Università degli studi di Udine

GIS, Big Data e Social per l'analisi di sistemi territoriali complessi

Original

Availability:
This version is available http://hdl.handle.net/11390/1125882 since 2018-11-22T11:54:36Z

Publisher:
Associazione dei Gegrafi Italiani (AGEI)

Published
DOI:

Terms of use:
The institutional repository of the University of Udine (http://air.uniud.it) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

27 April 2024

GIS, Big Data e Social per l'analisi di sistemi territoriali complessi

Salvatore Amaduzzi
Dipartimento di Lingue e Letterature, Comunicazione, Formazione e Società dell'Università degli Studi di Udine
salvatore.amaduzzi@uniud.it

Abstract

Il fenomeno dei Big Data è uscito dai laboratori di ricerca ed è diventato un argomento centrale nel dibattito sull'innovazione, nel settore privato e pubblico.
Il progetto che il laboratorio di Geomatica dell'Università di Udine sta sviluppando ha come obiettivo la ricostruzione della dinamica di sistemi complessi quali città, sistemi di trasporto, … utilizzando dati anonimi ed aggregati acquisiti da social e operatori telefonici e modelli matematico/statistici ad hoc predisposti.

Le domande alle quali si vuole dare risposta sono del tipo:

- Quanti turisti/persone sono presenti in un luogo ora o ad una certa ora?
- Da dove arrivano?
- Dove andranno?
- Quanto si fermano mediamente?
- Quali luoghi visitano?
- Dove sono oggi quelli arrivati ieri con un certo volo?
- Come cambiano le abitudini in funzione della provenienza?
- In che modo è percepito il territorio?
- Di cosa si discute sui social?
- …

I dati a supporto di queste analisi sono di due tipi:

- dagli operatori di telefonia cellulare presenze nello spazio e nel tempo
- dal web tweet e dati postati sui social

Parole chiave: Big Data, GIS, Cell phone data

1 Introduzione

Analizzare e comprendere gli spostamenti degli umani all'interno di un'area geografica (città, territorio, paese, …) è cruciale in diversi ambiti quali la pianificazione urbana e dei trasporti, la pianificazoine e gestione di un evento, la gestione delle emergenze, le strategie di marketing, ecc.
Con il rapido sviluppo delle applicazioni web 2.0 le persone sono in grado di pubblicare messaggi e informazioni multimediali sulla propria vita quotidiana, condividere opinioni ed emozioni on-line.
I social media coprono un'ampia varietà di argomenti, da informazioni semplici su eventi e prodotti a questioni più complesse legate alla finanza, alla cultura, alla politica, alla religione, al cibo, alle epidemie, alle carestie, ecc.
Twitter è uno dei più popolari strumenti della rete sociale. Ogni giorno circa 140 milioni di utenti attivi pubblicano oltre 400 milioni di Tweet da 140 caratteri. La velocità e la facilità della pubblicazione hanno

reso Twitter un mezzo di comunicazione importante per le persone indipendentemente dalla classe sociale di appartenenza.
Twitter viene utilizzato per condividere informazioni e report sui danni alle popolazioni durante grandi catastrofi naturali, così come per dare informazioni su viaggi di svago o lavoro.
Un ulteriore fonte di informazioni per studiare i movimenti e il comportamento sono i dati che provengono dalla rete della telefonia cellulare. I dati del telefono cellulare rappresentano con precisione la mobilità umana. Il roaming assicura che un dispositivo wireless in viaggio sia collegato a una rete senza interrompere la connessione.
I dati del telefono cellulare sono anche molto rilevanti per le statistiche di mobilità e trasporto. Lo studio delle dinamiche turistiche in arrivo e in uscita, il numero di viaggi, il numero di giorni e notti trascorse, il numero di visitatori unici, il paese di residenza, la destinazione principale, la destinazione secondaria, i transiti, le visite ripetute, …
Alcune delle applicazioni realizzate con i dati della telefonia cellulare sono per applicazioni di business turistiche, il monitoraggio degli eventi: Mobile Landscapes: Graz in tempo reale, diretto da C. Ratti, A. Shevtsuk ed altri MIT & Ratti Associates Real Time Rome, condotta da MIT per studiare un concerto di Madonna a Roma e le partite di calcio di Milano analizzate dall'Istituto Frauenhofer nel 2008.
In questo contributo descriviamo il processo di acquisizione e strutturazione della banca dati e i risultati delle prime elaborazioni che abbiamo predisposto per analizzare le potenzialità di queste informazioni.

2 Descrizione dati

Durante questo primo periodo di lavoro abbiamo avuto la conferma, se ce ne fosse stato bisogno, di quale sia la fondamentale differenza tra dati e informazioni.
Abbiamo acquisito e stiamo acquisendo in continuo una marea di dati che dobbiamo strutturare e dai quali dobbiamo estrarre informazioni. Siamo arrivati alla conclusione che spesso la semplicità nella definizione e gestione delle informazioni è inversamente proporzionale alla quantità di dati disponibili.
Qui di seguito descriviamo i dati che sono stati utilizzati per queste analisi.

2.1 Dati Presenze Operatori telefonici

Gli operatori di telefonia cellulare sono in grado di fornire il numero di persone presenti all’interno di una cella di 150x150 metri. Il dato è disponibile con cadenza di 15 minuti suddiviso per nazionalità di provenienza.

Oltre al dato di presenza sono disponibili anche le matrici origine destinazione in grado di analizzare lo spostamento di plotoni di persone predefinendo l’origine, la data di presenza presso l’origine e l’orario di presenza nelle destinazioni.

2.2 Dati Social Network

Dai social network oltre alle informazioni di posizione/presenza si possono acquisire ulteriori informazioni: testo e quindi sentiment nel caso di twitter, foto e le informazioni multimediali postate dagli utenti in Flikr ed Instagram, ….
Nell’ambito di questo progetto è stato utilizzato Twitter che rende disponibile TUTTI i tweet tramite la propria API (https://dev.twitter.com/streaming/overview).
Il DB che abbiamo scaricato è di circa 270.000 record che corrispondono ai tweet georiferiti postati in un anno sul territorio del Friuli Venezia Giulia dalle SIM straniere in roaming.

Il prossimo step progettuale sarà quello di utilizzare le immagini per poter riconoscere

- luoghi, esistono applicazioni in grado di confrontare una immagine caricata con le immagini caricate sul web per riconoscere in automatico il luogo in cui una foto è stata scattata
- situazioni, se viene caricata la foto di una discarica a cielo aperto con associato il testo *non se ne può più* è possibile in automatico riconoscere la tipologia di situazione e capire dal testo se il commento è negativo o meno
- …..

Alla analisi delle presenze si possono quindi aggiungere quelle, ben più complesse, delle immagini e dei testi.

2.3 Dati Twitter

La connessione con Twitter rende disponibile in tempo reale TUTTI i tweet postati sulla piattaforma. Considerando che sono circa 5.000 al secondo si può facilmente immaginare che, se non opportunamente filtrati, il Data Base sul quale verranno salvati crescerà in breve a dismisura. Per questo motivo è stata sviluppata una applicazione in grado di prefiltrare e scaricare sul DB locale, solo i tweet che corrispondono ai filtri impostati.

3 Applicazioni

La definizione delle fonti/fornitori di informazioni e la predisposizione delle modalità automatiche di integrazione con i relativi portali sono stati il primo ambito di sviluppo.
Quindi ci siamo focalizzati sulla scelta degli strumenti informatici per storicizzazione, organizzazione e retrieve dei dati in modo da avere tempi di risposta adeguati.
È evidente che la mole di dati da organizzare ed analizzare assume velocemente una dimensione molto importante per cui la tematica della gestione dei dati è un tema fondamentale in questi progetti.
Si rende poi necessario lo sviluppo di una piattaforma WebGIS che renda disponibile una interfaccia user friendly alle diverse tipologie di utenti in modo da consentire loro di rappresentare ed analizzare le informazioni disponibili.

3.1 Scarico e analisi dei tweet

La prima applicazione sviluppata è relativa all’uso dei dati da Twitter.
All’attivazione l’applicazione si connette a Twitter ed attende di ricevere in input i parametri con i quali filtrare e salvare i tweet.
L’interfaccia consente di inserire delle chiavi di ricerca per tutti gli attributi disponibile nell’API che sono:

- Testo
- Lingua utilizzata
- Coordinate
- Data e ora
- Nazione di provenienza
- Alias twitter
- Data creazione utente
- Numero follower
- Numero di seguiti

Queste condizioni possono essere composte in AND o OR.

A questo punto tutti i tweet che soddisfano le condizioni selezionate verranno scaricati nella banca dati locale. È quindi possibile attivare delle campagne di "ascolto" di giorni, settimane, mesi …. filtrando ad esempio i tweet che:

- contengono certi vocaboli
- sono stati effettuati dagli utenti di una certa nazionalità
- sono scritti in una certa lingua
- sono stati effettuati in una certa regione
- …..

In questo modo, per ogni analisi, saranno disponibili tutti e soli i tweet relativi al tema specifico.
Sono quindi possibili campagne per analizzare quando, chi, dove e cosa viene detto in merito, ad esempio, a discariche abusive, bontà del vino, le linee di alta tensione, ….

A questo si aggiunge la possibilità di mettere a punto degli algoritmi o di sfruttare piattaforme esistenti per la sentiment analysis in modo da poter analizzare i testi dei tweet per evidenziare situazioni anomale, punti di interesse non conosciuti, …

Figura 1: tweet che contengono la parola "wine" di Londinesi in FVG

3.2 Dati telefonia

Il primo problema nella gestione di queste informazioni è la quantità di dati
Consideriamo che il Friuli Venezia Giulia è composto da circa 350.000 celle di 150 per 150 metri. Generandosi una informazione di presenza per ogni 15 minuti per ogni cella, si generano circa 35 milioni di record ogni giorno e circa 12 miliardi di record all'anno.
Ci siamo trovati quindi a dover progettare una banca dati che consenta prestazioni adeguate durante la fase di ricerca delle informazioni.
In questa fase è in test una piattaforma PostgreSQL on line in grado di storicizzare e rendere disponibili questi dati alle applicazioni che verranno sviluppate per analizzare i dati acquisiti dalle varie fonti.

Un esempio di utilizzo di queste informazioni è il seguente:

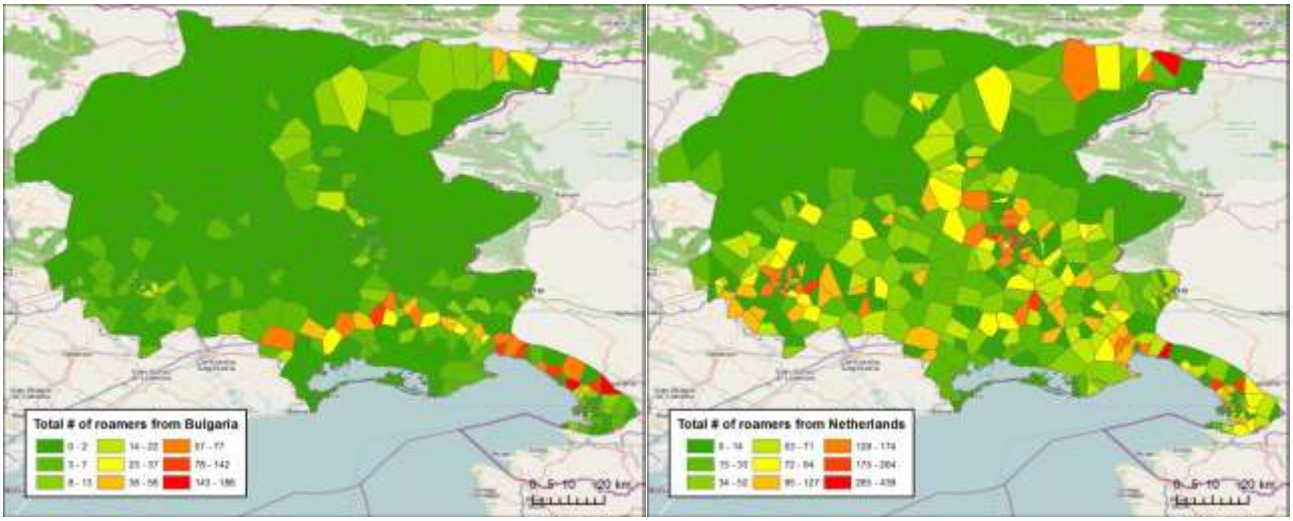

Figura 2: esempi di presenze per nazionalità in FVG

In queste due cartografie si rappresenta una fotografia ad un certo istante del numero di SIM bulgare (a sx) e olandesi (a dx) presenti sul territorio regionale.
Si evidenzia chiaramente che i Bulgari sono per lo più di passaggio sulla rete autostradale sia nella direttrice Est-Ovest, con entrata dalla Slovenia, sia nella direttrice Nord-Sud, con entrata da Tarvisio.
È quindi evidente che sono, per lo più, autotrasportatori che attraversano il Friuli V. G. in autostrada.
Gli Olandesi invece sono distribuiti sul territorio regionale e, da una prima analisi, sono molto presenti sulle più frequentate piste ciclabili per cui è molto probabile che siano turisti con una particolare attenzione ad una fruizione "slow" del territorio.

3.2.1 Presenze su base comunale

Per comprendere le potenzialità di queste informazioni si è deciso di fare una prima analisi aggregando i dati a livello comunale.
È stato predisposto un primo data base che contiene circa 2 milioni di record con le seguenti caratteristiche:

- Periodo dall'1 marzo 2016 al 30 settembre 2016
- 4 dati giornalieri per comune alle ore 00, 06, 12, 18
- Suddivisione tra italiani e stranieri
- Gli stranieri in base al paese di provenienza
- Gli italiani suddivisi per
    - Tipo di presenza
        - Residente
        - Visitatore occasionale
        - Visitatore regolare
    - Regione di provenienza

Sono stati utilizzati diversi strumenti GIS per predisporre le analisi che verranno descritte nei paragrafi successivi.

3.2.2 Matrice OD

Nei dati forniti dai provider telefonici sono disponibili anche informazioni in grado di analizzare lo spostamento di plotoni di persone.
Queste sono state, per adesso, strutturate in un DB oltre di 4.5 milioni di record che raccolgono il totale degli spostamenti per mese e per giorno della settimana da ogni comune ad ogni altro comune suddiviso tra Italiani e stranieri.
Con questo tipo di informazione sarà possibile rispondere a domande quali:

- dove sono oggi le persone con SIM russa che ieri alle 9 erano in aeroporto a Trieste
- quante persone entrano a Udine tra le 7 e le 8 e da quali comuni provengono
- …

3.3 Strumenti GIS

In questa fase di analisi e prototipazione delle potenzialità e di test della qualità delle informazioni sono stati utilizzati alcuni prodotti standard in funzione delle competenze presenti nel gruppo di lavoro ed in particolare ArcGIS, QGis e Mapinfo.
Non appena definiti gli obiettivi di dettaglio verrà poi sviluppato un WebGis che renderà disponibili tutte le funzioni necessarie dallo scarico ed organizzazione dei dati per all'analisi e rappresentazione degli stessi.

4 Casi studio

Sono state effettuate le prime analisi utilizzando i dati della telefonia cellulare analizzando alcuni eventi e situazioni specifiche del territorio regionale, in particolare: Friuli DOC e giornata di pioggia a Grado.

4.1 Friuli DOC

Friuli DOC è un evento che vede per quattro giorni la città di Udine diventare una vetrina per presentare e far conoscere il meglio della produzione enogastronomica, artigianale, artistica e culturale regionale.
Negli stand e nei chioschi, nelle piazze e nelle vie del centro, si possono gustare i prodotti tipici friulani, il prosciutto di San Daniele e di Sauris, il formaggio Montasio, il frico, i cjarsons, i vini, le grappe e molti altri.
L'occasione di assaggiare i tesori dell'enogastronomia friulana si coniuga ad appuntamenti con mostre, spettacoli, convegni, esposizioni artistiche e produzioni artigianali, sempre nel rispetto delle più antiche tradizioni locali.
Questa manifestazione attrae visitatori dall'hinterland, da altre regioni italiane e dall'estero. Nel 2016 si è tenuta dall'8 all'11 settembre.
Si è deciso di confrontare la giornata di sabato 10 settembre 2016, giornata più partecipata dell'evento, con il sabato 3 settembre della settimana precedente. Queste due giornate sono, in situazioni normali, più o meno coerenti tra loro e, in questo caso, hanno avuto anche una situazione metereologica simile.
Sono state analizzati gli orari delle 18 e delle 24 per le seguenti categorie di presenze:

- Italiani non residenti
- Italiani residenti
- Stranieri
- Stranieri provenienti dall'Austria

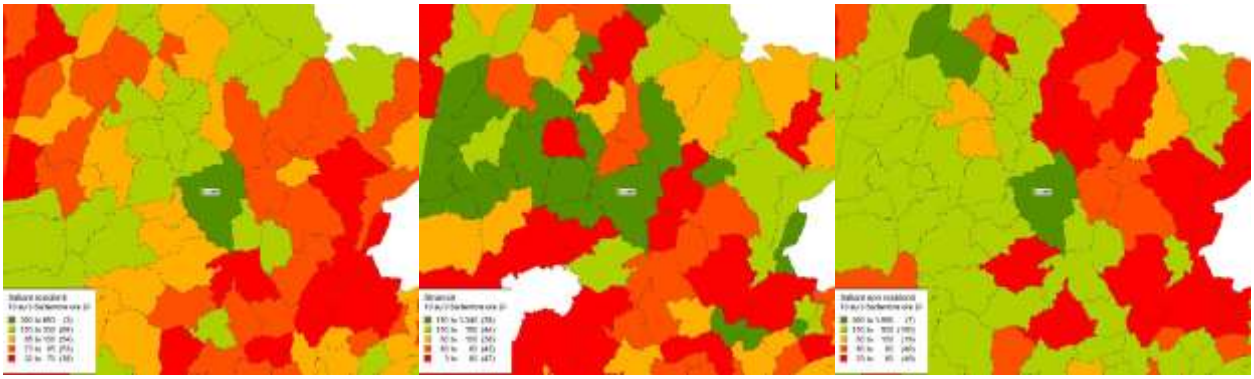

Figura 3: dati di presenze a Friuli DOC alle ore 24 comparati, sabato e sabato precedente

Le evidenze sono molto interessanti. Sono state analizzate e confrontate le percentuali di presenza e i valori assoluti. Per tutte le categorie le presenze nel comune di Udine aumentano drasticamente il 10 rispetto al 3 anche in termini assoluti. Eclatante è la differenza tra le due giornate alle ore 18, il 3 settembre sono presenti 67.000 friulani a Udine mentre il 10 settembre sono 107.000, contro i 67.000 delle situazioni normali. Dopo 6 ore, alle 24, i Friulani in comune di Udine sono 179.000 contro i 59.000 standard.

Di seguito una tabella che descrive i valori assoluti

	Friulani		Italiani Extra Regione		Stranieri	
	ore 18	ore 24	ore 18	ore 24	ore 18	ore 24
03-set	67.463	59.249	3.336	2.791	856	499
10-set	107.006	178.701	9.936	16.200	1.238	804

Interessante verificare che il calo che si verifica di solito tra le 18 e le 24 per gli Italiani è in totale controtendenza durante l'evento dove c'è un incremento del 400%, gli stranieri invece hanno una certa flessione molto probabilmente per il fatto che una parte degli Sloveni ed Austriaci rientrano a casa in tarda serata.
Si può inoltre verificare come alle 18 ancora molti Italiani non friulani siano nell'hinterland di Udine in particolare nei comuni del Collio e tra le 18 e le 24 si concentrano poi in città.
Analizzando con maggior dettaglio le diverse fasce orarie, ricordiamo che si può arrivare ad una granularità temporale di 15 minuti ed una granularità territoriale di 150x150 metri, si potrà capire dove e chi si ferma a dormire nel territorio, quali sono i percorsi utilizzati per accedere alla città e per il relativo deflusso, etc.

4.2 Grado Isola del Sole

Grado, l'Isola del Sole, assieme a Lignano e alla costiera Triestina è una delle destinazioni balneari più attrattive della Regione Friuli Venezia Giulia.
Nel 2016 ha visto circa 400.000 arrivi, di cui 250.000 stranieri, per un totale di circa 1.600.000 presenze.
Qui si è voluto analizzare quale il comportamento dei turisti Italiani e Stranieri durante una giornata di pioggia nel mese di agosto.
Utilizzando il sito https://www.ilmeteo.it/portale/archivio-meteo/Grado/2016/Agosto/9?refresh_cens si è scelto martedi 9 agosto 2016, giornata di sole, e mercoledi 10 agosto giornata di pioggia per analizzare

il comportamento dei turisti. Queste 2 giornate infrasettimanali non sono influenzate dai cambi settimanali che di solito avvengono di venerdi o sabato e nemmeno dalle chiusure delle attività commerciali.

I risultati sono molto interessanti, durante la giornata di pioggia la presenza delle 3 categorie analizzate (Stranieri, Friulani, Italiani provenienti da altre regioni) nel comune di Grado si riduce mediamente del 50% e si distribuisce nel seguente modo:

- Italiani non residenti in FVG e Italiani residenti in FVG
    - Villesse, dove è situato un grande centro commerciale con la presenza di Ikea
    - Medea dove è situata l'Ara Pacis, un famoso complesso monumentale che ricorda simbolicamente i caduti di tutte le guerre
    - Comuni limitrofi
- Stranieri
    - Villesse
    - Visitano il comune di Aiello che è famoso per la presenza di oltre 100 meridiane
    - Si spingono fino a visitare la città di Udine che vede un incremento di presenze del 60%
    - Non sono attratti dall'Ara Pacis.

Interessante notare che anche la vicina località balneare di Lignano subisce un esodo durante le giornate di pioggia ma non della stessa dimensione in quanto Lignano, rispetto a Grado, offre una maggior varietà di negozi e locali in cui "rifugiarsi" nelle giornate uggiose.

5 Conclusioni

Risulta evidente, nonostante questi siano solo i primi test di utilizzo, quanto queste informazioni siano precise e quanto possano essere utili per analizzare le dinamiche delle presenze e degli spostamenti in un territorio delle varie categorie di persone a fronte di eventi organizzati, eventi metereologici, …

Le prossime attività prevedono, oltre alla già citata strutturazione di una banca dati on line contenete tutti i dati resi omogenei ed aggiornati real time, lo sviluppo di un WebGIS che diventerà l'interfaccia di interrogazione ed analisi delle informazioni ed inevitabilmente un confronto con alcuni attori presenti sul territorio che necessariamente potrebbero utilizzare questa infrastruttura a supporto delle proprie decisioni.

I primi due "clienti" individuati con i quali si inizierà a breve una interlocuzione sono:

- PromoTurismoFVG che è l'Agenzia Regionale per la pianificazione e la promozione dell'offerta turistica della Regione Friuli Venezia Giulia e che evidentemente potrebbe trarre un notevole vantaggio da questa piattaforma
- Direzione Regionale per le Infrastrutture e i Trasporti che potrebbe pianificare l'offering del trasporto pubblico su gomma e su rotaia con miglior cognizione di causa.

L'accessibilità alle informazioni di presenze e spostamenti con alta granularità spaziale e temporale implica una analisi della tematica relativa alla riservatezza e sicurezza del dato che verrà approfondita prima della messa a disposizione delle informazioni on line ai vari attori.

6 Bibliografia

Ahas, R., Silm, S., Järv, S., Saluveer, E., (2010) "Using mobile positioning data to model locations meaningful to users of mobile phones", *Journal of Urban Technology*, 17, 3-27

Andrienko, G., Andrienko, N., Bak, P., Bremm, S., Keim, D., von Landesberger, T., Poelitz, C., Schreck, T., (2010) "A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage." *Journal of Location Based Services*, 4, 200-221

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C., (2011) "Real-time urban monitoring using cell phones: a case study in Rome", *IEEE Transactions on Intelligent Transportation Systems*, 12, 141–151

Furletti, B., Gabrielli, L., Renso, C., Rinzivillo, S., (2013), Analysis of GSM calls data for understanding user mobility behavior. IEEE International Conference on Big Data,

Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., Trasarti, R., (2011), "Unveiling the complexity of human mobility by querying and mining massive trajectory data", *The VLDB Journal*, 20, 695–719

Kumar, S., Morstatter, F., Liu, H., (2013), *Twitter Data Analytics*, Springer, New York

Ratti, C., Sevtsuk, A., Huang, S., Pailer, R.: *Mobile Landscapes, (2005), Graz in Real Time*. Location Based Services and TeleCartography, Springer, Berlin, 433-444

Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.-L., (2011), "Human mobility, social ties, and link prediction.", *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD 11*, 1100-1108

Sitografia

http://blogs.lse.ac.uk, London School of Economics and Political Science, http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/10/social-media-research-tools-overview/, Using Twitter as a data source: An overview of current social media research tools, (2017/05/31)

http://faculty.ce.berkeley.edu, Berkley University , http://faculty.ce.berkeley.edu/pozdnukhov/lbsn15/files/a1-abbasi.pdf, Utilising Location Based Social Media in Travel Survey Methods: bringing Twitter data into the play, (2017/05/31)

http://ec.europa.eu/eurostat/, Eurostat, http://ec.europa.eu/eurostat/documents/747990/6225717/MP-Task-1-report.pdf/1e81fd37-e9d9-4c2f-ab39-3528e6c1237c, Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics, (2017/05/31)

https://www.friuli-doc.it/en/, Municipality of Udine, (2017/05/31)

http://www.turismofvg.it/Locality/Grado PromoTurismoFVG, (2017/05/31)