

BIG DATA, TRACCE DIGITALI E LE CITTÀ

IL PROGETTO URBANSOPE DEL POLITECNICO DI MILANO

Michela Arnaboldi, Giovanni Azzone, Marco Brambilla, Paolo Ciuccarelli, Piercesare Secchi, Simone Vantini
Politecnico di Milano (polimi)

Il termine “Big data” è sulla bocca di molti. Ma cosa sono questi Big Data? A cosa servono, da dove vengono? Quali sono le loro potenzialità, quali sono i loro rapporti con la statistica pubblica?

In questo articolo viene presentato un interessante progetto svolto dal Politecnico di Milano grazie ai Big Data, che ha coinvolto imprese private, enti pubblici, università e centri di ricerca con l'obiettivo di valorizzare dati oggi inutilizzati e proporre analisi e visioni innovative. Sulla scorta dell'esperienza milanese, l'articolo offre l'occasione per una riflessione su un ambito innovativo e a tratti ancora difficile da identificare per gli statistici pubblici.

Introduzione

La citazione di De Rosnay introduce il termine macroscopio, a cui si è ispirato il progetto Urbanscope del Politecnico di Milano. Il progetto vuole essere un macroscopio dell'urbano inteso in senso lato come spazio fisico, sociale, economico, istituzionale e caratterizzato da una sempre maggiore complessità. La città del XXI secolo è composta di attori e di relazioni che si interconnettono stabilmente, sistemi fisicamente anche molto distanti ma strettamente interdipendenti tra loro, dove i meccanismi relazionali diventano cruciali e dove i confini tra individuale e sociale, tra locale e globale, tra mondo reale e digitale, diventano sempre più sfumati.

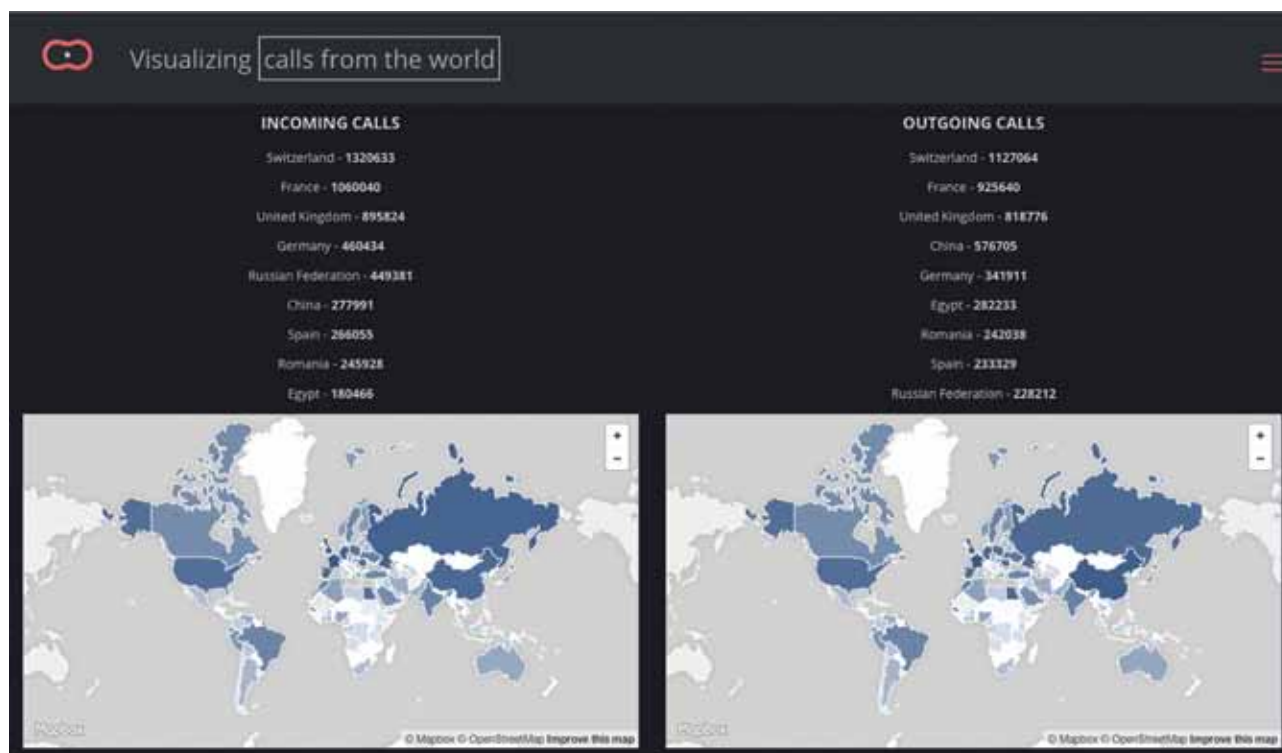
Urbanscope rivolge le sue lenti alla dimensione digitale, dinamica e complessa dei *Big data*. Un'immensa mole di dati sta crescendo in tempo reale, e si sta modificando insieme al più visibile strato fisico che definisce la città: il suo territorio, i suoi edifici e le sue infrastrutture. Il lavoro degli analisti inizia dall'osservazione delle tracce digitali che lasciamo sempre più, catturando segnali deboli generati ad esempio dalla rete di telefonia mobile o dalla disponibilità di *open data*. Urbanscope seleziona, raccoglie, analizza e visualizza sistematicamente e metodicamente queste tracce. Il metodo utilizzato mette al centro il decisore e gli *stakeholder* della città (cittadini, lavoratori, turisti,...), capendone le esigenze in modo diretto o indiretto. Con l'obiettivo di preservare la complessità del fenomeno osservato, vengono adottati

“Microscopio, telescopio. Le parole evocano le grandi esplorazioni della scienza verso l'infinitamente piccolo e verso l'infinitamente grande. Il microscopio ha permesso un tuffo vertiginoso nelle profondità del vivente, la scoperta della cellula, dei microbi e dei virus, il progresso della biologia e della medicina. Il telescopio ha aperto le menti all'immensità del cosmo, tracciato la rotta dei pianeti e delle stelle ed avviato gli uomini alla conquista dello spazio. Ci confrontiamo oggi con un altro infinito: l'infinitamente complesso. Senza strumenti questa volta. Nulla: solo un cervello nudo, una intelligenza ed una logica senz'armi di fronte all'immensa complessità della vita e della società. [...] Ci serve uno strumento, prezioso come lo furono il microscopio ed il telescopio per la conoscenza scientifica dell'universo, ma destinato, questa volta, a tutti coloro che tentano di capire il senso ed il posto della loro attività. [...] Chiamerò questo attrezzo il macroscopio (macro, grande; skopein, osservare)”
[De Rosnay, Joël (1978). *Il Macroscopio*. Dedalo libri, 7-8]

ti metodi statistici che colgono e mantengono la variabilità di dati complessi per poi interpretarli alla luce di nuovi modelli concettuali. La complessità viene veicolata attraverso visualizzazioni innovative che consentono sia una più semplice esplorazione dei fenomeni, sia un'analisi più approfondita. Questi risultati forniscono elementi utili tanto ai decisori e *stakeholder* quanto al pubblico, che ha la possibilità di ampliare la rappresentazione del mondo in cui vive. Sinteticamente, Urbanscope fornisce una visione inedita di come viviamo e usiamo la città, potenziando la nostra conoscenza, rinnovando lo spettro di domande che ci poniamo e di conseguenza permettendo di sondare vie nuove per decisioni ed azioni.

F.1

City and the world, chiamate in entrata e in uscita da Milano, per origine e destinazione, nel bimestre aprile-maggio 2015



Fonte: www.urbanscope.polimi.it

Riquadro 1 – Il progetto Urbanscope

Le città e più in generale l'urbano, sono un ambito di grande potenzialità per i *Big data*, grazie alla possibilità di raccogliere ed elaborare dati georeferenziati. Il progetto Urbanscope del Politecnico di Milano si pone in questo contesto e ha l'obiettivo di sperimentare metodologie per la valorizzazione dei *Big Data* per, e della, città. Le sperimentazioni Urbanscope hanno permesso di ottenere due risultati:

- 1) Metodo, aprendo la "scatola nera" dei dati, dei loro problemi di rilevazione ed elaborazione.
- 2) Output, ossia i dati rielaborati, visibili nella parte pubblica attraverso il sito: www.urbanscope.polimi.it.

Il sito è stato sviluppato sulla città di Milano, come esemplificativo di un città dinamica e complessa. La scelta di sviluppare il sito in lingua inglese è stata fatta esclusivamente per favorire un accesso più ampio a ricercatori, decisori e policy maker di tutto il mondo.

Nel seguito di questo articolo si illustrerà il progetto Urbanscope nei suoi due elementi principali: iniziando dall'output, ossia i dati rielaborati, visibili nella parte pubblica attraverso il sito Urbanscope¹ [Riquadro 1]; per poi presentare il metodo e infine, nelle conclusioni, si proietteranno i prossimi passi e gli scenari che iniziano a profilarsi all'orizzonte.

Output: Urbanscope una struttura a più livelli

Urbanscope è composto da quattro lenti principali: *City and the world*, *Cities within cities*, *City magnets*, *Top Venues*.

City and the world

City and the world è la lente che osserva con chi parla Milano, attraverso gli *open data* di un operatore telefonico, relativi alle chiamate da rete radiomobile tra Milano e il resto del mondo. I dati analizzati si riferiscono ad un bimestre ed escludono le chiamate interne all'Italia. Pur essendo l'analisi ristretta a due mesi, si entra qui nel mondo dei *Big Data* [Riquadro 2], visualizzando milioni di chiamate in entrata (incoming calls) e in uscita (outgoing calls) [F. 1]. Le chiamate sono poi caratterizzate più precisamente per Paese di destinazione o, nel caso delle chiamate in entrata, per Paese di origine.

I dati visualizzati, sia numericamente che dall'intensità dei colori, mostrano i Paesi con i ¹ www.urbanscope.polimi.it.

Riquadro 2 – Big data, definizione

Il termine *Big Data* è ormai sovrautilizzato e non esistono definizioni univoche di cosa si intende per *Big Data*. Ricorre tuttavia spesso il riferimento alla definizione di Gartner IT per i cui “*I Big data*” sono informazioni ad alto volume, alta velocità e alta varietà, che richiedono forme di elaborazione efficienti ed innovative per avere nuove visioni e migliori decisioni (tradotto da Gartner IT Glossary (n.d.)).

La definizione richiama le note “3V” dei *Big Data*:

- *Volume*, riferendosi all’ampiezza dei dati ed è il concetto maggiormente associato all’etichetta *Big Data*. Sia a livello accademico che pratico si dibatte su cosa sia “big”, introducendo unità di misura sempre maggiori come soglia (terabyte, petabyte, ecc.). Un campo in cui questa prima proprietà è visibile ed entrata preponderatamente è la genomica, favorendo la nascita di centri di ricerca ad esso dedicati (si veda ad esempio l’Oxford Genomics Centre - <http://www.well.ox.ac.uk/ogc/home>)
- *Velocità*, che denota la sempre maggiore frequenza con cui i dati vengono generati e potenzialmente processati. I social media e i dati telefonici sono esemplificativi di questa proprietà in quanto vengono generati in modo continuo (real-time). La sfida è mettere in campo metodologie sostenibili in termini di costi e infrastrutture che permettano, ove necessario, non solo la raccolta ma anche l’elaborazione “real time”.
- *Varietà*, indicando la diversità della tipologia di dati. Si pensi ad esempio alla combinazione tra i dati social media sulle lingue e i dati tradizionali sulle lingue delle popolazioni: sono due tipi di dati diversi che osservano lo stesso fenomeno. Pur essendo questa terza proprietà la meno nuova, essa pone molte sfide ai ricercatori e agli analisti che devono affrontare il problema dell’integrazione dei dati.

Altre “V” sono state aggiunte nel tempo, come Valore, Veridicità, che sono tuttavia meno peculiari dei *Big Data*, ma caratteristiche fondamentali di ogni dato utilizzato a fini di monitoraggio o decisionale.

quali Milano è maggiormente connessa telefonicamente. Domande per future analisi nascono quasi immediate, per capire i motivi, o anche più facilmente per controllare la costanza dei fenomeni osservati nel bimestre in esame. In questo caso, con i dati riferiti al bimestre aprile-maggio del 2015: Asia e Africa parlano con Milano più che il Nord-America; dopo i maggiori paesi UE e la Svizzera, i paesi che più parlano con Milano sono la Russia, la Cina e l’Egitto. Anche la diversa direzionalità dei flussi di chiamate è interessante; se tipicamente il numero di chiamate verso Milano è maggiore del numero di chiamate da Milano, fenomeno vistoso con la Russia dove il numero di chiamate entranti è il doppio di quelle uscenti, il verso si inverte con Cina, Egitto e Senegal. Per questi paesi il numero di chiamate da Milano è notevolmente superiore a quelle entranti; il fenomeno non è irrilevante, se si considera che nel bimestre considerato il Senegal ha generato un traffico telefonico verso Milano della stessa dimensione di quello prodotto dall’intero Benelux.

Questa fotografia apre molte domande, alcune già in fase di investigazione (sempre attraverso dati anonimizzati, rispettando i vincoli di privacy). Ad esempio qual è la proporzione tra “chiamate per affari” e “chiamate private”? Analizzando i due mondi separati cambiano i flussi da e verso Milano? Un altro aspetto interessante è il cambiamento dei flussi durante la giornata. Questo secondo elemento è in fase di esplorazione, cercando di comprendere l’esistenza di “ricorrenze” e “percorsi” emergenti, all’interno della città di Milano, nelle diverse ore del giorno e della notte.



foto: T. Press / Gauriele Putzu

Cities within cities

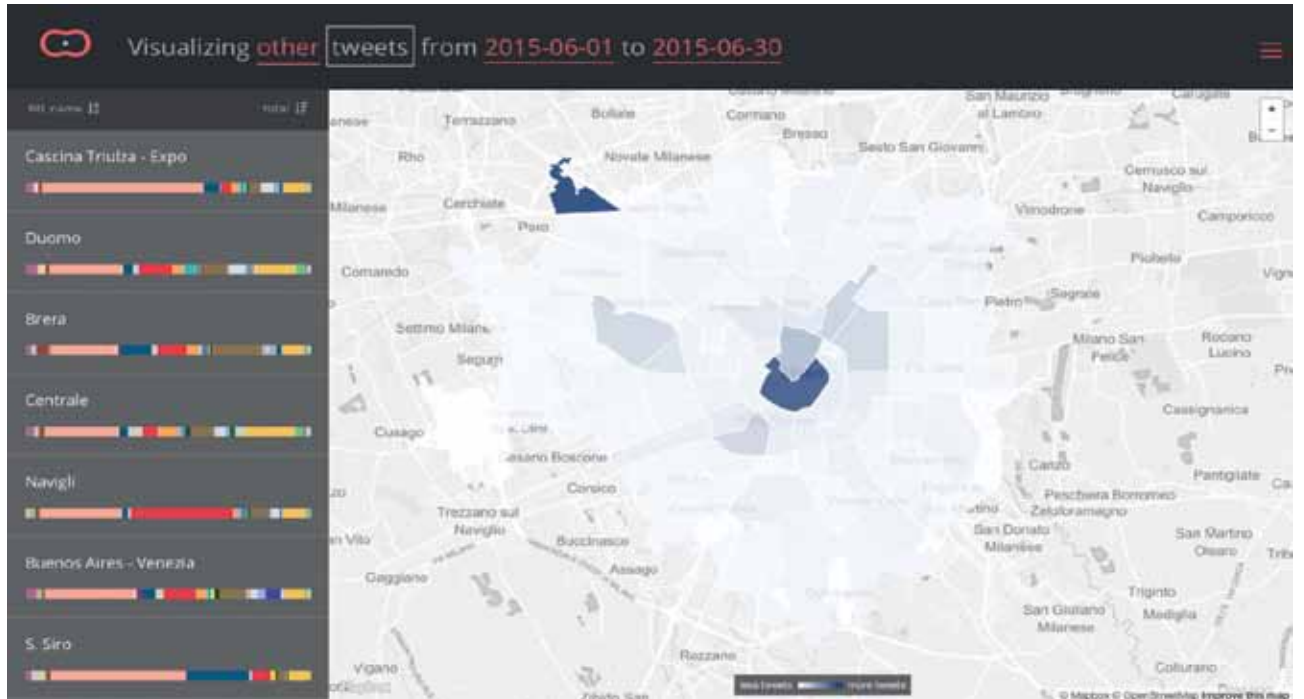
Cities within cities visualizza la città esplorando i messaggi che vengono scambiati a Milano attraverso Twitter [Riquadro 3]. Questa lente, che segue la lingua dei messaggi, rivela tre città digitali:

- la Milano che parla in italiano;
- la Milano internazionale, che parla in inglese;
- e la Milano multi-etnica, ossia quella che utilizza lingue diverse dall’italiano e l’inglese.

In particolare sono evidenziati i Nuclei di Identità Locale (NIL) relativamente più rilevanti rispetto ai *tweet* scritti in italiano, oppure in inglese, oppure in altre lingue, attraverso due sezioni: “Explore” e “Analyse”.

F.2

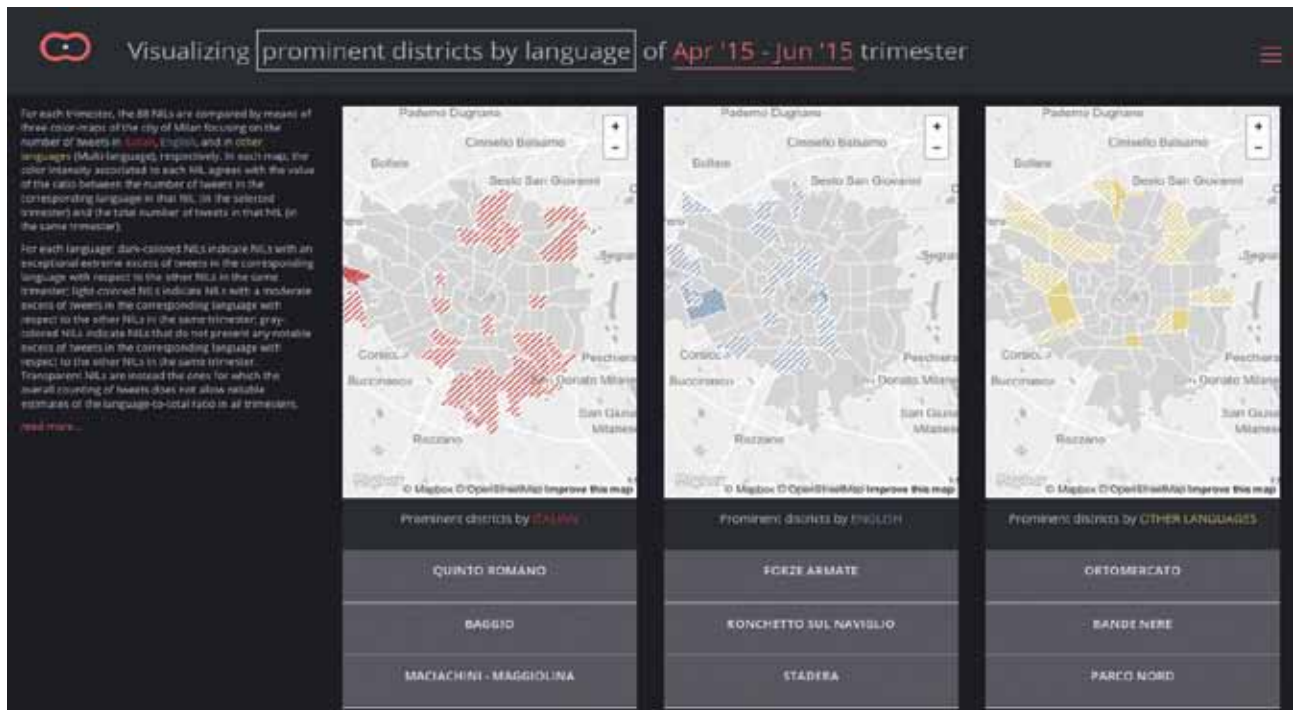
Cities within Cities – Sezione Explore, numero di Tweet per NIL (in migliaia) e lingua utilizzata (in %), nel mese di giugno 2015



Fonte: www.urbanscope.polimi.it

F.3

Cities within Cities – Sezione Analyse, variazione di Tweet per NIL (in %), per lingua, per il secondo trimestre 2015



Fonte: www.urbanscope.polimi.it

La sezione “Explore” consente una prima perlustrazione delle lingue [F. 2]; grazie alle visualizzazioni interattive sussiste la possibilità di navigare attraverso la mappa di Milano (sulla destra), in questo caso i NIL si colorano di blu più

o meno intenso in base alla numerosità dei *tweets*. A sinistra, l’utente più curioso può scoprire le diverse lingue presenti nei quartieri digitali di Milano attraverso una barra colorata. Scopriamo così che esistono NIL ad alta entropia linguistica,

Riquadro 3 – Le fonti del Progetto Urbanscope

Il progetto Urbanscope ha attinto a quattro tipologie di fonti dati.

- 1) La prima fonte di tipo tradizionale è la suddivisione amministrativa in 88 NIL - Nuclei d'Identità Locale della Città di Milano. Come riportato dal Comune di Milano i NIL sono "definibili come quartieri di Milano, in cui è possibile riconoscere quartieri storici e di progetto, con caratteristiche differenti gli uni dagli altri. Vengono introdotti dal PGT (Piano di Governo del Territorio) come un insieme di ambiti, connessi tra loro da infrastrutture e servizi per la mobilità, il verde. Sono sistemi di vitalità urbana: concentrazioni di attività commerciali locali, giardini, luoghi di aggregazione, servizi; ma sono anche 88 nuclei di identità locale da potenziare e progettare ed attraverso cui organizzare piccoli e grandi servizi" (www.comune.milano.it)
- 2) La seconda fonte dati sono i dati di Twitter. Twitter è una piattaforma gratuita di social networking e microblogging. Ogni persona o organizzazione può attivare una pagina e "twittare" (dall'inglese cinguettare) attraverso messaggi della lunghezza massima di 140 caratteri. I dati sono stati raccolti da agosto 2014 ad oggi attraverso le Application Programming Interface (API) messe a disposizione da Twitter. Le API sono interfacce dedicate a sviluppatori, appassionati e ricercatori che possono essere interrogate attraverso la costruzione di programmi (si veda ad esempio: <https://dev.twitter.com/overview/api>).

In particolare sono stati utilizzati soltanto i Tweets georeferenziati e riferibili alla città di Milano.

- 3) La terza fonte sono i dati di telefonia mobile. Il dataset reso disponibile da un operatore telefonico italiano, include i dati telefonici per i mesi di Aprile e Maggio 2015. I dati originali hanno una frequenza temporale giornaliera, visualizzata tuttavia in Urbanscope in modo aggregato per il bimestre oggetto dell'analisi. Sono in corso ulteriori analisi.
- 4) La quarta fonte dati proviene da Foursquare. Foursquare è un social network attraverso cui gli utenti si possono gratuitamente registrare ed eseguire un "check-in" che notifica e registra la loro posizione in un determinato luogo. Il "check-in" ha generalmente un'accezione positiva. Sempre tramite le API pubbliche, Urbanscope ha raccolto i dati da Foursquare per l'area milanese da agosto 2014 ad oggi.



foto: T. Press / Gabriele Putzu

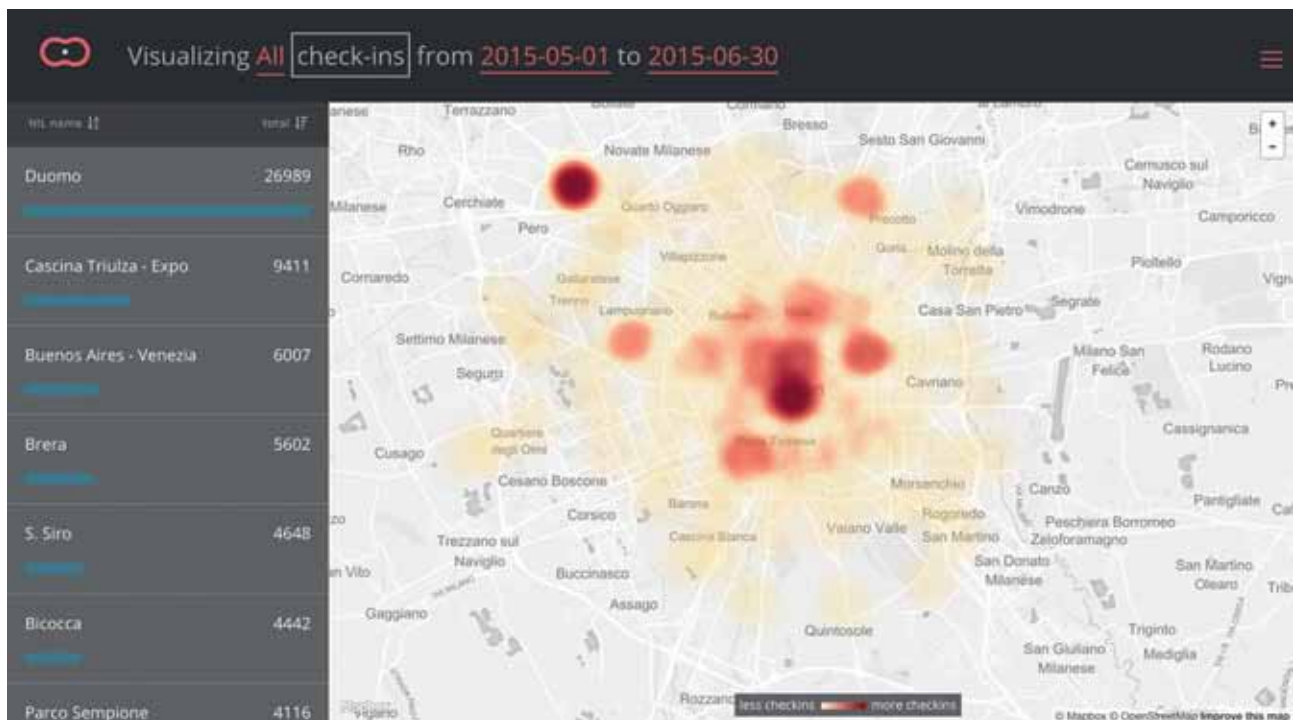
come Cascina Triulza-Expo, dove l'eterogeneità linguistica è elevata, ed altri NIL di media entropia linguistica, come Buenos Aires e S. Siro.

La sezione "Analyze" ha l'obiettivo di veicolare in modo semplice analisi più complesse [F. 3]. A cadenza trimestrale gli 88 NIL sono confrontati attraverso tre mappe colorate: rossa per l'italiano, blu per l'inglese e gialla per le altre lingue. In ogni mappa l'intensità del colore indica la variabilità di

quella lingua per quel NIL rispetto agli altri NIL nello stesso trimestre. I NIL colorati più intensamente possono essere considerati "anomali" in quanto hanno, nel trimestre oggetto dell'analisi, un numero maggiore di tweets per una determinata lingua. Questa sezione più analitica apre le porte all'uso dei dati dei *social media* come segnali deboli di tendenza e la necessità di cogliere variabilità e non solo numeri assoluti.

F.4

City Magnets, variazione di check-in con foursquare, per luogo, (in %), nel bimestre maggio-giugno 2015



Fonte: www.urbanscope.polimi.it

City Magnets

La terza sezione di Urbanscope, denominata *City Magnets*, identifica i luoghi attrattori della città attraverso *Foursquare*. Qui si mostrano i luoghi più frequentati e condivisi dagli utenti di questo *social network*, che come *Twitter* e altri mette a disposizione del pubblico degli *open data* [Riquadro 4]. Le mappe fornite da questa lente illuminano i luoghi della città con intensità crescente (rosso) in base al numero di persone che hanno scelto di rappresentarsi per mezzo di un check-in digitale in quei luoghi [F. 4]. *City Magnets* si è dimostrato sensibile agli eventi, partendo da Expo nel 2015. Nei mesi di maggio e giugno 2015 la zona dell'Expo appare nella zona rossa, mentre prima di quella data era quasi inesistente. Expo 2015 tuttavia, non ha oscurato i luoghi tradizionali di Milano, con il NIL Duomo sempre in vetta nei check-in virtuali.

Top Venues

L'ultima area di esplorazione è *Top Venues* (letteralmente principali luoghi frequentati). Dove l'idea è di rappresentare, mese per mese e in modo un po' giocoso, i luoghi di Milano nei quali si sono osservati più check-in (sempre con *Foursquare*), suddividendoli per categorie. Ad esempio dal confronto tra i mesi marzo e aprile 2016 della *Top Venues* della categoria "*Event, Arts & Entertainment*", emerge che lo Stadio San Siro continua a svettare [F. 5]. Osservando lo stesso periodo, si evidenzia come altri luoghi salgono e scendono – in alcuni casi – inaspettatamente, come il quinto posto ad aprile 2016 di "*SuperStudio Più*", luogo del design. In realtà la sua ascesa riflette un evento imponente a Milano, che richiama specialisti e appassionati di design: cioè il Fuori Salone Design Week.

Riquadro 4 – Big data e Open data e dati di Urbanscope

Non tutti i *Big Data* sono *Open Data*. Con *Open Data* infatti ci si riferisce a dati pubblicamente e gratuitamente disponibili. Si pensi ad esempio ai dati telefonici; alcuni selezionati insiemi di dati (come quelli utilizzati in Urbanscope) sono resi disponibili pubblicamente ad uso dei ricercatori o per competizioni. Normalmente questi dati non sono “open” ed anzi sono soggetti a forti restrizioni nell’uso in base alle regolamentazioni sulla privacy dei diversi paesi.

Viceversa alcuni dati *Open* non sono *Big*, si pensi ad esempio ai dati sui bilanci delle amministrazioni pubbliche: sono reperibili ma la loro mole è ridotta.

Quest’ultimo esempio può tuttavia far comprendere come i confini tra *Big* e *non Big* siano sfumati. Se ad esempio i bilanci delle aziende sanitarie venissero utilizzati come fonte integrativa di dati, insieme ai dati social media relative alle stesse amministrazioni ed insieme ad altre fonti di dati sui pazienti, attraverso il “data integration” si entrerebbe lentamente nel mondo *Big Data*.

F.5

Top Venue, classificazione dei check-in con foursquare, per luogo e categoria di evento (esempio: “Events, arts and Entertainment), nel mese di aprile 2016



Fonte: www.urbanscope.polimi.it

Metodo: il decisore al centro

Come si è arrivati a definire le lenti e il loro dettaglio? Come è possibile declinare Urbanscope in altri contesti?

In questa sezione si illustra il metodo utilizzato, sperimentato sia sul progetto Urbanscope visibile pubblicamente, sia su altri progetti sviluppati per esigenze specifiche. Lo schema metodologico di riferimento del gruppo [F. 6] prevede due categorie di fasi: le fasi più ricorrenti nella filiera di analisi dei *Big Data* (in blu nella figura) e le fasi distintive dell’approccio Urbanscope (in giallo).

Il punto di partenza sono le tre aree concentriche: *decisori*, quali attori pubblici e privati che agiscono sugli ambiti di indagine; *stakeholder*, gruppo che rappresenta gli altri attori interessati, seppur non coinvolti in prima persona nelle decisioni; *contesto*, cioè i confini dell’analisi e i

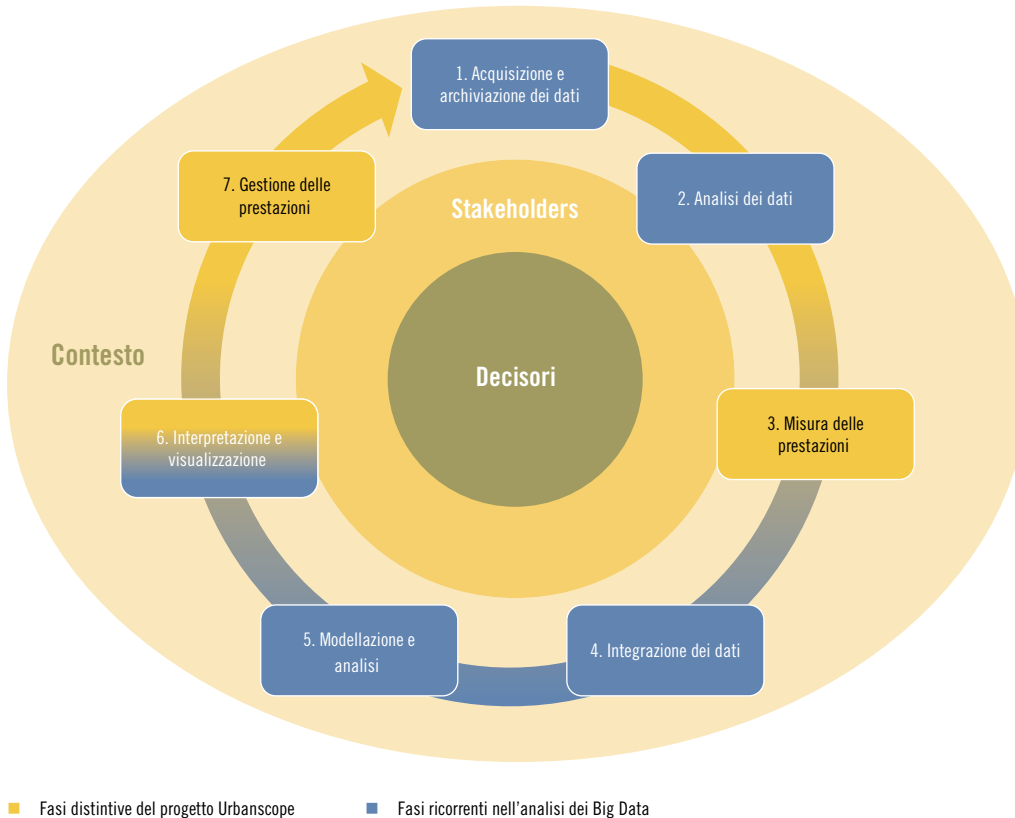
relativi macro-trend evolutivi, i quali sono utili all’interpretazione dei dati.

Il progetto Urbanscope, nato su iniziativa del Politecnico, ha coinvolto nei due anni di sviluppo diverse categorie di decisori e *stakeholder*, sia pubblici che privati. Il coinvolgimento, avvenuto attraverso interviste dirette, incontri ristretti e presentazioni plenarie ha permesso di identificare domande di ricerca più specifiche partendo da un macro tema: “Quanto è internazionale Milano?”.

La definizione di domande più specifiche, come riscontrato in altri progetti di *Big Data*, è emersa solo dopo le prime fasi di raccolta e analisi dei dati (fasi 1 e 2 nella figura). Ad esempio l’osservazione dei dati Twitter ha permesso di fare emergere come interessante non solo l’internazionalità di Milano vista come unica unità di analisi, ma anche la diversità dei suoi quartie-

F. 6

Schema di sintesi del metodo di ricerca



ri. Questa scoperta è stata condivisa con alcuni decisori che hanno portato a definire la lente “*Cities within cities*” e a scegliere come unità geografica di analisi i NIL (metrica riconosciuta e già utilizzata dai decisori milanesi).

La definizione di domande più puntuali consente di passare alla fase 3, dove valutare le prestazioni dei dati: quanto i dati sono precisi nel rappresentare il fenomeno? Quanto sono completi? Quanto sono tempestivi? Un esempio sono i dati telefonici. La posizione dell'utente viene determinata attraverso metodologie proprietarie e riservate dell'operatore telefonico che non indicano perfettamente la posizione dell'utente ma la posizione in un'area di alcuni metri quadrati. Ciò deve essere considerato per il tipo di utilizzo. Nel caso di Urbanscope data l'aggregazione geografica scelta l'errore è diventato marginale.

Tutti i dati monitorati in Urbanscope hanno il pregio di un aggiornamento in tempo reale, tuttavia il grado di precisione e completezza è inferiore ai dati tradizionali. Per comprendere la differenza con i dati tradizionali si pensi ad esempio ai dati di censimento della città confrontati con i post di Twitter. Il censimento per sua natura vuole essere completo e preciso includendo tutti i cittadini in una finestra temporale specifica di raccolta dati. I dati di Twitter invece non possono esserlo, in quanto sono una voce spontanea di chi usa questo strumento; inoltre le *policy* dei *social media* non sempre chiariscono se quanto estratto

dalle interfacce pubbliche sia la totalità dei post. Questa situazione apre due sfide che coinvolgono, in uno sforzo congiunto, ricercatori e decisori: i primi non devono abbandonare il rigore, ma cercare nuovi modi per garantire la qualità del dato; i secondi devono comprendere che i *Big Data* non possono essere letti come i dati tradizionali.

Le fasi 4 e 5, relative all'integrazione e alla modellazione dei dati, sono il luogo della ricerca di relazioni, correlazioni e variabilità. La sezione analitica di *Cities within cities* è l'esempio più esplicativo, dove i dati Twitter sono stati integrati con i dati di geolocalizzazione dei NIL, analizzati rispetto alla totalità dei dati disponibili avendo definito la loro variabilità attraverso metodi statistici. Anche in questo caso l'interazione con gli *stakeholder* è stata importante per verificare l'allineamento delle scelte con le esigenze, per spiegare alcune scelte (come i valori soglia per la variabilità), ma anche per verificare la comprensibilità della analisi fatte.

In queste interlocuzioni, le fasi 4 e 5 si sovrappongono con le ultime due di visualizzazione e gestione delle prestazioni (6 e 7). L'elaborazione di una grande quantità di dati con molte dimensioni sottostanti, rende poco funzionali i metodi di rappresentazione tradizionali. Un esempio è la lente *City and the world*. I primi risultati furono visualizzati attraverso una rete, con NIL e Paesi del mondo come nodi e le linee relazioni per tracciare i flussi telefonici. La rappresentazione, pur



foto: Il Press / Francesca Agosta

essendo completa, rendeva confusa la visualizzazione, minando il processo conoscitivo di ricercatori e decisori. Si è quindi optato per semplificare, in una prima fase, la visualizzazione, utilizzando una mappa geografica. È ora in fase di sperimentazione una grafica più articolata che possa arricchire le informazioni (maggiore completezza) senza minare la fruibilità da parte del decisore.

Conclusioni e sviluppi futuri

Il progetto e l'approccio Urbanscope si pongono come voce non solo nuova ma anche critica per gestire in modo integrato la valorizzazione dei *Big Data* per contesti complessi come quello urbano. I due anni di sperimentazione e sviluppo del progetto hanno permesso di raggiungere alcuni risultati, ma soprattutto di aprire nuove sfide e porre nuove domande.

Un primo risultato è quello relativo al consolidamento del metodo, e in particolare a due scelte fondamentali:

- 1) decisori e *stakeholder* devono essere messi al centro del processo, sia nella definizione del fabbisogno che nella fase di modellazione;
- 2) è necessario dotarsi di un team multidisciplinare² che si incontra regolarmente per confrontarsi.

Questi due punti implicano tempi di realizzazione dei progetti più lunghi, ed un approccio diverso ai problemi, dove la scomposizione in attività e suddivisione del lavoro non sempre è possibile. Tuttavia proprio questa nuova modalità di lavoro permette a ricercatori e decisori di accrescere le proprie competenze e senso critico.

Un secondo risultato è l'uso di visualizzazioni grafiche alternative, sia durante le fasi di ricerca sia per veicolare i risultati finali. Le competenze del team di Design hanno consentito di condensare un numero di informazioni elevate

in modo semplice e di sperimentare elementi di azione (come la scelta delle lingue, e degli intervalli temporali) che rendono il sito più interattivo e, potenzialmente, accrescono la consapevolezza dei decisori e degli *stakeholder*.

Le sfide per Urbanscope, come per tutto il mondo *Big Data*, sono ancora molte. Si pone qui l'accento su due temi spesso trascurati. Il primo è il rigore scientifico: i *Big Data* (in primis quelli provenienti dai *Social Media*) vengono spesso associati all'idea di essere immediatamente pronti all'uso. Invece, tanto i dati estratti dai social media così come i dati telefonici seguono delle metodologie complesse di preparazione dei dati. Fasi necessarie per, ad esempio, nascondere alcune caratteristiche dei dati e assicurare la protezione della privacy. Oppure, come già accennato in precedenza, occorre verificare la qualità dei dati: la geolocalizzazione delle telefonate dipende dalla metodologia di estrazione dell'operatore, che nel caso di Urbanscope, ha una "imprecisione" di alcuni metri per la città di Milano. È compito dei ricercatori effettuare controlli e verifiche delle metodologie in modo da fornire ai decisori informazioni sulla precisione e sulla completezza dei dati. L'altra importante sfida, in particolare per gli attori pubblici, è utilizzare dati e mezzi nuovi per creare, in alcuni casi recuperare, un legame con i cittadini, favorendo una maggiore partecipazione alla vita politica, intesa nella sua accezione originaria di sfera pubblica e comune.

Urbanscope ha deciso di accettare queste sfide e partendo dai risultati ottenuti sta ora lavorando per fornire nuovi spunti di riflessione ai decisori. Un prima area di sviluppo si basa sui dati telefonici, preziosi per la loro "velocità" e per ora analizzati con un'aggregazione temporale molto alta. Una seconda area di indagine apre all'utilizzo dei dati Instagram³. Infine un'ultima area di indagine è l'integrazione dei dati accoppiando dati digitali ai dati statistici più tradizionali

² Il team Urbanscope è composto da: un ricercatore senior e due ricercatori junior del dipartimento di Design; due ricercatori senior e due junior del dipartimento di Matematica; due ricercatori senior e due junior del dipartimento di Ingegneria Gestionale; due ricercatori senior e tre ricercatori junior del dipartimento di Ingegneria Informatica.

³ Instagram è un social network che permette di visualizzare, caricare e condividere fotografie da parte degli utenti.