

Le informazioni geografiche dei social network (SMGI) a supporto della pianificazione del turismo. L'esempio di Cagliari.

Roberta Floris ^(a), Pierangelo Massa ^(b), Michele Campagna ^(c),

^(a) Università degli Studi di Cagliari-DICAAR, via Marengo 2, tel 070 675 5210, roberta.floris@unica.it

^(b) Università degli Studi di Cagliari-DICAAR, via Marengo 2, tel 070 675 5210, pmassa@unica.it

^(c) Università degli Studi di Cagliari-DICAAR, via Marengo 2, tel 070 675 5203, campagna@unica.it

Abstract

Il contributo propone una discussione sul possibile utilizzo delle informazioni geografiche provenienti dai social network nell'ambito della pianificazione del turismo su scala regionale e locale. Il crescente utilizzo dei *social network* da parte degli utenti in tutto il mondo ha fatto sì che queste piattaforme, e soprattutto l'informazione derivante da esse, sia utile per i turisti, che possono facilmente ottenere dati adeguati sulle destinazioni e sui servizi offerti, grazie alle recensioni liberamente condivise dagli altri utenti, e per gli operatori turistici, i quali possono utilizzare queste informazioni per migliorare le proprie strategie di marketing e promozione. Questi dati, comunemente definiti *Social Media Geographic Information (SMGI)*, possono inoltre offrire nuove opportunità per supportare le fasi decisionali nella pianificazione del turismo. Purtroppo la ricchezza di contenuti sulle percezioni e sulle opinioni degli utenti, resa disponibile dalle SMGI, non viene ancora adeguatamente utilizzata dai pianificatori per le analisi territoriali.

Lo studio propone un nuovo approccio per analizzare, qualitativamente e quantitativamente, attraverso l'uso delle SMGI, le relazioni che insistono tra il gradimento dei turisti, le località geografiche e l'offerta turistica in Sardegna. La metodologia adottata include la raccolta di dati da *Booking.com* e *TripAdvisor*, la loro integrazione ed elaborazione con i dati ufficiali in ambiente GIS, e l'applicazione di tecniche di analisi di statistica spaziale per identificare e valutare i fattori che possono determinare il successo di una destinazione turistica.

La metodologia viene applicata alla scala regionale, individuando le aree maggiormente apprezzate dai turisti, ed alla scala locale per la destinazione di Cagliari, per la quale vengono identificati e valutati i principali fattori che ne determinano il successo come destinazione turistica. I risultati ottenuti possono essere utilizzati come base conoscitiva per guidare ulteriori specifiche analisi e per sviluppare strategie di sviluppo sostenibile nell'ambito della pianificazione territoriale e del turismo tramite processi decisionali informati.

Introduzione

Nell'ultimo decennio, i continui progressi nell'*Information Communication Technology (ICT)* e nelle tecnologie *Web 2.0* hanno interessato diversi domini di interesse (Grabner et al. 2012) e favorito l'ingresso dell'informazione geografica nella vita quotidiana di un crescente numero di utenti. Questo fenomeno sta portando ad un cambiamento nel paradigma della produzione e diffusione di dati georiferiti (Elwood et al., 2012), così come nelle caratteristiche e nei contenuti che vengono condivisi, favorendo una rinascita dell'informazione geografica ed offrendo ai pianificatori una vasta gamma di informazioni..

In Europa lo sviluppo delle Infrastrutture di Dati Territoriali (IDT), definite come l'insieme di tecnologie, metodi, politiche ed accordi istituzionali tesi a facilitare la disponibilità, l'omogeneità e l'accesso ai dati geospaziali (Craglia e Campagna, 2010), ha consentito il pubblico accesso ai dati geospaziali ufficiali e il loro riutilizzo.

Inoltre la diffusione dei *geo-browser* e della *mobile technology* hanno permesso ai cittadini di agire come sensori volontari (Goodchild, 2007) per la raccolta e la disseminazione di informazioni

geografiche in tempo reale, note come *Volunteered Geographic Information* (VGI). Queste informazioni racchiudono in se sia il sapere tecnico dei professionisti, sia quello esperienziale delle comunità locali, rafforzando il *know how* del contesto territoriale di riferimento, imprescindibile punto di partenza della pianificazione regionale e urbana. Queste informazioni possono inoltre rappresentare una nuova fonte di informazione geografica, liberamente accessibile e universalmente fruibile (Sui and Goodchild, 2012), che può essere integrata con gli attuali dati geospaziali disponibili per supportare processi decisionali informati al contesto sociale e geografico, favorendo pertanto più democratiche pratiche di pianificazione.

Diversamente dalle informazioni geografiche ufficiali, le informazioni derivanti dai social media (*Social Media Geographic Information*, SMGI) possono riguardare processi dinamici che avvengono sulla Terra o la percezione, da parte degli utenti, di questi processi (Campagna, 2014), in uno specifico lasso temporale. La produzione e la condivisione delle SMGI ricadono nel dominio dei *Big Data*, e pertanto occorrono specifiche tecniche di *Big Data analysis* e *Data Mining* per estrarre e gestire adeguatamente tale ingente mole di informazioni. In tal senso, le tecniche di *Computational Social Science* (Lazer et al., 2009) rappresentano un nuovo paradigma emergente che ambisce a sviluppare nuovi metodi per far fronte a questa problematica.

Il turismo nell'era dell'informazione geografica

Il market turistico appare oggi fortemente rivoluzionato dallo sviluppo del *Web 2.0* (Buhalis and O'Connor, 2005), che ha modificato il modo di pubblicizzare le diverse destinazioni turistiche tramite, ad esempio, l'offerta di formule di pacchetti vacanza su misura per qualsiasi tipologia di turista (Miguens, et al., 2008). Il sempre crescente trend di utilizzare il web come strumento di pianificazione delle proprie vacanze ha portato alla nascita del cosiddetto *Travel 2.0 World* (Conrady, 2007) e dell'*e-Tourism* (Biassoulis, 2002). La scelta della location e i giudizi espressi dagli viaggiatori sulle differenti strutture ricettive diventano gli argomenti più popolari all'interno delle *chat room* delle *Tourists on line Communities*.

TripAdvisor e *Booking.com* rappresentano i più popolari *Social Network* in ambito turistico, e assumono dunque un ruolo fondamentale nella produzione di SMGI in ambito turistico. Dati di questo tipo costituiscono la fonte principale per lo studio dei livelli di soddisfazione dei turisti in merito alle singole destinazioni, come dimostrato dal successo dei servizi di rating di tali piattaforme. Inoltre, i dati multimediali provenienti da *Instagram*, *Flickr* e *Panoramio*, consentono di individuare i luoghi e i percorsi che affascinano gli utenti, identificando tra questi non solo la bellezza delle spiagge e la limpidezza del mare, ma anche i luoghi della cultura e della biodiversità del territorio. Nel dominio della pianificazione turistica spaziale, conoscere le preferenze dei viaggiatori rappresenta un importante target di riferimento per i pianificatori che, oltre ad acquisire le conoscenze sulle strutture ricettive che compongono l'industria turistica, possono comprendere la psicologia dei turisti e delle loro esperienze pregresse (Benitez et al., 2007).

Per la gestione, l'analisi e la distribuzione di dati geografici sono disponibili numerosi strumenti (*web-based maps*, *Geographic Information Systems*, *newsgroups* e *Social Network*), che annoverano anche la partecipazione degli utenti. Tali strumenti rappresentano elementi concreti per lo sviluppo di un ambiente di pianificazione digitale innovativo; tuttavia un loro diffuso e accettabile livello di utilizzo non è ancora stato raggiunto, né da parte dei professionisti, né da parte delle comunità locali.

L'approccio metodologico adottato

Nella definizione della metodologia, il punto di partenza è stato considerare il turismo come fenomeno spaziale (Peroni, 2008), quantificando il modello concettuale proposto da Cooper (2008):

Preferenze turistiche = localizzazione geografica + qualità dell'industria turistica [1]

al fine di individuare le relazioni che intercorrono tra gli attrattori turistici principali del territorio sardo, le strutture ricettive e i turisti, e il rapporto che questi ultimi hanno con l'offerta turistica, sia a livello regionale che locale.

Nello specifico, gli obiettivi della metodologia sono i seguenti:

- i. studiare la distribuzione dell'offerta turistica regionale e i luoghi geografici di interesse;
- ii. rilevare i *clusters* significativi delle preferenze turistiche e le aree geografiche di interesse strategico (livello locale);
- iii. analizzare le aree di interesse strategico, sviluppando un modello, geograficamente pesato, che determini quanto la distribuzione delle preferenze turistiche in quelle aree sia spazialmente influenzata dalla presenza di altre variabili socio-economiche nelle stesse.

Le posizioni geografiche sono state espresse mediante l'integrazione di SMGI e dati ufficiali, mentre la qualità del servizio mediante SMGI. L'estrazione di tale informazione da *TripAdvisor* e *Booking.com*, riferita al periodo Maggio 2012-Maggio 2013, ha consentito di costruire un *database* spaziale contenente 1992 strutture ricettive localizzate in Sardegna, e quasi 1 milione di commenti testuali ed esse correlati e rappresentanti i feedback dei turisti sul loro soggiorno nell'isola. Il *database* include anche un giudizio globale dato dalla media dei punteggi che ogni utente ha assegnato a ciascuno dei 6 attributi principali, ossia 1) pulizia, 2) cordialità del personale, 3) rapporto qualità prezzo, 4) localizzazione geografica, 5) presenza di servizi (negozi, ristoranti, fermate dell'autobus, etc.) e 6) comfort percepito. Tale *database* è stato poi utilizzato per le successive analisi.

I dati estratti sono stati gestiti tramite metodi di analisi statistica spaziale in ambiente GIS e tecniche di analisi testuale, per identificare i motivi principali delle preferenze turistiche espresse e individuarne i *cluster* spaziali rilevanti alla scala regionale. Successivamente, a livello locale, con riferimento ad un particolare *cluster* o area di interesse, sono state sviluppate ulteriori analisi per comprendere le possibili cause dei *cluster* spaziali identificati e le singolarità.

Inizialmente è stata calcolata la distanza incrementale di autocorrelazione. La misura della distanza di Autocorrelazione Spaziale, calcolata considerando il valore dell'indice di Moran (Moran, 1950) considera una serie di distanze via via crescenti, misurando l'intensità dei *cluster* di preferenze turistiche per ciascuna distanza. L'intensità di raggruppamento è determinata dallo *z-score* calcolato (deviazione standard). A una certa distanza particolare si presenterà un picco significativo che rappresenta il valore di riferimento mediante il quale relazionare ogni singolo *cluster*. Maggiore è l'intensità del *cluster* e quindi lo *z score*, maggiore è la possibilità che a quella determinata distanza si manifesti un *pattern* significativo. Successivamente, l'analisi *Hot Spot* (Getis and Ord, 1995) ha consentito di visualizzare su mappa i *cluster* spaziali delle preferenze turistiche (*hot spot*) e le aree a bassa preferenza (*cold spot*) e, successivamente, interpretare i risultati ottenuti per entrambe le scale. Successivamente, l'analisi testuale effettuata per ogni singolo *cluster* di interesse, ha permesso di individuare indizi significativi nascosti tra i commenti lasciati dai turisti, relativi ad ogni singola struttura alberghiera.

Infine, la regressione geografica pesata ha permesso di esaminare le relazioni spaziali (Fotheringham et al., 2002) sussistenti tra i giudizi degli utenti, la localizzazione geografica e l'industria turistica, integrando i dati estratti dai social media con l'informazione proveniente da fonti ufficiali (Infrastruttura dei dati territoriali della regione Sardegna). L'obiettivo è testare l'affidabilità delle ipotesi derivanti dalla precedente analisi testuale, al fine di prendere decisioni appropriate in termini di politiche regionali.

Analisi dei risultati

Dalle analisi condotte si evince come le strutture ricettive presenti in Sardegna e maggiormente recensite dagli utenti, ricadano nelle seguenti categorie, secondo la classificazione dell'Istituto di

Statistica Italiano - ISTAT: campeggi (6,0%), bed and breakfast (15,7%), alberghi (42,0%), altri alloggi (29,0%), residence e resort (7,3%). La prevalenza di una di queste categorie in una determinata zona geografica denota la dominanza di un particolare modello di turismo nella stessa.

Per quanto riguarda la localizzazione spaziale delle reali preferenze dei turisti, l'87% di queste sono concentrate nelle aree costiere, mentre il restante 13% nelle aree interne dell'isola. Questo risultato va a confermare il generale trend regionale del settore turistico, denotando come anche le politiche di settore diano una risposta adeguata alla forte domanda dei turisti di soggiornare nelle aree più conosciute dell'isola. Al contrario, i territori interni, impregnati di una profonda identità storico-archeologica e peculiarità ambientali notevoli, risultano meta di un turismo elitario, forse perché poco pubblicizzato.

Le analisi *Hot Spot* applicate alla scala regionale denotano quali sono i territori ampiamente apprezzati dai turisti (Figura 1), fornendo così spunti significativi su future politiche di sviluppo turistico territoriale, particolarmente indicate nelle aree a valenza strategica, o perché spazialmente localizzate in prossimità di aree ad alta preferenza turistica, oppure perché ancora poco sviluppate dal punto di vista dell'offerta ricettiva. Dall'immagine è evidente come nelle aree Nord-Est, Nord-Ovest e nel Sud dell'isola vi sia una maggiore concentrazione di cluster di preferenze turistiche. Questo denota come l'offerta turistica di quelle aree, nonché la loro localizzazione geografica, soddisfa le aspettative dei visitatori.

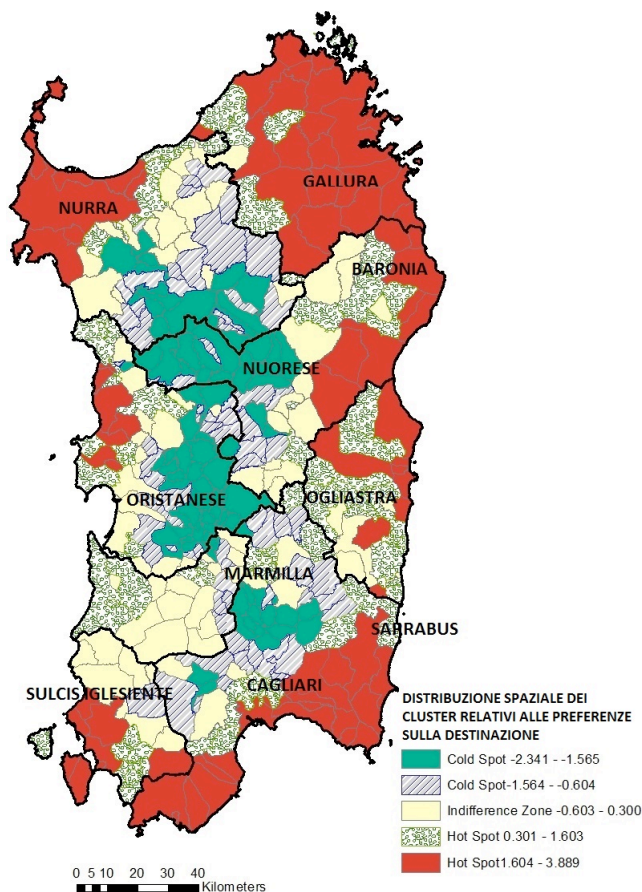


Figura 1. Cluster spaziale della preferenze turistiche sul territorio regionale

Analizzando singolarmente i singoli comuni, il comune di Cagliari rappresenta un particolare *spot* di interesse. Per questo motivo le successive analisi sono state implementate per questa destinazione turistica. I *cluster* spaziali delle preferenze turistiche sono identificate nelle aree in cui sono localizzate le strutture ricettive. La Figura 2 mostra la distribuzione spaziale della preferenze turistiche nel comune di Cagliari. Nelle aree in rosso, localizzate per lo più nel centro storico, la concentrazione delle preferenze degli utenti è maggiore, mentre nelle aree in blu, localizzate nelle periferie urbane, tale concentrazione è modesta.

Il passo successivo concerne l'analisi testuale, tramite l'applicazione di tecniche di analisi *spatio-temporal textual* (STTx, Campagna, 2014) relativa al set di commenti associati alle singole strutture ricettive localizzate all'interno dei *cluster* spaziali, con l'obiettivo di comprendere cosa i turisti pensano davvero della città di Cagliari. Il *tag cloud* risultante, riportato nella Tabella 1, mostra chiaramente che la maggior parte delle parole contenute nei post degli utenti si riferiscono ad aspetti fisici e spaziali della città, come ad esempio il "centro storico", la "città", e le "Chiese". I risultati includono anche termini relativi ai luoghi dello svago, quali "ristoranti" e "negozi", e ai alla dotazione di servizi delle strutture ricettive, come "personale" e "camera". Infine, il *tag cloud* indica che l'elevato livello di soddisfazione è legata anche all'accessibilità: parole come "vicinanza" e "camminare" potrebbero essere riferiti alla localizzazione spaziale dei servizi, delle risorse naturali o dei monumenti storici. Quindi, muoversi facilmente da un sito ad un altro genera nell'immaginario del turista un'opinione positiva. Questo non è sicuramente il tipo di informazioni contenute di solito nei documenti ufficiali o nei piani urbanistici, ma diviene un potente mezzo di supporto alla progettazione e al processo decisionale.

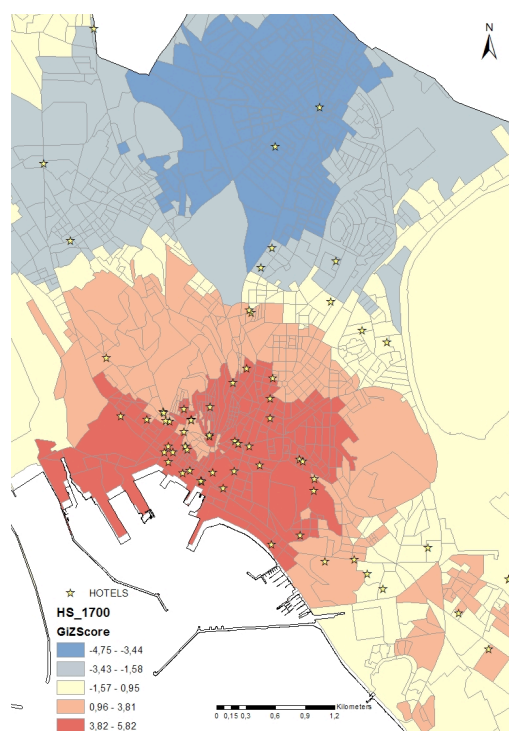


Figura 2. Cluster spaziale della preferenze turistiche nell'area di Cagliari

Categoria	Parole [frequenza]
Localizzazione geografica	localizzazione [1214]; centro storico [476]; vista[212]
Servizi	staff [649]; breakfast [564]; shopping [539]; ristoranti [443]; camera [358]; cibo [169]
Accessibilità	minuti [450]; prossimità [164]; camminare [146]
Componenti naturali e non	città [526]; spiaggia [375]; Chiese [132]

Tabella1. Risultati del Tag Cloud

L'ultima parte delle analisi riguarda l'applicazione di un modello di regressione geografica pesata, che ha permesso di esaminare le relazioni spaziali esistenti tra il valore delle preferenze espresse dagli utenti, considerata come una variabile dipendente, e altre caratteristiche fisiche del territorio, tra cui la aree naturali, le infrastrutture di trasporto e altre caratteristiche socio-economiche,

esaminate come variabili esplicative e quindi contribuenti al raggiungimento del valore di preferenza turistica in una data area. Tale analisi è stata implementata sia per modellare nello spazio le preferenze espresse sia per testare l'affidabilità delle ipotesi derivanti dai risultati dell'analisi testuale. Il modello è stato applicato ad un campione di 150 servizi ricettivi spazialmente distribuiti su oltre 100 delle 1359 zone censuarie di Cagliari. Il valore della variabile dipendente è stato normalizzato come frazione dei commenti positivi appartenenti ad una precisa zona censuaria. Per ciascuna di queste poi è stato calcolato il valore di un set di variabili indipendenti, identificate come segue:

1. numero di edifici storici;
2. numero di ristoranti e servizi turistici;
3. ettari delle aree naturali protette;
4. distanza dai principali nodi trasportistici;
5. vicinanza al centro storico della città.
6. distanza dalla spiaggia più popolare.

Se il valore, normalizzato, che la variabile dipendente assume in una data unità spaziale, è simile ai valori che la stessa assume nelle unità spaziali più vicine, essa è caratterizzata da autocorrelazione spaziale, rilevata attraverso la misura dell'indice di Moran. Tale valore risulta essere abbastanza significativo per il secondo ordine di contiguità rispetto al valore ottenuto con una distanza di soglia (*threshold distance*) di 2.500 metri: il valore di *R-squared* è inferiore al 40%, il valore del *p-value* del coefficiente della variabile dipendente è molto significativo (1,0 10⁻⁸), mentre il valore dell'indice di Moran è 0.024. Un valore del *p-value* molto basso indica che il valore della variabile dipendente è caratterizzata da autocorrelazione spaziale significativa. I risultati relativi alla bontà del modello di regressione spaziale sono mostrati in Tabella 2: *R-squared* raggiunge un valore dell'86%, il che indica che l'86% della varianza delle preferenze degli utenti viene spiegata dalle variabili incluse nel modello.

Variabile	Coefficiente	Std. Error	z-value	Probabilità
W_Pref_users	0.0662904	0.0306595	2.162143	0.0306071
Costante	0.0030834	0.010238	0.301150	0.7632998
N_ristoranti	-0.039035	0.0299971	-1.301316	0.1931503
Proxy_centro_storico	0.4747516	0.0585060	8.114574	0.0000000
N_ed_storici	-0.0273241	0.0313221	-0.872359	0.3830122
H_area_naturali	0.0027261	0.0096194	0.283397	0.0776872
Distanza_aero	0.7660687	0.0339702	22.55116	0.0000000
Distanza_spiagge	0.5470476	0.0382759	14.29221	0.0000000

Tabella 2. Risultati del modello di regressione spaziale applicato

I risultati del modello di regressione spaziale descrivono in maniera significativa la distribuzione spaziale delle preferenze dei turisti. I coefficienti delle variabili “vicinanza al centro storico della città”, “distanza dall'aeroporto” e “vicinanza alla spiaggia”, che riguardano la localizzazione geografica della strutture ricettive, assumono quasi sempre valori rilevanti (*p-value* inferiore al 5%) e segno positivo. Le variabili “numero di edifici storici” e “numero di ristoranti e servizi turistici” non vengono ritenuti importanti a causa dei valori di *p-value* superiori al 10%, mentre la variabile “ettari di aree naturali protette” mostra un coefficiente significativo per l'8% e segno positivo. Nel complesso, tali risultati suggeriscono che il valore che la preferenza dei turisti assume spazialmente è quantitativamente influenzata dalle variabili esplicative contenute nel modello. I valori dei coefficienti, legati alle variabili che descrivono la posizione geografica delle strutture ricettive e alla

presenza dei servizi, risultano i maggiori responsabili delle variazioni delle preferenze turistiche nello spazio.

Discussione e conclusioni

Lo studio condotto dimostra come le informazioni provenienti dai social media rappresentino un valido supporto per processi decisionali in ambito della pianificazione turistica. L'analisi è stata condotta su due dimensioni e due scale con l'obiettivo di esplorare le preferenze turistiche sulla destinazione e sull'offerta turistica. La metodologia sviluppa analisi e tecniche spaziali per dimostrare come l'informazione volontaria proveniente dai social media può essere integrata con i dati ufficiali e gestita in ambiente GIS.

I risultati ottenuti attraverso questo nuovo e integrato approccio possono offrire interessanti prospettive di studio verso lo sviluppo di analisi più specifiche, che mirano alla percezione degli utenti sui luoghi e sull'offerta turistica. Questi forniscono infatti indizi importanti sulle dinamiche del turismo sardo, individuando le destinazioni turistiche più popolari, le peculiarità territoriali maggiormente apprezzate dai turisti e quale ruolo assumono i servizi offerti. Reperire tali informazioni mediante il solo utilizzo dei dati ufficiali è complicato e temporalmente dispendioso. Acquisite tali informazioni, si possono definire delle linee strategiche preferenziali su cui orientare la futura pianificazione turistica, fortemente contestualizzata grazie al valore aggiunto di tale conoscenza.

Infine lo studio proposto fornisce diversi spunti di riflessione sul tema della partecipazione nei processi di pianificazione con riferimento al turismo. La raccolta e l'analisi degli SMGI rappresentano una nuova ed importante opportunità per integrare il sapere tecnico con la conoscenza esperienziale delle comunità locali, permettendo di supportare le fasi progettuali, analitiche e decisionali dei processi di pianificazione e *governance*. Le opinioni che gli utenti hanno sui luoghi turistici favoriscono nuove metodologie di pianificazione e supportano processi decisionali e futuri scenari di sviluppo territoriale, aprendo inoltre la strada a nuove opportunità di ricerca.

Riferimenti bibliografici

- Benitez, J. M., Martín, J. C., & Román, C. (2007). Using fuzzy number for measuring quality of service in the hotel industry. *Tourism management*, **28**(2): 544-555.
- Briassoulis, H. (2002). "Sustainable tourism and the question of the commons," *Annals of Tourism Research* **29**: 1065-1085
- Buhalis, D., & O'Connor, P. (2005). "Information communication technology revolutionizing tourism", *Tourism recreation research*, **30**: 7-16.
- Conrady, R. (2007). "Travel technology in the era of Web 2.0", *Trends and Issues in Global Tourism 2007*: 165-184, Springer Berlin Heidelberg.
- Cooper, C. (2008). *Tourism: Principles and practice*. Pearson education.
- Craglia, M., & Campagna, M. (2010). "Advanced Regional SDIs in Europe: comparative cost-benefit evaluation and impact assessment perspectives," *International Journal of Spatial Data Infrastructures Research*, **5**: 145-167.
- Elwood, S., Goodchild, M.F. & Sui, D.Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, **102**(3): 571-590. doi: 10.1080/00045608.2011.595657
- Fotheringham, A.S., Brunson, C., Charlton, M.E. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley and Sons: West Sussex, United Kingdom.
- Getis, A. and Ord, J. K. (1992). "The analysis of spatial association by use of distance statistics." *Geographical analysis*, **24**: 189-206.

- Goodchild, M. (2007). "Citizen as voluntary sensors: spatial data infrastructure in the world of Web 2.0," *International Journal of Spatial Data Infrastructures*, **2**: 24-32.
- Gräbner, D., Zanker, M., FliedL, G., & Fuchs, M. (2012). *Classification of customer reviews based on sentiment analysis*. na.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Jebara, T. (2009). "Life in the network: the coming age of computational social science". *Science* (New York, NY), 323(5915), 721.
- Miguéns, J., Baggio, R., Costa, C.(2008). Social media and tourism destinations: TripAdvisor case study. IASK ATR 2008 (*Advances in Tourism Research*), Aveiro, Portugal, May 26-28, 2008: 26-28.
- Moran, P.A.P.(1950). Notes on continuous stochastic phenomena. *Biometrika*, 14, 17–33.
- Peroni, G., Formisano, M., Matiddi, M., & Sfodera, F. (2008). *Fondamenti di marketing turistico territoriale*. Milano: Franco Angeli
- Sui, D., & Goodchild, M. (2011). "The convergence of GIS and social media: challenges for GIScience," *International Journal of Geographical Information Science*, **25**: 1737–1748.