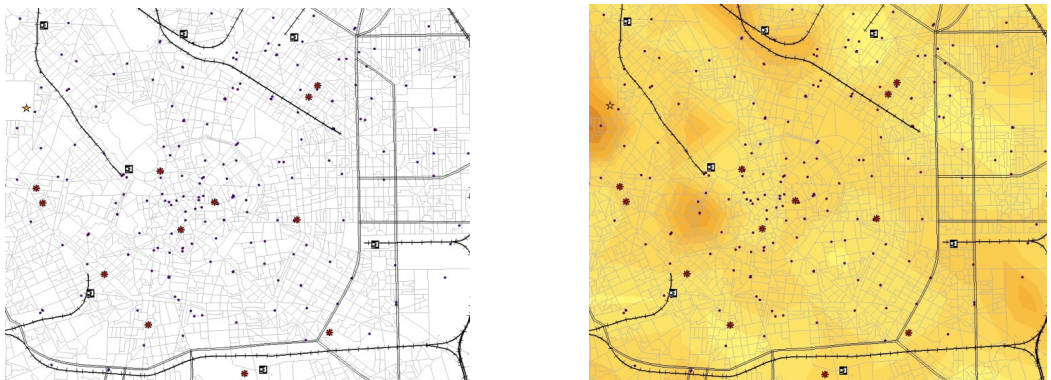


Figura 2 – A sinistra la localizzazione delle antenne; a destra il risultato dell'interpolazione che identifica le rispettive aree di influenza in proporzione ai valori per ciascuna antenna.

Le elaborazioni di seguito proposte restituiscono, per un ambito discreto della regione Lombardia, particolarmente significativo, poiché comprendente la parte centrale della Regione urbana milanese, l'andamento del rapporto Handover/Erlang per alcune fasce orarie nel giorno festivo tipo e in un giorno feriale tipo.

E' utile premettere che il valore di Erlang è molto differente in termini di valori assoluti tra le due giornate: si registra un picco di valore oltre 3.500 di Erlang per il giorno feriale a fronte di un picco di 435 di Erlang domenicale. La lettura del dato Erlang in riferimento al giorno festivo ed al giorno feriale mette in evidenza i seguenti fenomeni e le seguenti differenze nell'andamento:

	ERLANG DOMENICA	ERLANG FERIALE
VALORI	Valori relativamente più bassi	Valori relativamente più elevati
ANDAMENTO	<ul style="list-style-type: none"> Aumento dei valori nella fascia oraria compresa tra le ore 9.00 e le ore 13.00 L'andamento e l'assetto territoriale non è influenzato dai flussi pendolari (rimane costante sino alla tarda serata (ore 21.00)) 	<ul style="list-style-type: none"> Aumento dei valori nella fascia oraria compresa tra le ore 9.00 e le ore 13.00 L'andamento e l'assetto territoriale presenta valori relativamente più bassi nel centro città a partire dalle ore 17.00 (fenomeno di progressivo spopolamento)
FENOMENI PARTICOLARI	Presenza di luoghi urbani con valori alti: fiera, strutture sportive, strade commerciali, ecc.	Presenza di luoghi urbani con valori alti: fiera, strutture sportive, strade commerciali, ecc.

Figura 3 – Sintesi del confronto fra giorno festivo e giorno feriale.

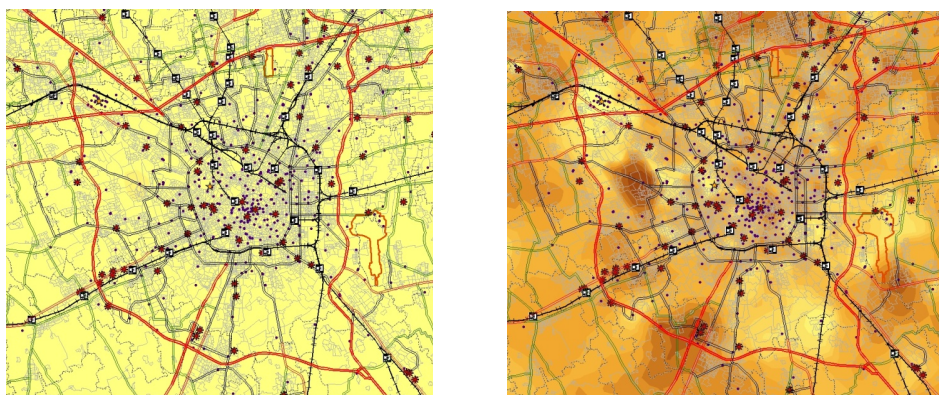


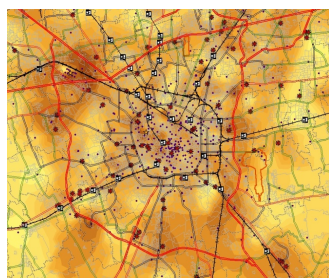
Figura 4 – Rappresentazione dei valori di Erlang nella fascia oraria 13.00 – 17.00; a sinistra la domenica e a destra il giorno feriale. I colori più scuri rappresentano un valore più elevato.

Nella costruzione dell'indicatore Handover/Erlang e nella sua rappresentazione spaziale, gioca un ruolo fondamentale anche la localizzazione delle antenne; la concentrazione di queste individua già di per sé alcuni luoghi urbani o sistemi contraddistinti da densità di presenze e da potenziale *motilità*, come ad esempio il centro di Milano e la Fiera (dove ne sono presenti una quindicina), così come lungo le principali infrastrutture rispetto alle quali le antenne si collocano linearmente (come ad esempio lungo il sistema tangenziale).

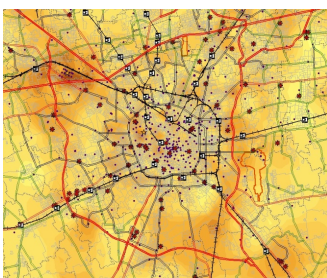
Dalle mappe tematiche elaborate di seguito è possibile vedere quali sono le aree a maggiore propensione alla mobilità. L'andamento del rapporto Handover/Erlang mostra un andamento decrescente dei valori durante la giornata, fenomeno dovuto probabilmente al maggior numero di telefoni presenti e alle telefonate (componente Erlang) a partire dalle prime ore del mattino. Tale fenomeno è meno riscontrabile nell'andamento dell'indicatore nella domenica, giorno nel quale l'area urbana risente meno delle presenze esterne e del pendolarismo.

Rispetto alla struttura territoriale è possibile individuare maggiore mobilità sulle principali infrastrutture (come ad esempio il sistema tangenziale di Milano) e su alcuni nodi dove sono presenti grandi funzioni, come ad esempio il polo fieristico, alcune grandi piastre commerciali, aeroporto, stadio, grandi parchi urbani.

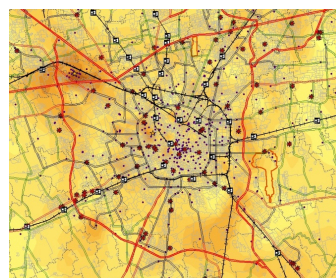
Di seguito è riportata la sequenza dei valori di Handover/Erlang relativi al giorno feriale per alcune ore significative. Il valore più elevato dell'indicatore è rappresentato in colore più scuro.



ore 6.00



ore 8.00



Ore 10.00

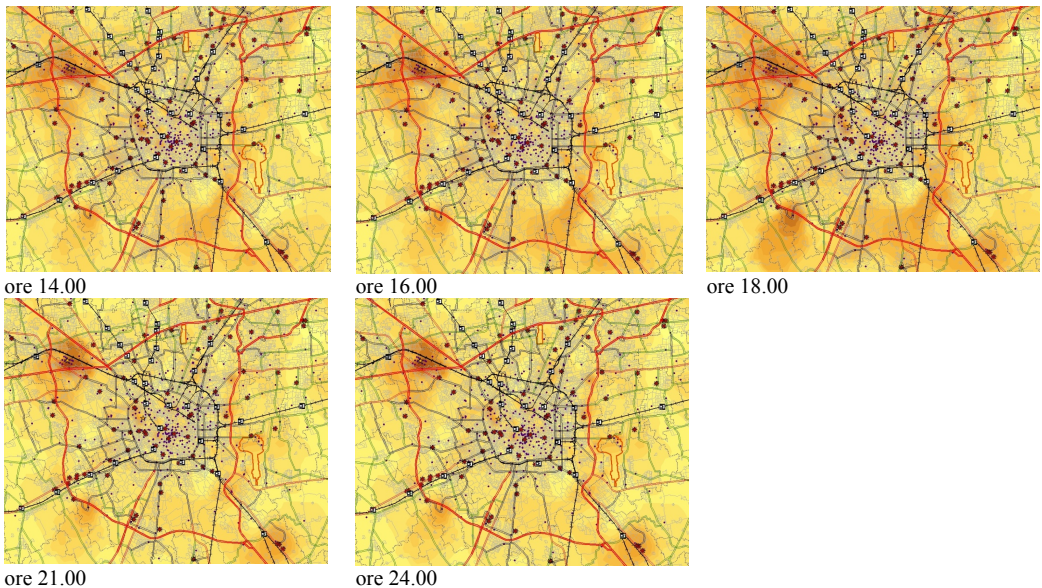


Figura 5 – Rappresentazione dell'indicatore Handover/Erlang nell'area metropolitana milanese per alcune fasce orarie di una giornata tipo ferialle.

L'andamento dell'indicatore evidenzia, in qualsiasi ora considerata, un ambito della regione milanese dove il rapporto Handover/Erlang risulta generalmente più elevato rispetto al resto del territorio. Si tratta della fascia dei comuni di prima cintura lungo l'anello del sistema tangenziale e di alcuni luoghi notevoli, in quanto attrattori di traffico e di mobilità, come il polo fieristico, l'aeroporto di Linate.

Gli alti valori del rapporto Handover/Erlang possono essere ricondotti tanto a valori di Handover elevati, quanto a valori di Erlang bassi; in questo senso, l'andamento dell'indicatore alle ore 6.00 che presenta valori tra i più elevati rispetto alle altre ore considerate, a fronte di valori di Erlang bassi, esprimerebbe una maggiore propensione alla mobilità in quella fascia oraria, accomunabile peraltro alla situazione riscontrabile per le 24.00, ora in cui la componente Erlang risulta significativamente bassa.

La significatività dell'indicatore nel descrivere le forme d'uso del territorio, che va comunque correlata con gli andamenti del dato in Erlang, è restituita anche osservando l'andamento del rapporto Handover/Erlang in Milano, città dove si può notare come l'andamento del dato restituisca il profilo d'uso del capoluogo, con valori cioè relativamente bassi tra le 10.00 e le 16.00, in incremento alle 18.00 (uscita dal lavoro di popolazioni che non abitano nel comune), per poi decrescere alle 21.00.

Per concludere, si può affermare che gli andamenti dell'indicatore, nelle diverse ore considerate, restituiscono un'immagine per certi versi inedita del territorio che completa la mappa sulle densità d'uso restituita attraverso i soli dati di Erlang, facendo emergere cioè quegli ambiti con una più intensa propensione alla motilità.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008), "Understanding individual human mobility patterns", *Nature*, 453, 779-782.

Reades, J., Calabrese, F., Sevtsuk, A., & Ratti, C. (2007), "Cellular Census: Explorations in Urban Data Collection", *IEEE Pervasive Computing*, 6 (3).

Kaufmann, V. (2002). *Re-thinking mobility*. Aldershot: Ashgate.