



RUMORS E SENTIMENT DELLA RETE: GEOLOCALIZZARE L'OPINIONE PUBBLICA SUI SOCIAL MEDIA¹

Le mappe dell'odio in rete

Qualche settimana fa molti organi di informazione ([L'Espresso](#), il [Corriere della Sera](#), televisioni e agenzie di stampa) hanno diffuso i risultati della ricerca "Le mappe dell'intolleranza", un progetto promosso da Vox Diritti sull'*hate speech* attraverso i *social media*. Nei giorni immediatamente successivi, la redazione italiana di Wired, noto magazine di tecnologia, [riprende la notizia](#) sollevando alcuni [dubbi sui risultati della ricerca](#). Le obiezioni - espresse da Alberto Longo e supportate da esperti del settore - chiamano in causa i limiti metodologici dell'*opinion mining*, soprattutto rispetto a quella che è la cifra distintiva della ricerca, ossia la geolocalizzazione dei tweet sottoposti ad analisi.

La ricerca di Vox, realizzata in collaborazione con le Università di Milano, Roma e Bari, aveva l'obiettivo di creare una mappa geografica dei discorsi d'odio nei confronti di immigrati, omosessuali, disabili, donne ed ebrei attraverso la raccolta dei tweet discriminatori. Il problema del *flaming*, il linguaggio eccessivo, violento e aggressivo degli utenti, è uno degli aspetti più controversi dell'uso dei *social*: l'anonimato e la distanza fisica dalla vittima dell'aggressione spesso liberano le ultime inibizioni delle persone². Non è quindi raro leggere sulle bacheche dei principali social network, ma anche nello spazio commenti degli articoli dei giornali online, scambi di battute razzisti, omofobi, antisemiti; ancor più di recente, come ciclicamente accade, sono stati i Rom a essere stati il bersaglio di numerose "micro-aggressioni" online, brevi scambi comunicativi contenenti messaggi

¹ Il paper è il risultato della collaborazione tra Alfredo Borrelli (Estrogeni), Davide Magini (Estrogeni), Olivier La Rocca (Europartners), Gianfranco Zucca (Europartners).
Per informazioni: a.borrelli@estrogeni.net
o larocca@europartnersnetwork.eu

² Un *flame* è un messaggio deliberatamente ostile e provocatorio inviato da un utente alla comunità o a un altro individuo specifico, lo scambio di insulti che spesso ne consegue, paragonabile a una "rissa virtuale"; cfr. P. WALLACE, *The psychology of the Internet*, Cambridge (UK), Cambridge University Press, 1999

volti a denigrare alcuni individui a causa del loro gruppo appartenenza³. Per i *provider* di servizi come Facebook e Twitter la discriminazione digitale e, più in generale, il *flaming* è un grosso problema, al punto che di recente il CEO della società ha dichiarato che Twitter sta lavorando a degli strumenti per contrastare questo genere di comportamenti, la dichiarazione di Dick Costolo è molto onesta: "[we suck at dealing with abuse and trolls on the platform](#)". Sotto l'aspetto sostantivo, l'iniziativa di Vox diritti va quindi a colmare un vuoto di conoscenza significativo

Meno convincente, almeno stando alle critiche ricevute, è l'approccio metodologico. Le critiche si appuntano sul fatto che la geografia dell'intolleranza è basata su un numero di tweet molto basso. Le discussioni che hanno seguito la pubblicazione di questa ricerca danno l'opportunità di fare una riflessione sulle potenzialità di geolocalizzare l'analisi dell'opinione pubblica. In questo articolo cercheremo di offrire alcune indicazioni a riguardo. L'elemento che contraddistingue il contributo è dato dal fatto che non ci limiteremo a fare riflessioni di metodo, ma proporremo un confronto basato su una rilevazione realizzata *ad hoc* per verificare le differenze tra il "metodo Vox" e un metodo basato su un approccio alla localizzazione geografica che assicura basi dati più ampie e robuste. Occorre precisare che l'analisi proposta ha l'obiettivo di avviare una discussione e non certo di giudicare la qualità di altre ricerche.

Cosa funziona e cosa non funziona nelle hate map

Le *hate map* sono una rappresentazione geografica dei messaggi a carattere discriminatorio rinvenuti sui *social*, realizzate attraverso le coordinate geografiche degli utenti (latitudine e longitudine). Su Twitter ad esempio è possibile abilitare la funzione di geolocalizzazione dei tweet, permettendo così agli utenti di dare e avere informazioni sulla località dalla quale proviene il messaggio⁴. Questa opzione apre delle grandi possibilità alla *sentiment analysis*, lo studio dell'"umore" della rete attraverso l'analisi di grandi quantità di messaggi. La ricerca di Vox sulla geografia italiana dell'intolleranza sfrutta questa possibilità, replicando la metodologia di un'indagine realizzata nel 2013 sugli Stati Uniti⁵. Sotto il profilo metodologico, l'analisi di Vox è basata sulla raccolta oltre 2 milioni di tweet nel periodo gennaio-agosto 2014. Il *retrieving* dei dati è stato realizzato usando parole chiave collegate al sentimento di intolleranza nei confronti di stranieri, omosessuali, disabili, donne ed ebrei: ad esempio, per esaminare il razzismo sono stati raccolti i tweet contenenti parole come "negro", "zingaro" e "terrone". Il passaggio successivo è stato l'individuazione dei messaggi contenenti le coordinate geografiche dell'utente al momento della pubblicazione: nel complesso, i tweet geolocalizzati sono circa 43.000 (il 2,3% del totale). Sulla base delle coordinate geografiche sono state realizzate delle mappe termografiche che indicano le zone dell'Italia dove i tweet discriminatori sono più frequenti.

La metodologia dell'indagine sollecita alcune domande.

Prima domanda. Di quale popolazione sono "rappresentative" le mappe? La risposta è banale. Le mappe dell'odio rappresentano una parte della popolazione che usa un linguaggio discriminatorio su Twitter, quel segmento che decide di geolocalizzare i propri messaggi. In senso stretto, il campione analizzato è autoeletto, ossia contiene solo coloro che usano tag geografici quando twittano. Quindi le mappe sono inutili? No, tutt'altro. Le

3. Cfr. D.W. SUE, *Micro-aggressions in Everyday Life. Race, Gender, and Sexual Orientation*, John Wiley & Sons, Hoboken [N.J.] 2010.

4. In linguaggio tecnico questo tipo di dati viene definito *Volunteered Geographical Information* (VGI).

5. Il progetto è denominato [Geography of Hate. Geotagged Hateful Tweets in the United States](#) è realizzato della Humboldt University di Arcata in California

hate map rappresentano una fonte di conoscenza preziosissima, soprattutto se collegate ad altre informazioni geografiche e socio-economiche. Di norma, nell'analisi dell'opinione pubblica, si usano variabili territoriali dal potere esplicativo molto limitato: dire che nel Nord-Est le persone la pensano in un certo modo, significa poco anche perché in quest'area geografica rientrano sia Belluno che Riccione, due comuni tra loro molto differenti.

Seconda domanda. Il 2,3% di messaggi geolocalizzati è tanto o è poco? In assoluto è tanto, perché ad oggi, almeno in Italia, la geolocalizzazione nell'uso dei *social* non è così diffusa, anzi gli *user* sembrano ancora avere una certa diffidenza. D'altronde è pur vero che se si vogliono usare tecniche statistiche avanzate, la numerosità dei casi è importante, soprattutto quando si ha l'esigenza di risultati con un dettaglio territoriale molto fine. Non a caso, nel commento delle mappe i ricercatori di Vox si guardano bene dal citare comuni o province specifiche, ma si limitano a discutere le differenze tra macro-ripartizioni o tutt'al più regioni.

Un confronto tra due approcci di geotagging

Pur con i suoi limiti, la ricerca realizzata da Vox ha l'indubbio pregio di introdurre almeno nel panorama italiano un approccio nuovo. Combinare gli strumenti di *opinion mining* con l'attenzione alla dimensione territoriale dei fenomeni, è un approccio molto promettente che merita di essere approfondito e applicato in numerosi contesti di analisi.

Con questo obiettivo, presentiamo i risultati di una rilevazione sperimentale che, partendo dagli stessi presupposti delle mappe dell'intolleranza, introduce un'importante differenza. La sperimentazione è collegata allo sviluppo di uno strumento che integra statistica territoriale e *sentiment analysis*, denominato DEEP (Drawing Elements for Effective Politics)⁶.

Così come fatto dai ricercatori di Vox, abbiamo raccolto tutti i tweet contenenti le parole: "negro", "zingaro", "muso giallo", "mangiabanane". Il periodo di riferimento della rilevazione è, invece, molto inferiore: le prime tre settimane di febbraio 2015.

Nel complesso, la base dati a disposizione è di 11.655 tweet. Per comparare i risultati ottenuti tramite DEEP con quelli delle mappe dell'intolleranza, è necessario esplicitare il processo di *geotagging* usato. Sotto questo profilo la differenza è sostanziale: al posto di usare le coordinate geografiche - scelta che comporta un netto restringimento della base dati disponibile - DEEP lavora sulla *user location* (UL), un campo previsto dal modulo di registrazione su Twitter, all'interno del quale l'utente può specificare la località nella quale risiede. I vantaggi di usare la UL sono almeno due, uno di ordine tecnico, l'altro teorico.

Gli utenti che rendono disponibile la propria UL sono molto più numerosi degli utenti che geolocalizzano i tweet: nella rilevazione pilota realizzata a febbraio 2015, il 61,4% dei tweet raccolti conteneva un'indicazione della UL, una percentuale non confrontabile con il 2,3% di tweet geolocalizzati analizzati da Vox.

Usare la UL offre anche un vantaggio di tipo teorico. Decidere di inserire una località all'interno del proprio profilo twitter significa rivendicare un'appartenenza territoriale: sono romano, sono napoletano, sono milanese. Si tratta di un'informazione all'apparenza banale, ma che nell'analisi di molti fenomeni di opinione può risultare importante. Inoltre, a differenza delle coordinate geografiche, l'UL è stabile: se un tweet è taggato con

6. DEEP è un prodotto realizzato da Europartners ed Estrogeni con la collaborazione di Aster. Su richiesta è disponibile un paper di presentazione dello strumento, contenente anche un'applicazione alla regione Marche. Per informazioni contattare: Olivier La Rocca (o.larocca@europartnersnetwork.eu) o Alfredo Borrelli (a.borrelli@estrogeni.net).

coordinate geografiche ricadenti nel centro città di Milano, non è possibile sapere se la persona vive effettivamente in città oppure se è un pendolare o un turista.

La scelta di usare la UL comporta però un lavoro supplementare per la preparazione della base dati. La tabella 1 illustra numericamente i tre passaggi di *cleaning* necessari alla creazione di una database di messaggi geolocalizzati sulla base dell'UL.

Tabella 1 - Processo d'identificazione dei tweet geolocalizzati

STEP 1			STEP 2			STEP 3		
<i>Presenza della User Location (UL)</i>	<i>N</i>	<i>%</i>	<i>Pertinenza del tweet</i>	<i>N</i>	<i>%</i>	<i>Usabilità della UL</i>	<i>N</i>	<i>%</i>
Tweet con UL	7157	61.4	Tweet con UL pertinenti	3839	53.6	Tweet pertinenti con UL utilizzabile	1547	46.6
Tweet senza UL	4498	38.6	Tweet con UL non pertinenti	3318	46.4	Tweet pertinenti con UL inutilizzabile	1771	53.4
Tweet totali	11655	100.0	Totale tweet con UL	7157	100.0	Totale tweet con UL pertinenti	3318	100.0

Fonte: elaborazioni Europartners/Estrogeni su dati propri (febbraio 2015)

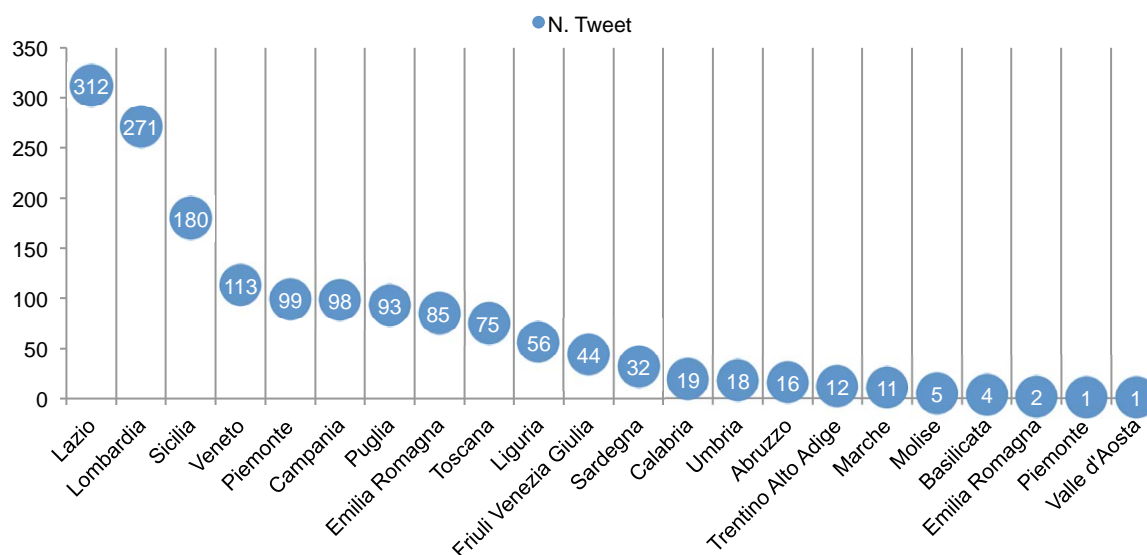
Le operazioni di pulizia dei dati comportano una perdita d'informazione che tuttavia non arriva ai livelli che si ottengono usando le coordinate geografiche. Come detto, il primo step (eliminazione dei tweet senza UL) comporta una perdita d'informazione del 38,6%. Nel secondo step, si analizza la pertinenza del contenuto del tweet, eliminando i messaggi che pur contenendo una o più *keywords* non rientrano tra quelli d'interesse per l'analisi: nel caso specifico si è trattato soprattutto di messaggi che rimandavano a materiale pornografico e che, spesso, contenevano la parola "negro".

Nel secondo step, è stato eliminato il 46,4% dei messaggi risultanti dalla prima operazione di *cleaning* dei dati. C'è da precisare che, sebbene nella metodologia della ricerca Vox questo passaggio non sia esplicitato, è probabile che sia stato realizzato. Infine, nel terzo step si è provveduto a identificare i messaggi pertinenti contenenti una UL utilizzabile ai fini della geolocalizzazione. Questo passaggio è necessario per eliminare i tweet di quegli utenti che non valorizzano il campo UL usando un'indicazione geografica. Siccome nel form di registrazione usato da Twitter il campo UL è libero, alcuni utenti inseriscono indicazioni fittizie, ironiche o semplicemente lacunose. Si precisa che il livello di *geotagging* usato da DEEP è molto fine perché si considerano solo i tweet riconducibili a uno degli oltre ottomila comuni italiani. Questo passaggio ha comportato la perdita del 53,4% dei tweet pertinenti con indicazione di UL.

Nel complesso, l'uso della *user location* in luogo delle coordinate geografiche ha garantito una percentuale del 13,3% di tweet utili all'analisi (la ricerca Vox dichiara una percentuale di messaggi utile del 2,3%): in pratica da una base di partenza di oltre 11mila messaggi, si è ottenuto un *corpus* di 1547 messaggi geolocalizzati a livello comunale.

A scopo illustrativo si riporta la distribuzione regionale dei tweet (grafico 1), corredata dalla classifica dei primi dieci comuni per numero di tweet (tabella 2).

Grafico 1 - N Tweet geolocalizzati



Fonte: elaborazioni Europartners/Estrogeni su dati propri (febbraio 2015)

Trattandosi di una rilevazione pilota, la numerosità di alcune regioni è decisamente esigua. Tuttavia, emergono alcune informazioni interessanti come ad esempio il fatto che i tweet provengono da 275 diversi comuni italiani e che in testa alla classifica per regioni oltre alla Lombardia e al Lazio (indicate anche da Vox), ci sia anche la Sicilia.

Tabella 2 - Primi dieci comuni per numero di tweet
(N comuni = 275)

Comune	N. Tweet	%
1 Roma	304	19.7
2 Milano	169	10.9
3 Palermo	107	6.9
4 Napoli	68	4.4
5 Torino	49	3.2
6 Genova	41	2.7
7 Bologna	38	2.5
8 Firenze	24	1.6
9 Padova	22	1.4
10 Verona	22	1.4
Totale primi 10 comuni	844	54.6
Altri comuni (265)	703	45.4
Totale	1547	100.0

Fonte: elaborazioni Europartners/Estrogeni su dati propri (febbraio 2015)

Venti giorni di antiziganismo in rete

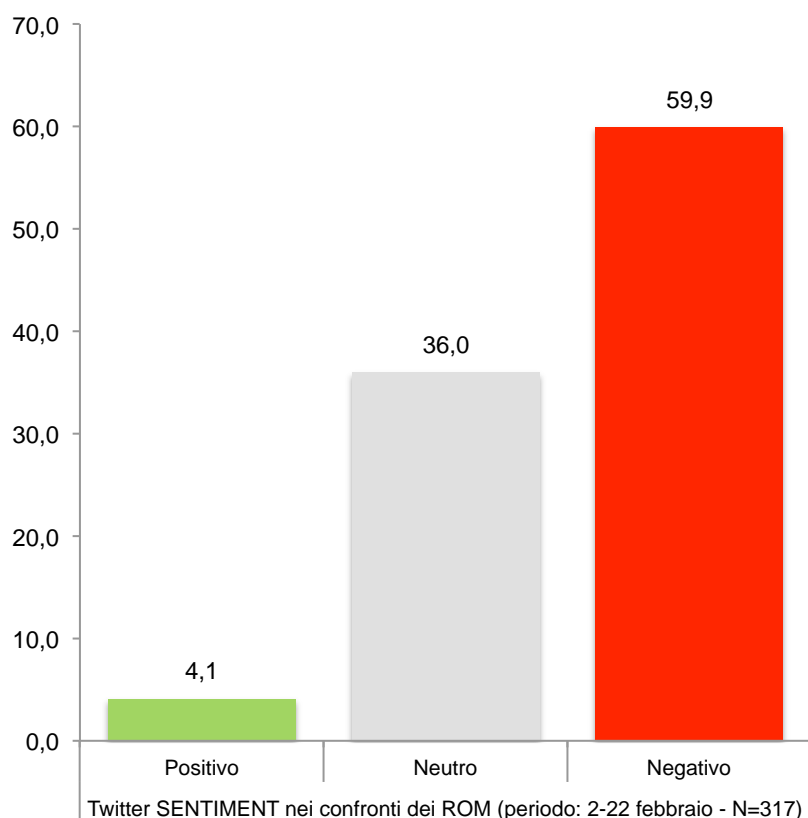
Nel periodo compreso tra il 2 e il 22 febbraio abbiamo analizzato il *sentiment* di Twitter nei confronti dei rom. All'interno della rilevazione pilota, realizzata per testare l'efficienza di DEEP e confrontarne le prestazioni con i dati della ricerca Vox, era presente anche la *keyword* "zingaro". La raccolta ha originato, 317 tweet geolocalizzati sulla base della UL.

Si tratta del 20,5% del totale di messaggi utili. Benché la rilevazione fosse un test dello strumento, pensiamo sia utile analizzare i dati di *sentiment* poiché mettono in luce elementi interessanti sul pregiudizio e l'ostilità che circola su Twitter⁷.

Il grafico 2 rappresenta il *sentiment* generale nei confronti dei rom: nelle tre settimane di riferimento il dato è abbastanza inequivocabile poco meno del 60% dei tweet analizzati conteneva messaggio negativi, denigratori, aggressivi o apertamente razzisti nei confronti dei rom. Un terzo dei tweet, invece, presentava un contenuto neutro. Infine, solo il 4,1% dei messaggi aveva un contenuto positivo.

Grafico 2 - Il sentiment di twitter verso i rom

Fonte: elaborazioni Europartners/Estrogeni su dati propri (febbraio 2015)

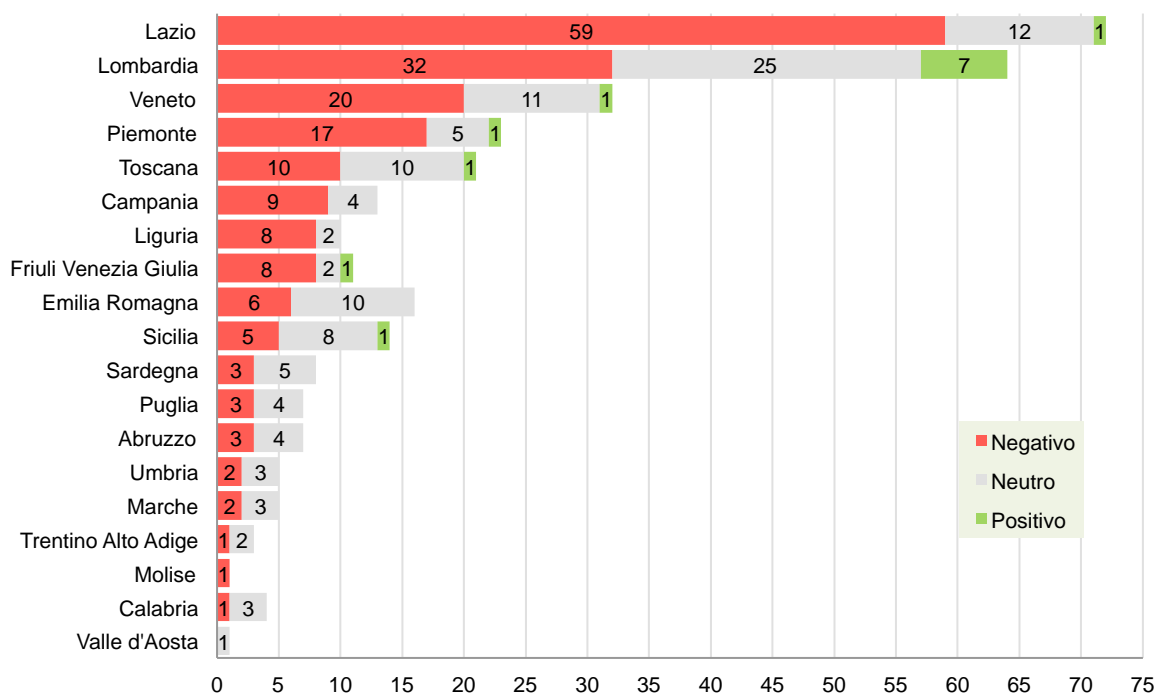


Il secondo dato d'interesse si trova nel grafico 3, nel quale si presenta il *sentiment* per regione. Non si usa la localizzazione comunale, pur disponibile, poiché l'esiguità numerica

7. DEEP analizza il *sentiment* con un approccio qualitativo, ossia la classificazione dei tweet non avviene attraverso procedure automatiche o semi-automatiche, ma attraverso l'analisi del contenuto tweet per tweet.

dei dati raccolti durante il test non permette statistiche robuste. Le regioni sono ordinate sulla base del numero di tweet negativi.

Grafico 3 - Sentiment verso i rom per regione (N)



Fonte: elaborazioni Europartners/Estrogeni su dati propri (febbraio 2015)

Lazio e Lombardia, pur avendo circa lo stesso numero totale di messaggi, presentano una differente composizione interna. Dal Lazio provengono soprattutto tweet negativi, mentre in Lombardia messaggi negativi e neutri quasi si equivalgono: rimane fermo lo scenario di sostanziale negatività nei confronti dei rom. Purtroppo i dati della rilevazione pilota non permettono altri approfondimenti. C'è comunque da ricordare che il periodo analizzato è coinciso con un fatto di cronaca che ha avuto una larga eco: in Veneto, il benzinaio Graziano Stacchio ha ucciso Albano Cassol, vicentino di origini rom, mentre stava tentando una rapina in una gioielleria. Nei giorni successivi, i media hanno dato spazio e amplificato i sentimenti di antiziganismo suscitati dall'episodio.

Twitter non ha fatto eccezione: leggendo i tweet raccolti per questa sperimentazione (molti riferiti all'hashtag #iostoconstacchio) si ha l'idea di quanta violenza e ostilità ci possano essere in 140 caratteri. Molti messaggi infatti riportano commenti sprezzanti e denigratori, basati su pregiudizi e stereotipi; in alcuni casi vengono formulate vere e proprie minacce. Non di rado, infine, la molla che fa scattare le aggressioni verbali via Twitter sono i talk show televisivi: i tweet contro i rom che contengono *hashtag* riferiti a trasmissioni come "La gabbia" (La7) o "Quinta colonna" (Rete4) sono numerosi.

Twitter diventa così la cassa di risonanza di un discorso pubblico esacerbato e polarizzato: come in quei talk show dove non ci sono argomentazioni da discutere, ma solo fazioni da sostenere o avversare.

Analizzare l'opinione pubblica sulla rete: alcune riflessioni

Sulla scorta delle riflessioni proposte in precedenza si possono porre tre questioni sulle quali occorre lavorare per sfruttare appieno le potenzialità della geolocalizzazione nell'analisi dell'opinione pubblica sui *social*.

- **Migliorare i sistemi di *geotagging***: al momento gli utenti che usano le coordinate geografiche non sono abbastanza numerosi, peraltro la localizzazione latitudine-longitudine può essere fuorviante se si vuole considerare l'appartenenza territoriale di una persona come una variabile socio-culturale della quale tenere conto nel corso dell'analisi. Nel caso di Twitter la soluzione basata sull'uso della *user location* ci sembra, al momento, soddisfacente.
- **Associare ai dati di opinione, informazioni socio-economiche**: geolocalizzare le opinioni delle persone permette di agganciare all'analisi una serie di informazioni accessorie che possono risultare molto utili in sede di interpretazione. Le statistiche territoriali sono sempre più ricche e dettagliate. L'obiezione che si può fare all'analisi del *sentiment* è che le opinioni espresse su internet galleggino in un "vuoto socio-economico", si può ovviare a questo problema collegando i dati di opinione con le statistiche territoriali, anche comunali. Ad esempio, in un recente test di DEEP - lo strumento di analisi sul quale è basata l'analisi presentata in questo paper - abbiamo collegato, l'analisi dei tweet, con statistiche comunali e dati elettorali. Questo è solo un esempio di come la combinazione tra dati di diversa fonte e origine sia una prospettiva metodologica molto promettente.
- **Usare tecniche quali-quantitative**: l'analisi quantitativa del *sentiment* è ormai uno standard, tuttavia pensiamo che per alcuni fenomeni specifici non sia l'approccio migliore. Le metriche quantitative in alcuni casi non sono applicabili, in altri non colgono aspetti importanti dei fenomeni. La ricchezza informativa dei *social media* può essere sfruttata anche usando tecniche di analisi qualitativa, basate sull'interpretazione dei messaggi e sulla capacità di creare collegamenti tra i comportamenti online e la vita offline delle persone.