Fuori Luogo

Rivista di Sociologia del Territorio, Turismo, Tecnologia

Numero 5 - Giugno 2019
FedOA - Federico II University Press
ISSN (on line) 2723-9608 - ISSN (print) 2532-750X

Fuori Luogo
Rivista di Sociologia del Territorio, Turismo, Tecnologia
Numero 5 - Giugno 2019

SOMMARIO

Turismi e consumi culturali LGBT, dati, flussi e resilienza
Fabio Corbisiero

Perché devo dire qual è il mio orientamento sessuale se voglio farmi semplicemente una vacanza.
L'esperienza dei giovani turisti non eterosessuali italiani
Roberta Bartoletti e Lorenzo Giannini

I giovani LGBT+ nell'Italia contemporanea tra consumi culturali e tecnologie relazionali
Fabio Corbisiero e Carmine Urciuoli

Una questione di matrici: dal questionario cartaceo alle web-survey
Cleto Corposanto e Beba Molinari

Reti resilienti e qualità della vita: formale e informale nelle pratiche di comunità
Luigi delle Cave

Campania university students' motivations to migrate
Francesco Santelli

Rubriche

Interstizi. Letture a 3 T

Incontro Fuori Luogo. Intervista a Salvatore Monaco
Carmine Urciuoli

ISSN (on line) 2723-9608
ISSN (print) 2532-750X

FUORI LUOGO

Numero 5
Giugno 2019

Sommario

6. Turismi e consumi culturali LGBT, dati, flussi e resilienza
Fabio Corbisiero

8. Perché devo dire qual è il mio orientamento sessuale se voglio farmi semplicemente una vacanza. L'esperienza dei giovani turisti non eterosessuali italiani
Roberta Bartoletti e Lorenzo Giannini

22. I giovani LGBT+ nell'Italia contemporanea tra consumi culturali e tecnologie relazionali
Fabio Corbisiero e Carmine Urciuoli

41. Una questione di matrici: dal questionario cartaceo alle web-survey
Cleto Corposanto e Beba Molinari

54. Reti resilienti e qualità della vita: formale e informale nelle pratiche di comunità
Luigi delle Cave

71. Campania university students' motivations to migrate
Francesco Santelli

Rubriche

86. Interstizi. Letture a 3 T

94. Incontro Fuori Luogo. Intervista a Salvatore Monaco
Carmine Urciuoli

DIRETTORE
Fabio Corbisiero (Università degli Studi di Napoli Federico II)
✉ direttore@fuoriluogo.info

CAPOREDATTTORE
Carmine Urciuoli
✉ caporedattore@fuoriluogo.info

COMITATO SCIENTIFICO
Fabio Amato, Enrica Amaturo, Biagio Aragona, Elisabetta Bellotti, Erika Bernacchi, Kath Browne, Gilda Catalano, Manuela Cipri (†), Matteo Colleoni, Domenica Farinella, Mirella Giannini, Mariano Longo, Noureddine Harrami (†), Mara Maretti, Giuseppe Masullo, Antonio Maturo, Khalid Mouna, Pierluigi Musarò, Katherine O'Donnell, Giustina Orientale Caputo, Gaia Peruzzi, José Ignacio Pichardo Galán, Cirus Rinaldi, Elisabetta Ruspini, Lello Savonardo, Roberto Serpieri, Sarah Siciliano, Annamaria Vitale

COMITATO DI REDAZIONE
Francesco Antonelli, Francesco Calicchia, Amalia Caputo, Linda De Feo, Teresa De Rosa, Monica Gilli, Rosanna Marino, Ilaria Marotta, Pietro Maturi, Dario Minervini, Salvatore Monaco, Santina Musolino, Mirella Paolillo, Emanuele Rossi, Francesco Santelli, Carmine Urciuoli, Anna Maria Zaccaria

English text editor: Pietro Maturi

✉ redazione@fuoriluogo.info
tel. +39-081-2535883

English text editor: Pietro Maturi
Impaginazione a cura di Michele Adriano Brunaccini
Grafica di copertina di Michele Adriano Brunaccini

EDITORE

FedOA - Federico II University Press
Centro di Ateneo per le Biblioteche "Roberto Pettorino"
Università degli Studi di Napoli Federico II

Responsabilità editoriale
Fedoa adotta e promuove specifiche linee guida in tema di responsabilità editoriale, e segue le COPE's Best Practice Guidelines for Journal Editors.

Autorizzazione del Tribunale di Napoli n. 59 del 20 dicembre 2016
Direttore responsabile: Carmine Urciuoli
ISSN 2723-9608 (pubblicazione on line)
ISSN 2532-750X (pubblicazione cartacea)

Nella valutazione dei lavori proposti, la rivista segue una procedura di peer review. Gli articoli vengono proposti alla valutazione di due referee anonimi dopo aver eliminato ogni eventuale elemento che possa identificare l'autore.
I lavori della rivista sono consultabili su www.fuoriluogo.info

I diritti di traduzione, di memorizzazione elettronica, di riproduzione e di adattamento anche parziale, con qualsiasi mezzo, sono riservati per tutti i Paesi. Non sono assolutamente consentite le fotocopie senza il permesso scritto dell'Editore.

Una questione di matrici: dal questionario cartaceo alle web-survey*

Cleto Corposanto e Beba Molinari[1]

1. Introduzione

La pervasività con la quale le indagini online sono presenti non solo nei social, ma anche all'interno di siti di diversa natura, è impressionante. È sempre più facile navigare e incontrare un sondaggio d'opinione di testate giornalistiche più o meno conosciute, una customer satisfaction, un'indagine conoscitiva proposta online, indifferentemente dal device (notebook, smartphone, tablet) che si sta utilizzando. È indubbio che il legame che ci vincola ai dispositivi mobili è tale per cui è sempre più complesso allontanarsi dal loro utilizzo: per tale motivo, nel momento in cui un'applicazione mobile e/o una pagina web ci negano l'accesso se non tramite la compilazione di un "breve sondaggio d'opinione", l'utente medio risponde con estrema facilità; potremmo in realtà chiederci anche con quale livello di attenzione siano fornite tali risposte. La rete oggi non è più un luogo virtuale al quale accedere da una postazione fissa, ma è sempre più occasione di incontro, ed in alcuni casi anche di reciproco scambio di informazioni: da semplice vetrina a luogo di costruzione sociale dal basso di opinioni, condivisione di file di qualsiasi genere e natura (Corposanto, Molinari, 2015b), un esempio tra tutti è l'*e-democracy* che pervade discussioni in agorà digitali (Molinari, 2017) in cui molti dei commenti discussi sono riportati anche all'interno dei media tradizionali (Boccia Artieri, 2015).
Ci siamo quindi interrogati in merito alle nuove opportunità offerte dal web e sulle nuove tecniche di approfondimento messe a disposizione dalla cosiddetta *web research* (Bethlehem e Biffignandi, 2012). Nel presente contributo ci siamo posti l'obiettivo di approfondire da un punto di vista prettamente operativo in quali formati i big data siano disponibili per il ricercatore. Riferendoci quindi all'estrapolazione dei dati, ci interrogheremo attorno ai seguenti quesiti: quali forme hanno le matrici, sono le stesse di un tempo, cambiano il loro formato, è possibile effettuare ugualmente le analisi che potremmo definire "tradizionali" di analisi bivariata e multivariata? (Marradi, 1992-1997).
Al fine di poter discutere attorno a tali interrogativi, saranno presi a titolo d'esempio alcuni dati scaturiti da diversi formati di big data: web survey somministrata attraverso il supporto di una piattaforma online, web survey alla quale si può accedere esclusivamente attraverso un'Applicazione Mobile (App) ed i relativi dati di natura secondaria generati dell'App stessa, analisi di rete svolta attraverso i social network, mappatura dei siti di maggior interesse degli internauti ed infine dati generati da una mappa di geo-localizzazione (cartografia).

2. Matrici e big data: quali scenari si svelano agli occhi del ricercatore

* Data ricezione 26/04/2019. Versione finale approvata 6/06/2019.
[1] Università di Catanzaro.

La presentazione di differenti studi ci permette di discutere assieme in merito alle forme che i big data possono assumere a seconda dello strumento di rilevazione utilizzato sul web, che sia esso considerato di natura 1.0 o 3.0: spazieremo quindi dai *data warehouse* di organizzazioni come ISTAT e Organizzazione Mondiale della Sanità, considerate parte del mondo del Business Intelligence and Analytics 1.0, ossia strumenti risalenti ai primi anni duemila, fino a discutere di web survey veicolate attraverso le App, considerate la frontiera della web research 3.0, grazie al Mobile and sensor-based content (Chen, 2012).
La chiave interpretativa che adotteremo verte attorno alle informazioni fornite dal web nelle sue diverse forme d'interrogazione e quanto queste siano riconducibili a quelle che solitamente siamo abituati ad analizzare dai dati forniti dalla somministrazione di un questionario paper & pencil, ossia matrici casi per variabili (CxV), oppure variabile per variabile o valore per valore (Delli Zotti, 1985).
Se consideriamo la matrice come "una qualsiasi disposizione di informazioni (numeriche e non) ordinate per riga e per colonna" (Di Franco, 2005, p. 34) potremmo dire che sono innumerevoli le matrici di cui potremmo disporre attraverso il web, così come i software che potremmo utilizzare per lo svolgimento delle analisi, dovremmo anche confrontarci con un *ordine* della matrice molto più ampio, dove il prodotto delle righe per le colonne raggiunge una numerosità tale per cui i vettori si presentano spesso con *n* elementi. La differenza riguarda sostanzialmente, a nostro avviso, alcuni aspetti tradizionali (Di Franco, 2005, p. 37) che andremo a discutere nei prossimi paragrafi:

1. l'operativizzazione dei concetti oggetto di studio,
2. il progetto della matrice dei dati,
3. la codifica delle informazioni,
4. l'immissione dei dati in matrice,

a cui noi aggiungiamo altri due aspetti riferiti a:

5. l'estrapolazione dei dati,
6. l'analisi.

I primi cinque punti saranno oggetto di approfondimento attraverso alcuni esempi concreti di matrice, l'ultimo punto relativo alle analisi sarà oggetto di discussione in fase conclusiva: cercheremo quindi di "coprire" le aree dei big data maggiormente utilizzate ai giorni nostri, dove per big data consideriamo database di grandi dimensioni che rispondano ai primi tre criteri volume, velocità, varietà (Laney, 2001; Burrows e Savage 2014), ai quali si sono aggiunti altri ad alcuni anni di distanza altri due, valore e veridicità (Kitchin 2014).
Quello che ci preme sottolineare, e che ci interessa in particolar modo in relazione alla discussione del presente contributo, è che all'interno di tali database non si trovano esclusivamente variabili in formato numerico, bensì all'interno della definizione da noi adottata, ed ampiamente condivisa a livello internazionale (Ward, Barker 2013; Intel 2012, IBM 2013, Bennato 2015), i big data possono essere organizzati anche in formato immagine, metadato, testo e suono. Ciò ovviamente va a modificare la configurazione di quella che solitamente siamo abituati a conoscere come la matrice CxV (Corposanto Molinari, 2015a).

3. Dalle matrici ottenute con le Web survey, passando per le applicazioni mobili fino ai data warehouse delle Istituzioni Pubbliche

Iniziamo con le matrici che a nostro avviso assomigliano in modo piuttosto evidente alla matrice CxV, ossia le matrici derivanti da web survey.
Punto di inizio dei prossimi paragrafi sarà quello di confrontare i cinque aspetti sopramenzionati, riservando l'area dedicata alle analisi alla discussione conclusiva.

	Operativizzazione	**Progetto della matrice**	**Codifica**	**Immissione dati**	**Estrapolazione**
Web survey	=	=	≠	≠	=
Web survey con APP	=	=	≠	≠	=

Tabella 1: corrispondenza tra matrici derivanti da Web survey e Web survey con Applicazioni Mobili rispetto alla matrice classica CxV.

Partendo dalla fase di operativizzazione dei concetti, possiamo evidenziare che non si riscontrano grandi differenze: il ricercatore fin da principio "costruisce" il questionario attraverso la formulazione delle domande che ritiene essere più idonee ad approfondire l'oggetto di studio. In tal modo, di riflesso, anche la progettazione della matrice è già inserita in questa fase, in quanto la costruzione del questionario all'interno della piattaforma definisce in tempo reale anche la matrice nella quale confluiranno i dati. La codifica invece cambia, perché varia a seconda della piattaforma online che utilizziamo per la rilevazione: le piattaforme che prestano maggiore attenzione agli aspetti metodologici codificano i dati in modo piuttosto strutturato, cercando di rispecchiare il *codebook* definito in fase di progettazione del questionario stesso. Altre, invece, non utilizzando indicatori numerici per definire lo stato sulla variabile corrispondente, creano invece delle matrici nelle quali si presentano le risposte in forma lessicale (Corposanto Molinari, 2014).
Discutendo tali specificità siamo quindi già passati a illustrare uno degli aspetti che caratterizzano la fase di immissione/inserimento dei dati, l'altro aspetto riguarda il fatto che non sarà più il ricercatore che inputa i dati, bensì sarà il software che, nel momento in cui riceve le risposte, trasporta i dati direttamente in matrice. Estrarre i dati dal software è piuttosto semplice, la procedura cambia a seconda della piattaforma utilizzata, alcune di queste permettono l'estrapolazione di una matrice in Excel, altre in SPSS, altre ancora in Access, Stata, R, etc.
Fino ad ora, quindi, non vi sono grandi difficoltà, sempre che il ricercatore conosca bene la piattaforma che intende utilizzare e predisponga nel modo più opportuno la formulazione delle domande, senza dimenticare la tipologia di analisi che intende effettuare successivamente, così come d'altronde abbiamo sempre fatto per i questionari paper & pencil (Corposanto Molinari, 2014).
Poter dire di estrarre i dati in un formato di file piuttosto che in un altro è semplice a dirsi, ma non è detto che la matrice che appare ai nostri occhi sia facile da utilizzare attraverso i software di elaborazione dei dati. Solitamente il formato file maggiormente scaricato per l'estrapolazione, e consigliato dalle piattaforme stesse, è il formato Excel, che può essere facilmente trasformato in formati accettati da SPSS, STATA, R, etc. ed è proprio in questa fase che cominciano le difficoltà: si rischia infatti di avere una non semplice trasposizione della matrice a causa di valori cosiddetti "nascosti" come punti, virgole, spazi, etc. Tali

problematiche non inficiano l'utilizzo dei dati, ma sicuramente rendono più complessa la procedura: diviene necessaria una pulizia minuziosa della matrice che spesso richiede un dispendio di tempo ragguardevole, ci sarebbe quindi da chiedersi se il tempo che si è guadagnato nel non dover editare i dati in matrice, non si perda in seconda battuta per la pulizia degli stessi.
Oltre alla pulizia del dato è necessario fare una particolare attenzione anche al passaggio relativo alla classificazione assegnata alla variabile, che non sempre rispecchia il volere del ricercatore. Il più classico tra gli esempi è quello di dover trasformare una variabile da stringa, con la quale si riconoscono le risposte aperte, ad una nominale, ordinale o cardinale.
Nella tabella precedente è stata effettuata una distinzione tra web survey e web survey veicolate attraverso le App: entriamo quindi ora nel merito delle specificità di quest'ultima forma alternativa di somministrazione.
La prima grande differenza riguarda la numerosità delle domande che possono essere inserite all'interno della survey: la piattaforma utilizzata è direttamente connessa alla programmazione dell'applicazione mobile che, spesso, ha uno spazio dedicato alla survey piuttosto limitato, ma ciò non significa che non si possano avere a disposizione grandi quantità di dati.
Il questionario somminsitrato tramite applicazione mobile appare a video all'utilizzatore finale nel momento in cui accede all'applicazione per un numero di domande che spazia attorno alla ventina. I dati quindi a nostra disposizione saranno di due tipi: un primo derivante direttamente dalle domande del questionario e un secondo riconducibile a una serie di informazioni che l'applicazione mobile incorpora autonomamente e che un tempo era considerato "rumore statistico".

Figura 1: Esempio di Dashboard di Google Analytics. Fonte: nostra elaborazione su studio di caso

Per l'estrapolazione dei dati dovremo avvalerci di Google Analytics dove, dopo l'inserimento di diversi parametri attraverso la dashboard, avremo modo di accedere e scaricare la matrice. Anche in questo caso, potranno verificarsi una serie di problematiche riguardanti i simboli

nascosti che richiedono, come già precedentemente descritto, un'opportuna e accurata pulizia del dato.

	Operativizzazione	Progetto della matrice	Cod.	Immissione dati	Estrapolazione
App dati di natura secondaria	n.p.[2]	n.p.	n.p.	n.p.	=
Banche dati da Fonti istituzionali	n.p.	n.p.	n.p.	n.p.	=

Tabella 2: corrispondenza tra matrici derivanti da banche dati rilevate attraverso applicazioni mobili e banche dati di natura secondaria provenienti da fonti istituzionali rispetto alla matrice classica CxV.

Differenti invece sono i dati di natura secondaria derivanti dall'uso di applicazioni sul web: nel momento in cui si accede a un'Applicazione mobile in qualità di usufruitori del servizio, il sistema registra infatti una serie di informazioni riguardanti la tipologia di connessione, l'IP con il quale si accede, la geolocalizzazione e le consultazioni effettuate all'interno dell'applicazione stessa, che molto ci possono dire dei comportamenti/servizi per i quali è stato effettuato l'accesso.
Questi dati però non sono organizzati in forma matriciale, sarebbe d'altronde estremamente difficile analizzarli attraverso i software che siamo abituati ad utilizzare, perché siamo nell'ambito di una vera e propria *data explosion*: consideriamo che un'applicazione mobile di media consultazione registra mediamente più di 100 accessi al giorno. Il ricercatore può conoscere quindi una serie di informazioni che non ha richiesto in prima persona, ma che possono essere di estremo interesse a supporto di quanto emerge da una *web survey* veicolata proprio attraverso l'applicazione mobile in questione, a completamento di una visione complessiva.
Per poter accedere a questi dati il ricercatore dovrà far sì che colui che ha predisposto l'applicazione mobile, ossia il programmatore informatico, gli dia pieno accesso a tali informazioni. Dopodiché il ricercatore sarà accreditato per accedere a Google Analytics ed effettuare tutte le interrogazioni che considera più opportune.
Proprio perché stiamo discutendo di estrapolazione dei dati, questa tipologia d'informazioni è accessibile esclusivamente attraverso la dashboard, che poco ha a che fare con una matrice CxV. Non avere a disposizione la matrice non significa però non poter analizzare i dati: la dashboard ci permette di effettuare analisi bivariate, costruire grafici, definire sotto-aree di interesse predisponendo tali criteri di approfondimento nel tempo, di modo che al prossimo accesso alla dashboard siano già presenti i parametri precedentemente impostati. Il limite è riferito al fatto che Google Analytics non permette analisi multivariate, ma di fatto i big data di natura secondaria, ossia derivanti da API, non permettono analisi di questo tipo se non opportunamente *"trattati"* con algoritmi specifici (Neresini, 2015).

[2] La non pertinenza è riferita al fatto che le matrici derivanti dalle applicazioni mobili e dalle banche dati provenienti da fonti istituzionali sono già state predisposte ex-ante dal software specifico e dai ricercatori per altri scopi a fini conoscitivi del fenomeno oggetto di studio.

4. Web structure mining: la matrice quella sconosciuta

Se fino a questo momento abbiamo discusso di matrici molto simili a quelle che siamo abituati ad utilizzare, tranne che per i dati derivanti dall'utilizzo delle applicazioni mobili, ora la situazione è differente, così come il procedimento.
Entriamo nel mondo degli Application Programming Interface (API) (Monovich, 2011) rispetto al quale andremo ad approfondire il *web structure mining,* con il quale è possibile svolgere un'analisi di rete dalla quale emerge l'analisi dei reticoli prodotti dalle relazioni tra gli utenti ed i siti web (Hansen, Shneiderman e Smith, 2011).
Prendiamo a titolo d'esempio l'ego rete derivante da Facebook: la prima cosa che ci appare evidente è che effettueremo un procedimento molto differente da quello precedentemente discusso: la matrice sussiste già precedentemente ai fini della programmazione del social, può essere richiesta dal ricercatore, attraverso un apposito software (*netvizz*), al fine di effettuare delle analisi particolareggiate su alcuni aspetti piuttosto che altri (differenze di genere, nazionalità, interesse verso un determinato post, definizione della ego rete, etc.). La matrice diviene quindi soltanto mezzo attraverso il quale ottenere i focus di studio che ci interessano.
I dati così come siamo abituati ad analizzarli, siano esse variabili nominali, ordinali o cardinali, sono visibili solo alla fine di un procedimento piuttosto complesso e, potremmo aggiungere che, di fatto non sono stati richiesti direttamente dal ricercatore. Ma andiamo con ordine e discutiamo tale procedimento con riferimento ai nostri 5 aspetti che caratterizzano la definizione di una matrice CxV.

	Operativizzazione	**Progetto della matrice**	**Cod.**	**Immissione dati**	**Estrapolazione**
Facebook ed ego rete	n.p.	n.p.	n.p.	n.p.	≠

Tabella 3: corrispondenza tra una matrice proveniente da Facebook rispetto alla matrice classica CxV.

Dalla tabella appare evidente che il ricercatore ha la possibilità di intervenire solo nella fase relativa all'estrapolazione dei dati, che non consiste nell'ottenere un file formato Excel, ma avere in restituzione un file estraibile esclusivamente attraverso programmi di analisi di rete. Il primo passaggio da effettuare per ottenere tale file è quello di accedere all'interno di Facebook utilizzando un'area dedicata alla "programmazione", ossia *netvizz* nel quale inseriremo attraverso la sintassi una serie di comandi in cui sarà richiesta la tipologia di informazioni di nostro interesse: l'ego rete a partire da una determinata pagina e/o da un determinato utente, sempre che l'oggetto di nostro interesse non abbia apportato limitazioni specifiche nell'area *Privacy*. In tal caso la possibilità di elaborazione delle informazioni è fortemente limitata, ma questo è un problema che può sorgere per qualsiasi social media per il quale sia prevista la possibilità di limitazione degli accessi esterni.
Richiesti i giusti parametri, sarà scaricato un file contenente una serie di informazioni a cui non è possibile accedere se non trasportando il file all'interno di un software specifico tra quelli esistenti (es. Ghepi, NodeXL Pajek). Elaborato il file attraverso il software avremo accesso ad una matrice dalla quale sarà possibile effettuare le dovute analisi, organizzato secondo la configurazione n*p. (nodi per proprietà) Ciò che differenzia questa forma

matriciale dalle altre è riferito al fatto che i dati in essa contenuti non sono stati effettivamente definiti dal ricercatore, se non per i pochi parametri richiesti nella sintassi. La matrice in realtà è sempre esistita già a partire dal procedimento impostato su *netvizz*, ma a meno che non si abbiano le dovute conoscenze informatiche di programmazione, non è possibile aprirle in un formato leggibile fino a quando, la matrice non è ulteriormente elaborata con un software che abbia un'interfaccia grafica.

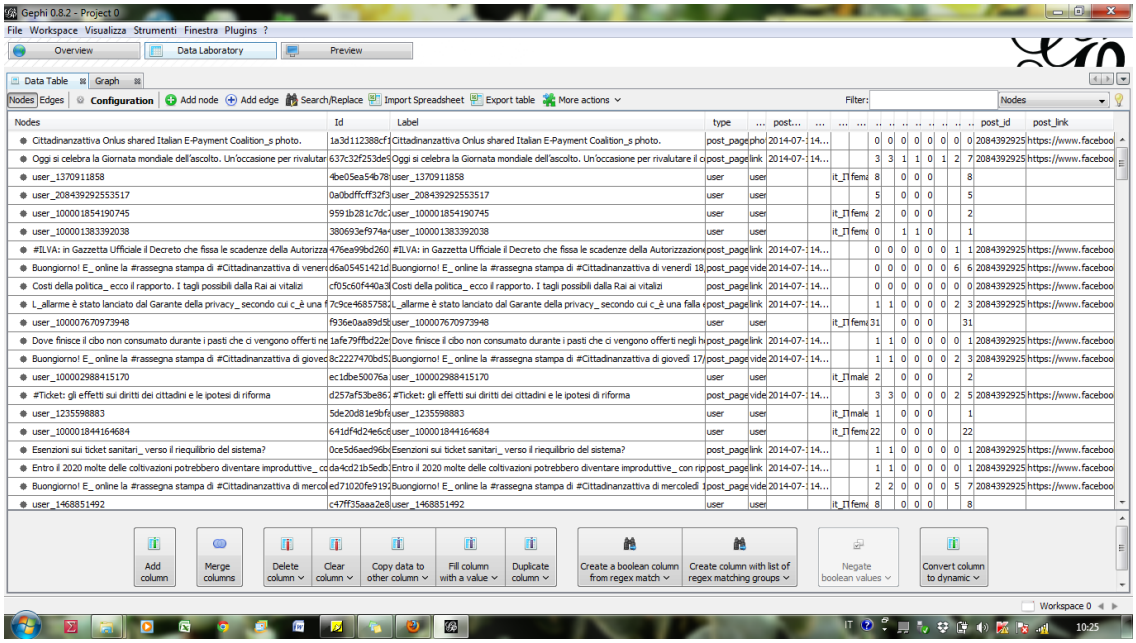

Figura 2: esempio di matrice di una ego-rete di una pagina facebook. Fonte: nostra elaborazione su studio di caso

Resta comunque il fatto che le variabili che avremo modo di osservare sono piuttosto complesse da riconoscere: l'etichettatura è differente da quella che siamo soliti avere a disposizione, non siamo infatti noi ad averle editate; potremo quindi distinguere tra variabili stringa, etichette riferite alle pagine visitate, alla direzione dei legami, ai valori booleani di questi ultimi, ai commenti, ai like, etc. tutto *mescolato* assieme in una matrice di affiliazione, molto simile ad una matrice CxV.

Oltre ad avere le dovute conoscenze in merito all'analisi di rete, è necessario essere altresì a conoscenza degli algoritmi più idonei per elaborare i dati e fare emergere da una data moltitudine di informazioni una rete chiara e facilmente comprensibile. Quindi, in ultima istanza, dopo attenti procedimenti e dovute conoscenze anche di natura informatica, avremo a disposizione una serie di variabili di natura quantitativa che richiede conoscenze riferite all'analisi di rete: «*l'impostazione può essere definita a tutti gli effetti di sociologia matematica, in quanto l'obiettivo è quello di modellizzare strutture sociali dotate di differenti proprietà partendo dalla teoria matematica dei grafi e utilizzando l'algebra delle matrici. L'approccio è quindi astratto e deduttivo, piuttosto che sostanziale ed applicativo ai problemi di interpretazione sociologica. L'elemento che unifica i diversi contribuiti (è) [...] il metodo della network analysis*». (Chiesi 1999, p. 46).

5. Web usage mining: possiamo ancora parlare di matrice?

Il web usage mining ci permette di conoscere le preferenze di ricerca nella rete dove, anche in questo caso, giocano un ruolo di rilievo gli API, senza i quali non potremmo conoscere tali preferenze.
Il Web usage mining è quello che tra i tre è maggiormente conosciuto anche dalla gran parte degli utenti della rete: le analisi più estreme si basano su modelli predittivi come *Google Trend*[3] ideato da Google al fine di poter prevedere con un certo anticipo l'andamento di alcuni fenomeni.
In questo paragrafo si discute della mappatura dei siti internet a partire da un lemma a nostra scelta inserito all'interno di un'area di un motore di ricerca qualsiasi, che nel nostro caso specifico fa riferimento a Google, ma rispetto al quale avremmo potuto utilizzare yahoo, safari, etc. Al riguardo è possibile tracciare il percorso dell'utente medio, ossia tracciare le consultazioni dei vari siti, a partire dal lemma oggetto di studio. Tale percorso prende la veste di un grafo, sfruttando le proprietà della *network analysis*.
Il procedimento però, per l'estrapolazione dei dati è ancora più complesso del precedente, ma il risultato è una matrice molto simile a quella del paragrafo sopradescritto, ossia avremo modo di lavorare su di una matrice di affiliazione.
Anche in questo caso il ricercatore non interviene se non nell'ultima fase riferita ai 5 aspetti della definizione/costruzione della matrice, spariscono dal processo quindi l'operativizzazione, il progetto della matrice, la codifica e l'immissione dei dati[4], resta l'estrapolazione che si configura come un'operazione piuttosto complessa che richiede buone conoscenze di programmazione.

	Operativizzazione	**Progetto della matrice**	**Cod.**	**Immissione dati**	**Estrapolazione**
Mappatura di siti internet con Web crawler	n.p.	n.p.	n.p.	n.p.	≠

Tabella 4: corrispondenza tra una matrice proveniente dalla mappatura di siti internet rispetto ad una matrice classica CxV.

Al riguardo viene usualmente utilizzato il web crawler[5] (fig. 3) che ci permette di ottenere informazioni attraverso alcuni parametri da noi impostati. Possiamo quindi decidere il tempo medio di consultazione all'interno di un sito, prima che questo sia abbandonato per la consultazione in un altro, possiamo escludere i ping, le sotto-aree delle pagine web, ossia la navigazione del sito attraverso i menù a tendina, etc., il file così ottenuto può essere inserito all'interno del software *Gephi* e risulta pronto per l'elaborazione.
I file contenente i dati che si trasformeranno in matrice saranno anche in questo caso stringhe ed etichette, in modo molto simile al caso precedente, ma con un numero inferiore di variabili: scompariranno le variabili dedicate ai like, ai commenti che lasceranno lo spazio

[3] https://trends.google.it/trends/?geo=IT
[4] La fase relativa alla codifica dei dati è svolta dal programmatore in fase di definizione del software.
[5] A dispetto di quanto si possa pensare i programmi di web crawler nascono già nei primi anni 90, la loro diffusione alla grande massa avviene però solo negli ultimi anni al fine di effettuare le analisi sopradescritte.

alle variabili riguardanti i diversi canali di ricerca, le piattaforme attraverso le quali si accede alla ricerca, la forza dei legami tra le pagine, ossia variabili comprensibili solo ad un occhio esperto che abbia conoscenze di *network analysis* e di linguaggio di programmazione.

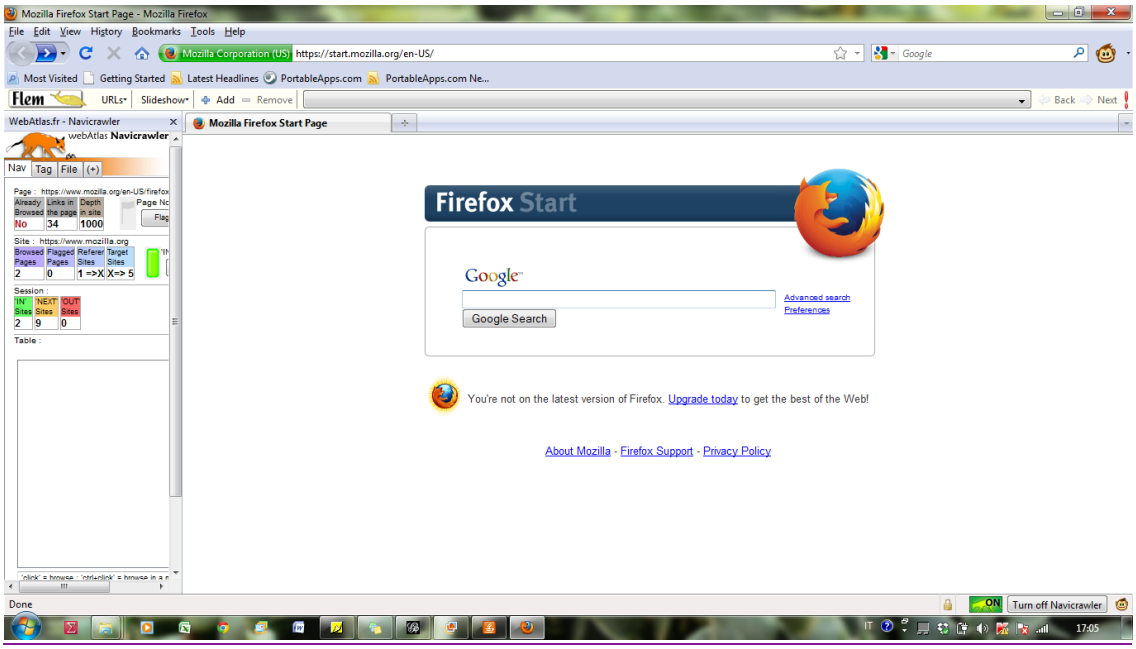

Figura 3: esempio di pagina di web crawler. Fonte: nostra elaborazione su studio di caso

6. Web Contest Mining: la poliedricità delle analisi online

Abbiamo lasciato per ultimo il *web contest mining* perché in questo caso il ricercatore non avrà alcun tipo di matrice da elaborare ed inserire in un software di analisi dei dati, ma il procedimento avverrà tutto all'interno del web, grazie alle competenze del ricercatore, che sempre più si avvicinano a quelle di un programmatore informatico.

	Operativizzazione	**Progetto della matrice**	**Cod.**	**Immissione dati**	**Estrapolazione**
Rappresentazione geo-localizzata / Cartografia	n.p.	n.p.	n.p.	n.p.	n.p.

Tabella 5: corrispondenza tra una matrice proveniente da una rappresentazione geo-localizzata rispetto alla matrice classica CxV.

I cinque aspetti – operativizzazione, progetto della matrice, codifica, immissione dati, estrapolazione – non sono in alcun modo considerati, utilizziamo infatti dati che non sono stati operativizzati, né codificati, né tantomeno estrapolati, è possibile invece salvare l'output, ossia la cartografia nei formati più diversi, pdf, immagine, ma non la matrice, che resta a noi ignota. Questa tipologia di analisi ricorda molto più una restituzione in forma grafica dell'analisi, piuttosto che un'analisi *tout court*, anche in questo caso gli API sono la porta di accesso a tali elaborazioni.

Su alcuni siti le analisi sono possibili all'interno del sito stesso, che diviene in tal modo strumento statistico; altre volte invece si passa da un sito ad un altro, dove in una prima pagina web si scaricano le informazioni di interesse, sempre attraverso gli API, in un altro sito invece si elaborano tali informazioni geo-localizzando l'oggetto di studio: in questo processo il web crawler diviene fondamentale tanto quanto gli API stessi.

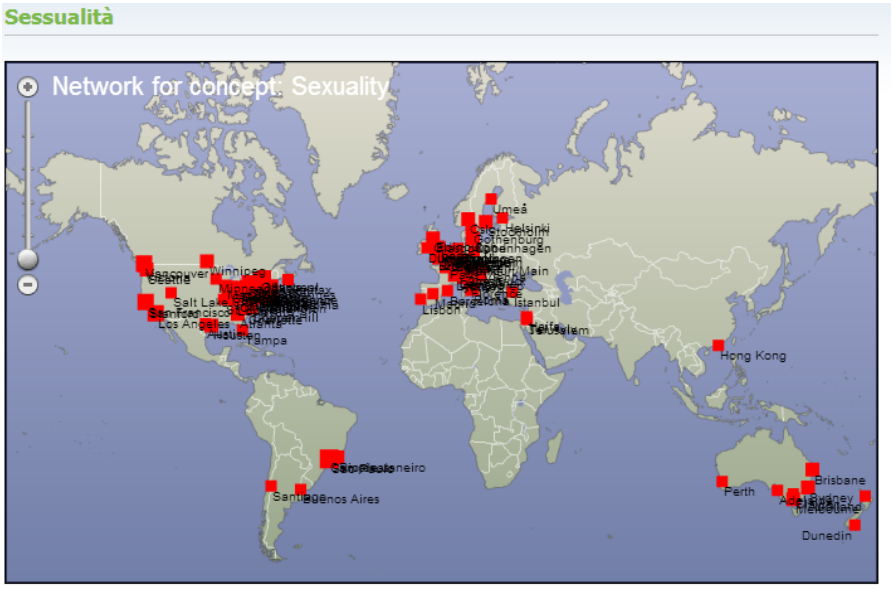

Figura 4: esempio di rappresentazione geo-localizzata.
Fonte: nostra elaborazione su studio di caso

Anche in questo caso è necessaria più competenza di programmazione che non capacità di analisi: il ricercatore dovrà quindi avere buone competenze informatiche, ma sapere anche cosa ricercare e attraverso quali canali, perché come abbiamo accennato nel terzo paragrafo, non sempre le informazioni hanno accesso illimitato a causa di eventuali restrizioni apportate dal singolo utente dagli API. In quest'ultimo caso si discute di un vero e proprio mashup di dati (Corposanto, Molinari, 2015b).

7. Considerazioni conclusive

	Operativizzazione	Progetto della matrice	Cod.	Immissione dati	Estrapolazione
Web survey	=	=	≠	≠	=
Web survey App	=	=	≠	≠	=
Mappatura di siti internet con Web crawler	n.p.	n.p.	n.p.	n.p.	≠
Banche dati da fonti istituzionali	n.p.	n.p.	n.p.	n.p.	=
Facebook e ego rete	n.p.	n.p.	n.p.	n.p.	≠
App dati di natura secondaria	n.p.	n.p.	n.p.	n.p.	=
Cartografia	n.p.	n.p.	n.p.	n.p.	n.p.

Tabella 6: corrispondenza tra diverse forme di matrice discusse nei paragrafi precedenti rispetto alla matrice classica CxV.

In fase conclusiva cerchiamo di "tirare le fila" evidenziando quanto fino ad ora descritto in merito alle differenze tra le matrici di derivazione paper & pencil rispetto a quelle scaturite dal web: grazie a tale distinzione avremo modo di discutere in merito alle analisi che tali matrici ci permettono di svolgere.
Come abbiamo modo di vedere dalla tabella riassuntiva sotto riportata, comprensiva di tutti gli esempi d'interrogazione svolti sul web presentati in questo contributo e dei cinque aspetti che caratterizzano la matrice dei dati "classica", solo le web survey e le web survey somministrate attraverso le App sono quelle che presentano una maggiore vicinanza in termini di paragonabilità: l'operativizzazione e il relativo progetto della matrice sono due tra i cinque aspetti che le accomunano. All'estremo opposto abbiamo le informazioni provenienti dalle Applicazioni mobili di natura secondaria e quelle scaturite dalle cartografie che non permettono l'estrapolazione della matrice, nonostante questa sia presente nella fase di elaborazione dei criteri di approfondimento richiesti dal ricercatore, senza la quale gli algoritmi che permettono la configurazione grafica, così come ci apparare, non sarebbero possibili. Tranne questi ultimi due casi ci teniamo a sottolineare che le matrici estrapolate dal web sono facilmente riconducibili a *items* quantitativi ed in alcuni casi, quando direttamente richiesti dal ricercatore, anche qualitativi.

Oltre alla fase di estrapolazione della matrice vi è la fase di analisi dei dati, ossia il sesto aspetto di cui non abbiamo ancora discusso.
L'estrapolazione dei dati, cui segue la pulizia dei dati stessi, e l'analisi sono la diretta conseguenza l'una dell'altra, comprese quindi le differenze che le diverse forme di estrapolazione dei dati comportano sulle rispettive matrici, cerchiamo di andare oltre e capire attraverso quali software e tipologia di analisi statistiche è possibile confrontarsi attraverso il web.
Le web-survey somministrate su piattaforme online oppure tramite applicazioni mobili e le banche dati ci permettono di svolgere le analisi così come da sempre siamo abituati a effettuarle, i software sono quelli più comunemente utilizzati per le analisi statistiche: non vi sono quindi limitazioni, per cui sarà possibile analizzare i dati sia attraverso analisi mono e bivariate, sia multivariate.

WEB 2.0	**Estrazione**	**Analisi**
Web survey (Piattaforma)	Estrazione standard – Matrice – dati quali e quanti	Tutte SPSS/SAS/STATA/R
Web survey (APP)	Google analytics – Matrice – dati quali e quanti	Tutte SPSS/SAS/STATA/R
Banche dati	Estrazione standard – Matrice – dati quali e quanti	Tutte SPSS/SAS/STATA/R
Dati (APP) natura secondaria	Google analytics – Dati quantitativi	Clustering da Google analytics
Mappatura dei siti web	API – Web crawler	Network analysis e Google analytics
APP – Mobile and sensor-based content	API – Web crawler MASH-UP	Carte topografiche, etc.

Tabella 7: estrapolazione e analisi del web 2.0

Le differenze più marcate si riscontrano con i dati di natura secondaria provenienti dalle Applicazioni mobili, le mappature dei siti web, le ego reti svolte attraverso i social media, che ci ricordano quanto sia utile e fondamentale, in questi casi, svolgere interrogazioni attraverso *Google Analytics*, senza il quale non avremmo modo di effettuare analisi e approfondimenti.
Per concludere nella web research giocano un ruolo fondamentale la presenza degli API, del web semantico e la possibilità di individuare e ricostruire le informazioni attraverso il *web crawler*. Da tali forme matriciali è possibile effettuare la *network analsysis* grazie a WEKA, Pajek, NodeXl, UCINET, etc. Infine API e *web crawler* ci permettono raffigurazioni prettamente grafiche di carte topografiche, geolocalizzazioni in formato immagine "*rappresentativo*" di una serie di altre immagini raccolte su social media (es. Pinterest) con riferimento ad oggetti di interesse specifico, etc. In questo caso tali configurazioni aprono il campo al web 3.0, ossia a quello ad oggi conosciuto come *Mobile and sensor-based content*, dove la matrice e le analisi divengono parte integrante dall'applicazione mobile stessa e il fruitore dell'applicazione mobile non è più solo colui che utilizza un servizio, bensì diviene esso stesso analizzatore della realtà di suo interesse.

Riferimenti bibliografici

Agnoli, S., Parra-Saiani, P. (2016). Sulle tracce dei big data. Questioni di metodo e percorsi di ricerca. *Sociologia e Ricerca Sociale*. n.109, pp. 1-3.
Angus, D., Rintel, S., Wiles J. (2013). Making sense of big text: a visual-first approach for analyzing text data using Leximancer and Discursis. *International Journal of Social Research Methodology*, vol. 16, n. 3, pp. 261-267.
Burrows, R., Savage, M. (2014). After che Crisis? Big Data and the Methodological Challenges of Empirical Sociology. *Big Data and Society*, n. 1, pp. 1-6.
Bethlehem J., Biffignandi S. (2012). Handbook of Web Surveys, Wiley Handbooks in Survey Methodology. New Jersey: John Wiley & Sons.
Berry D. (2011). The computational turn: thinking about the digital humanities. *Culture Machine*, vol. 12, [Online] Available at: http://www.culturemachine.net/index.php/cm/article/view/440/470 (26 April 2019).
Bennato D. (2015). *Il computer come macroscopio. Big data e approccio computazione per comprendere i cambiamenti sociali e culturali*. Milano: FrancoAngeli.
de Boeck P., Wilson M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. NY: Springer.
Boccia Artieri, G. (2015). *Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online*. Milano: FrancoAngeli.
Cardon, D. (2016). *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data*. Milano: Mondadori Università.
Chiesi A.M. (1999). *L'analisi dei reticoli, Metodologia delle scienze umane*. Milano: FrancoAngeli.
Chen, H., Chiang, R.H.L., Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly* Vol. 36 No. 4, pp. 1165-1188/December 2012
Corposanto, C., Molinari, B. (2014). "Survey e questionari online?". In Corposanto, C. Valastro, A. (a cura di). *Blog, fb e twitter*. Milano: Giuffrè.
Corposanto, C., Molinari, B. (2015a). Chasing a dragonfly on the lawn. *Science Innovation*. Vol. 3, No. 4, 2015, pp. 39-45. ISSN: 2328-7861 (Print); ISSN: 2328-787X (Online) doi: 10.11648/j.si.20150304.11
Corposanto, C., Molinari, B. (2015b). "Rilevare dati sul web: la cassetta degli attrezzi 2.0". In Sannella, A. Toniolo, F. (a cura di). *Le sfide della società italiana tra crisi strutturali e social innovation*, Venezia: Edizioni Ca'Foscari.
Corposanto C., Molinari B. (2016). Say it with un App. *Journal of Advanced Statistics.*Vol. 1, No. 2, June 2016, pp.52-56.

Corposanto C., Molinari B. (2017). Le politiche pubbliche per l'alimentazione e l'uso dei Big Data:
percorsi di valutazione condivisa. In Corposanto, C. (a cura di). *Sociologia dell'alimentazione. Diete, culture, rischi*. Napoli: Rogiosi Editore.
Corposanto C., Molinari B. (2017). Big Data and evaluation of policy. *Rassegna Italiana di Valutazione*, a. XXI, n. 68, DOI: 10.3280/RIV2017-068006
Corposanto C., Molinari B. (2018). Analizzare dati di microblogging con la Sentiment Analysis. Quale rappresentatività? *Sociologia Italiana*, n. 11, Egea.
Creswell, J.W., Plano, Clark V.L. (2011). *Designing and Conducting Mixed Methods Research*, Thousand Oaks (CA): Sage.
Delli Zotti G. (1985). Tipologia delle matrici usate nella ricerca sociale. *Rassegna Italiana di Sociologia*. XXVI, 2, pp.141-168
Di Franco G. (2005). *EDS: Esplorare, descrivere e sintetizzare i dati. Guida pratiche all'analisi dei dati nella ricerca sociale*. Milano: FrancoAngeli.
Grimmer J., Stewart B.M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(3):267–297.
Gross, JL., Yellen, J. (2004). *Handbook of graph theory*. London: CRC Press.
Harman, H.H. (1967). *Modern Factor Analysis*. Chicago: University Press of Chicago.
Hansen D.L., Shneiderman, B., Smith, M.A. (2011). *Analyzing Social Media Networks with NodeXL. Insights from a connected world*. Burlington: Morgan Kaufmann.
Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, I, n. 1, pp. 1-12.
Kosinski, M., Stillwell, D.J., Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *National Academy of Sciences of the United States Of America*, vol. 110, n. 15, pp. 5802-5805.
Kosala, R., Blockeel, H. (2000). Web Mining Resarch: a Survey. *SIGDKK Explorations*, vol. 2 n. 1. Pp. 1-15.
Kaushik, A. (2010). *Web analytics 2.0. Misurare il successo online nell'era del web 2.0*. Milano: OEPLI.
Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*, file n. 949 [online], disponibile in: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (23 April 2019).
Lombi, L. (2015). *Le web survey*. Milano: FrancoAngeli.
Manovich, L. (2011). Trending: the promises and the challenges of big social data. *Debates in the Digital Humanities*, Minneapolis: The University of Minnesota, Press. [Online] Available at:
http://www.manovich.net/DOCS/Manovich_trending_paper.pdf (26 April 2019).
Marradi, A. (1992). *L'analisi monovariata*, Milano, Franco Angeli.
Marradi, A. (1997). *Linee guida per l'analisi bivariata dei dati nelle scienze sociali*. Milano: Franco Angeli
Molinari B. (2017), Partecipation. In Lombi, L., Marzulli, M. (a cura di). *Theorising Sociology in the Digital Society*. Milano: FrancoAngeli.
Neresini, F. (2015). Quando i numeri diventano grandi: che cosa possiamo imparare dalla scienza. *Rassegna Italiana di Sociologia*, n. 3-4, pp. 405-431.
Palumbo, M., Garbarino, E. (2006). *Ricerca sociale: metodo e tecniche*. Milano: FrancoAngeli.
Papacharissi, Z. (2016). Remaking Events, Storytelling, and News. In Boczkowski, P., Anderson, C. (Eds.). *Remaking the News*. Cambridge: MIT Press.
Trobia, A. (2014). Web Mining e Application programming interfaces per la ricerca sociale: caratteristiche, strumenti, prospettive e limiti. In Corposanto, C., Valastro, A. (a cura di). *Blog, fb e twitter*. Milano: Giuffrè.
Twitter (2011) 'One hundred million voices', Twitter Blog, [Online] Available at:
http://blog.Twitter.com/2011/09/one-hundred-million-voices.html (26 April 2019).
Ward, J. S., Barker, A. (2013). Undefined By Data: A Survey of Big Data Definitions. *arXiv* 1309.5821 [online], disponibile in http://arxiv.org/abs/1309.5821.
Witten, I.H., Franck, E., Hall, M.A. (2011). *Data Mining. Practical Machine Learning Tools and Techniques*, Burlington: Morgan Kaufmann.