



Analisi dei post sui social network per l'individuazione di social event

Università di Pisa
Anno Accademico 2014/2015

Tesi di laurea triennale di Alessio Vaccai
Relatore: Maurizio Tesconi
Correlatori: Mirko Tivosanis, Fabio Del Vigna

**Nuova tendenza:
pubblicare online
ciò che
riteniamo
interessante**





**Spesso anche i
social event
vengono condivisi
online**



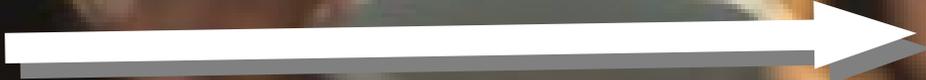
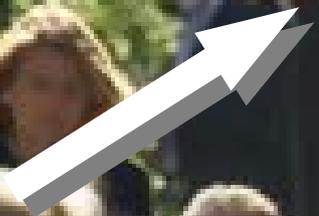
Google
Maps



Language
Detection
API



Instagram





Instagram

- **Social network fotografico**
- **Permette di inviare foto online istantaneamente**
- **Utilizzato attraverso la Search API**



Language Detection API

- Servizio web per l'identificazione di lingue
- Utilizzato per individuare la lingua di un post



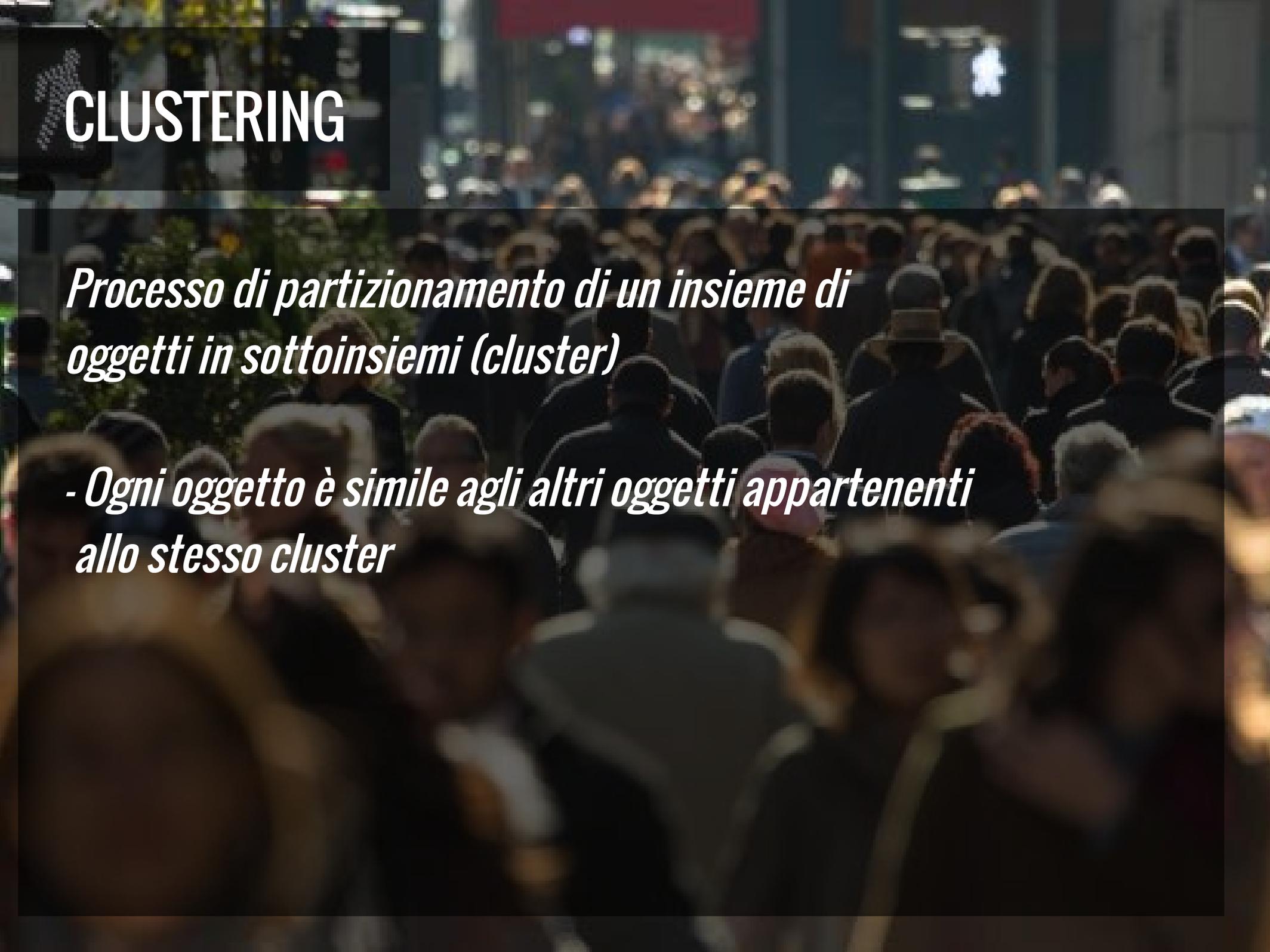
Google
Maps

*Libreria di funzioni JavaScript per la
visualizzazione ed interazione con la mappa.*

Marker



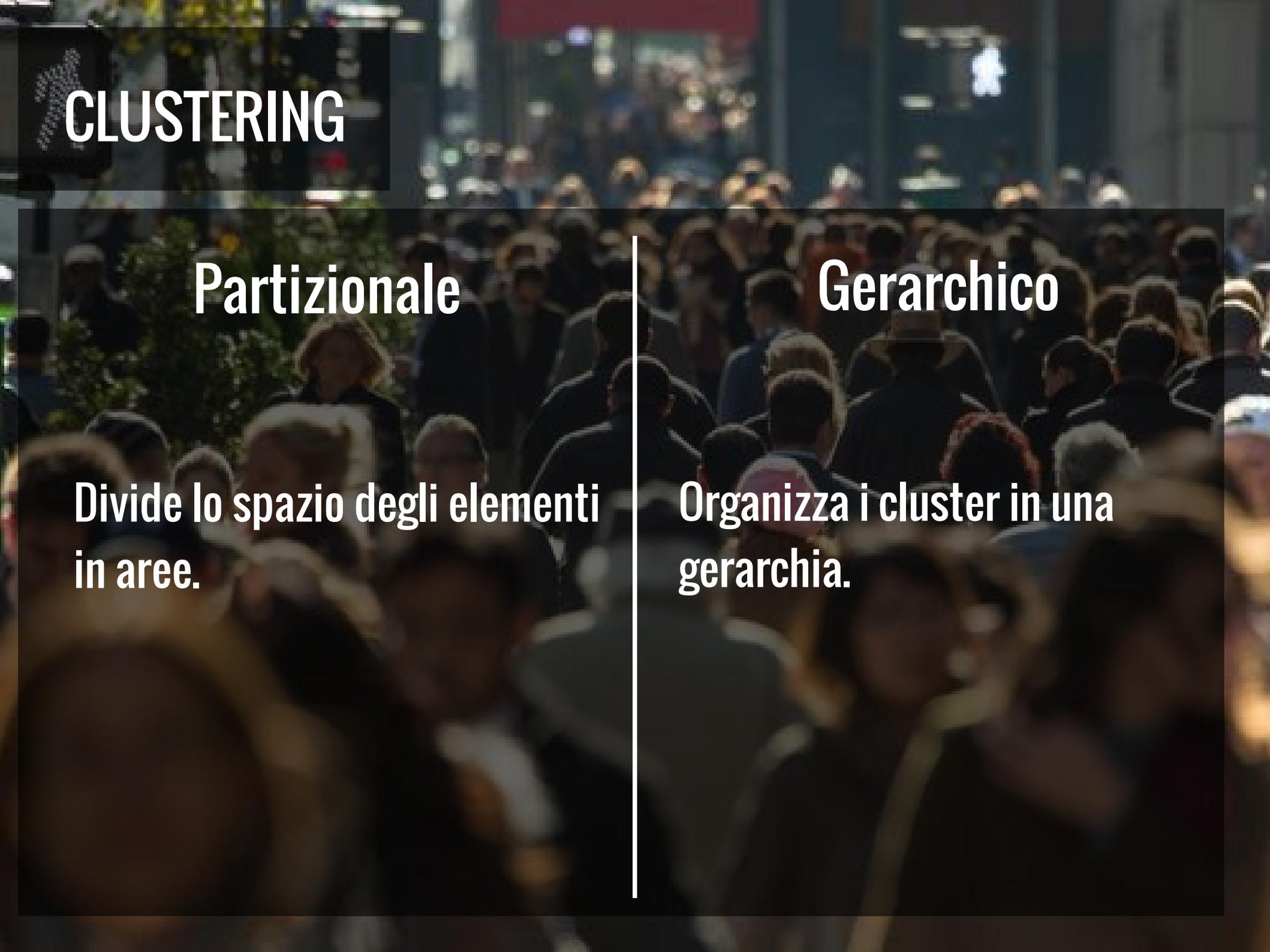
CLUSTERING



Processo di partizionamento di un insieme di oggetti in sottoinsiemi (cluster)

- Ogni oggetto è simile agli altri oggetti appartenenti allo stesso cluster

CLUSTERING



Partizionale

Divide lo spazio degli elementi in aree.

Gerarchico

Organizza i cluster in una gerarchia.

CLUSTERING PARTIZIONALE

K-Means

- Preso in input un insieme di punti, lo divide in k sottoinsiemi ed assegna ogni punto al sottoinsieme appropriato.

Svantaggi

- Conoscenza a priori di k
- Possibilità di creare solo cluster di forma sferica o ellittica
- Clusterizzazione troppo approssimativa

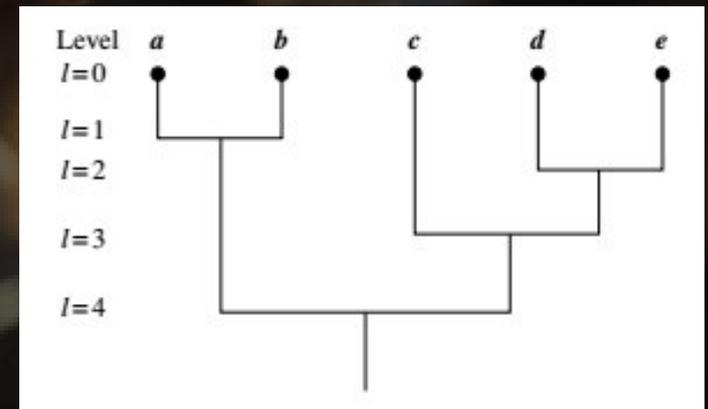
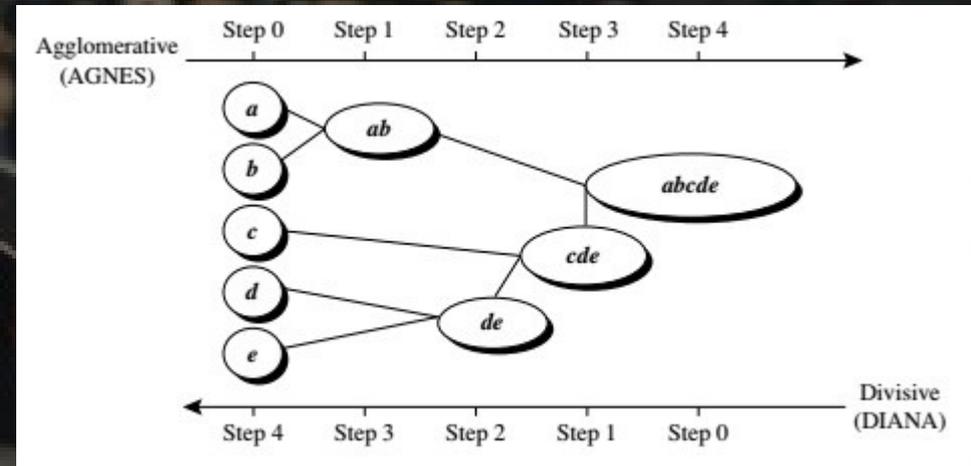
CLUSTERING GERARCHICO

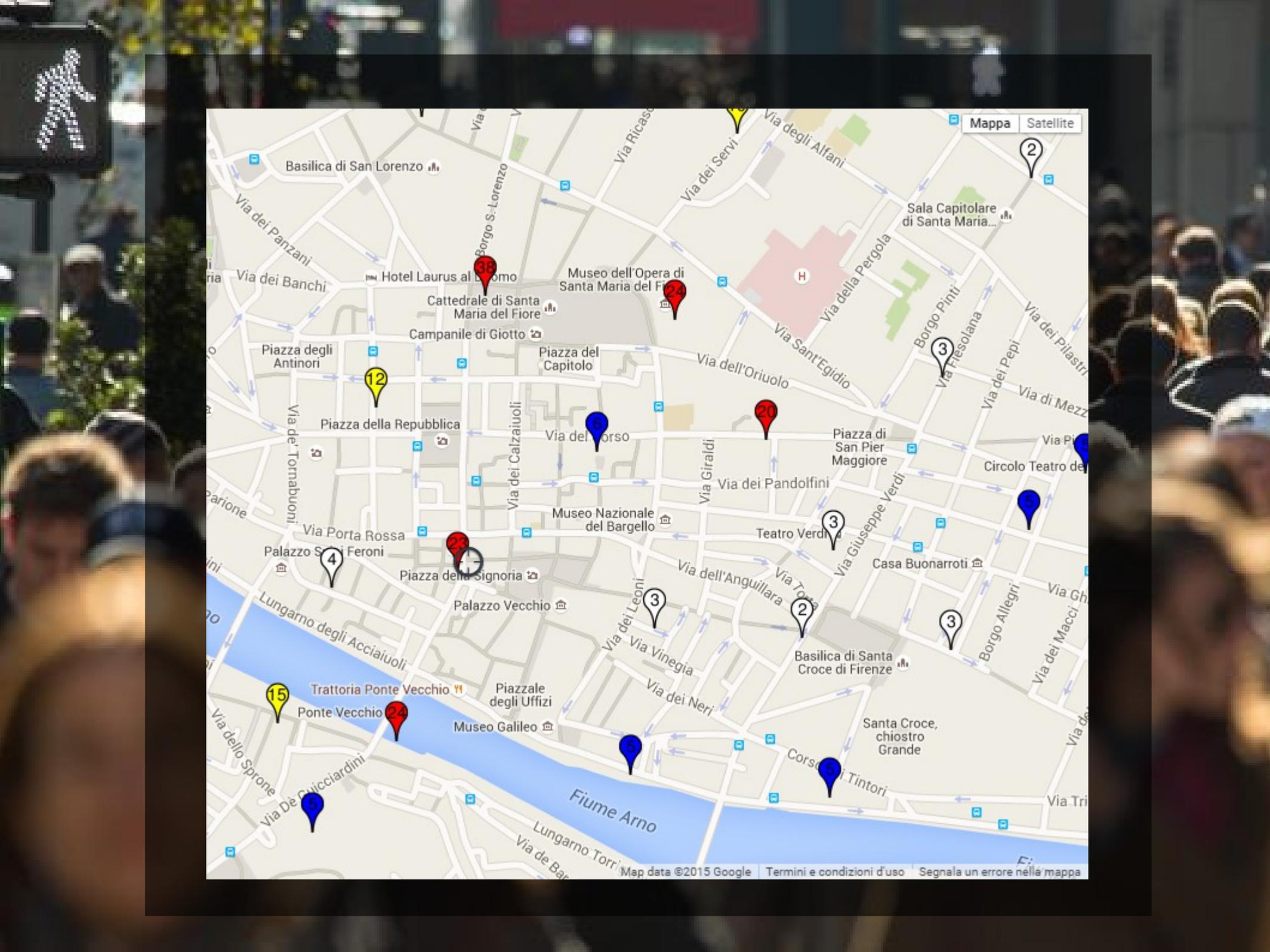
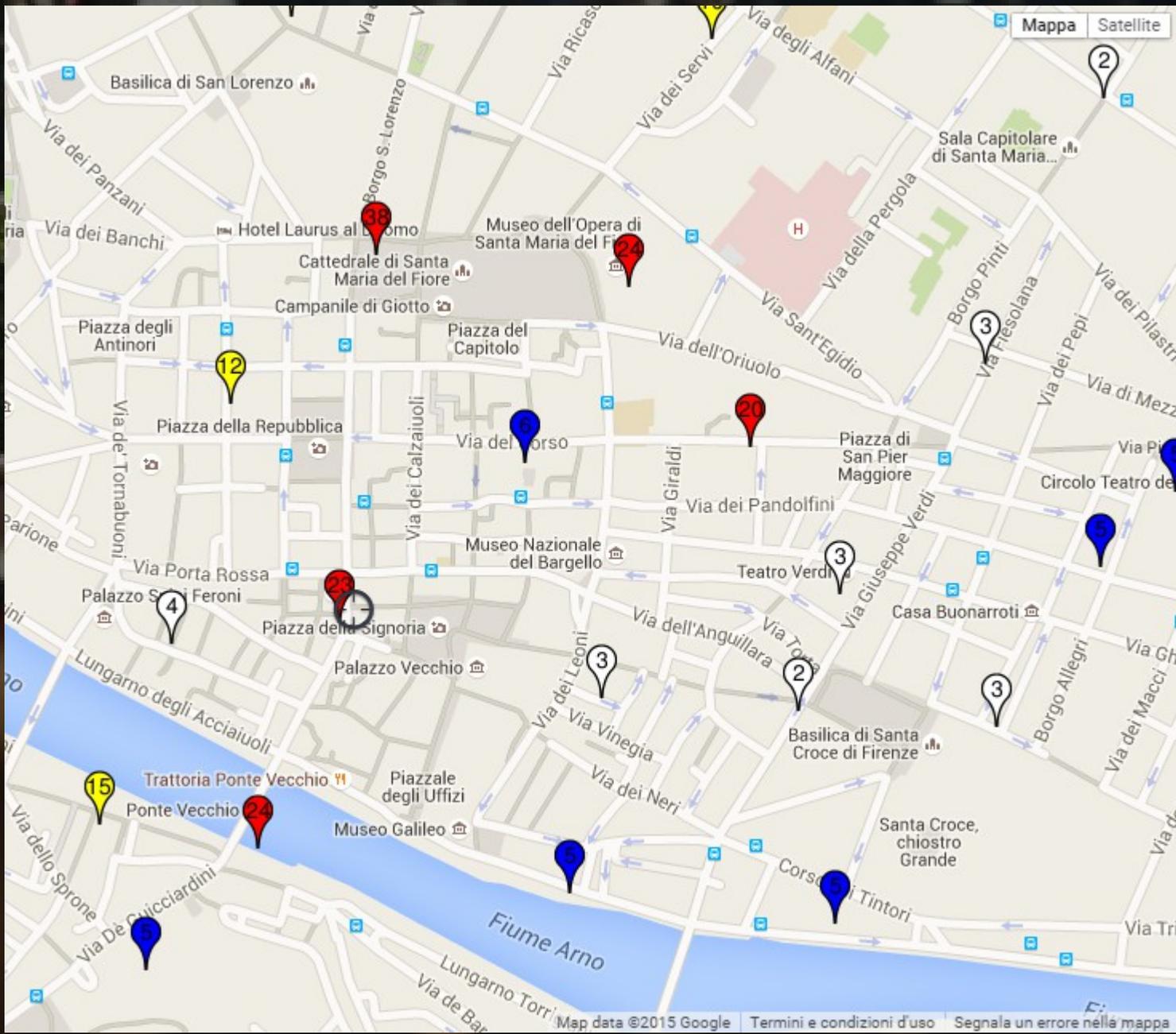
- Agglomerativo o Divisivo

- Struttura risultante:
Dendrogramma

- Vantaggi: Possibilità di navigare
il dendrogramma

- Svantaggi: algoritmo rigido





INDIVIDUAZIONE DELLA LINGUA

Categorizzazione mediante n-grammi

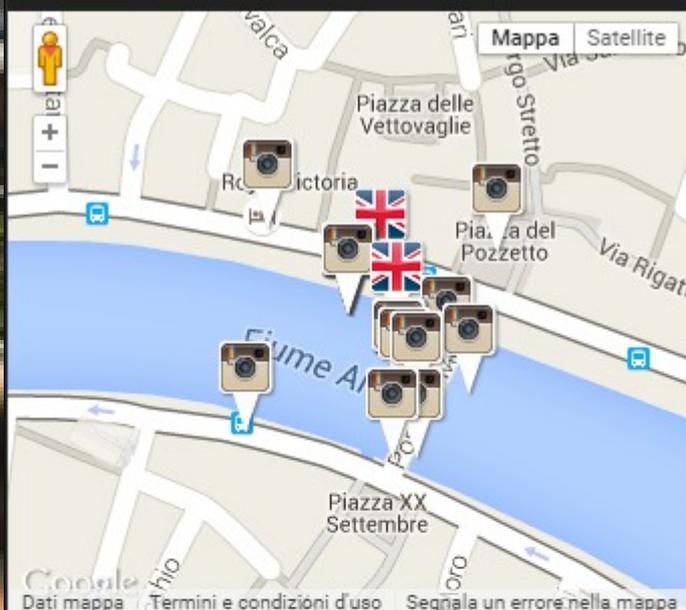
Bigrammi: _c, ci, ia, ao, o_

Trigrammi: _ci, cia, iao, ao_, o__

Tetragrammi: _cia, ciao, iao_, ao__, o___

Pisa, dalle 21.00 del 16/6/2014 alle 01.00 del 17/06/2014

Dettagli



Rank: 29

Lingue individuate



(5)



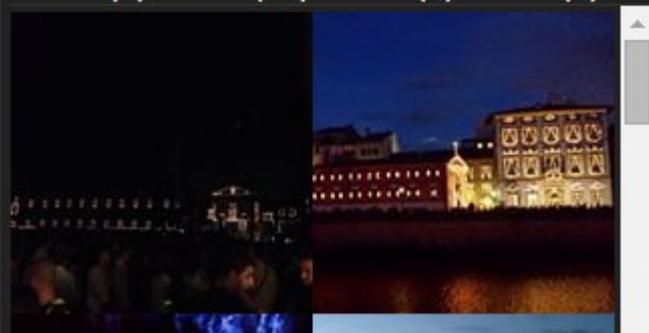
(21)



(1)



(2)

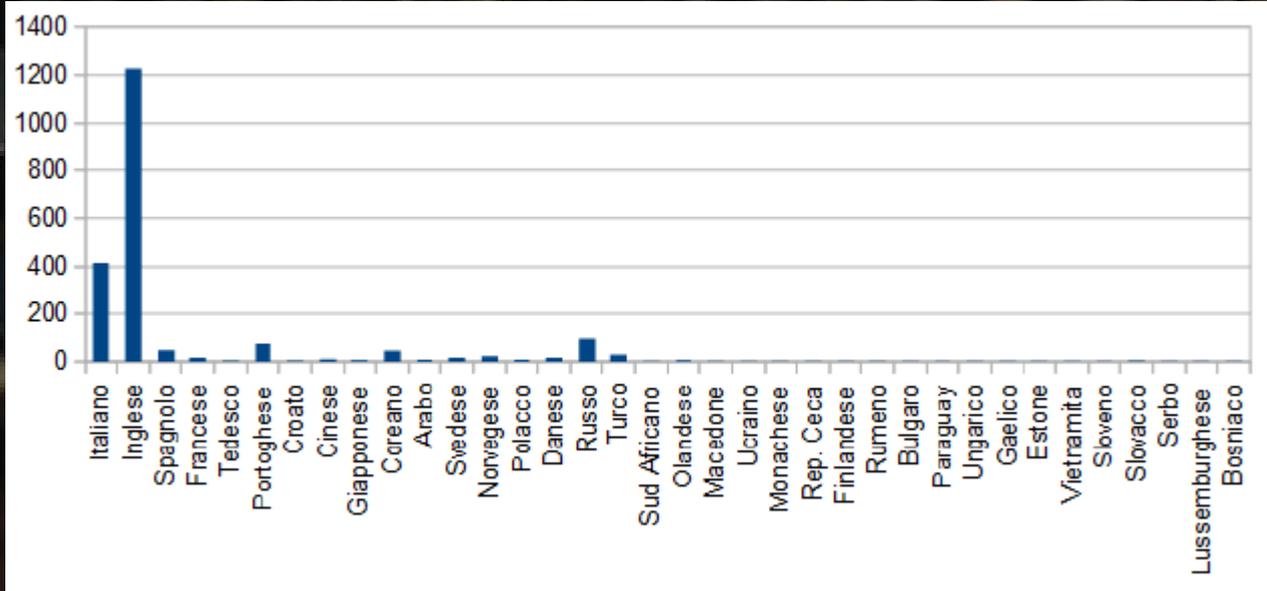


Tag in questo Cluster

pisa(19) luminara(11) italy(5) sanranieri(3)
luminara2014(3) igerspisa(3) festa(3) lungamo(2) luminaria(2)
night(2) instagood(2) 2014(2) igersitalia(2) luminare(2)
sanranier2014(2) picoftheday(2) ingerstoscana(1) exposure(1)
capture(1) pictures(1) all_shots(1) pics(1) snapshot(1) (1)
beautiful(1) photooftheday(1) picture(1) art(1) ranieri(1)
san(1) color(1) with(1) friends(1) out(1) light(1)
itsabeautifulife(1) photo(1) fun(1) party(1) june(1) lights(1)
ingerspisa(1) ingerstuscany(1) italia(1) bridge(1) igerslucca(1)
visitpisa(1) fireworks(1) sky(1) pic(1) focus(1) igerstoscana(1)
s(1) photos(1) summer(1) moment(1) visittuscany(1)
visititaly(1) tangled(1) adult(1) legal(1) lanternfestival(1)
voldygreen(1) composition(1) nassavera(1) potooftheday(1)
lungami(1) top(1) tuscany(1) luminaria2014(1)
luminiera2014(1) fuochi(1) inamo(1) instapisa(1) colors(1)
fuochidiartificio(1) nofilter(1) nice(1) selfiegrou(1)
lafestapiubelladellatostana(1)

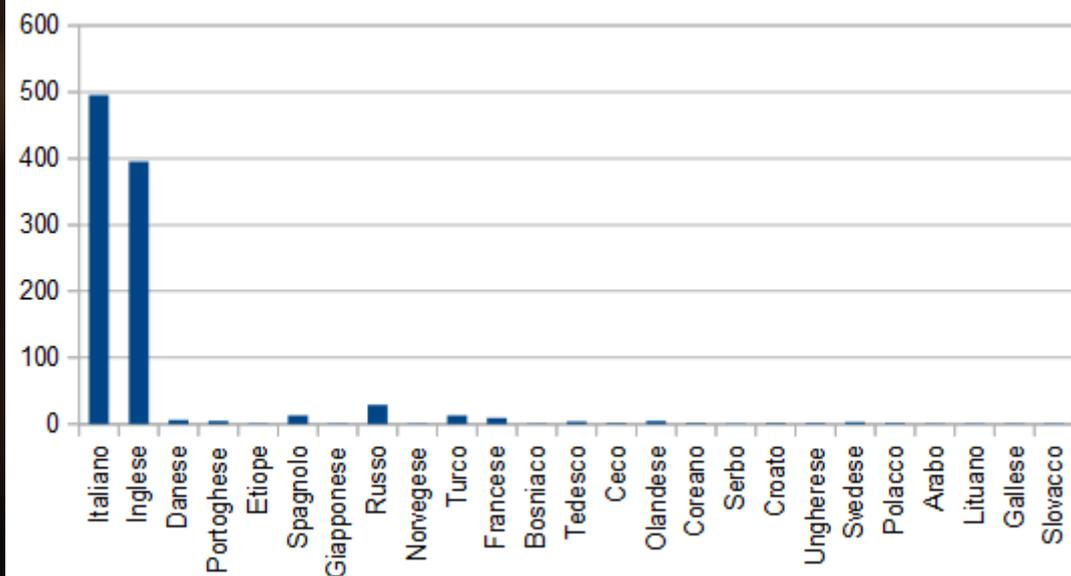
RISULTATI

Aprile - Maggio 2015



Zona Torre di Pisa

Non Classificati:
2090

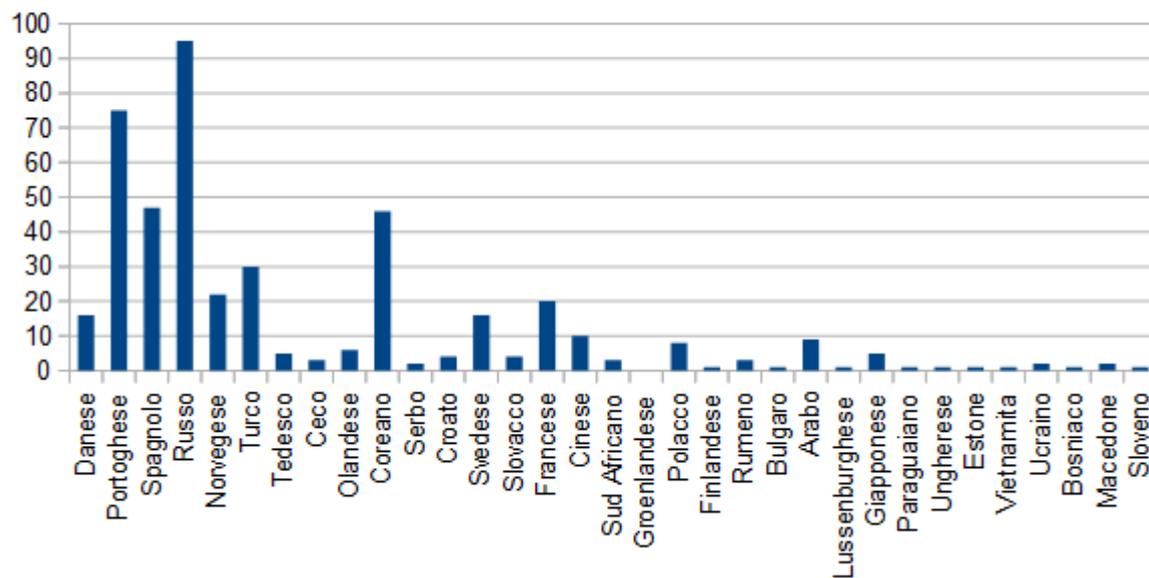


Montecatini

Non Classificati:
1033

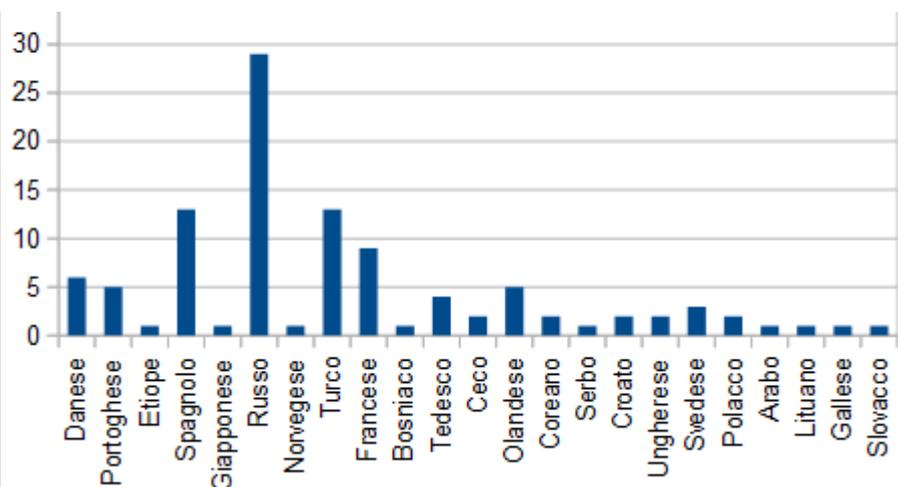
RISULTATI

Aprile - Maggio 2015



Escludendo Italiano e Inglese

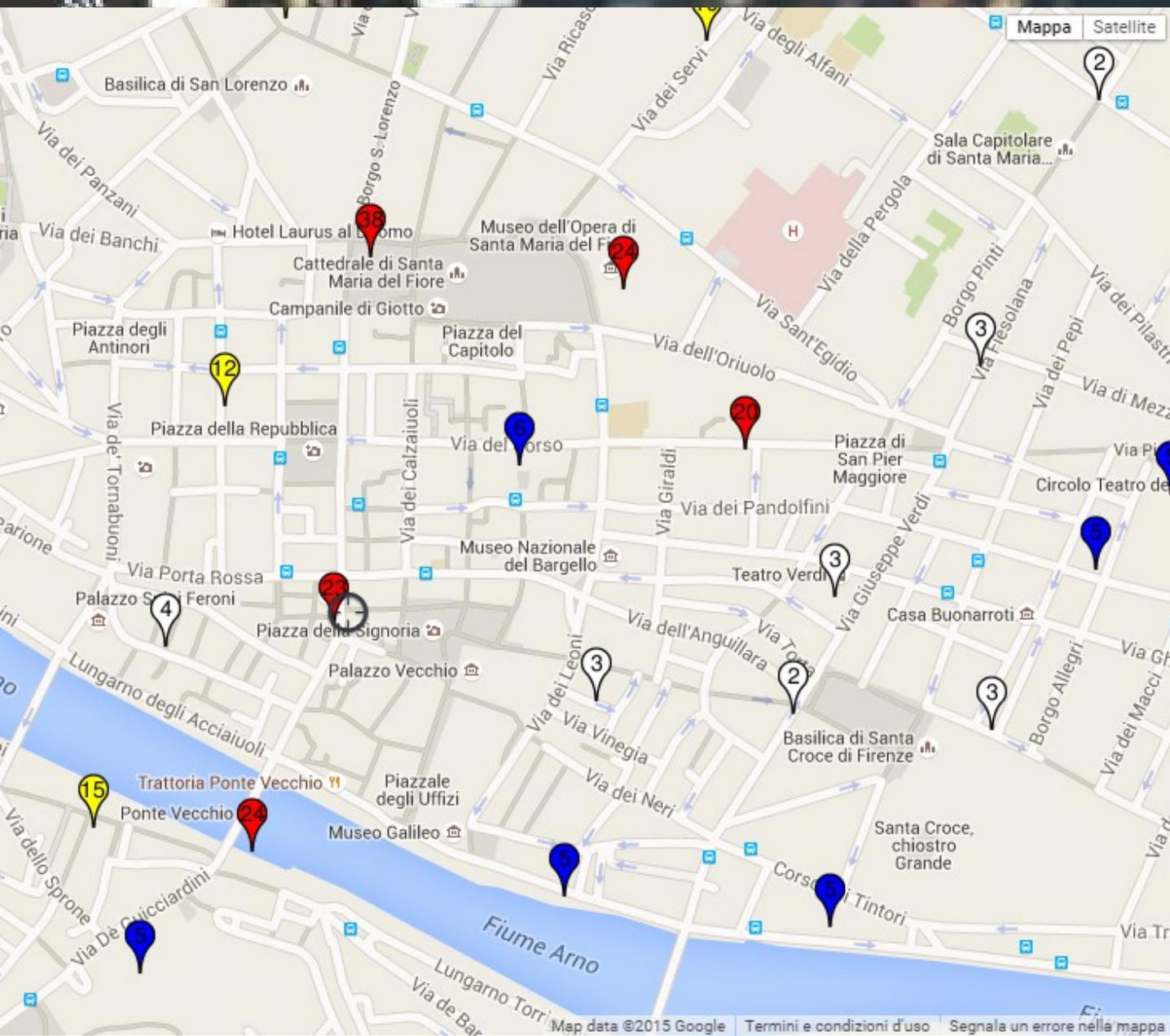
Zona Torre di Pisa



Montecatini

FUTURE WORKS

- Filtrare i cluster rimuovendo i POI delle città.
- Utilizzare una combinazione di metadati per migliorare il clustering.
- Aggiungere altri social network alle fonti di dati.



What's the Buzz

Ricerca

Dal
26/05/2015 13:00

Ai
26/05/2015 17:00

Raggio (metri)
3000

Cerca Qui

Risultati

Visualizza cluster più grandi di 0

Dettagli

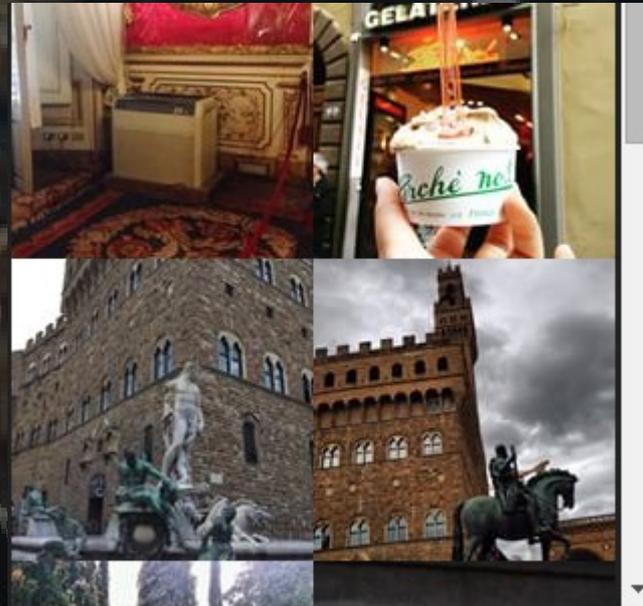


Dettagli



Rank: 16

Lingue individuate



Tag in questo Cluster

firenze(3) palazzovecchio(2) florence(2) gelato(2)
кругофаллосы(1) (1) nomnom(1) good(1) streetart(1)
luminibitch(1) bitches(1) dark(1) tag(1) graffiti(1)
fluminibitch(1) urbanart(1) italy(1) lumini(1) tags(1) flumini(1)
bad(1) black(1) night(1) italia(1) bitch(1) loverp(1)
rpfors(1) rpwithme(1) rpforgirls(1) straightrp(1) hmrp(1)
dirtyrp(1) single(1) cleanrp(1) openrp(1) rp(1) couple(1)
rpforguys(1) satisfaction(1) toscana(1) clouds(1)
cutecatsofinstagram(1) cute(1) eyes(1) catstagram(1) kitty(1)
blackcatsrule(1) ohnoes(1) blackcat(1) freyja(1) rescuecat(1)
a_world_of_cats(1) yourdailypets(1) meow(1) dailycat(1)
stelios(1) catsofinstagram(1) excellent_cats(1) strangerdanger(1)
cat(1) cats(1) cat_features(1) catlounge(1) adorable(1)
foreverkitten(1) piazzasignoria(1) enjoy(1) summer(1) life(1)
love(1) pausa(1) moments(1) sweet(1) droga(1) chocolate(1)
ciao(1) yummy(1) eat(1) italya(1) floransa(1)