



Odi et amo in Twitter

Cristina Bosco, Viviana Patti
Dipartimento di Informatica
Università degli studi di Torino

bosco@di.unito.it, patti@di.unito.it

T4T

3 Dicembre 2016

Team



Giancarlo RUFFO



Cristina BOSCO



Viviana PATTI



Emilio SULIS



Mirko LAI



Marco STRANISCI



Delia I. HERNANDEZ FARIAS

Competenze e aree

- **Natural Language Processing** - Opinion Mining & Sentiment Analysis (OM&SA), sviluppo di risorse linguistiche
- **Data Analysis** applicata ai social media per rilevare dinamiche di comunicazione
- **Data Visualization** - visualizzazione interattiva di informazioni complesse per consentire l'accesso ai dati a vari livelli, e viste sul problema analizzato utili a **prendere decisioni**, **pianificare azioni**, o **mostrare in ambito didattico**

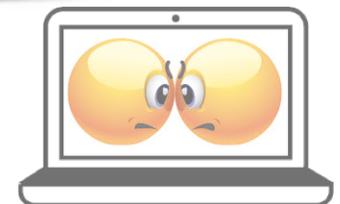


NLP & OM&SA & geo-informatica

- Mettere insieme **approcci diversi** per sviluppare efficaci tecniche di analisi e visualizzazione di big social data con particolare attenzione **all'impatto sociale**
- Piattaforme per il **monitoraggio delle espressioni di felicità/infelicità o odio con finalità didattiche.**
- Tre progetti principali basati su Twitter
- **FELICITÀ** > misurare la **felicità** in Italia
- **FELICITÀ meets SWELLFER** > rilevare **sentimenti** e loro **polarità** rispetto al tema della **fecondità** in Italia
- **HATE SPEECH** > monitorare la diffusione di contenuti di **odio** razziale, omofobico, di genere...



Felicità

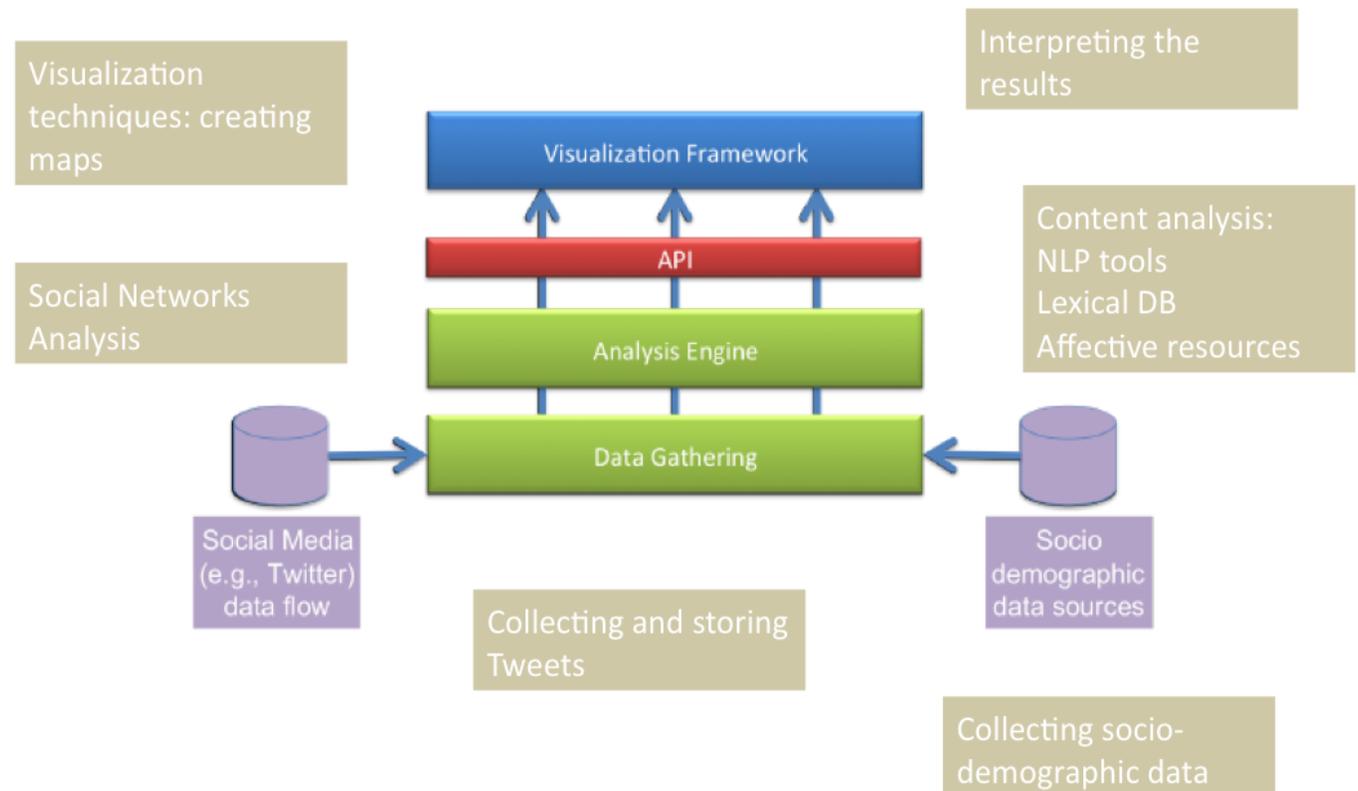


Felicitattà

Raccogliere i dati sulla felicità

- I **social media**: il luogo in cui molte persone comunicano tra loro
- **Twitter**: 313 milioni monthly active users e circa 500 milioni di post al giorno

- Felicità **estrae** i post geolocalizzati e **rileva il grado di felicità** associabile ai luoghi



Felicità

Raccogliere i dati sulla felicità

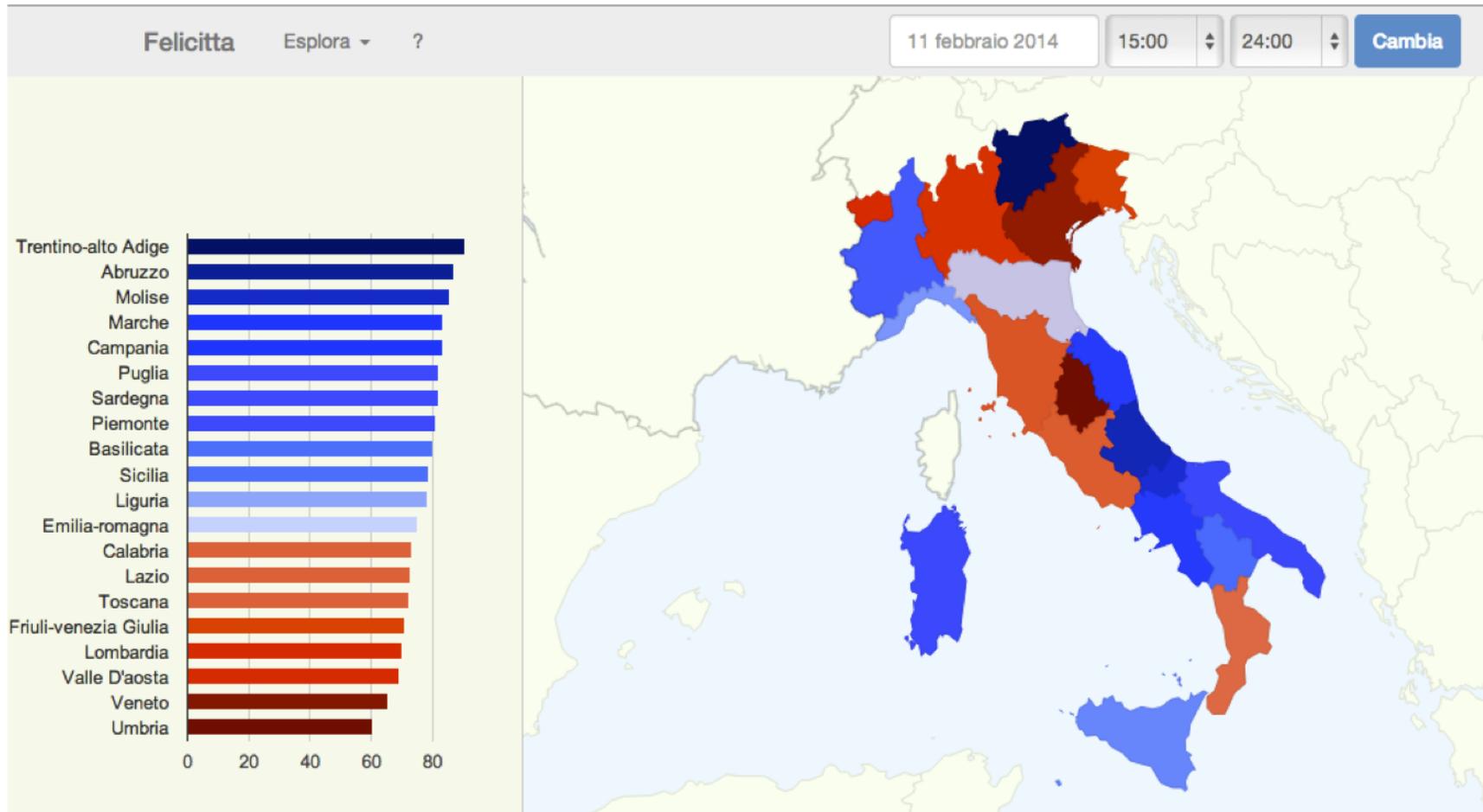
- Nei post si esprimono **opinioni, sentimenti e emozioni** su **eventi, trasmissioni televisive, prodotti di mercato, politici, rock star, sport ...**
- Felicità applica tecniche di **sentiment analysis** e **opinion mining** per estrarre e classificare automaticamente le emozioni e la valenza positiva o negativa espressa nei post, ed aggregarle secondo la distribuzione geografica



Leonardo Allisio, Valeria Mussa, Cristina Bosco, Viviana Patti, Giancarlo Ruffo: Felicità: Visualizing and Estimating Happiness in Italian Cities from Geotagged Tweets. ESSEM@AI*IA 2013: 95-106

Felicità

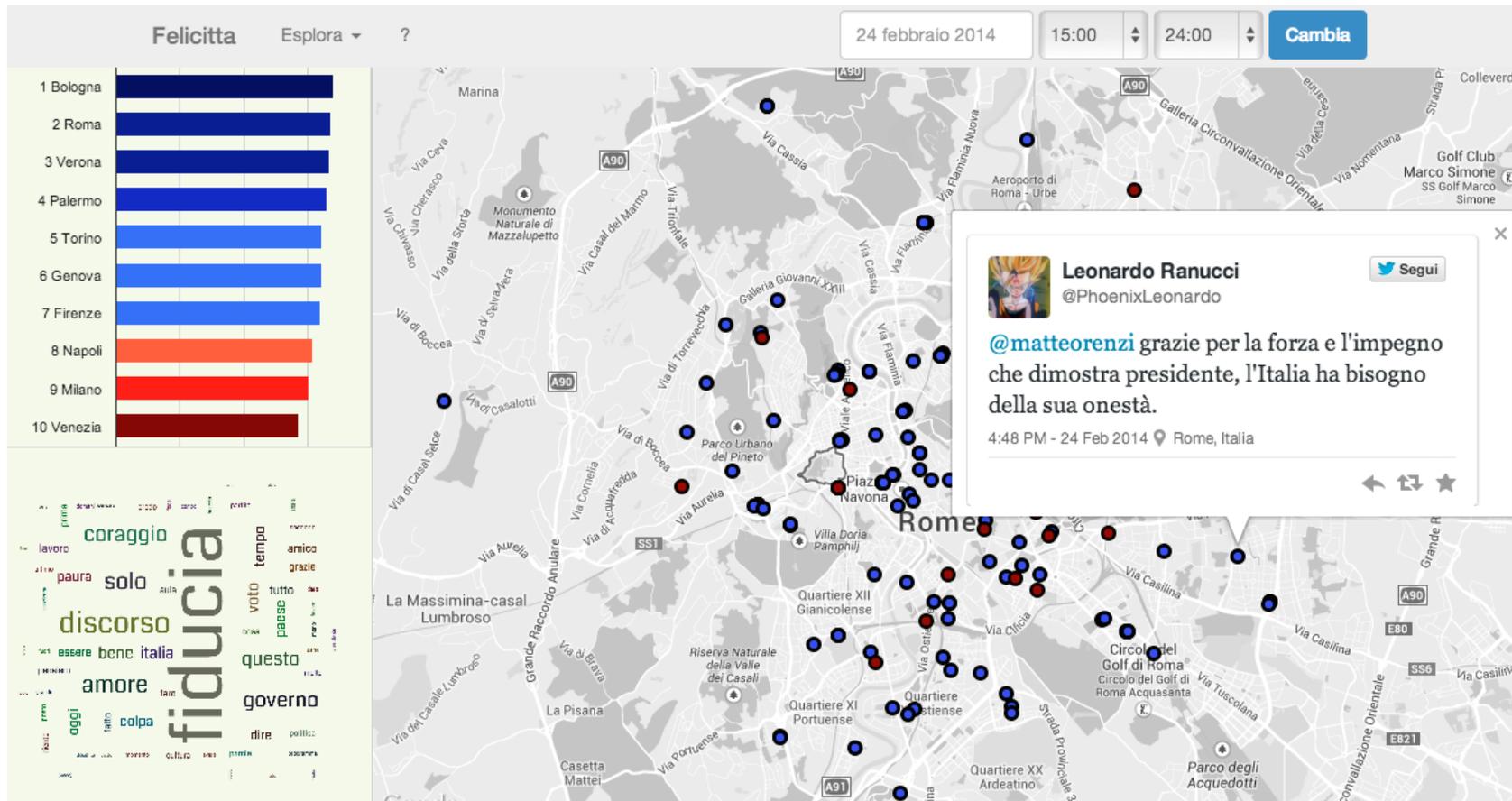
Visualizzare
la felicità



nelle regioni italiane e ...

Felicitto

Visualizzare
la felicità



... nelle città italiane

Felicitto

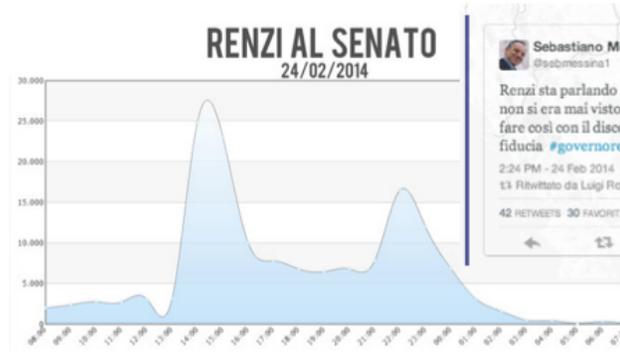
Analisi di Twitter e opportunità

- Per le **imprese** significa poter conoscere le **preferenze dei consumatori**
- Rilevare gli **orientamenti politici** e il **grado di benessere percepito**, o informazioni cruciali in situazioni di **emergenza**

- **Fact checking:**

- reaction to Matteo Renzi's call for a new Italian government
- opinion mining on Renzi's speech 24/02/2014: many tweets conveying a **clear expression of sentiment, mostly positive**

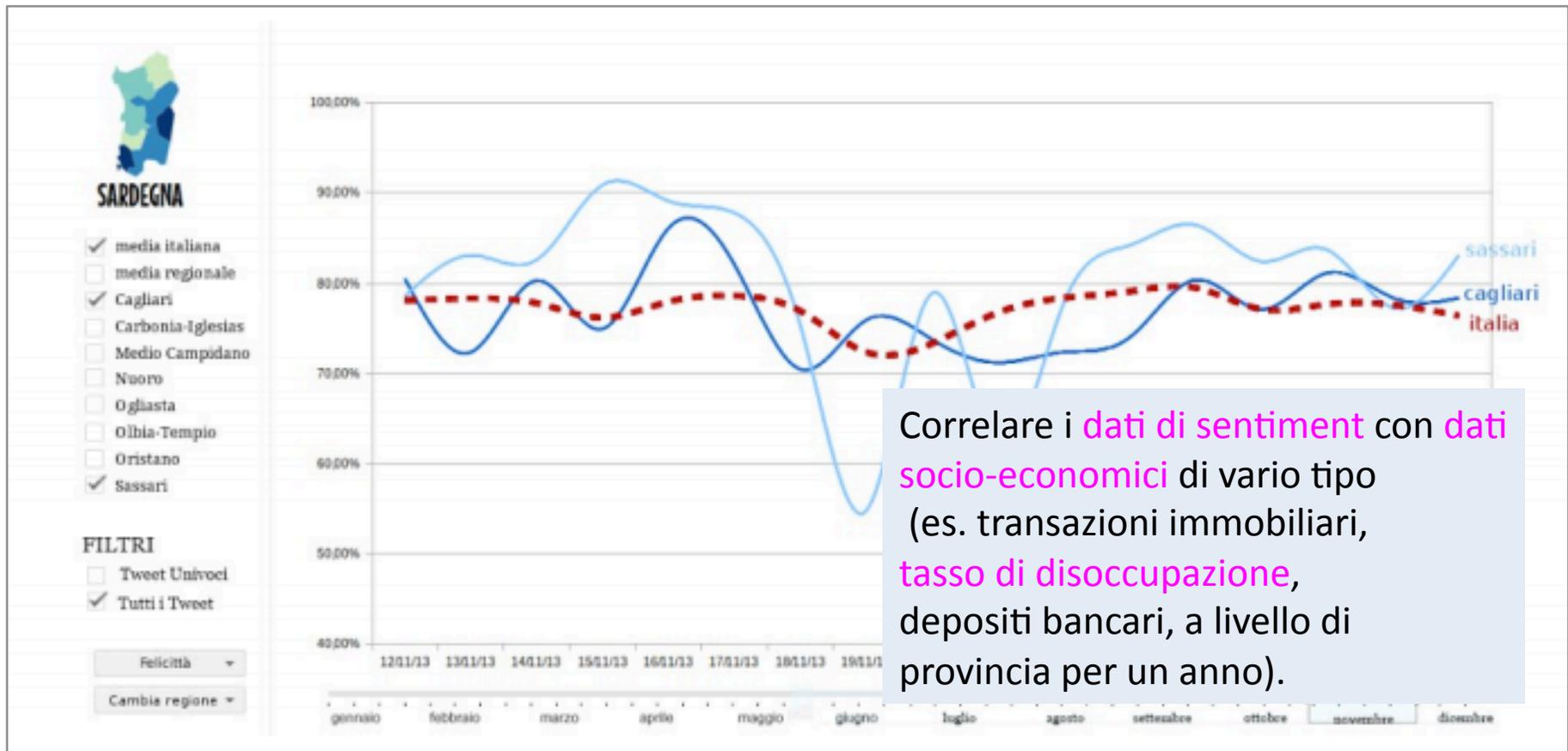
- **Data journalism:**
fact checking



Ironia!

Felicitto

...si evolve



Emilio Sulis, Mirko Lai, Manuela Vinai, Manuela Sanguinetti:
Exploring **Sentiment** in Social Media and **Official Statistics**:
a General Framework. ESSEM@AAMAS 2015: 96-105

Felicità 4 SWELLFER



- **SWELLFER**: ERC starting grant project
Host institution: Centre for Research on Social Dynamics, Bocconi University, Milan, Italy
- Key research questions in SWELLFER: **how subjective wellbeing drives fertility trends (and viceversa)?**
- Idea: usare Twitter data per sviluppare misure di subjective wellbeing (SWB)/ benessere soggettivo + data da analisi tradizionali (es. tassi di fecondità)

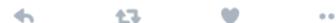


Spinoza.it

virgilio natola



Silvia Calabrese @silvi_cala · 2 set
#fertilityday ...il responso by @spinozait



Felicità 4 SWELLFER



- Happy Parents?
 - Twitter data:
 - Misura di benessere soggettivo costruita in modo indiretto;
 - Possibilità di incorporare – via Twitter data – misure di atteggiamenti nei confronti dei figli e dell'essere o diventare genitori
 - Produrre una copertura geografica molto maggiore che nel caso di normal survey information;
 - Aggregazione geografica della misura di benessere soggettivo a livello regionale: ci sono relazioni con regional fertility levels?

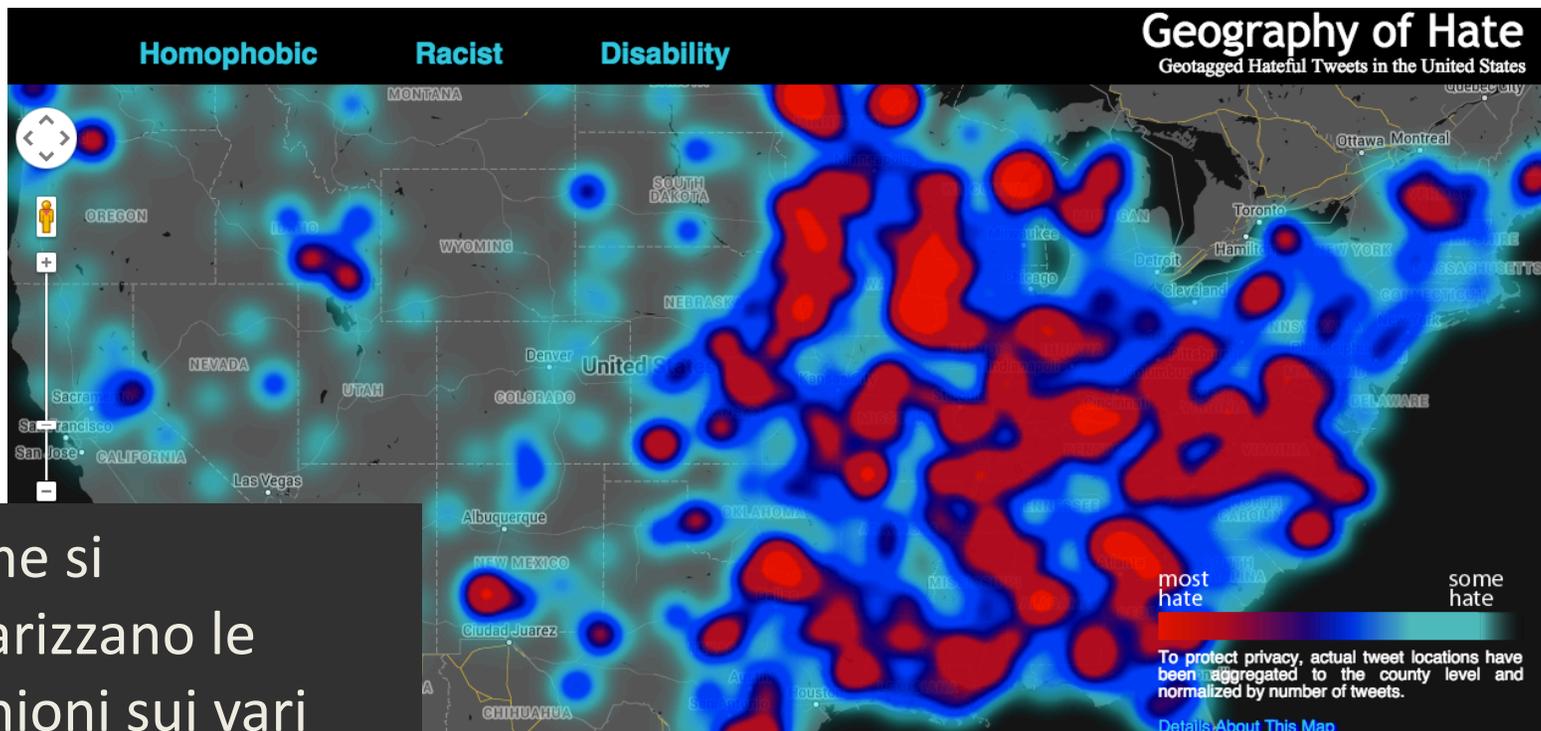
Felicità 4 SWELLFER



- Workshop su [Social Media and Demographic Research: Applications and Implications @ ICWSM 2016 \(Colonia\)](#)
- Versione estesa submitted to SMDR special issue in "Demographic Research" journal [under review]
- [Subjective Well-Being and social media. A semantically annotated Twitter corpus on fertility and parenthood. To appear at CLIC-it 2016.](#)

Monitorare e mappare l'odio

- Analisi di hate speech online
 - http://users.humboldt.edu/mstephens/hate/hate_map.html



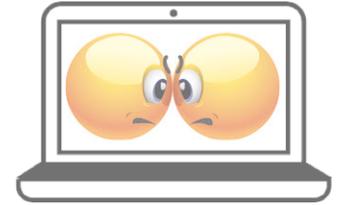
Come si polarizzano le opinioni sui vari target dell'odio?

Monitorare e mappare l'odio

- Italian Hate Map (Twitter data):
 - <http://www.voxdiritti.it/>



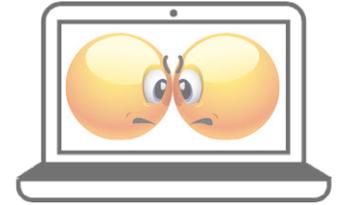
Progetto Hate Speech & Social Media



Obiettivo: rilevazione dell'HS online tramite tecnologie linguisticamente informate, e conseguentemente:

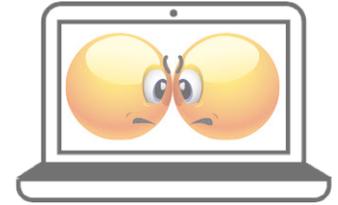
- sviluppo di un'**infrastruttura computazionale per la raccolta, analisi e visualizzazione di big data**, prodotti negli ambienti digitali dei social media, largamente frequentati dagli adolescenti
- **utilizzo di tale infrastruttura a scopo didattico** per la sensibilizzazione degli studenti piemontesi sulle dinamiche di diffusione dell'HS

Focus specifico di attività con la scuola



- Conoscere meglio le **dinamiche della comunicazione** e della **socialità** analizzando testi da social media
- Sviluppo di una piattaforma di **monitoraggio delle espressioni di odio** (razziale, omofobico, di genere) con **finalità didattiche**.
- Analisi degli **hate speech** online come fulcro di un **percorso di educazione** rivolto alla cittadinanza digitale che potrà essere messo a disposizione di tutte le scuole piemontesi
- Sperimentazione con **scuole secondarie di secondo grado**

Impatto sulla ricerca



Avanzamento Sentiment Analysis (SA):

- applicazioni di SA consolidate per marketing e politica
- accuratezza di analisi linguistica anche per interpretazione del linguaggio figurato come **ironia e sarcasmo**
- espansione degli strumenti e risorse per la lingua italiana e per contenuti orientati al sociale

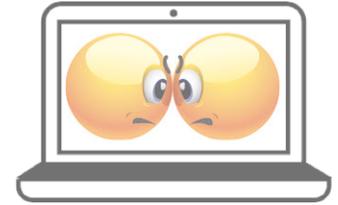
Avanzamento delle tecniche di visualizzazione:

- **dinamicità** dell'applicazione rivolta a visualizzare dati in continua evoluzione

Trattamento di big data:

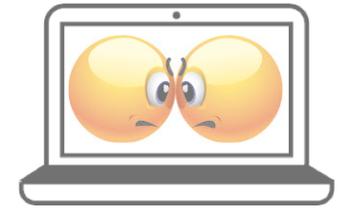
- scalabilità ad altre **aree geografiche**, focus di indagine e tipologie di dati

1: definizione dell'HS in prospettiva socio-linguistica



- **Identificazione dell'HS** nei contesti comunicativi online e raccolta delle espressioni linguistiche che caratterizzano l'HS in lingua italiana
- **Selezione** e pre-processing di **testi** da includere in un corpus di riferimento annotato con polarità di sentimento e sviluppo di risorse lessicali semantiche per l'italiano (PoS tagging di social a Evalita 2016)
- Il gruppo ha esperienza nella costruzione di corpora di contenuto socio-politico (SentiTUT, Felicità, #mariagepourtous, #buonascuola, #independencia, SWELLFER)

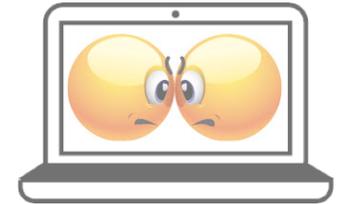
1: definizione dell'HS in prospettiva socio-linguistica



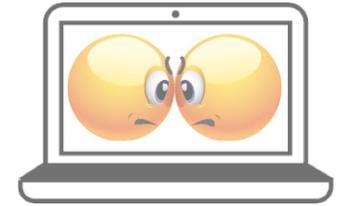
- Possibile attività da svolgere nelle scuole
- Coinvolgere studenti delle scuole secondarie superiori nell'attività di annotazione dei testi per l'individuazione del sentiment sui target dell'odio
 - Espressioni gergali tipiche degli adolescenti
 - Dentro l'hate speech: anche chi ha poco contatto con questi temi può toccare da vicino.
- Esperimento incoraggiante da questo punto di vista: attività nell'ambito dell'alternanza scuola-lavoro
 - 20 ore con studenti del liceo scientifico
 - Analisi di dati di social media
 - Annotazione di dati
 - Utilizzo di piattaforme di annotazione collaborative (Crowdfunder)
 - Corpus #labuonascuola



2: Rilevazione automatica dell'HS

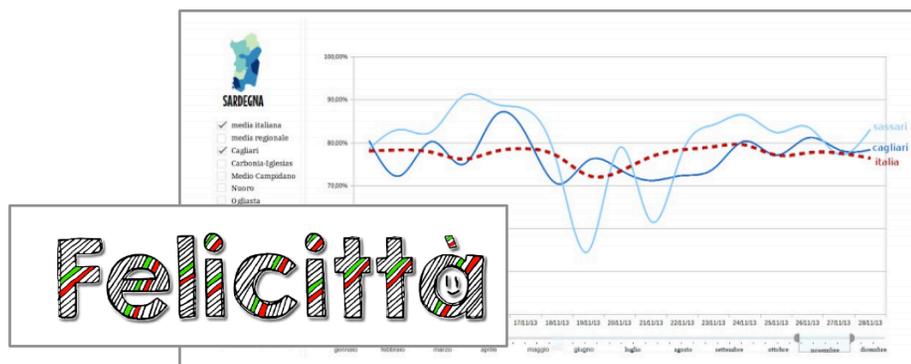


- **Raffinamento di un motore di Sentiment Analysis** (IRADABE) sviluppato in collaborazione con l'Universitat Politecnica di Valencia (UPV) seguendo un approccio di **apprendimento supervisionato** che sfrutta informazioni derivanti da corpora annotati e **risorse lessicali semantiche**
- Il gruppo ha esperienza nello sviluppo di diversi strumenti di SA in collaborazione con UPV:
 - Sistema ValenTO: SemEval 2015-Task 11, Sentiment Analysis of Figurative Language in Twitter
 - Sistema IRADABE: Sentipolc 2014
 - Organizzazione di campagne di valutazione sull'Italiano per la SA su Twitter (Sentipolc @ Evalita 2014 e 2016) e per il Part of Speech tagging di Twitter (POSTWITA @ Evalita 2016)

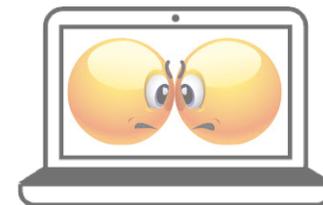


3: Visualizzazione dell'HS

- Sviluppo dell'interfaccia di visualizzazione che soddisfi le domande degli utenti, fornendo i dati in diversi formati di aggregazione statistica, mostrando dinamicamente la diffusione e distribuzione dell'HS in Piemonte e in Torino anche con riferimento a eventi contingenti come fatti di cronaca o talk show, “overview first, zoom and filter, then details on demand”
- Il gruppo ha esperienza nello sviluppo di tecniche di visualizzazione (Felicittà)



4: test e disseminazione: attività con la scuola



- Perché l'interfaccia diventi fulcro di un percorso di educazione sarà oggetto di test con classi pilota (in 10 scuole scelte), di disseminazione grazie alla collaborazione degli insegnanti del **Centro di Iniziativa Democratica (CIDI)**, per poi essere pubblicizzato tramite una campagna apposita e reso disponibile online
- Il gruppo ha una rete di contatti con le scuole di Torino e Piemonte e con varie associazioni legate alla scuola del nostro territorio
 - Acmos (Marco Stranisci, borsista@Dipinfo Master dei Talenti)
 - Commissione scuole Dipinfo UniTO
 - Comitato Unico di Garanzia, UniTO

Contatti

- Cristina Bosco e Viviana Patti
- bosco@di.unito.it, patti@di.unito.it



Alcuni riferimenti bibliografici

- Bosco C., Patti V., Bolioli A. (2013) *Developing corpora for sentiment analysis and opinion mining: the case of irony and Senti-TUT*, IEEE Intelligent Systems
- L. Allisio, V. Mussa, C. Bosco, V. Patti, G. Ruffo (2013) *Felicittà: Visualizing and estimating happiness in Italian cities from geotagged Tweets*. In Proceedings of ESSEM@AI*IA13.
- Bosco C., Allisio L., Mussa V., Patti V., Ruffo G. (2014) *Twitter-based Sentiment Analysis and Sentiment Visualization in Felicittà*, ES³LOD@LREC 2014.
- M. Lai, C. Bosco, V. Patti, D. Virone (2015) *Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization*, IEEE Int. Conf. on Data Science and Advanced Analytics 2015.
- D. I. Hernández Farías, E. Sulis, V. Patti, G. Ruffo, C. Bosco (2015) *ValenTo: Sentiment Analysis of Figurative Language Tweets with Irony and Sarcasm*, SemEval@NAACL-HLT 2015
- C. Bosco, M. Lai, V. Patti, D. Virone (2016) *Tweeting and Being Ironic in the Debate about a Political Reform: the French Annotated Corpus TWitter-MariagePourTous*, LREC 2016.
- M. Stranisci, C. Bosco, D. I. Hernández Farías, V. Patti (2016) *Annotating Sentiment and Irony in the Online Italian Political Debate on #labuonascuola*, LREC 2016.
- E. Sulis, M. Lai, M. Mozzachiodi, C. Bosco, V. Patti, L. Mencarini, D. Vignoli, D. I. Hernández Farías (2016) *Subjective Well-Being and social media. A semantically annotated Twitter corpus on fertility and parenthood*. CLIC 2016.
- E. Sulis, D. I. Hernández Farías, P. Rosso, V. Patti, G. Ruffo. (2016) *Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not*. Knowledge-Based Systems Vol. 108. Elsevier.